Emerging Technologies in Engineering

Emerging Technologies in Engineering

Editors-in-Chief

Mahesh P K
Don Bosco Institute of Technology, India

Su-Qun Cao
Huaiyin Institute of Technology, China

Ghanshyam Singh
Malaviya National Institute of Technology, India

McGraw Hill Education (India) Private Limited
CHENNAI

McGraw Hill Education Offices

Chennai New York St Louis San Francisco Auckland Bogotá Caracas
Kuala Lumpur Lisbon London Madrid Mexico City Milan Montreal
San Juan Santiago Singapore Sydney Tokyo Toronto

McGraw Hill Education (India) Private Limited

Published by McGraw Hill Education (India) Private Limited,
444/1, Sri Ekambara Naicker Industrial Estate, Alapakkam, Porur, Chennai-600 116, Tamil Nadu, India

Emerging Technologies in Engineering

ISBN (13): 978-93-5260-317-6
ISBN (10): 93-5260-317-6

Managing Director: *Kaushik Bellani*

Director: Products (Higher Education and Professional): *Vibha Mahajan*

Manager—Product Development (B&E–BGR): *Hemant K Jha*
Senior Specialist—Product Development (B&E–BGR): *Laxmi Singh*
Senior Coordinator—Product Development: *Kunal Adhikari*

Head—Production (Higher Education and Professional): *Satinder S Baveja*
Assistant Manager—Production: *Jagriti Kundu*

General Manager—Production: *Rajender P Ghansela*
Manager—Production: *Reji Kumar*

Typeset at Text-o-Graphics, B-1/56, Aravali Apartment, Sector-34, Noida 201 301, and printed at

Cover Printer:

Committees

Honorary Chair
Dr. Shuvra Das (University of Detroit Mercy, USA)
Dr. Jiguo Yu (Qufu Normal University, China)
Dr. Pawel Hitczenko (Drexel University, USA)
Dr. Lisa Osadciw (Syracuse University, USA)
Dr. Jiguo Yu (Qufu Normal University, China)
Prof. Harry E. Ruda (University of Toronto, Canada)

Technical Chair
Dr. Anooj P K (Al Musanna College of Technology, Sultanate of Oman)
Dr. Urmila Shrawankar (G H Raisoni College of Engineering, India)
Dr. Ching-Chih Tsai (IEEE SMCS TC on Intelligent Learning on Control Systems)
Dr. A G Hessami (IEEE SMCS TC on Systems Safety & Security)
Dr. Xuelong Li (Birkbeck College, University of London, U.K.)
Dr. R Vijayakumar (MG University, India)
Dr. N.Nagarajan (Anna University, Coimbatore)

Chief Editors
Dr. Mahesh P.K (Don Bosco Institute of Technology, India)
Dr. Su-Qun Cao (Huaiyin Institute of Technology, China)
Dr. Ghanshyam Singh (Malaviya National Institute of Technology, India)

Technical Co-Chair
Dr. Natarajan Meghanathan (Jackson State University, USA)
Dr. Hicham Elzabadani (American University in Dubai)
Dr. Pingkun Yan (Philip Research North America)

General Chair
Dr. Janahanlal Stephen (Matha College of Technology, India)
Dr. Yogesh Chaba (Guru Jambeswara University, India)

General Co-Chair
Prof. K. U Abraham (Holykings College of Engineering, India)

Finance Chair
Dr. Gylson Thomas (Thejus Engineering College, India)
Dr. Zhenyu Y Angz (Florida International University, USA)

Publicity Chair
Dr. Amit Manocha (Maharaja Agrasen Institute of Technology, India)

Publicity Co-Chair
Dr. Ford Lumban Gaol (University of Indonesia)
Dr. Amlan Chakrabarti (University of Culcutta, India)

Publication Chair
Dr. Vijayakumar (NSS Engg. College, India)
Dr. T.S.B.Sudarshan (BITS Pilani, India)
Dr. KP Soman (Amritha University, India)

Program Committee Chair
Dr. Harry E. Ruda (University of Toronto, Canada)
Dr Deepak Laxmi Narasimha (University of Malaya, Malaysia)
Dr.N.Nagarajan (Anna University, Coimbatore, India)
Prof. Akash Rajak (Krishna Institute of Engg. & Tech., UP, India)
Dr. Javed Vassilis Khan (Academy for Digital Entertainment, The Netherlands)

International Advisory Committee
Dr. Pawel Hitczenko (Drexel University, USA)
Dr. Kristian J. Hammond (Northwestern University, USA)
Dr. Long Que (Louisiana Tech University, USA)
Dr. Peter Vadasz (Northern Arizona University, USA)
Dr. Hamid Bagheri (Sharif University of Technology, Iran)

National Advisory Committee
Dr. Saurabh Chaudhury (IIT Kharagpur, India)
Dr. R.Kumar (SRM University, India)
Dr. M Kumar (SIRT, Bhopal, India)

Program Committee Members
Dr. Shu-Ching Chen (Florida International University, USA)
Dr. T.S.B.Sudarshan (BITS Pilani, India)
Dr. Habibollah Haro (Universiti Teknologi Malaysia)
Dr. V. K. Bhat (SMVD University, India)
Dr. Keivan Navi (Shahid Beheshti University, Tehran)
Prof. Anil K Jacob (MES College of Engineering, India)
Dr. B. Kannan (CUSAT, India)
Dr. Liviu Vladutu (Dublin City University, Ireland)
Dr. Malabika Basu (Dublin Institute of Technology, Ireland)
Dr. Pritimoy Sanyal (West Bengal University of Technology, India)
Prof. J.Arunadevi (Thiagarajar School of Management, India)
Dr. R.D. Sudhaker Samuel (S J College of Engineering, India)
Dr. Amitabha Sinha (West Bengal University of Technology, India)
Prof. Ananta Ojha (ICFAI University, India)
Dr. A K Sharma (YMCA Institute of Engineering, India)
Dr. Debasis Giri (IIT Kharagpur, India)
Prof. Suparna Biswas (WBUT, India)

Foreword

It's a great privilege and pleasure for me to serve as the Editor-in-chief for the IDES Joint International Conferences. Innovative ideas and research in two streams are extremely important for the current Electronics industry to be in support to "Digital India" and "Make in India" an initiative by Government of India. This Joint International Conference provides a rostrum to the researchers from the academia and industries all around the world to share their research results, novel ideas as well as the improvements over the existing methodology.

This conference covers a wide variety of topics in Control Systems and Power Electronics, to name a few—Mobile Communication Technology, Natural Language Processing, Algorithm/Protocol Design and Analysis, VLSI Systems, Intelligent Systems and Approach, Data Communication, Embedded System, Digital Security, Data Compression, Data Mining, Databases, Digital Signal Processing, Telecommunication Technologies, Control Theory and Application, Computational Intelligence, Robotics, HVDC, MEMS-Related Technology. The response from the researchers for the research papers is staggering.

I would like to convey my heartfelt gratitude and appreciation to the members of the following committees—Honorary Chair, Technical Chair, Technical Co-Chair, General Chair, General Co-Chair, Publicity Chair, Publicity Co-Chair, Publication Chair, Finance Chair, National Advisory Committees, Program Committee Chair, International Advisory Committee, Review Committee, Program Committee Members for contributing either their precious time in reviewing the papers or their effort in monitoring and making the conference a grand success. I would also like to acknowledge the support of IDES, Matha College of Technology, Association of Computer Electrical Electronics and Communication Engineers (ACEECom), ACEE and AMAE for organising such a platform to welcome the future technology. I also wish to convey my gratitude to McGraw-Hill Education for publishing the registered papers.

Mahesh P K
Don Bosco Institute of Technology, India

Preface

The IDES Joint International Conference "Emerging Technologies in Engineering" attracted several hundreds of submissions. After rigorous reviews, 63 high-quality papers are included in the proceedings.

The goal of the joint conference is to promote research, developmental activities and scientific information interchange between researchers, developers, engineers, students, and practitioners working in India and around the world in the fields of Computer Science, Information Technology, Computational Engineering, Communication, Electrical Measurements, Instrumentation Engineering, Electronic Devices, Digital Electronics, Circuits, Control and Instrumentation, Communication system, Robotics, Power Electronics, Civil Engineering and Power Engineering.

The conference is jointly organised by the ECE Department of Matha College of Technology, the IDES, ACEECom, ACEE and AMAE. I thank the members of the Organizing Committee and the Programming Committee for their hard working in the past several months. I wish to express my heartfelt appreciation to the keynote speakers, session chairs, reviewers and student helpers. Finally, I thank all the authors and participants for their great contributions and exchanging the experiences.

Su-Qun Cao
Huaiyin Institute of Technology, China

Contents

Non-Word Error Detection for Luganda

Robert Ssali Balagadde * and Parvataneni Premchand **
* ** Department of Computer Science & Engineering, University College of Engineering, Osmania University
Hyderabad, 500007, TS, India
baross.kla@gmail.com, Profpremchand.p@gmail.com

Abstract: Editing or word processing Luganda text has been an uphill task mainly because of lack of a system in this environment which could detect spelling errors in Luganda. In this context, this research paper presents a model for non-word error detection for Luganda (LugDetect) which comes in handy to address this gap and consequently provide a more user friendly environment for editing Luganda text. To the best of our knowledge LugDetect is the first of this kind of system developed for Luganda. Experimentation results show that LugDetect detects non-word errors with an accuracy (A_P) of 100% for all the five categories of Luganda words at an average speed of 1471 Hz (number of words per second) so long as the erroneous word is not a real word. Experimentation on the Luganda corpus which was used in this research work shows that 19% of Luganda text is composed of *Clitics Host Word Combination (CHWC),* while the other part (81%)- *Real Luganda Words (RLW).*

Keywords: non-word error detection, Luganda error detector, dictionary look-up technique, clitic-host word combination (CHWC), spelling detector.

Introduction

Editing or word processing Luganda text has been an uphill task mainly because of lack of a environmental system which could detect spelling errors in Luganda. In this context, this research work presents a model for *Non-Word Error Detection For Luganda (NWEDL- LugDetect)* which comes in handy to address this gap, and consequently, provide a more user friendly environment for editing Luganda text. To the best of our knowledge, LugDetect is the first of this kind of system developed for Luganda.

One challenge encountered while developing a model for Luganda spell checking is dealing with the infinite number of clitic-host word combinations (CHWCs) which makes Luganda distinct from other foreign languages especially non-Bantu languages. In this research work, three types of CHWC are identified in respect to the use of *inter-word apostrophe (IWA*) defined in subsection on ”The Error Detection Mechanism"

Type one - CHWC_1, bolded in Example 1 - are created as a result of compounding a modified monosyllabic word (MMW) or modified disyllabic word with an initial vowel (MDWIV) with another succeeding word which begins with a vowel. Compounding in Luganda may involve two to three words. MMW or MDWIV- referred to as clitic - is formed by dropping the ending vowel of the mono-syllabic word (MW) or disyllabic word with an initial vowel (DWIV) and replacing it with an apostrophe resulting into a long sound that is not represented by a double vowel.

Example 1

- *omwenge* ***n'ennyama*** *(alcohol and meat) [conjunctive form]*
- ***n'otema*** *omuti (and you cut the tree) [narrative form]*
- ***ew'omuyizzi*** *(at the hunter's place) [locative form]*
- *Minisita omubeezi* ***ow'ebyobusuubuzi*** *ne tekinologiya (Minister of State for Trade and Technology)*
- ***n'obw'embwa*** *sibwagala (even that for the dog, I don't like)*
- ***n'olw'ensonga*** *eyo, sijja kujja (and for that reason, I will not come)*

Example 2 shows some MWs and DWIVs (bolded) used in compounding to form CHWC_Is. Not all MWs and DWIVs are used in compounding, and Example 3 shows some of these.

Example 2

- *ne, na, nga [conjunction]*
- *ne, nga [narratives]*
- *be, ze, ge. gwe , bwe, bye, lwe, lye, kye, bye [object relatives]*
- *kya, ya, za. lya ba, bya, ga, gwa, ka, lwa, wa [possessives]*
- ***ekya, eya, eza. erya, owa, aba*** *[possessives with initial vowel]*
- *kye, ye, be, ze, twe [copulatives]*
- *e [locatives]*

Example 3

- *era, ate, nti, so [conjunction]*
- *atya, oti [adverbs]*
- *ggwe, nze, ye, yo. bo, zo, bwo [emphatic pronouns]*
- *bba [nonn]*
- *si [negation]*

Type II - CHWC_II, shown in Example 4 - are formed as a result of using clitic ng' and nng' to represent "ŋ" character and double "ŋŋ" character respectively in words which are not initially CHWC_1.

Example 4

- ***ng'ang'****ala (to whimper like a dog)*
- *bbiri****ng'****anya (egg plant)*
- *agaku****ng'****anyizzaamu (to collect in it something)*
- *e****nng'****anda (relatives)*
- ***nng'****oma (dram)*

Type III - CHWC_III, shown in Example 5 - are formed as a result of using clitic ng' and nng' to represent "ŋ" character and double "ŋŋ" characters respectively in words which are initially CHWC_1.

Example 5

- ***n'a****gaku****ng'****anyizzaamu (and to collect in it something)*
- ***ng'e****ng'****anda (like relatives)*
- ***ng'e****nng'****oma (like a dram)*

In view of this, the first task is disambiguating the three types of CHWCs. The approach adopted to disambiguate these words in LugDetect is discussed in subsection on "Disambiguating the CHWCs"

The second task is how to detect errors in each type of CHWCs. The approach adopted to provide this functionality in LugDetect is discussed in subsection on "Spell Checking Luganda Clitic Host Word Combination Type I (CHWC_I)". CHWC_IIs, which can be easy converted into real Luganda words (RLW) by substitution, are dealt with in subsection on "Spell Checking RLW".

It is worthwhile noting that the CHWC_1 akin to Luganda are also found in other Bantu languages like Runyankore-Rukiga, Kinyarwanda, among others. Example 6 and Example 7 shows samples of CHWC_I extracted from Runyankore-Rukiga and Kinyarwanda text respectively.

Example 6

- *ky'okutunga*
- *nk'eihanga*
- *n'ekya*
- *g'ebyemikono*

Example 7

- *n'uw'Umukuru (and His Excellence)*
- *Nk'uko (like that)*
- *y'Umukuru (His Excellence)*
- *w'Igihugu (of a country)*

The third challenge is deciding when to invoke LugDetect or the Error Detection Mechanism (EDM) during interactive processing and during batch processing. For interactive processing, should LugDetect be invoked when the space bar (SB) is struck or at every key-stroke? The former approach has a shortfall in that a modified word can miss out on spell checking through the use of insertion bar (IB) movement keys - a list of which is shown in Table 1 - which move the position of IB from the modified word to some other word without striking the SB. The latter approach would eliminate this problem but its main problem is that it is strenuous on the system due to the numerous EDM invocations. A solution to this is the development of an Algorithm for Interactive Invocation of EDM (IIEDM) discussed in subsection on "Interactive invocation of EDM"

For batch processing, how should EDM be invoked? A solution to this is discussed in form of an algorithm. A summary of the modules making up LugDetect is presented in subsection on "The Error Detection Mechanism"

LugDetect can detect numerical errors and word level punctuation errors, however, in this paper, we only mention them. Details of the modules dealing with these errors can be obtained from "unpublished" [2].

The scope of LugDetect is limited to Luganda non-word detection. Named entities, abbreviations, e-mail addresses, uniform resource locator (URL) strings are not handled in LugDetect, and neither are the real word errors.

Literature Survey

Errors

According to Peterson [15], spelling errors are basically of two types. Firstly, cognitive errors which are errors due to lack of knowledge of the language and are errors often ignored on the assertion that they are infrequent. In this context, their frequency is evaluated to be between 10% and 15% [17]. Secondly, typographical errors are 80% of "typos" (also called single character errors) and are one of the following four types: one exceeding character, one missing character, a mistaken character, or the transposition of two consecutive characters [4]. This means that 80% of errors are within edit distance of one and almost all errors - within edit distance of two.

Error Detection

There are two techniques used in detecting non-word errors. First, the Dictionary Lookup Technique (DLT), in which each word in the input text is lookup in the dictionary. In case a word is not found in the lexicon that word is a non-word or incorrect, and therefore, flagged or entered into the list of erroneous words. The larger the lexicon the better is the results. Shinghal and Toussaint [18] noted that DLT has low error rates, but is disadvantaged by large storage demands and high computational complexity. Most current spelling correction techniques rely on DLT for non-word error detection.

Second, a technique that uses a non-positional bi-gram 26 by 26 matrix which captures information on the existence of the bi-grams. A bi-gram is assigned a value of one if it exists in the corpus of the language in consideration, otherwise it is assigned zero. All the bi-grams making up the input text are checked for their existence in the matrix. if any of the bi-grams is non-existence than the word is flagged or a non-word. This technique is appropriate for Optical Character Recognition (OCR) errors and it has proven to be less accurate for detecting human-generated errors.

Lexicon

User-lexicon can be interactively enriched with new entries enabling the checker to recognize all the possible inflexions derived from them. A lexicon for a spelling correction or text recognition application must be carefully tuned to its intended domain of discourse. Too small a lexicon can burden the user with too many false rejections of valid terms; too large a lexicon can result in an unacceptably high number of false acceptances. The relationship between misspellings and word frequencies is not straightforward.

Peterson [16]. recommend that lexicon for spelling correction be kept relatively small based on the fact that approximately half a percent of all single error transformations of each of the words on a 350,000-item word list result in other valid words on the list However, Damerau and Mays [5] challenge this recommendation by using a corpus of over 22 million words of text from various genres and they found that by increasing the size of their frequency rank-ordered word list from 50,000 to 60,000 words; they were able to eliminate 1,348 false rejections while incurring only 23 additional false acceptances. Since this 50-to-l differential error rate represents a significant improvement in correction accuracy; therefore, they recommend the use of larger lexicons.

Dictionaries alone are often insufficient sources for lexicon construction. Walker and Amsle [19] observed that nearly two-thirds (61%) of the words in the Merriam-Webster Seventh Collegiate Dictionary did not appear in an eight million word corpus of New York Times news wire text, and conversely, almost two-thirds (64.9%) of the words in the text were not in the dictionary.

On the topic of construction of the lexicon for the spell program, McIlroy [13] provides some helpful insights into appropriate sources that may be drawn upon for general lexicon construction.
An article by Damerau [6] provides some insights into and guidelines for the automatic construction and customization of domain oriented vocabularies for specialized NLP (Natural Language Processing) applications.

LugDetect

LugDetect, developed using Python programming language, identifies the type of Luganda word or token which need be spell checked, and consequently, invokes the appropriate module for detecting the error. LugDetect works in tandem with LugCorrect to provide detection and correction for Luganda words. In other words, LugDetect is required to detect the error before LugCorrect can be invoked for correction purposes. The mechanism for LugCorrect, which provides correction candidate list (CCL) or an explanation or hint on the type of error committed, is shown in Algorithm 1 and more details are provided in [3].

Algorithm 1 : Error Correction Module

```
Get Erroneous Word
IF{punctuation Error}
Give feedback on punctuation error
ELSE IF{Numeric Error }
Give feedback on numeric error
ELSE IF (Word Error)
Generate correction candidate list (GCCL)
ENDIF

GCCL
SubDic=Get all words beginning with 1st character of erroneous word from
       lexicon
L=Number of words in SubDic

IF {L>10}
Get 1st 10 words from SubDic load in Dictionary with their corresponding
       Jaccard coefficient(JC)
ELSE
Get all words in SubDic load in Dictionary with their corresponding JC
ENDIF

REPEAT
swap if JC of word in SubDic is greater than minimum JC of word in
       dictionary
UNTILL all words in SubDic have been checked

Sort words in Dictionary in descending order of JC

Present the word to the user for selection

IF{user selects  Dictionary word and gives consent}
Change word in text
ELSE
Ignore
ENDIF
```

The two models form LugSpell, a model for an interactive spell checker providing spelling feedback while the end user is word processing their document in the editor. However, in this article, more emphasis is directed towards LugDetect.

The Luganda corpus which was experimented upon in the development of LugDetect was developed from Luganda text obtained from an online local news paper, Bukedde.

Disambiguating the CHWCs

Analysis of the task of disambiguating the three types of CHWC zeros down on disambiguating between the use of inter-word apostrophe (IWA) - which is defined in subsection on "The Error Detection Mechanism" - in Luganda compounding resulting into creation of CHWC_1 and the use of IWA in representing the character "ŋ" whether in single or double character which results into formation of CHWC_II and CHWC_III. At the bottom line of the analysis, we finally arrive at disambiguating between clitic ng' used in the representation of character 'ŋ' in CHWC_II - shown in Example 4 - and in

presentation of MMW which is formed from the "nga" MW during Luganda compounding, which process results into formation of "nga" CHWC_1 shown in Example 8. In other words, the task is how do we disambiguate between the use of clictics ng' in Example 4 and in Example 8.

Example 8

- ***ng'****amagi (like eggs)*
- ***ng'****oggyeko (after removing)*
- ***ng'****onoobufunamu (If you will benefit)*
- ***ng'****era (and also)*

In a bid to address this ambiguity, two groups of words were compiled from the Luganda corpus as well as Luganda dictionary - namely, Group I which contained the "nga" CHWC_1 and Group II which contained CHWC_II, that is, words with the character ng' (ŋ) - and studied them for common feature extraction. The result of the study showed that:

- Words in Group I always begun with the proclitic ng' and there is no situation in which a word in Group II began with the same proclitic - that is to say, with a single 'ŋ' character - except with the verb "ng'ang'ala" (ŋaŋala) meaning "to whimper like a dog".
- Words in Group II if they begin with character 'ŋ' then the character 'ŋ' is always in double 'ŋ' (proclitic nng') except in one case, "ng'ang'ala" [ŋaŋala]..
- Clitic nng' is purely a Group II characteristic and so is clitic ng' after the second position.

The approach adopted was, firstly, to extract "ŋaŋala" word and treated it as a special case; and secondly, to take the three observations into consideration and implement them in LugDetect to disambiguate these two groups by substitution. It is worthwhile noting that during the process of substitution, the CHWC_IIIs are automatically converted to CHWC_Is, while the CHWC_IIs are converted to *Real Luanda Words (RLWs).* The algorithm developed is shown in Algorithm 2.

Algorithm 2 : CHWC Disambiguation Module

```
Get CHWC Token

Substitute 'nng' with 'ŋŋ' in word if exists

IF{Token = CHWC}
    Check whether the word is "ng'ang'ala" or "ng'aŋala"
    IF{Token != "ng'ang'ala" or  Token != "ng'aŋala" }
        Substitute 'ng' with 'ŋ' in word at 2nd position upward if exists
    ELSE
         Substitute "ng'ang'ala" or "ng'aŋala" with "ŋaŋala"
    ENDIF
ENDIF

IF{Token = CHWC}
     Token is CHWC_I
     Invoke CHWC_I Module
ELSE
     Token is CHWC_II and in this case, Token is RLW
     Invoke RLW Module
ENDIF
```

Spell Checking Luganda Clitic Host Word Combination Type I (CHWC_I)

One inevitable challenge while spell checking Luganda text is dealing with the infinite number of CHWC_I created as a result of compounding. Direct application of the *Dictionary Lookup Technique* (DLT) to detect non-word error in CHWC_Is is practically impossible because their number is infinitively large for developing the required lexicon or dictionary. Consequently, in search for a solution, CHWC_I were extracted from Luganda corpus for analysis. The results of the analysis showed that:

- The number of CHWC_I forms is enormous
- The maximum number of components in CHWC_1 is three. In this context, there are two types of CHWC_1, namely two component (CHWC_1_2) and three component (CHWC_1_3)

- In CHWC_1_2, the *first component* (C_{11}) is MMW or MDWIV or locative 'e', a form of MW; and the *second component* (C_{21}) is a *real Luganda word* with an *initial vowel* (RLWIV)
- In CHWC_1_3, the *first component* (C_{12}) is either MMW or MDWIV or 'e'; the *second component* (C_{22}) is either MDWIV or 'e'; and the *third component* (C_{32}) is RLWIV

In view of these observations, a decomposer was developed to break the CHWC_1 into its constituent components. For detection purpose, DLT was applied on each component but using three different lexicons. The *first lexicon* (L_1) containing MMWs, MDWIVs and 'e'; the *second lexicon* (L_2) containing MDWIVs and 'e'; and the *third lexicon* (L_3) containing RLWIVs. The good news is that the number of MWs and DWIVs which are used in Luganda compounding is finite, therefore, simplifying the work of developing the corresponding lexicons. The three lexicons were kept separately because each has its own unique characteristics.

For CHWC_1_2, C_{11} is checked with L_1, and C_{21} - with L_3. For CHWC_1_3, C_{12} is checked with L_1, C_{22} - with L_2, and C_{32} - with L_3. Figure 1 show the flowchart for the detection and correction process for Luganda CHWC_Is. Note that the methodology begins with the checking of word level punctuation details and algorithm of which are discussed in "unpublished" [2]. The correction process - whose algorithm is shown in Algorithm 1 - involves generation of the *Candidate Correction List* (CCL).

Spell Checking RLW

The approach adopted to spell check *Real Luganda words* (RLWs) or Non CHWCs is shown in Figure 3 This same methodology is used with CHWC_IIs after they have been converted to RLW by substitution as earlier elucidated. Details of the word level punctuation checking referred to can be obtained in "unpublished" [2].

RLW lexicon was developed from the vocabulary captured from Luganda corpus whose lexical density (LD) is 1.5. LD defines the lexical richness of the text. This lexicon was cleaned of compound words, named entities, numbers in digital form, and foreign words.

The RLW lexicon is expected to be in hundreds of thousands of tokens and this may affect the performance of the system in terms of speed. In a bid to mitigate this shortfall, a dictionary clustering approach is proposed where the dictionary is split and clustered along the permissible alphabet of Luganda, which include the 23 characters of the 26 Luganda alphabet shown in Figure 2, excluding characters 'i', 'r', and 'u' which do not begin Luganda words. A similar approach has been used in the Punjabi checker [8], where the clustering is performed along POS in order to improve speed.

The Luganda alphabet, shown in Figure 2 , consist of 26 characters just like the English alphabet, but excludes the two consonants - 'q' and 'x' - and includes two consonants - 'ny' and 'ŋ' (sometimes represented by ng'). Letters q and x are used for spelling out loanwords from other languages, for instance, English. The sorting order is as show in Figure 2.

Algorithm for Invocation of Error Detection Mechanism

There are two types of invocation, namely, interactive and batch invocation.

Interactive invocation of EDM

The algorithm for Interactive invocation of EDM or LugDetect, shown in Algorithm 3, intends to mitigate the strenuous problem associated with invoking EDM at every key stroke, as discussed in the introduction, and we adopted an hybrid approach by taking the following issues into consideration:

- Invocation of EDM at every key stoke is limited to marked words (words which have been marked for errors)
- For unmarked new or changed words, EDM is invoked when the insertion bar (IB) leaves the word. This is done through monitoring the IB moving key strokes and the left button mouse click. Table 1 shows a list of IB moving keys
- For white space key strokes, EDM is invoked when:
  - the action results into formation of two words, that is, in situation when the IB is initially within the word
  - the IB is initially at the end of an altered word or new word.

On most systems whitespace includes the characters space, tab, linefeed, return, form-feed, and vertical tab. EDM is discussed in subsection on "The Error Detection Mechanism".

Batch Invocation of EDM

Batch invocation of EDM is useful when the end user has already some Luganda text which the user would like to spell check. The algorithm for this purpose is shown in Algorithm 4.

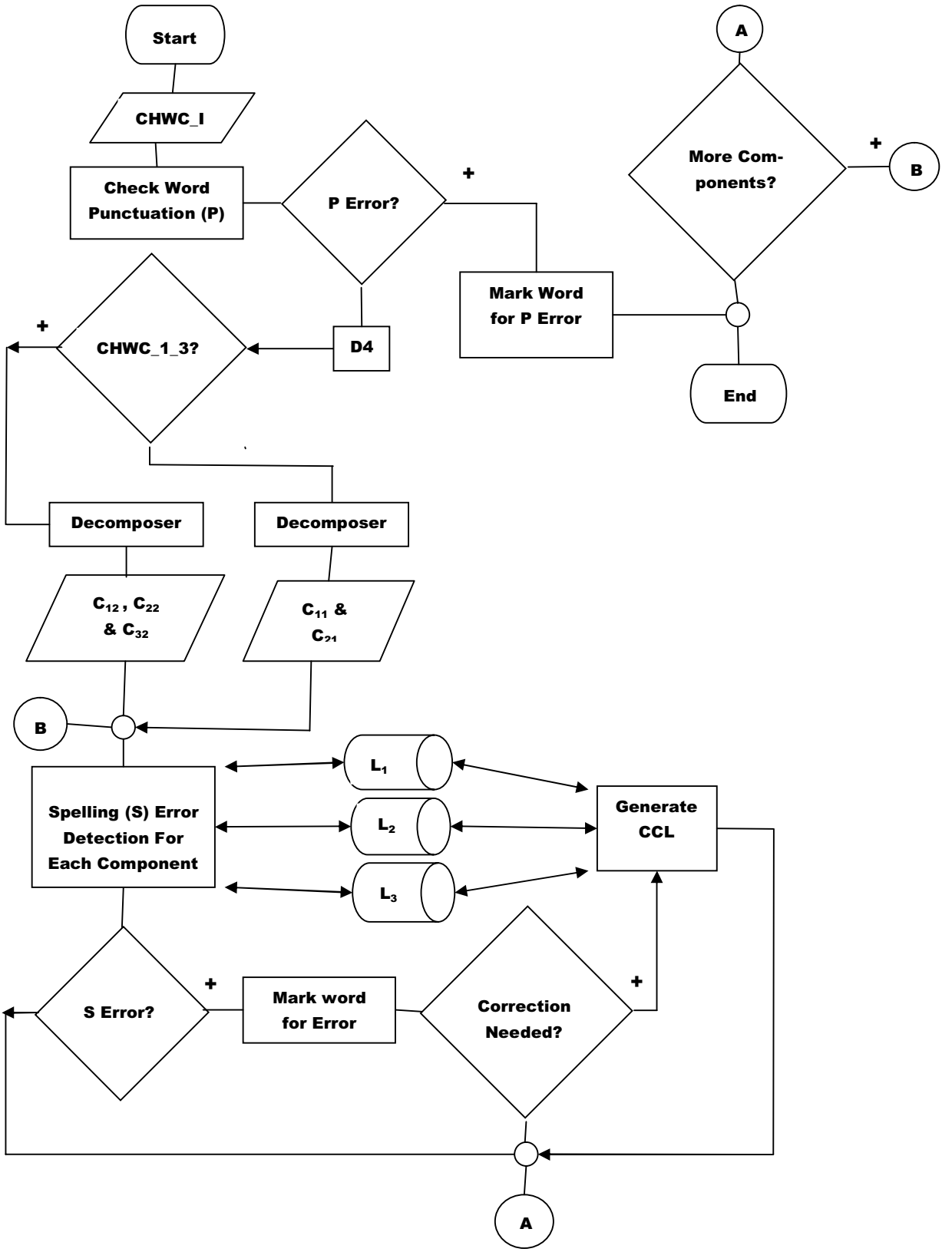


Figure 1 Flow Chart for Error Detection and Correction Algorithm for CHWC_1

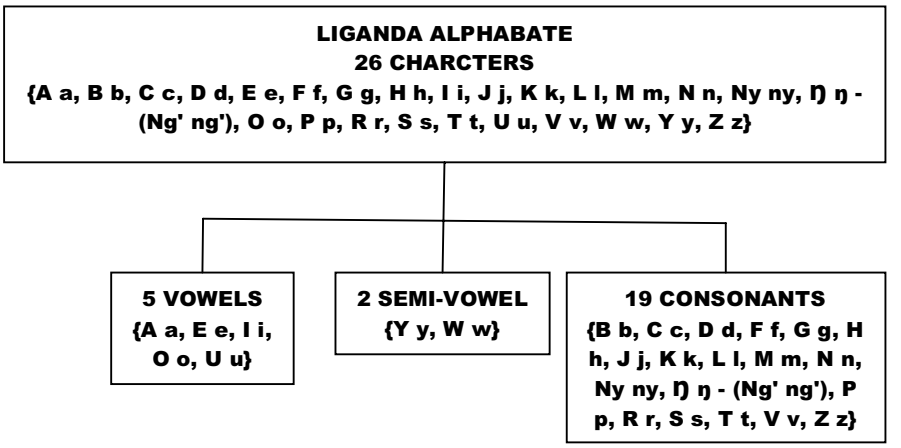


Figure 2 Categorisation of Luganda Alphabet composed of 26 characters sorted in the prescribed order

Table 1 Insertion bar (IB) movement keys with their respective description

Key_sym	Key_code	Description
Down	104	↓
End	103	end
Home	97	home
Left	100	←
Linefeed	54	Linefeed (control-J)
KP_Down	88	↓on the keypad
KP_End	87	end on the keypad
KP_Home	79	home on the keypad
KP_Left	83	← on the keypad
KP_Next	89	PageDown on the keypad
KP_Prior	81	PageUp on the keypad
KP_Right	85	→ on the keypad
KP_Up	80	↑ on the keypad
Next	105	PageDown
Prior	99	PageUp
Right	102	→
Up	98	↑

The Error Detection Mechanism (LugDetect)

The Error Detection Mechanism (EDM), shown in Figure 4, is invoked when a token is passed to the Type Detection Module (TDM) - whose algorithm is shown in Algorithm 5 and labeled as D1 in Figure 4 - which determines the type of token passes on to it as per the type rules elucidated below, and consequently, invokes the appropriate module.

The type rules used for identifying the different types of token are as follows:

i. A token is a word if it consists of only letter or a mixture of letters and punctuation marks
ii. A token is a number if it consists of any of the decimal alphabet. This is inclusive of a mixture of decimal digits and punctuation marks.
iii. A token is free punctuation if it consists of only punctuation marks.

The Number Module and Free Punctuation Module are just mentioned, however details of which can be obtained from "unpublished" [2].

The *RLW Detection Module (D2)*, whose algorithm is shown in Algorithm 6 differentiates between CHWCs and RLWs by, taking into consideration the following features which were captured through a compositional study of these two categories of words:

i. CHWCs are characterised by *inter-word apostrophes (IWA)*
ii. RLW are characterised by absence of IWA

Where, the IWAs are apostrophes found within a word, rather than at the end or beginning of a word.

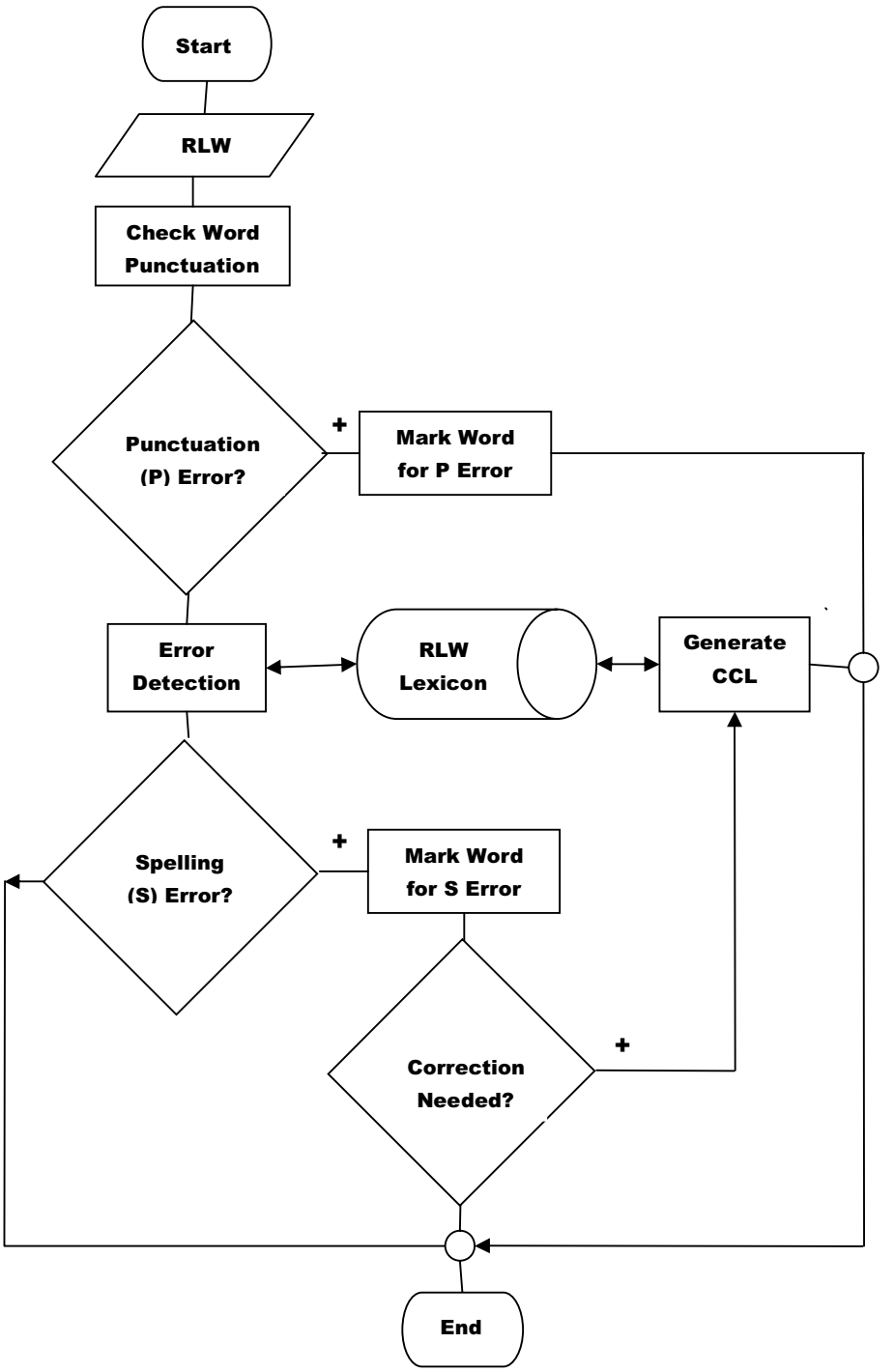


Figure 3 Flow Chart for Detection and Correction Algorithm for RLW

The *CHWC Disambiguation Module (D3)* and *RLW Module,* the two modules invoked by D2, are elucidated in subsection on "Disambiguating the CHWCs" and subsection on "Spell Checking RLW" respectively.

D3 determines the type of *CHWC* which has been passed to it, and then passes it on to the appropriate module as stipulated by the following rules:

i. If token is *CHWC_I* then pass it on to *CHWC_I Module*, discussed in subsection on "Spell Checking Luganda Clitic Host Word Combination Type I (CHWC_I)"
ii. If token is *CHWC_II* then convert it to *RLW* and pass it on to *RLW Module* aforementioned.

Algorithm 3 : Interactive Invocation of EDM

```
Char=keystroke \\ Capture Keystroke
Get Word
IF{Word = marked and Char = printable and Char != white space}
        \\ Marked for misspelling
        \\ The Word in question is the word which contains the insertion bar
    Invoke EDM at every keystroke \\ Error Detection Module (EDM)
ENDIF

IF{Word != marked and Char = printable and char != white space}
   IF{ insertion bar leaves Word and Word has changed}
        Invoke EDM
   ENDIF
ENDIF

IF{char = white space} \\ This results into splitting the word into two words,
                       \\ that is, left word and right word
   For left word
        Invoke EDM
   For right word
        IF{insertion bar leaves word}
            Invoke EDM
        ELSE
            IF (Word=marked}
                Word != marked
            ENDIF
        ENDIF
ENDIF
```

Algorithm 4 : Batch Invocation of EDM

```
Get Token

REPEAT
    Invoke EDM
UNTIL no more tokens to check
```

Component Detection Module (D4), which determines the number of components in the CHCW_I, and Decomposer are sub modules in the *CHWC Module*.

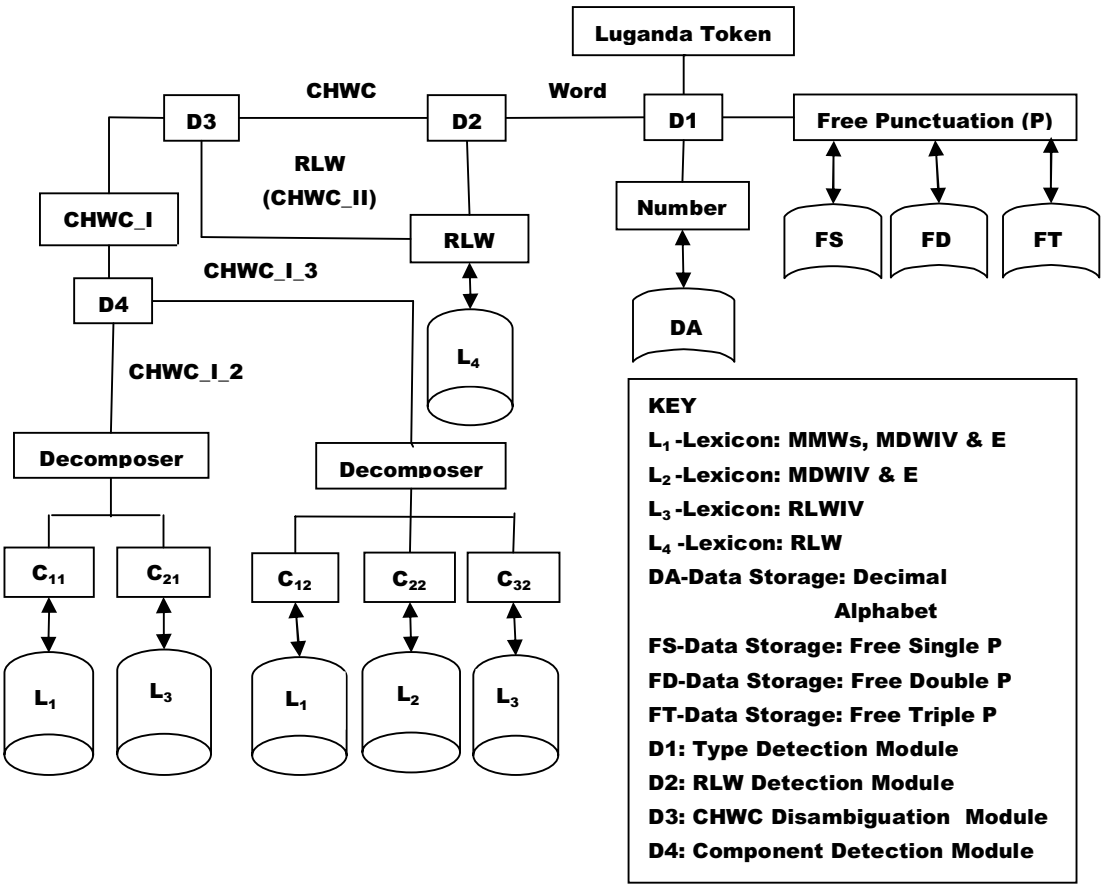


Figure 4 Summary of modules making up the Error Detection Mechanism for LugDetect

Algorithm 5 : Type Detection Module

```
Get Token

Determine the type of token as per the Type Rules

IF{Token=Word}
    Pass Token to RLW Detection Module (D2)
ELSE IF{ Token=Number}
      Pass Token to Number Module
ELSE
      Pass Token to Free Punctuation Module
ENDIF
```

Algorithm 6 : RLW Detection Module

```
Get Word

Check Word whether it is CHWC or RLW by checking
existence of Inter Word Apostrophe (IWA)

IF{ Word =CHWC }
   Pass Word to CHWC Disambiguation Module (D3)
ELSE
     Pass Token to RLW Module
ENDIF
```

Evaluation

Experimental Setup

Determining the Distribution of the Various Types of Luganda Words in the Corpus

The Luganda corpus which was used in the experimentation has 20,000 tokens collected from 50 articles from different subsection of an online local news paper, Bukedde. These articles were cleaned of any misspelling and their vocabulary was extracted and statistically analysed. The lexical density (LD), which defines the lexical richness of the text, was determined using Equation1. The results of the analysis are presented in Table 2, and diagrammatically, these results are shown in Figure 5, Figure 6, and Figure 7.

$$L_D = T_T / T_V \quad (1)$$

Where:

- T_T represents the total number token in the text
- T_V denotes total number of tokens making up the vocabulary of the text

Table 2 Distribution of various Luganda words in the Corpus

. Total Words	20,000
Total Vocubulary	5994 of 20,000 = 29.97%
Lexical Density (L_D)	3.33
RLW	4835 of 5994 = 80.66%
CHWC	1159 of 5994 = 19.34%
CHWC_I	1135 of 1159 = 97.93%
CHWC_I_2	1130 of 1135 = 99.56%
CHWC_I_3	5 of 1135 = 0.44%
CHWC_II	17 of 1159 = 1.47%
CHWC_III	7 of 1159 = 0.60%

Determining the Accuracy of LugDetect

In this experimentation, real words for each category were arbitrary collected from the Luganda corpus and other Luganda sources such as dictionaries, literature books. The words were mutated using a *mutator function* which works by randomizing the process of:

- Picking the type of operation which may be: deletion, insertion, substitution, or transposition (or inter-changing)
- Determining the character to insert or substitute
- Determining where to insert relative to the specified position; that is, either to the left or right of the position
- Determining the positions of the characters used in the operation
- Determining how many times to repeated the above four process

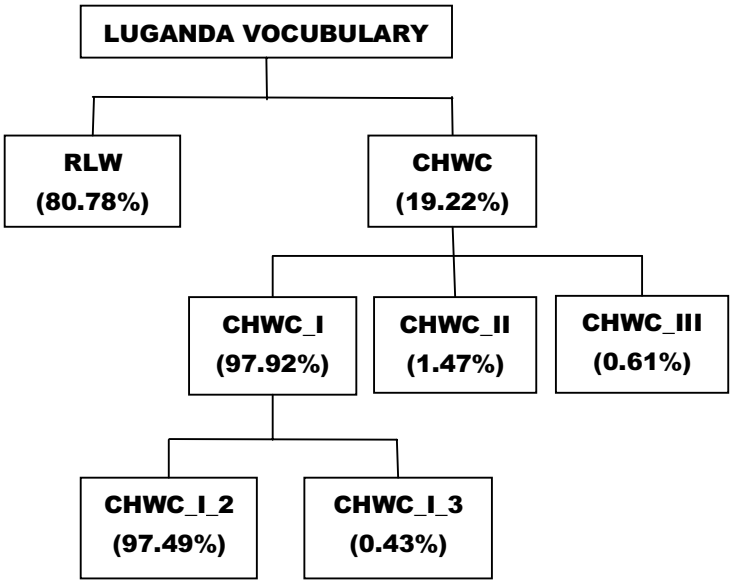


Figure 5. Categorisation and Distribution of vocubulary in a Luganda corpus

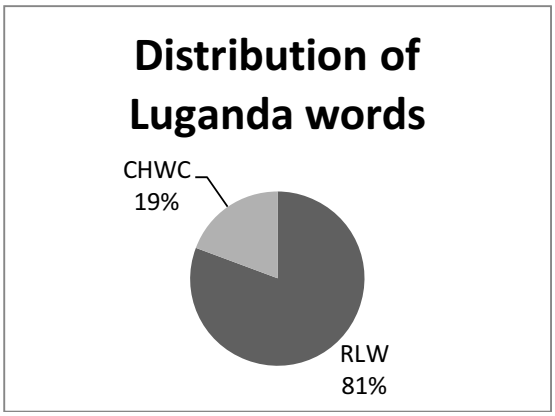


Figure 6 A pie-chart showing the distribution of Luganda words

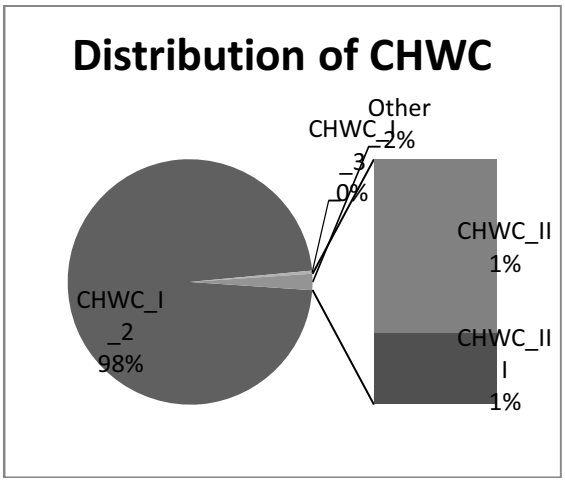


Figure 7 A pie-chart showing the distribution of CHWC

Table 3.Number of tokens per category for experimentation

. Total Words	5834
Total Errors	385 of 5834 ≈ 6.6%
RLW	250 of 385 = 64.9%
CHWC	135 of 385 = 35.1%
CHWC_I	90 of 135 = 66.7%
CHWC_I_2	75 of 90 = 83.3%
CHWC_I_3	15 of 90 = 16.7%
CHWC_II	25 of 135 = 18.5%
CHWC_III	20 of 135 = 14.8%

The output of the *mutator function* consists of:

- Non- mutated tokens (NMT), that is, tokens which have not been changed at all
- Group Tokens - tokens that fall within the group of interest and these fall under two categories: Real Word Error Tokens (RWET) - tokens that are changed into other words in the lexicon; and Non-Word Tokens (NWT) - tokens which are misspelt
- Non group tokens (NGT) - tokens that fall out of scope of the group of interest, and therefore, fall in other groups.

In the experimentation, NGTs were discarded as they fall out of scope of the group of interest. Table 3 shows the distribution of tokens used in the experimentation, which were passed to LugDetect to find out whether they can be detected for errors or not.

Accuracy (A_P) in percentage, used as a metric of evaluation in this experimentation, is evaluated using Equation 2. A_P is a measure of how accurate LugDetect detects erroneous word, in this case, non-words.

$$A_P = 100 * W_{CD} / (T_T - T_{RW} - T_{NM}) \quad (2)$$

Where:

- W_{CD} represents the total number of erroneous tokens correctly detected
- T_T denotes total number of tokens processed
- T_{RW} represents the total number of *Real Word Error Tokens* (RWET)
- T_{NM} denotes the total number of *Non- Mutated Tokens* (NMT)

The overall results of the experimentation are shown in Table 4

Determining the Speed of Detection for LugDetect

In this experimentation, two metrics are used to determine the speed of detection. The first is F_D, which is evaluated using Equation 3, and it defines the number of words that can be checked in one second. F_D, which depends on the processing power of the computer, is directly proportional to the speed of detection; that is, the bigger the value of F_D, the bigger is the speed of detection.

Table 4 Summary of results for the determination of A_P

	T_W	T_T	T_{NM}	T_{RW}	W_{CD}	A_P with RWET	A_P without RWET
RLW	250	250	5	3	242	98.78%	100%
CHWC_I_2	75	150	36	14	100	87.72%	100%
CHWC_I_3	15	45	14	2	29	93.55%	100%
CHWC_II	25	25	0	0	25	100.00%	100%
CHWC_III	20	41	15	4	22	84.62%	100%

$$F_D = T_T / T \quad (3)$$

Where:

- T_T denotes the total number of tokens
- T represents the time taken to process the tokens

The second is T_C, which is evaluated using Equation 4. T_C estimates the speed of detection by determining the number of comparisons made between the target token and the lexicon words. We presumed that the process of access is linear and considered the worst case scenario where all the words in the lexicon are compared. It is evident that the bigger the lexicon, the bigger is the number of comparisons (T_C). Also, T_C is inversely proportional to the speed of detection, that is, the bigger the value of T_C, the slower is the speed of detection.

$$T_C = S_1T_1 + S_2T_2 + \ldots + S_nT_n \quad (4)$$

Where:

- S_n denotes the size of lexicon L_n
- T_n represents the number of tokens checked with lexicon L_n

The computer used in the experimentation uses an ADM FX(tm) - 8150 eight core 3.6 GHz Processor running 64 bit Windows 7 Ultimate Operating System.

Files of various sizes and type were passed onto LugDetect and the time of processing for each type was captured in milliseconds (ms). F_D was determined for each category and ranked by speed of detection. The overall results of the experimentation are shown in Table 5. Information in Table 5 was used to determine T_C and ranked by speed of detection. The results are shown in Table 6.

Table 5 Summary of results for the determination of F_D (number of words per second)

	T_W	T_T	Time (ms)	F_D (Hz)	Rank
RLW	501	501	1271.061	394	5
CHWC_I_2	150	300	240.237	1248	3
CHWC_I_3	30	90	31.268	2878	1
CHWC_II	66	66	68.435	964	4
CHWC_III	40	82	43.756	1874	2

Table 6 Summary of results for determination of T_C (number of comparison)

lexicon	L_1	L_2	L_3	L_4	T_C	Rank
Lexicon Size - S_n	1.9	1	110	266		
RLW	0	0	0	501	133266	5
CHWC_I_2	150	0	150	0	16785	3
CHWC_I_3	30	30	30	0	3387	1
CHWC_II	0	0	0	66	17556	4
CHWC_III	49	1	39	0	4384	2

Discussion of Results

In reference to Table 2, Figure 5, Figure 6 and Figure 7, it is evident that most (81%) of the words in Luganda text are of type RLW, while the CHWC form only 19%. The most common words among the CHWC are of type CHWC_1_2 and the least common of type CHWC_1_3. Luganda text and more specifically the Bukedde News paper text have a lexical density (LD) of 3.3. LD is a measure of lexical richness of a given text. This distribution has been used to determine the number of tokens for each category of Luganda words used in the experimentation as shown in Table 3. Basil [22] used an error rate of 1%, however, in this research work, because of the smaller size of the corpus, an error rate of 6.6% is adapted.

In reference to Table 4, LugDetect shows a performance (accuracy - AP) of 100% on all the five categories of Luganda words - namely RLW and the four types of CHWC - so long as the target (the initial word before mutation) is in the lexicon and the erroneous word is not a real word. However, in the real world where real word errors are inevitable the performance of LugDetect falls as shown in column labeled "A_P with RWET" in the table.

In reference to Table 5 and Table 6, the slowest process is the RLW detection and the fastest is the CHW_I_3 detection. RLW detection has the slowest speed because it involves the use of lexicon L_4 which is the largest lexicon (that is, 266-fold bigger than L_2, the smallest lexicon). On the other hand, CHW_I_3 detection uses essentially small sized lexicons, and consequently, it has the highest speed. The results in Table 6 were used in this experimentation to justify the results in Table 5 and the results obtained in the two tables are corroborated by rank. The average F_D of 1471 Hz for LugDetect is good enough for interactive detection, given that the minimal speed of 10 Hz is required for interactive correction, [14].

Related Work

Kitaka's Luganda Detector (KLD) [11] is non-word error detector using a lexicon of 100,000 tokens. Kitaka made an attempt to capture all the Luganda CHWC but this is an uphill task because the number of Luganda CHWC, though they form 19% of Luganda text, is infinitively large exceeding half a million mark; therefore KLD cannot exceed an accuracy of 85%. KLD - developed using Python, a programming language - uses one lexicon and uses DLT for detection. However, our research work comes in handy to provide a solution to deal with this avalanche of CHWCs by, first, disambiguating them (refer to subsection on "Disambiguating the CHWCs"), then second, detecting errors in them (refer to subsection on "Spell Checking

Luganda Clitic Host Word Combination Type I (CHWC_I)"). The proposed corresponding algorithms can deal with any Luganda CHWC.

Youssef Bassil [22], proposes a parallel shared-memory spell-checking algorithm that uses rich real-world word statistics from Yahoo! N-Grams Dataset (YND) [21] to detect and correct non-word and real-word errors in electronic text. Youssef's system is divided into three sub-algorithms that run in a parallel fashion: The error detection algorithm (EDA) also referred to as *Parallel Error Detection Algorithm (PEDA)* that detects misspellings, the candidate generation algorithm (EGA) that generates correction suggestions, and the error correction algorithm (ECA) that performs contextual error correction. Here, focus is directed towards EDA. Experimentation conducted on a set of text articles containing misspellings, showed that a detection and correction accuracy of 99% on non-word was obtained for PEDA.

Aspell .31developed by Atkinson [1], is a Free and Open Source spell checker designed to eventually replace Ispell. Test results show that Aspell achieved a detection and correction accuracy of 93.1% at a speed of 137 Hz (words per second).
Integrated Scoring for Spelling error correction, Abbreviation expansion and Case restoration (ISSAC) is a mechanism developed by Wilson Wong et al [20] for cleaning dirty texts from online sources and the idea of which was first conceived as part of the text preprocessing phase in an ontology engineering project. Evaluations of ISSAC using 400 chat records reveal an improved accuracy of 96.5% over the existing 74.4% based on the use of Aspell only.

Kaur and Garg [9] proposed a hybrid system which checks both spelling and grammatical errors for Punjabi language. The system uses DLT for detection and the accuracy reported is 83.5%.

Conclusion and Future work

Conclusion

This research work provides models or methodologies of dealing with non word error detection for Luganda, a Bantu language and as a result LugDetect was developed.

Experimental results show that 81% of Luganda words are of type RLW and the remaining 19% is of type CHWC. Amongst the CHWC the most common are the CHWC_1_2 representing 97%; and the least common are the CHWC_1_3 which represent 0.43%.

CHWC_1 (Clictic-host Word Combination type 1) - which are akin to Luganda text and form 98% of Luganda CHWC - are also found in other Bantu Languages - for instance, Runyankore-Rukiga and Kinyarwanda; therefore, the approach / algorithms adopted to deal with CHWC in LugDetect can also be adopted to deal with the same in these other languages.
In addition, LugDetect - although has been initially developed for Luganda - can easily be adopted to spell check and detect errors for other Bantu languages, as most of them have a lot in common.

As per experimental results, LugDetect detects non-word errors at an accuracy - A_P of 100% so long as a the erroneous word is not a real word at an average speed of 1471 Hz (number of words per second) which is over and above the minimal speed of 10 Hz required for interactive correction. The speed of detection varies along the different types of Luganda words due to the use of different lexicons with different sizes; and it is re-emphasised that the speed of detection is inversely proportional to the lexicon size.

Future work

The scope of LugDetect has been limited to non-word errors, however this scope can be extended to cover real word errors. A number of approaches are in existence most of them use statistical approaches and therefore require an annotated corpus; In this context, to apply these approach to Luganda there is need to develop a Luganda annotated corpus.

The approach proposed by Mays et al., [12] is a statistical method which uses word-trigram probabilities for detecting and correcting real-word errors.

Another approach, proposed by Kernighan, et al, [10] is a real word error detector based on a noisy channel model.

Word Sense Disambiguation (WSD) algorithms are other approaches applied to detect and correct real word errors in context. WSD algorithm uses the context of the sentence to resolve semantic ambiguity. Ginter et al [7] presented a technique which uses WSD in a biological domain for correcting real word errors.

Consequently, there is a need to investigate the aforementioned approaches in bid to identify the most appropriate for the highly inflected Luganda language.

Another investigation on language and error models - statistical selection model for selecting correction candidate - is required in order to ascertain whether they can improve the performance of LugDetect in reference to real word detection.

Luganda is an agglutinative language which means that it is highly inflected language (that is, it has many form of a given word). This corresponds to a large lexicon to capture these forms; and therefore, a call and need for a language-dependent algorithm for handling the language morphology, in a bid to reduced on the size of the required lexicon.

The use of a language and error models in the selection process can be investigated upon in order to access the impact of the two models on the selection process.

References

[1] Atkinson, K. (2006), "Gnu Aspell 0.60.4", http://aspell.sourceforge.net/.

[2] Balagadde Robert.Ssali, " Designing a Tag-set and Models for Spell Checking Intelligent System with a bias in Luganda Language", PhD thesis, Osmania University, 2016.

[3] Balagadde Robert Ssali, and Parvataneni Premchand, "Non-Word Error Correction for Luganda", Proceeding of the Seventh International Conference on Advances in Communication, Network, and Computing (CNC 2016), Bengaluru, India, 2016.

[4] Damerau F. J., "A technique for Computer Detection and Correction of Spelling Errors", Communication ACM, 7:171–176, 1964.

[5] Damerau F J.; and Mays. E., "An Examination of Undetected Typing Errors", Information Process Management, 25, 6, 659-664, 1989.

[6] Damerau. F. J.' "Evaluating Computer-Generated Domain - Oriented Vocabularies", Information Process Management, 26, 6, 791–801, 1990.

[7] Ginter, Filip; Jorma, Boberg; Jouni, Jarvinen; and Tapio, Salakoski, ``New Techniques for Disambiguation in Natural Language and Their Application to Biological Text", J. Mach. Learn. Res., 5:605–621, 2004.

[8] Harpreet Kaur, Gurpreet Kaur, Manpreet Kaur, "Punjabi Spell Checker Using Dictionary Clustering". International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 7, July 2015 2369 ISSN: 2278 – 7798, 2015.

[9] J. Kaur and K. Garg," Hybrid Approach for Spell Checker and Grammar Checker for Punjabi", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 6, June 2014

[10] Kernighan; Mark D.; Kenneth W. Church; and William A. Gale, "A Spelling Correction Program Based on a Noisy Channel Model", Proceedings of COLING 1990, 205-210, 1990.

[11] Kitaka Felix, "Luganda Detector", https://github.com/kitaka/Luganda-Detector, 2015

[12] Mays Eric; Fred J. Damerau; and Robert L. Mercer, "Context Based Spelling Correction", Information Processing and Management, 27(5):517–522, 1991

[13] MCIlroy, M. D; "Development of a Spelling List", IEEE Transaction Communication COM-30, 1 (Jan.), 91–99, 1992.

[14] Peter Norvig, "How to Write a Spelling Collector", an essay, 2007

[15] Peterson J.L., "Computer Programs for Detecting and Correcting Spelling Errors", Comm. of ACM vol.23, issue 12, 1980.

[16] Peterson, J L, "A Note on Undetected Typing Errors", Commun. ACM 29, 7 (July), 633-637, 1986.

[17] Pollock J.; and Zamora A., "Automatic Spelling Correction in Scientific and Scholarly Text", Comm. of ACM vol.27 358- 368, 1984.

[18] Shinghal, R.; and Toussaint, G. T. A. "Bottom-up and Top-down Approach to Using Context m Text Recognition", International Journal Man-Machine Studies 11, 201-212,1979.

[19] Walker, D. E.; and Amsler, R. A., "The use of Machine-Readable Dictionaries in Sub-language Analysis", In Analysing Language in Restricted Domains: Sub-language Description and Processing. Lawrence Erlbaum, Hillsdale, N. J., 69-83, 1986.

[20] Wilson Wong, Wei Liu and Mohammed Bennamoun, "Integrated Scoring for Spelling Error Correction, Abbreviation Expansion and Case Restoration in Dirty Text", Proc. Fifth Australasian Data Mining Conference (AusDM2006), 2006

[21] Yahoo! Webscope dataset Yahoo! N-Grams, ver. 2.0, http://research.yahoo.com/Academic_Relation

[22] Youssef Bassil, "Parallel Spell-Checking Algorithm Based on Yahoo! N-Grams Dataset", International Journal of Research and Reviews in Computer Science (IJRRCS), ISSN: 2079-2557, Vol. 3, No. 1, February 2012.

Modern Approaches Towards Superior Dynamic Allocation

Snehal C. Ratnaparkhi*, Mithra Venkatesan** and A.V. Kulkarni***
*PG Scholar DYPIET SPPU Pune
Snehalratnaparkhi4@gmail.com
**Research Scholar JSPM'S RSCOE SPPU Pune
mithrav@rediffmail.com
***Dean R&D DYPIET SPPU Pune
anju_k64@yahoo.co.in

Abstract: As technology evolved increasing demands of new wireless devices and application increases need for wireless radio spectrum. Traditional spectrum allocation strategy unable to fulfill users demands so result in bottleneck condition. Cognitive radio offers good alternative to fill gap between traditional static allocation and user demands. It provides better solution over bottleneck condition. Dynamically management of spectrum is promising technique of Cognitive radio to fulfill needs of users .Cognitive radio senses vacant space which are called spectrum holes and reuse vacant spectrum by allocating it to secondary users dynamically. Utilizing unused spectrum Cognitive radio provide high throughput and low latency communication. So with its advanced features Cognitive radio overcomes problem of scarcity. This paper is a survey of various methods that are used for dynamic spectrum allocation in Cognitive radio. It also contains service oriented approach for dynamic spectrum allocation. It uses Graph Coloring Problem (GCP) and Binary particle swarm optimization (BPSO) spectrum allocation techniques. It overcome problem arises due to mismatch between the spectrum resources and the service demands.

Keywords: Cognitive radio, Dynamic spectrum allocation, Scarcity, Graph Coloring Problem, Binary particle swarm optimization.

Introduction

The development in wireless communication technology results in scarce resources. Available spectrum is insufficient to fulfill users' need, which increases demand of electromagnetic spectrum. There is necessity to fulfill need of users arise due to developing technique. Current static allocation strategy does not utilized spectrum properly, so there is underutilization of spectrum. . FCC (Federal Communication Commission) assigns spectrum to licensed holders known as primary user. Fundamental problem arise due to future wireless systems is to find suitable carrier frequencies and bandwidths for operations because most of spectrum bands are vacant have been already assigned by FCC. Most of bands are used for specific part of time and in some areas, so most of spectrum remains vacant. This results in insufficient use of spectrum. To utilize this vacant space which is known as spectrum holes researcher have proposed dynamic spectrum allocation approach which allocates electromagnetic spectrum dynamically and utilized this unused band for communication purpose. Dynamic spectrum allocation is a characteristic of cognitive radio. Cognitive radio provides high bandwidth to users through heterogeneous wireless architecture and dynamically assign spectrum by sensing it.

The cognitive radio [1][2][3] is artificial intelligent radio. The cognitive radio has a capability to sense and gather information like transmission frequency, bandwidth, power, modulation from surrounding environment. Secondary users which are called cognitive users with this feature can identify best spectrum available. It works on three features that is sensing, learning and adaptation. Operating parameters also depend on radio environment, primary and secondary user requirements, availability of spectrum. It is one of the feature of cognitive radio that CR systems is aware of spectrum use in their surrounding environment, according to channel availability information system makes decisions on the transmission channels and further improve their performance by intelligent learning techniques. In order to achieve the optimal performance CR has the ability to rapidly adapt the operational parameters based on sensed information. Cognitive radio technology encompasses four stages: spectrum sensing, spectrum management, spectrum sharing and spectrum mobility. When primary users are not using its spectrum cognitive user senses that hole and dynamically allocates that spectrum. When primary user again demands for that spectrum cognitive user vacates that spectrum dynamically. In sensing stage secondary users detects unused spectrums. Best available spectrum is used with fulfill all requirement of communication of uses. Spectrum band is shared with primary users. Spectrum is shared without any interference to other users. In spectrum mobility after finding better spectrum seamless communication requirements during transition are maintained. CR provides dynamically assigning spectrum to different

services depending on their needs. Cognitive radio is not only increases the dynamic spectrum efficiency and Bit Error Rate but also maximizes data throughput and minimize power consumption.
Markov Chain and Learning model these are the existing models to predict the available spectrum. Allocation is done based on these predicted spectrums. Graph coloring problem (GCP), Auction theory and game theory increase allocation efficiency of system, which also eliminate problems related with fairness, inference, and network rewards. In the present, biologically inspired optimization algorithms such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Bee Colony Optimization (BCO) have been used to solve allocation problems because such algorithms can efficiently solve the nondeterministic polynomial time (NP) hard problem and it can achieve better allocation efficiency and stability. This paper introduces some of basic techniques of spectrum allocation. With the analysis of available result derived from the superior allocation technique using PSO is incorporated in this paper.

Literature survey

After sensing specific spectrum bands, CR network allocates spectrum if it is vacant to meet the application requirements. Among the most of allocation strategies researchers have focused primarily on the following problems (i) improving spectrum allocation efficiency (ii) enhancing fairness among cognitive users (iii) reducing the time cost and computational complexity (iv)based on user demands managing spectrum allocation.
Optimization method based on GCP has been developed because its easiness to implement and adaptation to practical environment. This method is very easy and computationally less complex. Wang et al. [4] solve spectrum utilization, fairness and throughout problem in CR networks using GCP. Maan et al. [5] explored common framework for studying distributed frequency allocation in mobile cellular network. This paper concerns algorithmic problem for solving Maximal independent set of a graph in a distributed fashion and it reduces independent set to graph coloring. Yu CAO [6] et al. proposed in fairly good system utilization condition, a novel system method called approach of round based distribution was introduce, which results in better fairness in link degree and round based algorithm. Spectrum allocation transformed into an NP problem based on GCP presented by [26] Zhao et al. After that they applied the three traditional allocation rules for improving allocation efficiency. Canberk et al. [28] proposed GCP based on QOS grading method. In paper [7] behavior between cognitive users is model in the form of games in Game theory. For Optimal resource allocation game is used. Nie et al. [10] formulated channel allocation problem using game theory, but due to secondary users interference primary users were not explicitly protected. Quifen Ni et al. [8] authors provides common framework model based on game theory for flexible and efficient spectrum allocation of cognitive radio. It explains cooperative and non-cooperative models of game theory for spectrum allocation and its analytical scheme.
Recently, many biologically inspired optimization algorithms have been adopted because it provides better solution over NP problem. Liu et al. [11] proposed a Genetic algorithm based spectrum allocation in cognitive wireless networks. It achieves many objectives such as maximizing system utilization and fairness. Stimulation of it compressed the search space. Balieiro et al. [13] presents detail description on adaptive sensing optimization scheme for CR networks depend on multi objective GA. It maximizes spectrum opportunities and keeping the sensing overhead always within a user distinct maximum value. Jia et al. [15] investigate the issue of congestion avoiding in wireless cognitive radio mesh networks. Route scheduling and optimal path selection to find maximum bandwidth path GA approach is used. Liu et al. [16] proposed method using GCP and PSO for spectrum allocation and efficiently satisfy user service demands and achieve better utilization and fairness. Tang et al. [18] proposed solution to achieve significant throughput and efficiently solve the nonconvex optimization problem. Koroupi et al.[19] applied ACO algorithm with GCP for spectrum allocation problem. This result compared with classic PSO and generalized special congestion games (GSCG). Ahmadi et al. [14] proposed a novel GA and modified ACO algorithm to solve subcarrier allocation problem. It is applied to distributed OFDMA based spectrum access networks. In, [20] performances of different algorithms are compared. Their comparison based on convergence characteristic and statistical metrics. Farad Khozeimeh [21] described a novel dynamic spectrum management (DSM) scheme for cognitive radio networks, termed the self-organizing dynamic spectrum management (SO-DSM). By using these bio-inspired algorithm computational time and complexity of previous graph theory and game theory is reduced.
In the paper, we introduce various methods which are used for dynamic spectrum allocation. Based on various results of these methods we tried to collect advantages and disadvantages of these methods. This paper also review the service oriented approach for spectrum allocation. This paper is review of modern approaches that are available for dynamic spectrum allocation.

Allocation Strategies of various methods

Graph coloring problem

GCP [4][5][6][26] is effective and simple method to implement. Graph coloring is nothing but special case of graph labeling. It is nothing but assignment of labels traditionally called colors to elements of graph. Optimization problem is treated as

graph coloring problem. No two adjacent vertices and edges share same color and no two faces that share boundary have same color. Allocation problem is equivalent to painting vertex with colors available in color list to maximize system proportional fairness. Channel allocation problem can be formulated as GCP.
Graph coloring is composed of vertex coloring, edge coloring and region coloring. vertex coloring model used for many scheduling problems, optimal assignment of channels to radio stations and assignment of spectrum frequencies to mobile operations. Edge coloring is same as vertex coloring. Goal of it to color edges of a given graph using fewest colors such as no two edges incident to common vertex have same color. Region coloring is used to color the map to which graph is associated. Problem is divided in two sub problems. This technique solves spectrum utilization, fairness and throughput problem in CR network. Fig. 1 shows GCP representation of network.

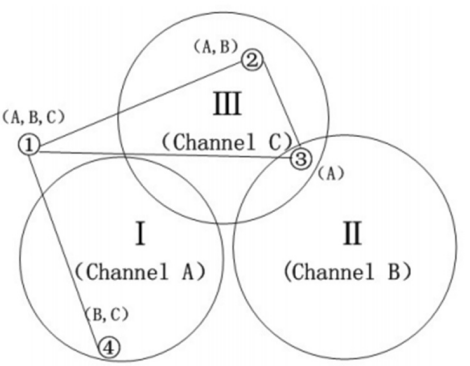


Fig 1: GCP representation of network

As shown in fig.1 three frequency bands or channels namely A,B and C are based on opportunity available to secondary users. Primary users represented by I, II, III. Due to sharing agreement if channel is used by primary users are not available for secondary users. Nodes 1-4 represents vertex and frequency bands assign to that nodes. Each frequency is assign by unique color. In this allocation nodes are assume within certain interference range of primary users. Secondary users cannot use same frequency use by primary users.

Game theory
This theory [7][8][10] was introduced in 20th century from der mathematical analysis of board games like chess, draughts etc. In this model wide variety of game situations can be stimulated, in which participants follow various goals and strategies and influence one others decision. In game theory every interaction between primary users and secondary users can be represented as game. On the method of cooperative, game theory is divided into cooperative game theory and non cooperative game theory. Matching game algorithm is very popular cooperative algorithm. In this, users and channel match bilateral market. Various non cooperative algorithms like Cournot, Bertrand, Stackelberg, Repeated, Super modular, Potential, Evolutionary, Auction game models are used. Adaptation and recursive interactive decision process is model with game theory in CRN network.
In spectrum allocation problem secondary and primary users include united game process. In primary users game process they are participants with spectrum authorization and their strategy is to choose the number of loan spectrum. In secondary users game process cognitive devices are participants of games and their task is to select demand spectrum based on decision making process. Each game theory has specific process, allocation problem is consider as that particular game model. Spectrum is allocated base on that particular process.

Genetic algorithm
Genetic algorithm [11][13][15] is bio-inspired process which achieves maximizing system utilization and fairness. As the number of subscribers increases computational complexity also increases. NP hard problem could not be liberalized and solved optimally. In the field of artificial intelligence GA is a process of natural selection. Heuristic provides useful solutions to optimization and search problems. GA generates solutions to optimization problems using techniques such as inheritance, mutation, selection and crossover. The main advantage of using GA over all other techniques is parallelism which increases speed of stimulation. To improve initial solution GA uses powerful tools crossover, mutation and selection.
Radio identity and modulation type are coded into binary chromosome and using fitness function optimality of solution is evaluated. In the selection of chromosomes based on chromosomes have the better level of fitness in current available population. Crossover is nothing but technique of taking more than one parent solutions and producing a child solution from them. Mutation changes a gene randomly for reaching better solutions and finally fitter solutions are selected for next

generation. Fitness function uses weighted sum approach. In addition of selection process focus is on best chromosome at each generation. Depending on the fitness value spectrum is allocated based on best fitness function value. This algorithm has its successful results in aircraft industry, chip design, computer animation, drug design, and computer creation.

Particle swarm Optimization

Particle swarm optimization [16][17][18] provides simple approach to find an optimal solution of channel assignment problem and for better results than other. Potential solution is considered as particles. Population candidate solutions are used to optimize problem. Particles are moved around search space and their movements are guided by their own best known position in search space according to simple mathematical formula. Based on its own position as well as best position of entire swarm each particle updates is position and velocity. While moving in free space each particle keeps track of its coordinates in problem space. These best values are called *pbest*. Global version of PSO tracked another best value those are overall best value and its location obtained by any particle in population called *gbest*. Each particles current fitness value is compared with previous *pbest* value and if it is better than previous then it is set as current value. Fix *gbest* of swarm as best of all *pbest*. Fitness value is calculated for each position value Position and velocity of particle updated and compare with previous value. Spectrum is allocated for that value which is having best fitness function. This algorithm efficiently address NP problem.

Ant colony system

This optimization technique inspired by way ants find shortest path between nest and food source. Ants deposit some liquid named as pheromone on path where they found food. Pheromone amount increases on path as more and more ants using that path which is shortest path. Longest path has less amount of pheromone some of it also evaporated hence number of ants passing on that ways decreases.

In ACO algorithm [14][19] all the possible paths are represented using complete graph. It is probabilistic technique for solving computational problems which can be reduced to finding good paths through graphs. Solution of problem is nothing but list of edges and vertices which are mostly used by ants to complete their paths. Visibility and trail intensity this factors help in choosing the path. Visibility is nothing but inverse of edge length and it is nothing but what ants can see. Trail intensity is number of ants which have passed through edge to complete the tour or amount of pheromone. Virtual ant selects path based on edges having different probability of being selected. Search is continuing until destination. During selection path with higher visibility and trail intensity have higher probability of being get selected.

Self-organizing dynamic spectrum management

Self-organization is nothing but spontaneous origination of special, temporal and functional order. It plays very important part in pattern formation. Self organization management [21,22]learns the spectrum utilization patterns in environment using self organizing maps. Self amplification, competition, correlation in input signals is fundamental elements for formation of SOM. Global orders arise from local interaction. Hebbins learning rule is used for learning. Tsigankov-Koulakov model is used for SOM map. Map formation is done by stochastic minimization process. SOM problem which minimize an energy function defined based on Hebbian rule.

Table 1 shows some key advantages and disadvantages of dynamic spectrum management methods.

By analysis of these methods and results obtained from various papers PSO is best methods among all. PSO [9][17] provides fewer disadvantages than other methods. It has multi objective mechanism to more efficiently utilize the spectrum and achieve better network rewards. This method converges rapidly and exhibits good interaction efficiency. It provides good advantages than GA. Graph coloring method is also simple but it is unable to provide solution over NP problem. PSO is used for continuous optimization. It is applicable for problems which are fuzzy in nature.

Service Oriented Spectrum Allocation methodology

GCP and PSO are very simple methods and efficient to allocate the spectrum so we are reviewing spectrum allocation strategy which uses this both methods. This service oriented spectrum allocation method uses GCP and Advance BPSO algorithm. Spectrum allocation is service oriented so it reduces signal degradation, poor communication and spectrum inefficiency resulting from mismatch between spectrum resources and service demand. Based on users demands and spectrum resources, multi attribute service oriented normalization model is developed which results in reduction in deterioration. It fulfills various service demands and better utilization achieves. Bandwidth, noise, and channel availability such parameters consider for separate services. Bearing in mind different properties of spectrum resources, channel is assign to match service demands. For better network reward channel characteristics are broadly consider. Graph model can be regarded as an NP problem. Allocation problem represented as graph problem. Spectrum is efficiently utilize using BPSO.

Table 1. Advantages and disadvantages of spectrum allocation strategies

Methods	**Key features**	**Disadvantages**
Graph coloring	1.Effective and adequate 2.Simple to implement	1.Ineffective over NP problem
Game theory	1.Easy reading of outcomes 2.Models agents behavior is situations of choice 3.Equilibrium solution to problem of spectrum	1.High cost 2.Does not make optimal choices
Genetic algorithm	1.Better spectral utilization 2.Higher throughput 3.Tackle large scale NP hard problem 4.Parallel processing 5.Simple calculation	1.Slow 2.Choice of parameters and operation is difficult
Particle swarm optimization	1.Conceptually simple and easy 2.Less number of parameter to adjust 3.Advantagious when population size is large 4.Local interaction solve global problem	1.Aggregate interference in order not to exceed interference temperature limit 2.Easily suffered from particle optimism
Ant colony system	1.Inherent parallelism 2.Positive feedback accounts for rapid discovery of good solution 3.Adapt to change such as new distances	1. Theoretical analysis is difficult one. 2.Probability distribution changes by iteration 3. Time to convergence is uncertain. 4.probability distribution changes by iteration
Self organization n management	1.Ability to self organize without centralized control or human intervention 2.enhance performance measure and cost reduction	1.Scalability is critical performance indicator in large network 2.May not leads to optimal solution 3.Tradeoff between the reliability and signaling overhead of SOM must be addressed 4.New theory of distributed self-control and management is required for reliable CRN

Primary functions of this multi objective based service oriented particle swarm optimization (MBSO-PSO) are divided into three modules: Initialize module, fitness module and update module. PSO provides optimal solution over NP problem. After updating function spectrum is allocated.

Initialization of Network parameters

Network Model as GCP and its related matrices

In graph coloring problem network is represent in the form of graph. Labeling is done with help of color. Elements of graph are color. Optimization problem treated as coloring problem. At initial stage network having primary users and secondary users are represented in the form of GCP. Primary means licensed users when unused channel for specific time period that channel are use for communication by cognitive users. Cognitive elements of network are representing by vertex and interference on channel represented by edge. Different colors are assigning for vertex and not two edges incident on same vertex have same color. In network suppose M primary and N secondary users are present. Cognitive users have range from 0 to N-1 and primary users have range from 0 to M-1. Range of primary users' represented by circle. Available spectrum resource is alienated into K orthogonal spectra.

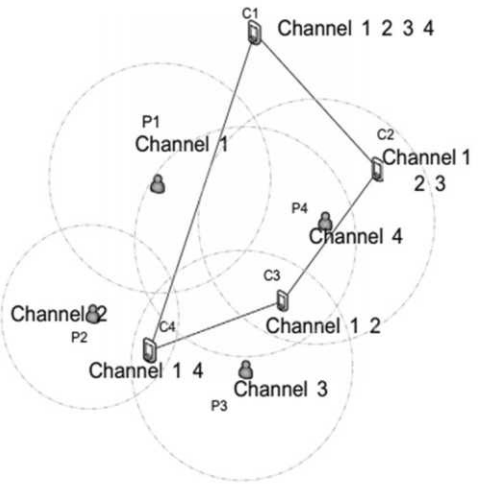


Figure 2: Topology of the network

Figure 2 represents GCP model of cognitive network. Primary users represented by P1, P2,P3 and P4. Cognitive users represented by C1, C2, C3and C4. Each primary user has its specific range represented by circle. Interference between different cognitive users is represented by edge. For cognitive users various channels are available calculated by graph theory. We are considering 10 primary users and 5 cognitive users so there are 10 channels available for each cognitive users'.

Several matrices are calculated using GCP

- *Channel availability matrix:*

Primary users are licensed users and cognitive are unlicensed of spectrum. When specific channel is assign to primary user that channel is not used by cognitive users within primary users power coverage. When CR use the channel it is represented by $l_{i,j}=1$. If primary user utilize channel it is represented by $l_{i,j}=0$. Channel availability matrix represented by L and describes as:

$$L=\{ l_{i,j} | \, l_{i,j} \in \{0,1\}\}_{N\times K} \quad (1)$$

Depending on primary users value availability matrix is calculated

- *Interference Matrix:*

As network contains number of users' same channel can be used by various users. It results in interference on that channel. Suppose user I and m use channel j then interference will be present on that channel. It denotes by $C_{i,m,j}=1$. If channel is occupied by single user then it is denoted by $C_{i,m,j}=0$. Interference matrix described by:

$$C=\{C_{i,m,j} | \, C_{i,m,j} \in \{0,1\}\}_{N\times N\times K} \quad (2)$$

Based on cognitive users and edge between vertex this matrix is calculates.

- *Non- interference allocation matrix:*

Non interference allocation records in matrix A. If $a_{i,j}=1$ then user i occupied channel j. If channel j is not occupied by user i then its denoted by $a_{i,j}=0$. Non interference matrix is described by:

$$A=\{ a_{i,j} | \, a_{i,j} \in \{0,1\}\}_{N\times K} \quad (3)$$

- *Reward matrices :*

Bandwidth (BW), Power efficiency (BD), Spectrum availability (BA), and Packet loss (BL) are feature characteristics of network reward. Using this features comprehensive matrix of channel is calculated.

M channels are available for N CR users, each describes by K quality characteristics. System reward reflects by this attributes. Different characteristics are available for diverse channel.

- *Bandwidth (BW):*

Spectral limit of channel have shown by bandwidth attribute. Limit for frequency passes through channel defines in terms of lower frequency and higher frequency of the signal. Bandwidth is vital key factor in spectrum allocation. Spectral buckle results without bandwidth. ISM frequency range selected in between 100-1000kbps for communication in industry, medical. These values are randomly generated. Bandwidth matrix is represented by:

$$BW=\begin{bmatrix} bw_{11} & \cdots & bw_{11} \\ \vdots & \ddots & \vdots \\ bw_{N1} & \cdots & bw_{NK} \end{bmatrix}_{N\times K} \quad (4)$$

- *Transmission power efficiency (BD):*

Transmission power efficiency is recompense of the users transmission power. Interference between different users changes with change in restriction of cognitive users which is transmission power. Power efficiency also depends on communication range.

$$BD=\begin{bmatrix} bd_{11} & \cdots & bd_{11} \\ \vdots & \ddots & \vdots \\ bd_{N1} & \cdots & bd_{NK} \end{bmatrix}_{N\times K} \quad (7)$$

- *Spectrum availability (BA):*

Availability probability of spectrum idle time length and frequency of channels depend on specific conditions. Spectrum switching gets lower when idle time is longer. It results in increase network reward. In time quantum spectrum utility described by 0and1. Low availability probability represents by 0 and high by 1. This reward is represented as follows:

$$BA=\begin{bmatrix} ba_{11} & \cdots & ba_{11} \\ \vdots & \ddots & \vdots \\ ba_{N1} & \cdots & ba_{NK} \end{bmatrix}_{N\times K} \quad (8)$$

- *Packet loss rate (BL):*

Packet loss rate is used for level of noise in channel. Based on service demands its range is set as 0.5-5%. Precise value is generated between this range. This reward is represented as:

$$BL=\begin{bmatrix} bl_{11} & \cdots & bl_{11} \\ \vdots & \ddots & \vdots \\ bl_{N1} & \cdots & bl_{NK} \end{bmatrix}_{N\times K} \quad (9)$$

For above features different users may have different preference. Normalizing this attributes with preference weight value gives comprehensive network reward matrix given in eq.(10). This above algorithm is service oriented spectrum allocation algorithm. Using eq.(10) can allocate spectrum but as PSO has good efficiency so it use for spectrum allocation. Advance PSO algorithm is used for spectrum allocation. These matrices are randomly generated in solution space.

Initialization of PSO parameter:

Activity of PSO in solution space is described by various parameters. Parameters like velocity, position, weight and the number of particles are initialized. Position matrix is initialized and this function is updated.

Estimation of fitness function

Optimization of solution space based on collision relation

Same channel cannot be used by different CR users Based on interference relation between users optimization function is calculated. This function checks user-channel pairs in the solution space of matrix L1. Two cognitive users uses same channel collision between them occurs in interference matrix C. User-channel pair which have larger reward receives position value 1 otherwise 0.

Comprehensive reward matrix calculation

Comprehensive reward matrix is representation of comprehensive features of channel. It is nothing but reward normalization function. Comprehensive reward matrix denoted by B. Matrix D describe as follows:

$$B=\{ b_{i,j} | b_{i,j} \in \{0,1\}\}_{N\times K} \quad (10)$$

This model focuses on service types of wireless communication, which highlights different indicators. For real time conferencing video users required high bandwidth and spectrum availability. Audio users required high bandwidth. Data users required low packet loss rates and low error rates.

For specific service type different preference weights of different spectrum attributes are assigning. Preference weight for audio users is W1, W2 assigns for video, for data users W3 are assign.

$$W1= \{w1_1, w1_2, w1_3, w1_4\} \quad (11)$$

$$W2= \{w2_1, w2_2, w2_3, w2_4\} \quad (12)$$

$$W3= \{w3_1, w3_2, w3_3, w3_4\} \quad (13)$$

Comprehensive matrix is calculated using these weights and sub rewards.

- *Matrix for audio user*

$$B_{i,j} = w1_1*bw_{i,j}+ w1_2*bd_{i,j}+ w1_3*ba_{i,j}+ w1_4*bl_{i,j} \quad (14)$$

- *Matrix for video users*

$$B_{i,j} = w2_1*bw_{i,j}+ w2_2*bd_{i,j}+ w2_3*ba_{i,j}+ w2_4*bl_{i,j} \quad (15)$$

- *Matrix for Data user*

$$B_{i,j} = w3_1*bw_{i,j} + w3_2*bd_{i,j} + w3_3*ba_{i,j} + w3_4*bl_{i,j} \quad (16)$$

Fuzzy comprehensive judgment is used to solve multifactor decision making issue. Spectrum characteristics of K channels available for cognitive users are indicated by performance indicators. S different reward matrices are used to describe rewards of different attributes.

Greater network reward obtained when values of indicators for bandwidth reward, transmitter power reward and spectrum availability rewards are larger. Lower network reward results from larger value of packet loss reward indicator. For different indicator various measurement standards are available. Normalization of this attributes values obtains. Values are worn to obtained results in 0-1 range and their relative importance.

- Large value preference:

$$V_{i,j} = \frac{b_{i,j}}{max \sum_{i=1}^{N} \sum_{j=1}^{K} b_{i,j}} \quad (17)$$

- Small value preference:

$$V_{i,j} = \frac{b_{i,j} - min \sum_{i=1}^{N} \sum_{j=1}^{K} b_{i,j}}{max \sum_{i=1}^{N} \sum_{j=1}^{K} b_{i,j}} \quad (18)$$

This preference gives value between 0 and 1. From this formula relative importance of each sub reward is obtained. Value between 1to10 is assigned to sub reward. Using this uniform measure for different reward types achieves.

3) Account of fitness function:

Using sum of the rewards for all positions it occupied fitness function is calculated. The Optimum fitness position is put as global optimum position and fitness as global fitness at the every iteration. Optimum fitness for the entire particle at every position are calculated and recorded. For PSO algorithm evaluation of fitness function is a vital problem. Network reward is used to calculate fitness. Individual particle fitness is nothing but sum of the rewards for all of the occupied position. Fitness function is describes as:

$$\text{Fitness} = \sum_{i=1}^{N} \sum_{j=1}^{M} (a_{i,j} x b_{i,j}) \quad (19)$$

Update function

Status Update

New velocities and positions for all particles are premeditated. At the t+1 moment using interference matrix new velocity of particle is calculated.

$$V_{id}^{t+1} = w*V_{id}^{t} + c_1 r_1 (p_{id}^{t} - x_{id}^{t}) + c_2 r_2 (p_{gd}^{t} - x_{idid}^{t}) \quad (20)$$

C1 and C2 are parameters for study. r1 and r2 are random values generated in range 0 to 1. The variable p_{id}^{t} is optimization position of each particle. p_{gd}^{t} is global parameter. We are considering C1 and C2 as 2 and W as 4. C1 and C2 are weighting of stochastic acceleration that pull each particle towards best position. Inertia controls impact of previous velocity on present.

Velocity and position is updated by change in last velocity and present position. Velocity is also updated when inertia coefficient w is updated for better solution. When it is updated iteration number should divided by 5 with no reminder and subtracted by 0.2 until minimum limit of w. update function for position as follows:

$$S_{(Vid)} = \frac{1}{1+e^{(-vid)}} \quad (21)$$

These positions are updated till iteration number. Inertia coefficient (W) also updated. Iteration number is divided by 5 with no remainder, and W=W-0.2 until W reaches to minimum limit. C1 and C2 we set as 2 and W as 4. Values of r1 and r2are within range 0 to 1.

Spectrum allocation

Allocation of array information obtained from graph coloring model and normalization model are done by PSO algorithm. High comprehensive reward for each solution is obtained and recorded. At each iteration velocity and position get updates for better solution. Result of this methodology is allocation. Taking into consideration channel characteristic suitable channel is allocated to cognitive users.

Entire network and fairness reward

Entire network reward represented by W_{max} . It is entire network reward based on users required. It measures network benefit resulting from allocation. It is describe by:

$$W_{max} = \sum_{i=1}^{N} \sum_{j=1}^{K} (a_{i,j} * b_{i,j}) \quad (22)$$

Fairness of network used for check network fairness in situation of entire network. It is describe as:

$$W_{fair} = (\prod_{n=1}^{N} (\sum_{m=1}^{M} (a_{i,j} * b_{i,j}) + 1 * 10^{-6}))^{1/N} \quad (23)$$

Conclusion and Future work

This paper mainly focuses on various modified techniques used for resolve this spectrum deficiency problem. This various methods are used for various problems in cognitive radio network. Among all Particle swarm provide better solution. This methods considers various parameter to satisfy user's demands. By analysis of these methods and results obtained from various papers PSO is best methods among all. PSO provides fewer disadvantages than other methods. It has multi objective mechanism to more efficiently utilize the spectrum and achieve better network rewards. This method converges rapidly and exhibits good interaction efficiency. It provides good advantages than GA. Graph coloring method is also simple but it is unable to provide solution over NP problem. PSO is used for continuous optimization. It is applicable for problems which are fuzzy in nature.

Service-oriented dynamic spectrum allocation comprehensively considers the bandwidth, the noise, the channel availability, and other spectrum characteristics to satisfy the users' service demands. It uses GCP and BPSO this algorithm which is having more advantages than others. Using this algorithm a better network reward and a better fairness reward is obtained viz a viz several other algorithms to satisfy cognitive users' service demands. It overcome problem arises due to mismatch between the spectrum resources and the service demands. It considers cognitive users service and different properties of spectrum resources and provide them match channel which fulfill users demands.

This service oriented approach we are implementing using incorporating learning concept which gives better allocation.

References

[1] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal", IEEE Personal Commun., Vol. 6, no. 4, Aug. 1999, pp. 13.

[2] Ian F Akyildiz ,Won-Yeol Lee Mehmet C Vuran and Shantidev Mohanty Georgia institute of technology, "A survey on spectrum management in cognitive radio networks", IEEE communications magazines, Vol 46 , April 2008, pp.40-48.

[3] Beibei Wang and K. J. Ray Liu ieee, "Advances in Cognitive Radio Networks: A Survey", IEEE Journal Of Selected Topics In Signal Processing, Vol.5, No.1, February 2011, pp. 5-22.

[4] W.Wang,X. Liu, "List-coloring based channel allocation for open spectrum wireless networks", proceedings of VTC05, 2005, pp.690-694.

[5] Vinod Maan, G.N. Purohit, "A distributed approach for frequency allocation using Graph Coloring in Mobile Networks", International journal of computer applications (0975-88870), volume 58, no. 6, November 2012.

[6] Yu CAO, Yibing LI, Fang YE, "Improved Fair Spectrum Allocation Algorithms Based on Graph Coloring Theory in Cognitive Radio Networks", Theory in CR networks journal of computational information systems,7:13(2011), pp.4694-4701.

[7] S. Buzzi, D. Saturnino, "A game- theoretic approach to energy- efficient power control and receiver design in cognitive CDMA wireless networks", IEEEJ.Signal Proccess 5(1), (2011), pp. 137-150.

[8] Quifen Ni, Rongbo Zhu, Zhenguo Wu, Yongli sun, Lingyun Zhou, Bin Zhou, "Spectrum allocation based on Game theory in cognitive radio networks", journal networks, vol. 8, no.3, March 2013.

[9] Zhu ji, K. J Ray Liu, University of Maryland, "Dynamic spectrum sharing: A Game Theoretical overview", IEEE communication magazine, Vol. 45, Issue 5, May 2007.

[10] N. Nie, C. Comaniciu, "Adaptive channel allocation spectrum etiquette for cognitive radio netwoks", IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, November 2005.

[11] Mengyi Liu, Huyin Zhang, Rongchao Fan, Zhenchen Duan, "The GA solution of dyanamic spectrum allocation in cognitive radio based on collaboration and fairness, Third Pacific-Asia Conference on Circuits, Communications and Systems(PACCS), July.2011, pp.1-4.

[12] V. Amudha, G.P Ramesh, "Dynamic spectrum allocation for cognitive radio using genetic algorithm", G.P. International journal of technology and engineering science [IJTES], Vol1970, October 2013 pp1092-1097.

[13] Andson Balieiro, Peterson Yoshioka, Kelvin Dias, Dave Cavalcanti, Carlos Cordeiro, "A multi-objective genetic optimization for spectrum sensing in cognitive radio," Expert Systems with applications, 41(2014), pp.3640-3650.

[14] H. Ahmadi, Y.H. Chew, "Evolutionary algorithm for orthogonal frequency division multiplexing-based dynamic spectrum access systems", Springer Computer Networks, Volume 56, 2012, pp. 3206-3218

[15] Jie Jia, Qiusi Lin, Jian Chen, Chunyu Li, Xingwei wang, "Congestion aware channel allocation with route scheduling in wireless cognitive radio mesh network," computers and Electrical Engineering, Volume 39, 2013, pp. 1346-1357.

[16] Quan Liu, Hongweri Niu, Wenjun Xu, Duzhong Zhang, "A service –oriented spectrum allocation algorithm using enhanced PSO for cognitive wireless networks", computer networks, Volume74, 2014, pp.81-91.

[17] Anindita Saha, Jibendu Sekhar Roy,"Dynamic spectrum allocation in cognitive Radio using particle swarm optimization", International Journal of Emerging Technology and Advanced Engineering (ISSN 2250=2459, Volume 4, Issue 4, April 2014.

[18] Meiqin tang, Chengnian Long, Xinping Guan, Xinjiang Wei, "Nonconvex dynamic spectrum allocation for cognitive radio networks via particle swarm optimization and simulated annealing," computer networks 56, 2012, pp. 2690-2699.

[19] F. Koroupi, S. Talebi, H. Salehinejad Scientia Iranica, "Cognitive radio networks spectrum allocation: An ACS perspective," Scientia Iranica, 19(3), 2012, pp. 767-773.

[20] Pyari Mohan Pradhan, Ganapati Panda, "Comparative performance analysis of evolutionary algorithm based parameter optimization in cognitive radio engine:A survey", Ad Hoc Networks, Volume 17, 2014, pp.129-146.

[21] Farad Khozeimeh, "Self organizing dynamic spectrum management: Novel scheme for cognitive radio networks," A thesis submitted to the school of graduate studies in partial fulfillment of the requirements for the degree doctor of philosophy.

[22] Zhongshan Zhang and keping long, university of science and technology Beijing jianping wang, city university of hong kong, "Self-organization paradigms and optimization approaches for cognitive radio technologies: a survey," IEEE wireless communication, April 2013, pp.1536-1284.
[23] Wen Chen, Tao Li, Tao Yang, "Intelligent control of cognitive radio parameter adaption: using evolutionary multi objective algorithm based on user preference," Ad Hoc networks, 26, 2015 pp.3-1 6,
[24] Dingde Jiang, Yuanting Wang Chunping Yao, Yang Han, "An effective dyanamic spectrum accesss algorithm for multihop cognitive wireless networks," Computer networks, 84, 2015 pp. 1-16.
[25] Natarajan Meghanathan, Yenumula B. Reddy, "Cognitive Radio Technology Applications for wireless and mobile Adhoc networks," Aadvance in wireless technologies and telecommunication (AWTT) book series, AGI Global, 30 Jun 2013.
[26] Zhao, Z. Peng, Z.Zheng, S. and J. Shang, "cognitive radio spectrum allocation using evolutionary algorithms", IEEE Transactions on Wireless Communications,8(9),2009, pp.4421-4425.
[27] Chunyi Peng, Haitao Zheng, Ben Zhao,"Utilization and fairness in spectrum assignments for opportunistic spectrum access", Mobile Networks Appl. , 11, 2006 pp.555-576.
[28] Berk Canberk, Ian F. Akyildiz, Sema Oktug, "A QOS- aware framework for available spectrum characterization and decision in cognitive radio networks", International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), September 2010, pp. 1533-1538.

Performance Analysis of AODV+ Routing Protocol for Wireless Ad-hoc Networks

Sanjeev Kumar Srivastava*, Ranjana D. Raut** and P.T. Karule***
*Research Scholar (Electronics Engineering), YCCE, Nagpur
sanjeevkumar.srivastava1@gmail.com
**Associate Professor, S.G.B. University, Amravati (M.S.)
rdr24164@rediffmail.com
*** Professor (Electronics Engineering), YCCE, Nagpur, Maharashtra (India)
ptkarule@rediffmail.com

Abstract: This paper dealt with critical analysis of AODV+ (modified) protocol with existing AODV and Dynamic Source Routing protocol and then we have analyzed our performance based on different parameters such as average throughput, average delay, packet drops and average energy consumption in mobility and non mobility environment. Finally, it has been observed that AODV+ routing protocol shows superior results in comparison of AODV protocol & dynamic source Routing.

Keywords: AODV+, Performance Analysis, Ad-hoc Networks, DSR, Mobility and Non Mobility etc.

Introduction

Standard ad-hoc wireless network (30 MHz – 5 GHz) is a group of multiple wireless enabled communications & networking capability devices. These devices able to communicate with another node immediately within respective radio range or one outside their radio range. An ad-hoc wireless network is adaptive and self-organised. Ad hoc nodes devices should be able to detect presence of other related devices for performing the required handshaking signals to permit transfer and the exchange of information- services. The Ad-hoc wireless networks could be used in military applications, emergency/rescue operations e.g. fire, floods or earthquake etc. There are several challenges facing by ad-hoc wireless networks such as limited bandwidth, overhead, physical range, packet loss due to transmission errors etc. in addition to it, routing aspect of ad-hoc wireless networks is major constraint due to increasing size of network, it is becoming more difficult and complicated day to day. Due to constant development in network related topologies, classification of routing protocols based on the network related behaviour. It is categorised into proactive (table driven) protocol & reactive (on demand) and several other protocols exist as hybrid protocols.

In this work, we have concentrated on AODV (Ad hoc on-demand distance vector) & DSR (Dynamic Source Routing) routing protocols only. However, this paper, AODV+ routing protocols are used for improving the performance of AODV routing protocol at various parameters.

Related Work

In this section, it has been discussed the related works of some research papers. The summary of those related literatures of the research work has been elaborated in a tabular form which emphasises on the work done by researchers in a specific way.

Sr. No.	Author/s Name	Title	Year of Publication	Proposed Work and Assumptions	Limitations
1.	R.K.Jha, and P. Kharga	Critical Performance Analysis for Routing Protocols MANET based using NS-3 Simulator	March 2015	A comparative analysis of four different protocols(routing) i.e. AODV, DSR, DSDV & OLSR the MANET based on the performance metrices as throughput, packet delivery ratio and end to end delay, packet dropped ending with jitter using NS-3.	Frequent change in network topology and large routing message overhead.
2.	R.Paulus,R.Garg, T.Kaur, and S.V.S.Rajput	Analyzing effects for varying CBR on AODV, DSR, IERP Routing Protocol in	October 2013	Evaluating the performance based analysis of AODV, DSR (Reactive) & IERP (Hybrid) routing protocol with variable Constant Bit Rate (CBR) on the	Variable Constant Bit Rate (CBR).

		MANET		performance metrics as average end to end delays, throughput, average jitter and packet delivery ratios.	
3.	D.Dhokal and K.Gautam	The Performance based upon the comparison of AODV & DSR Protocols in Mobile Ad-hoc Network: A Survey	May 2013	A survey to review the comparison of AODV & DSR protocols by considering different network based metrics as mobility pattern, average end-to-end delays, throughput, packets delivery ratio, packet drops and node densities of the ad-hoc networks.	In some condition, AODV better and in some DSR better. Not mentioned clearly which routing protocol better in case of most of the performance metrics.
4.	Aarti and S.S.Tyagi	MANET a study: Characteristics, Challenges, Applications & Security Attack	May 2013	Mobile ad-hoc networks its characteristics, challenges, applications, security goals & different type of security related attack (i.e. active and passive) at different layers.	Dynamic topology, felicitates distributed operations along with limited bandwidth, MANET highly vulnerable to many security based attack.
5.	S.Bhimla, and N.Yadav	Comparison between MANET based protocols AODV & DSR	March 2012	By characterizing of wireless ad hoc network's information-theoretic capacity scaling with randomly distributed (n) multiple nodes and using the exact channel model from Maxwell's equation, capacity scaling is given as minimum of the number of nodes and physical limit on Degrees of Freedom (DoF) given as the ratio of the network diameter and wavelength. It is also considered as channel model with a path-loss exponent (α) greater than two.	An upper bound not considered for both DoF limitations due to channel's correlation & limitations of the power due to continuously attenuating power over a distance.
6.	Q.Guan, and S.Jiang	Ad Hoc Mobile Wireless Networks: Protocols and Systems	April 2011	To improve the network capacity of MANETs with cooperative Communications (enhancement of transmission reliability and bandwidth efficiency in wireless networks), it is proposed a Capacity-Optimized Cooperative (COCO) topology control scheme by considering both upper layer network capacity and physical layer relay selection.	The impact of MAC layer between physical layer and network layer is not discussed.

7.	J.Li and C.Blake	Capacity of Ad hoc Wireless Networks	Oct. 2002	Depending on network size, traffic patterns and local radio interactions, it is examined the interactions of the 802.11 MAC and ad hoc forwarding and the effect on capacity for several simple configurations and traffic patterns.	No such network model to maximize the performance of throughput capacity in ad-hoc wireless network is discussed.

Routing Protocols

AODV (Ad hoc on-demand distance vector) routing protocol provides self-initialization, multiple-hop, and dynamic routing of mobile ad hoc network. The path is discovered without source routing & table is maintained instead of route cache. This is loop free using destination sequence numbers & nodes nodes of the mobile for responding to link breakages, changes in a network topology in timely manners. DSR (Dynamic Source Routing) is on-demand, simple & efficient routing protocol for multi-hop wireless ad hoc networks of mobile nodes [1]. DSR make use of source routing & protocol composed of mainly two mechanisms-'Route Discovery' & 'Route Maintenance', which together works, on demand [2]. AODV+ routing protocol is used for performing the analysis based on different parameters i.e. average throughput, average delay, packet drops and average energy consumption by considering link between different random nodes for change in routing scheme.

Performance Analysis

We have simulated the analysis and to do so, we have developed this model in NS-2 software. Therefore we have conducted Basic four steps for simulation as depicted in Figure (1).

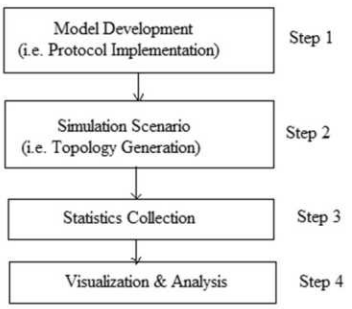


Figure (1) Simulation Steps

First step is the development of a model (i.e. implementation the protocol); step second being the creation of simulation scenarios (i.e. designing of the network topology and traffic scenarios); followed by the steps of choosing and data collection, and the final step being visualization and the simulation result analysis which might be carried out after (or during, in some cases) the simulation execution.

For enhanced performance of AODV routing protocol, there has been changed in the simulation results on different parameters i.e. Average throughput, Average delay, Drop Packets and Average energy consumption. Therefore, the performance of existing AODV routing protocol along with DSR routing protocol and modified AODV (AODV+) routing protocol simulation results have been compared at different simulation times i.e. 150 and 250 seconds. There will be change in routing scheme by considering link between different nodes. Nodes are assumed to be random with mobility and without mobility. By considering area of 400 × 400 square meters, 4 to 5 sources are taken into account. It is very important to contribute high scalability in terms of number of nodes. It has been observed in many references that simulations are carried out on less number of nodes. The summary of assumption of different parameters is listed.

Results and Discussions

This chapter is about simulating and discussing the performance of existing AODV, DSR and AODV+ (modified) on different parameters i.e. average throughput, average delay, packet drops and average energy consumption with and without mobility.

Table 1: Parameters Assumption

Sr. No.	Parameters	Values
1.	Traffic Model	CBR (Constant Bit Rate)
2.	Area of Experiment	400 meter × 400 meter
3.	Number of Nodes	25 ~100
4.	Routing Protocol	AODV and DSR
5.	Packet Size	1000 bytes
6.	Simulation Time	150 sec, 250 sec.
7.	Mobility Scenario	5 Km/hr.
8.	Mobility Model	Random Way Point
9.	MAC layer Protocol	IEEE 802.11 with 2.4 GHz

Table 2: Observation Table of AODV+ with Mobility

Nodes	Average Throughput	Average Delay	Packet Drops	Average Energy Consumption
25	10357	0.45956	157	0.653
50	12358	0.3537	197	1.127
75	13758	0.2755	212	0.725
100	12158	0.1232	225	0.325

Table 3: Observation Table of AODV+ without Mobility

Nodes	Average Throughput	Average Delay	Packet Drops	Average Energy Consumption
25	13785	0.3	87	0.553
50	15653	0.298	110	0.982
75	14083	0.23	127	0.617
100	13785	0.1232	143	0.125

Table 4: Observation Table of DSR with Mobility

Nodes	Average Throughput	Average Delay	Packet Drops	Average Energy Consumption
25	8100	0.260	112	0.957
50	10536	0.3900	200	1.327
75	11565	0.2898	219	1.25
100	9785	0.2135	230	0.625

Table 5: Observation Table of DSR without Mobility

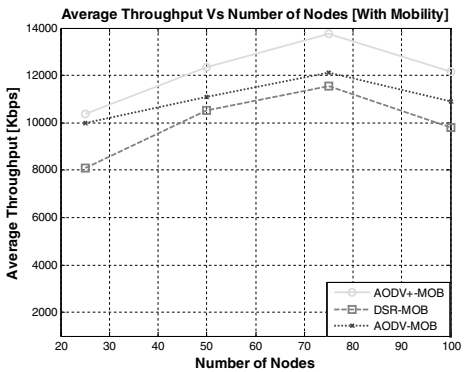
Nodes	Average Throughput	Average Delay	Packet Drops	Average Energy Consumption
25	11887	0.2	85	0.897
50	13875	0.3	115	1.207
75	12875	0.2120	122	0.987
100	10567	0.1897	149	0.527

Table 6: Observation Table of AODV with Mobility

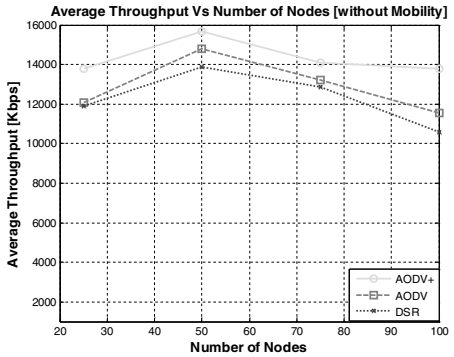
Nodes	Average Throughput	Average Delay	Packet Drops	Average Energy Consumption
25	9980	0.296	178	0.853
50	11087	0.406	212	1.237
75	12137	0.310	247	0.985
100	10898	0.2530	265	0.435

Table 7: Observation Table of AODV without Mobility

Nodes	Average Throughput	Average Delay	Packet Drops	Average Energy Consumption
25	12085	0.2700	98	0.653
50	14785	0.2537	120	0.982
75	13235	0.3007	137	0.617
100	11565	0.2037	157	0.125



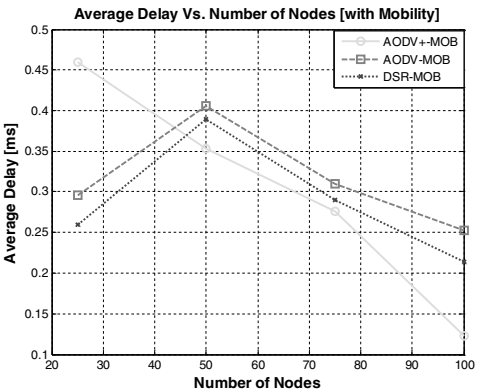
Graph (1): Simulation Results of Average Throughput with Mobility



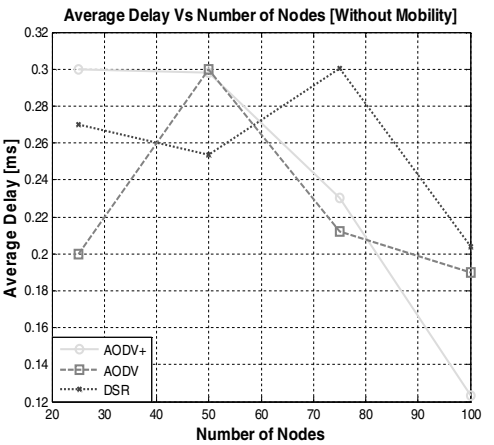
Graph (2): Simulation Results of Average Throughput without Mobility

By referring the observation tables of AODV+, AODV and DSR with mobility and without mobility and their respective graphs shown in graph (1) and (2) for average throughput,the performance of number of nodes vs average throughput in two scenarios are hereby discussed. It has been observed that initially at lower number of nodes DSR, AODV and AODV+ show lower throughput value as network resources are underutilized i.e. less number of packets are generated in the systems. As numbers of nodes are increased above lower value the average throughput starts increasing as network resources/ traffic are increased and utilize the channel capacity at full and as number of nodes is increasing performance starts degrading because

collision and competition to get channel access becomes very severe. From the graph we can observe that AODV+ is giving better results with mobility scenario as frequently updating routing table with link state is easy compared to traditional DSR and AODV protocols. However, in without mobility scenario DSR and AODV protocols show better performance compared to AODV + till medium number of nodes because stability of their routing table is better. As nodes are not mobile their routing table is better and static as compared AODV+.

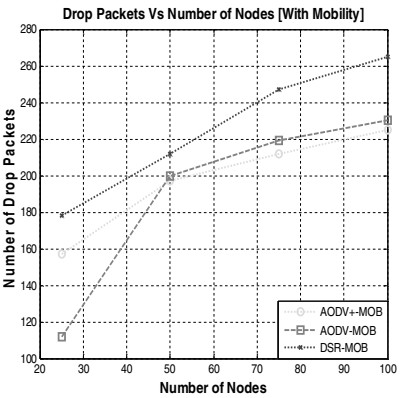


Graph (3): Simulation Results of Average Delay with Mobility

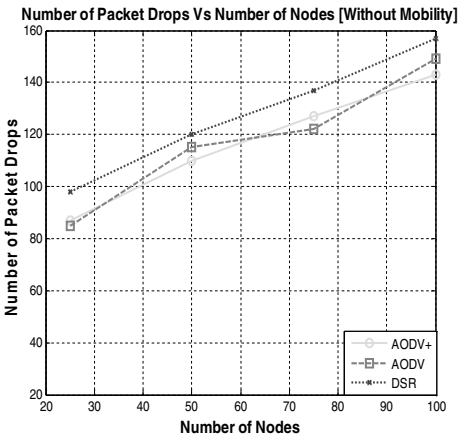


Graph (4): Simulation Results of Average Delay without Mobility

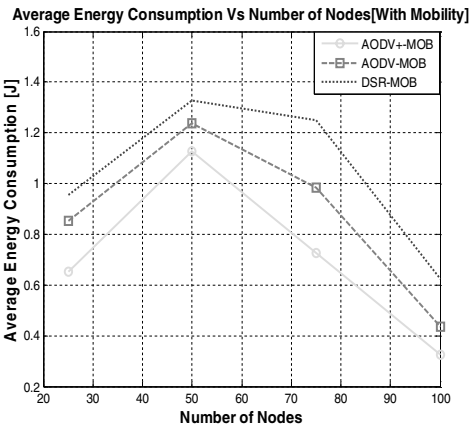
By referring the observation tables of AODV+, AODV and DSR with mobility and without mobility and their respective graphs shown in graph (3) and (4) for average delay, it is observed that as the nodes are increasing, the performance of AODV+ gives lower delay because the performance of DSR and AODV are fluctuating i.e. increasing and the suddenly decreasing whereas it is continuously decreasing in AODV+ routing protocol. Even though the performance of AODV+ routing protocol needs more detail analysis to check its lower performance but one of the reasons could be creating a heavy routing table by considering link quality for the same. This could be verified by checking the end to end delays performance.
Hence, in the end to end delay performance, AODV+ routing protocol gives the best results at higher node numbers as at lower number of nodes creating routing table could be a heavy exercise.
By referring the observation tables of AODV+, AODV and DSR with mobility and without mobility and their respective graphs shown in (5) and (6) for Packet Drops, it is observed that as the nodes are increasing, the performance of AODV+ routing protocol, DSR routing protocol and AODV routing protocol results are almost consistent with this analysis. But overall performance of AODV+ routing protocol is better as compared to DSR routing protocol and AODV routing protocol as number of packet drops are less.



Graph (5): Simulation Results of Drop Packets with Mobility

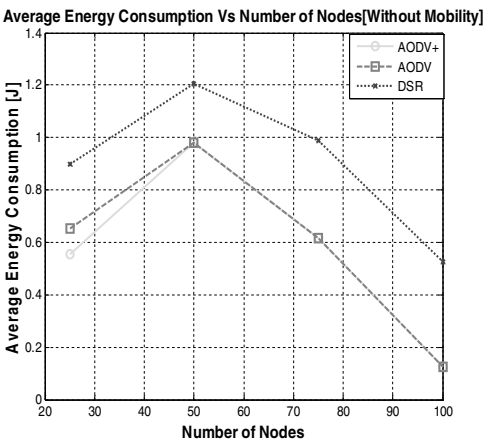


Graph (6): Simulation Results of Drop Packets without Mobility



Graph (7): Simulation Results of Average Energy Consumption with Mobility

By referring the observation tables of AODV+, AODV and DSR with mobility and without mobility and their respective graphs shown in (7) and (8) for average energy consumption, it is observed that as the nodes are increasing, the performance of AODV+ routing protocol, DSR routing protocol and AODV routing protocol results are almost consistent with this analysis. But overall performance of AODV+ routing protocol is better as compared to DSR routing protocol and AODV routing protocol as number of packet drops are less. Therefore, average energy consumption for AODV+ routing protocol is less.



Graph (8): Simulation Results of Average Energy Consumption without Mobility

Conclusions

Therefore, from the overall observation of various graphs, it is concluded that AODV+ routing protocol is giving almost 5 to 10 % improvements on various parameters such as average throughput, average delay, packet drops and average energy consumption as compared to AODV and DSR routing protocols in mobility and non mobility environment. Therefore, AODV+ routing protocol is giving better performance as compared to AODV routing protocol and DSR routing protocol in mobility and non mobility scenario.

Future work

It is suggested that we can include few more parameters such as QoS (Quality of Service) and priority to different load and data traffic in future.

Acknowledgment

We are very much thankful to all friends, family & colleagues for their valuable suggestions, constant support and cooperation.

References

[1] D.B. Johnson and D.A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks", Mobile Computing, vol. 353, Kluwer Academic, 1996, pp. 58-62.

[2] J. Broch, D.A. Maltz, D.B. Johnson, Y.C. Hue and J. Jetcheva, "A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols", Proc., MobiCom, 1998, pp. 112-120.

[3] C.E. Parkins and E.M. Royer, "Ad- Hoc on Demand Distance Vector Routing", Proc. Workshop Mobile Computing Systems and Applications (WMCSA' 99), Feb. 1999, pp. 140-155.

[4] Mohseni S., Hassan R., Patel A., and Razali R., "Comparative review study of reactive and proactive routing protocols in MANETs", 4th IEEE International Conference on Digital Ecosystems and Technologies, 2010, pp. 304-309.

[5] Humayun Bakht, "Survey of Routing Protocols for Moble Ad-hoc Network", International Journal of Information and Communication Technology Research, 2010,pp. 258-270.

[6] DevuManikatan and Shila Yu Cheng, "Ad-hoc wireless networks meet the infrastructure: Mobility, Capacity and Delay", 31st Annual IEEE International Conference on Computer Communications, 2012, pp. 65-72

[7] Salman Bhimla and NeenaYadav, "Comparison between AODV Protocol and DSR Protocol in MANET", International Journal of Advanced Engineering Research and Studies (IJAERS), 2012, pp. 8-20

[8] DependraDhokal, KiranGautam, "Performance Comparison of AODV and DSR Routing Protocols in Mobile Ad-hoc Networks: A Survey ", International Journal of Engineering, Science and Innovative Technology (IJESIT), vol. 2, Issue 3, 2013,pp. 258-265.

[9] Sanjeev Kumar Srivastava, Dr. (Mrs.) Ranjana D. Raut and Dr. P.T. Karule, "Evaluation of Performance Comparison of DSR and AODV Routing Protocols in Mobile Ad HOC Networks", International journal of Electronics, Communication, Soft Computing, Science and Engineering (IJECSCSE), 46th Midterm IETE Symposium, ISSN no.: 2277-9477,March 2015, pp. 82-88.

[10] Sanjeev Kumar Srivastava, Dr. (Mrs.) Ranjana D. Raut and Dr. P.T. Karule, "Evaluation of Performance Comparison of DSR and AODV Routing Protocols in Mobile Ad HOC Networks", International journal of Electronics, Communication, Soft Computing Science and Engineering (IJECSCSE), 46th Midterm IETE Symposium, DOI.: 10.1109/ICCN. 2015.50, November 2015 pp. 258-261.

Comparative Study of Various File Systems for Data Storage in Virtualized Cloud Environment

Rashmi Jogdand*, Mohan A. Gholap** and D.R. Gangodkar***
*Professor, Dept. of CSE, Gogte Institute of Technology, Belgaum, Karnataka
rmjogdand@git.edu
** System Analyst, Hirasugar Institute of Technology, Nidasoshi, Dist:Belgaum, Karnataka
mohangholap04@gmail.com
*** Professor, Graphic Era University, Dehradun, Uttarakhand
dgangodkar@yahoo.com

Abstract: Today numerous distributed file systems are available for data storage in the cloud. Let us have a brief comparative study of each of the five file systems (viz. ZFS, CEPH, GlusterFS, GFS and HDFS) which are used for data storage in the virtualized cloud. Zettabyte File System (ZFS) is a very simple file system which ranges from a small network of devices to a large network of devices. It is characterized with the two dominant features: Cooperative cache and Distributed transactions. CEPH file system is characterized with its metadata cluster architecture, which is highly adaptive and scalable. Gluster File system is a network file system which is also highly scalable. It has found its applications in media streaming services as well as cloud computing. Google File System (GFS) is proprietary in nature and characterized by its fault tolerance and high speed response to the client requests. Hadoop Distributed File System (HDFS) is mainly based on its Map/Reduce technique, using which it ensures higher level of fault tolerance as well as high throughput. HDFS is nothing but an open source version of GFS. In this paper we describe the architecture of each file system in brief, its advantages, limitations at last we compare these file systems in brief.

Keywords: Zettabyte File System, CEPH File System, Gluster File System, Google File System, Hadoop File System, Map/Reduce.

Introduction

The direct dependency of the performance of file systems on the overall performance of the applications has compelled many system designers to improve the file system performance. The scientific and high speed computing communities have done lot of advancements in the scalability and performance of distributed storage systems. Traditional systems were based on the client-server model which was very simple to implement and manage but main drawback was the dependency on central server. This led to the development of modern distributed file systems which uses intelligent Object Storage Devices (OSDs) replacing the traditional hard disks. The OSDs combine CPU, local cache and network interface with the underlying RAID or disk.

OSDs provide a more sophisticated and advanced interfaces by which the clients can perform read/write operations to much larger range and also for the variable sized named objects. Here, clients directly communicate with OSDs to perform metadata operations with Metadata Server (MDS).

In this paper we study four such distributed file systems: CEPH, GlusterFS, GFS and HDFS in brief. This paper also discusses one of the Disk file system the ZFS. This paper is organized as follows. Section II, III, IV, V, VI presents ZFS, CEPH, GlusterFS, GFS and HDFS respectively. In each of the above file systems, we discuss its simple architecture, advantages and disadvantages. In section VII, we present a comparison table. And in section VIII, we present our conclusion.

Zettabyte File System (ZFS)

Zettabyte File System (ZFS) is a very simple file system which ranges from a small network of devices to a large network of devices. It is characterized with the two dominant features: Cooperative cache and Distributed transactions. This is distributed in nature and distributes the file and storage management operations over the network of devices. ZFS integrates/combines the memory of all the communicating devices into a single block called a “Coherent cache”. Hence, for data read operations, it will not access the individual disk which contain that block, instead it fetches from the remote machine. Being a first 128-bit file system, ZFS is highly scalable.

This paper describes the high level architecture of ZFS and also how its goals achieved. We know that the two most prominent features of ZFS are its cooperative cache and distributed transactions. According to this architecture, it consists of six components:

Front End (FE), Cooperative Cache (Cache), File Manager (FMGR), Lease Manager (LMGR), Transaction Server (TSVR) and an Object Store (OSD). All these components work together to address the issues like fault tolerance, Security and backup/mirroring.

Architecture

In this section we describe the architecture of ZFS in detail which shows how ZFS components interact with each other to present to the applications a distributed file system. First we ZFS component interconnections followed by the protocols that describe how file system operation protocols are carried out in the ZFS architecture.

ZFS Component Interconnections

Figure-1 illustrates all the architecture of ZFS [1]. At the bottom, there are many object disks and at the top there are two hosts running ZFS components. The FE and Cache are situated inside the kernel while the LMGR, FMGR and TSVR are located in a single process in the user-space. Here, one very important observation is that, not all these components are active on all machines at all times.

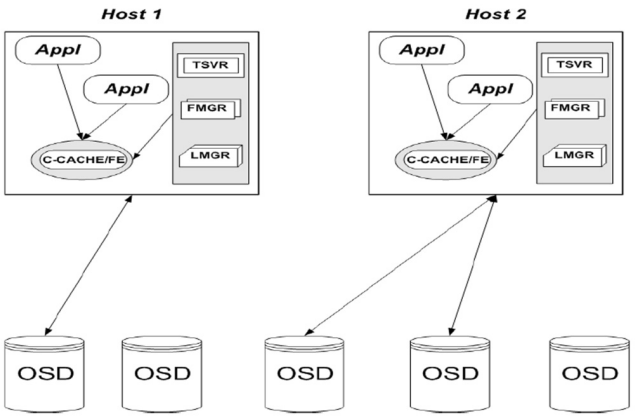


Figure -1: ZFS Components

If we consider the extreme case, only the TSVR may be active and all other components for the files used on this particular machine may be running on other nodes. A socket connects the in-kernel and out-of-kernel components, and OSDs are accessed directly by the hosts.

Let us consider the read() file operation to explore the interactions between the ZFS components.

File Read Protocol

Figure-2 shows the control and information paths for the read(file, ...) operation detailed below.

a. FE looks up through the fsptr for file.
b. If the read can be satisfied by locally cached file blocks (i.e., the data and read lease are locally cached) then the requested data is returned to the user and we return.
c. A read request is sent to the FMGR of the file and to the FE. Here, Cache will wait for the request to be satisfied.
d. The FMGR will check and if required creates a read-lease for the requested blocks.
e. The FMGR checks whether any other Caches hold the requested blocks of the file and does the following:
   If TRUE
   a. Forwards the above byte-range lease, the read request and the address of the requesting FE, the requester, to the Cache, FE on the host holding the requested blocks, host.
      ELSE
   b. Forwards the above byte-range lease, the read request and the address of the requesting FE, the requester, to the OSD holding the requested blocks.
f. The FE/Cache on host or the OSD send the requested data blocks and the read lease to the requester.

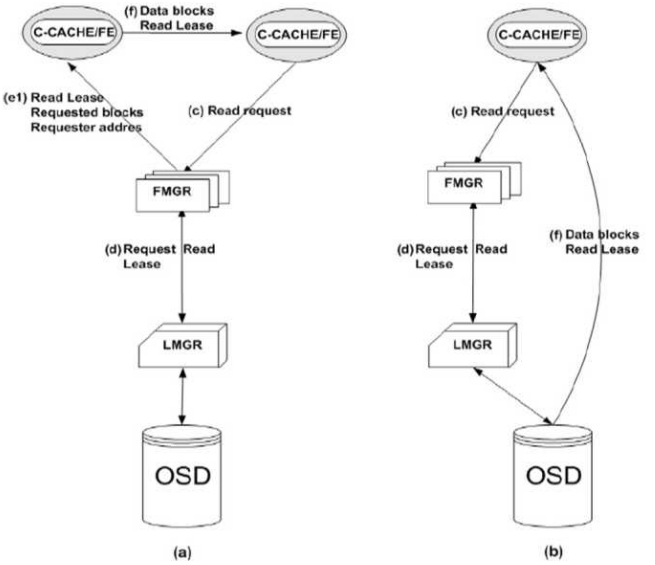


Figure-2: ZFS Operations (a) cache-to-cache read (b) OSD read

Advantages

1. ZFS guarantees Strong data integrity with the use of RAID 5.
2. Simple administration.
3. Efficient for file system compression.
4. Highly scalable (the first 128-bit file system).
5. Better Performance.

Limitations

1. Due to the absence of ZFS boot support, this is not a root file system.
2. ZFS lacks encryption option.
3. ZFS does not support Disk Quota facility (may be per-user or per-group).

CEPH File System

Ceph is an open-source distributed file system. Ceph's building blocks are OSDs (Object Storage Daemons). OSDs are responsible for storing objects on local file systems, as well as working together to replicate data, detect and recover from failures, or migrate data when OSDs join or leave the cluster. Hence, Ceph aims to ensure reliability and scalability by leveraging the intelligence of the OSDs. Each OSD uses a journal to speed-up the write operations by coalescing small writes and flushing them asynchronously to the backing file system when the journal is full. The journal can be a different device, partition or file [2].

Ceph File system separates metadata and data operations by using the generating functions unlike other file systems which uses file allocation tables. With this, it is possible to distribute the complexity of data operations such as data access, replication, failure detection and recovery.

Architecture

Ceph file system is characterized with its metadata cluster architecture, which is highly adaptive and scalable. The Ceph file system consists of three main components: Client, a Cluster of OSDs and a Metadata server cluster. Each instance of the client provides POSIX like file system interface to the hosts and processes. The Metadata server cluster manages the namespace (filenames and directories) and also coordinates the Security, Consistency and Coherence (See Fig-3) [3]. As already mentioned, the primary goals of this are Scalability, Performance and Reliability. Scalability may range beyond hundreds of petabytes and may be in terms of overall storage capacity, throughput of the system and performance.

Ceph's most interesting feature is the CRUSH algorithm (Controlled Replication Under Scalable Hashing) [12]. CRUSH is a pseudo-random placement algorithm that allows OSDs and clients to compute object locations instead of looking them up in a centralized table. Then clients can directly interact with the OSDs for I/O. This promises, in theory, extreme scalability. Ceph directly addresses the issue of scalability while simultaneously achieving high performance, reliability and availability

through three fundamental design features: decoupled data and metadata, dynamic distributed metadata management, and reliable autonomic distributed object storage [4].

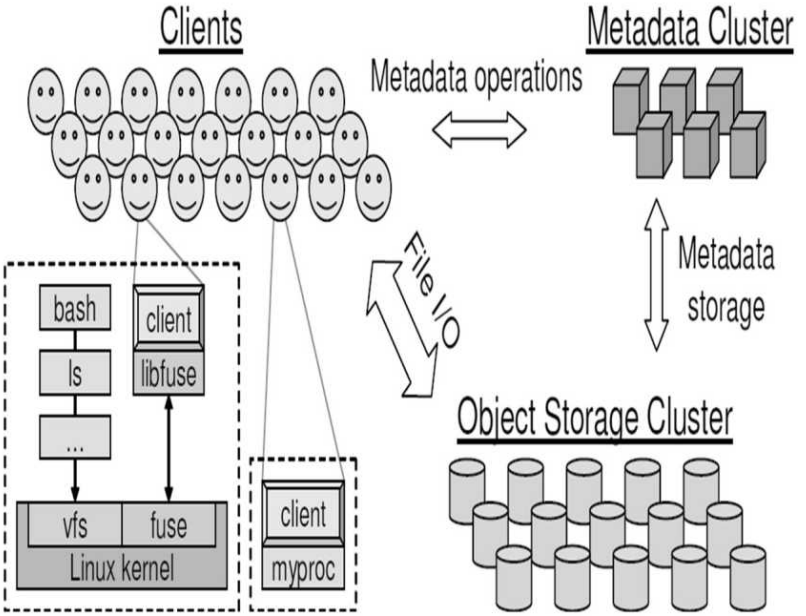


Figure-3: CEPH System Architecture

The OSDs in Ceph have peer-to-peer capability and manages replication and failure handling processes autonomously. The collection of OSDs and a number of highly scalable monitor nodes appear as a single large reliable store called “Reliable and Autonomous Distributed Object Store” (RADOS).

Advantages
1. It provides virtually unlimited storage to file systems.
2. Ceph automatically balances the file system to deliver maximum performance.
3. Ceph is highly reliable, easy to manage, and free.
4. Separates data and metadata operations. Dynamic distributed metadata management [4].

Limitations
1. Ceph dynamically replicates metadata, but this cannot be done with file data.
2. MDS and OSD are implemented in user-space.
3. Less Scalable compared to ZFS, as it’s a 64 bit file system.

Gluster File System

Gluster File system is a network file system which is also highly scalable. It has found its applications in media streaming services, content delivery networks and also in cloud computing. GlusterFS was developed by Gluster Inc. which is purchased by, Inc. in 2012. Red Hat Storage Server was announced as a commercially supported integration of GlusterFS [6] with Red Hat Enterprise Linux. With the acquirement of Inktank Storage (the company behind the Ceph distributed file system) in April 2014, RedHat re-branded GlusterFS-based Red Hat Storage Server to "Red Hat Gluster Storage".

Red Hat defines GlusterFS as a Scalable NAS and object store. Like in Ceph file system, Gluster also uses hashing algorithm for data storage in storage pools. This is key for scalability. The hashing algorithm is exported to all the servers, to allow them to decide where a particular data item should be kept. As a result, data replication becomes easily. The default block size of Gluster is 128KB. Hence, due to this larger size Red Hat was able outperform Ceph by three to one in benchmarking tests. Ceph and Gluster have similar data distribution capabilities.

We know that the default block size for Ceph is 64KB, which is very small and the data stream is fragmented into number of random IO operations. The disk drives (HDD) can perform upto 150 random IOs per second. The default block size for GlusterFS is 128KB. Because of this larger block size, the GlusterFS was able to outperform three times more than the Ceph

in benchmarking tests conducted by the RedHat. Deciding on the transfer size for Ceph may make it slower or faster than GlusterFS.

Architecture
GlusterFS [7] builds a larger parallel network by merging several storage servers over InfiniBand. This is open source in nature with some parts licensed under the GNUGeneral Public License..Figure-4 shows the architecture of GlusterFS which is nothing but a client-server model. As the architecture is based on Client-Server model, here Servers are deployed as storage bricks, with each server running a glusterfsd daemon. The glusterfs client process connects to servers with a custom protocol over TCP/IP, InfiniBand or Sockets Direct Protocol.

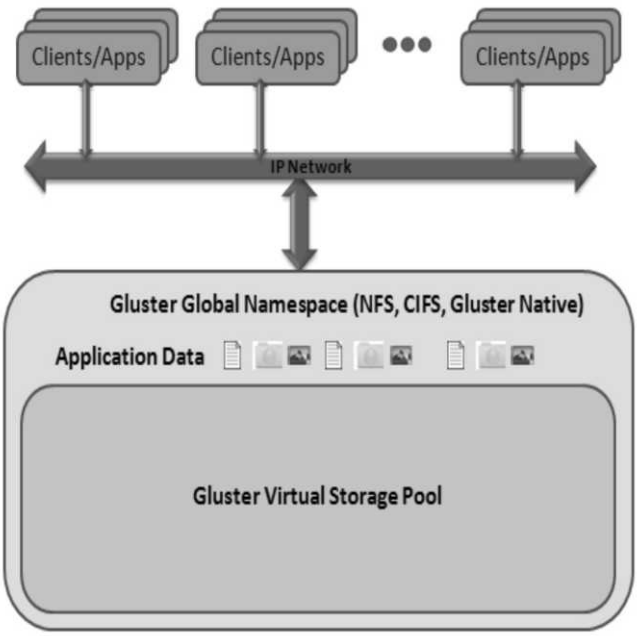


Figure-4: GlusterFS Architecture

Here, files are stored as whole by default, but the files are stripped across multiple remote volumes. After that, client host may mount the final volume by using the native protocol via the FUSE mechanism. In GlusterFS, server exports the directory as it is, so that a client side translator can decide on the structure of the store. The clients are stateless and do not communicate to each other, but keeps the translators configuration consistent with each other.

Advantages
1. Mirroring and Replication file.
2. Striping and Load balancing based on file.
3. Disk caching and Scheduling.
4. Highly Scalable, as it's a 128 bit file system.

Google File System (GFS)

Google File System is a distributed file system, designed and implemented by Google to cater for the rapid growing demands of massive data storage as well as concurrent client access. It is a typical distributed file system with the same goals of performance and scalability. Besides, GFS is regarded as an impactful file system for the sake of its fault tolerance and high performance. Google File System is built from large clusters of much inexpensive commodity hardware and of highly insensitive to failures since its framework leverages the multiple replicas, which is to distribute the replica chunks across the cluster. However, due to the complexity of the network, delay of communication and concurrent user access, the maintenance of data consistency grows sophisticated and unpredictable. As a result, the demands of formalization of GFS, especially the read/write behaviors and the consistency model guaranteed by GFS, are of pretty important and highly urgent. Thus the formal modeling and verification of GFS become the focus of attention in both academic circles and engineering fields.

Architecture

Google File System consists of three main parts: masters, clients and chunk servers [8]. As shown in Fig: 5 [9], a GFS cluster contains a single master and multiple chunk servers which are accessible by the number of clients.

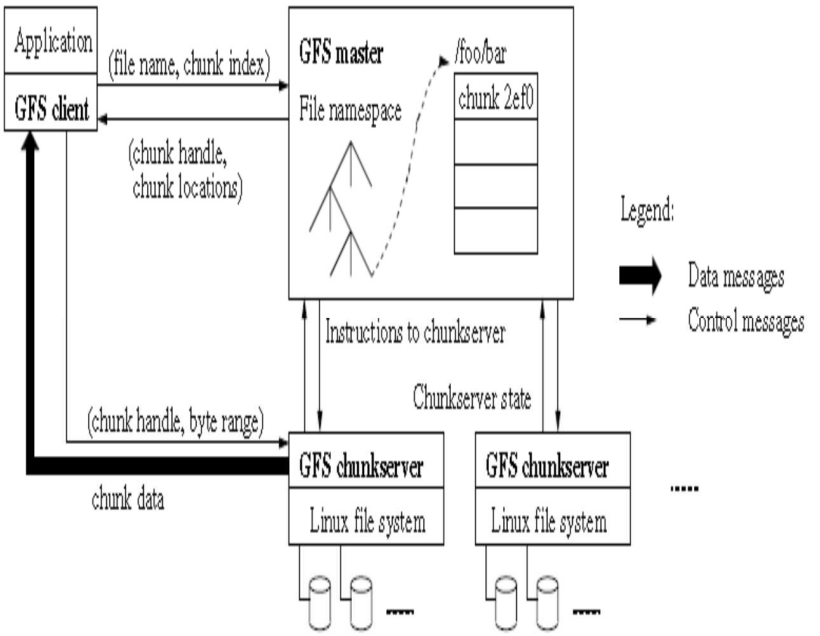


Fig: 5 Google File System (GFS) Architecture

Both a client and chunk server can be run on the same machine as long as the machine resources permit and the lower reliability caused is acceptable. Files are divided into number of fixed size chunks. At the chunk creation itself, master assigns 64-bit immutable handle which uniquely identifies each chunk. Chunk servers store the data on local disks as Linux files and chunks identified by unique handle. For reliability, each chunk is replicated on multiple servers and by default three copies of each chunk are placed [5].

Advantages

1. Efficient for huge files that are mostly appended and read.
2. File data is stored across different chunk servers thus reads come from different chunk servers.
3. Very high availability and fault tolerance through replication.
4. Simple and efficient centralized design with a single master. Delivers good performance for what it was designed for i.e. large sequential reads.
5. Batch operations like writing to operation log, garbage collection help increase the bandwidth.
6. Atomic append operations ensures no synchronization is needed at client end.
7. Decoupling of flow of data from flow of control allows using network efficiently.
8. Orphaned chunks are automatically collected using garbage collection.
9. GFS master constantly monitors each chunk server through heartbeat messages.

Limitations

1. Special purpose design is a limitation when applying to general purpose design.
2. Since a relaxed consistency model is used clients have to perform consistency checks on their own.
3. Performance might degrade if the numbers of writers and random writes are more.
4. Master memory is a limitation.
5. Single Master node

Hadoop Distributed File System (HDFS)

HDFS is much similar to the Google File System but unlike Google, it is open source. It is a very large distributed file system with two key features of high throughput and fault tolerance. It uses Map/Reduce technique [4] to place the data blocks. According to the developers, within a few years, it is likely that a half of the world's data will be stored in the HDFS. Here,

files are stored into the series of blocks and are replicated for fault tolerance. Even though there are many similarities with existing distributed file systems, they are very much different. HDFS is very efficient to the applications that deal with big data sets.

Architecture

As shown in Fig:6, HDFS architecture [10] consists of a single name node as the master node and a number of data nodes as the Slaves. Name node is responsible for access control of files by the potential clients and also to manage the file system namespace. The namenode maps data blocks to data nodes. The data nodes are distributed one per machine in the cluster, which is able to manage the data blocks attached to that machine. Upon the instructions from the namenode, the data nodes are responsible to serve the read/write requests from clients and perform block operations. HDFS also replicates the file blocks for reliability and fault-tolerance and the count replicas depends on the application, which can be changed any time. It also distributes the data chunks across the server for Load balancing, high performance and resiliency.

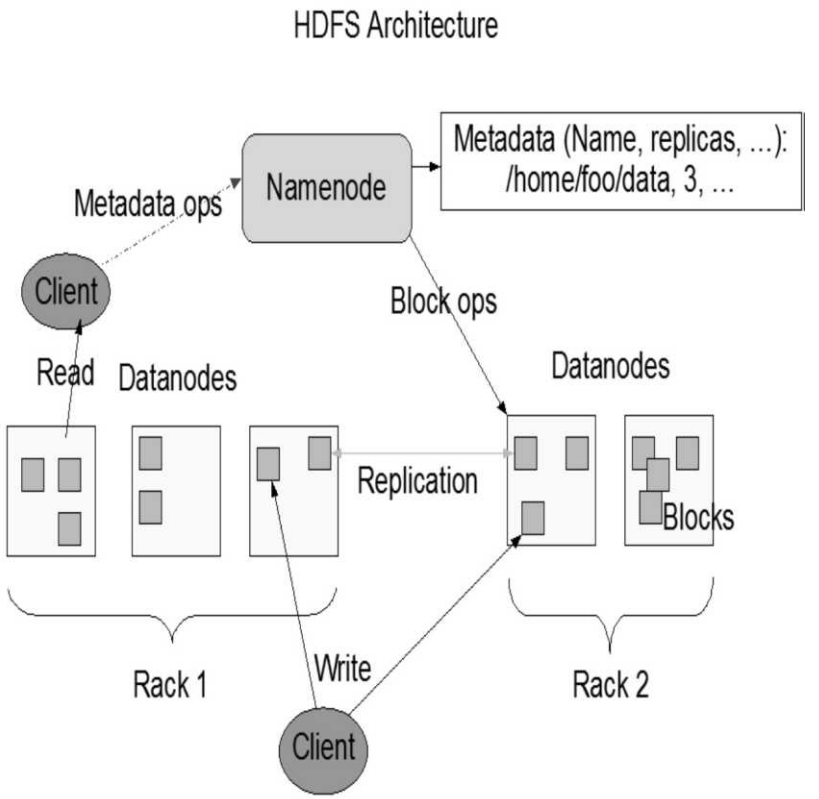


Figure-6: HDFS Architecture

Advantages

1. HDFS is block structured file system. Here, File is broken into number of blocks each with the fixed size and these blocks are stored across a cluster of one or more machines with data storage capacity
2. Distributes data access and computation operations across a cluster of computers.
3. This model is very simple as the end-user programmer itself writes map-reduce tasks.
4. Data will be written once and read several times.
5. Fault tolerance by fault detection and recovery [11].

Limitations

1. When numbers of data sets are joined, may become slow as the entire dataset may be copied in the process.
2. Dependency on single master node.
3. Less Scalable compared to ZFS & GlusterFS, as it's a 64 bit file system.

Comparison of the above file systems

Table 1. Comparison table

S.No	Criteria	ZFS	CEPH	HDFS	GFS	Gluster FS
1	Main feature	Provides cooperative cache and distributed transactions.	Decouples data and metadata operations.	Provides Fault tolerance and high throughput.	Supports fault tolerance & high aggregate performance.	Provide File-based mirroring, load balancing, etc.
2	Processes/ Components	FE, Cache, File Manager, Lease Manager, Transaction Server & Object Store.	Client, Cluster of OSDs and Metadata server cluster	Name node and Data node	Master and chunk server	Client and Server components (deployed as *storage bricks*)
3	Scalability	High (128-bit FS)	Limited (64-bit FS)	Limited (64-bit FS)	Limited (64-bit FS)	High (128-bit FS)
4	Storage Type	Intelligent object Storage devices (OSDs).	Semi-autonomous OSDs running a specialized local object file system.	GFS Clusters.	HDFC Clusters.	Scale-out NAS and object store.
5	Data block size	128 KB	64 KB	64 KB	64 KB	128 KB
6	Open Source?	Yes	Yes	Yes	No	Yes
7	Reliability	High as do not depend on Single node.	High as do not depend on Single node.	May be limited due to dependency on single Master node.	May be limited due to dependency on single Master node.	May be limited due to dependency on single Master node.

Conclusion

In this paper, in brief, we discussed the five major distributed file systems for storing data in the cloud. ZFS is the first decentralized file system that distributed all the functions of file/storage data access over all the communicating devices interconnected by a high-speed network. The ZFS is basically a Disk file system, having two important features: Distributed transactions and Cooperative cache and it fails to support for transparent encryption and disk quota. Hence ZFS can be used for the applications where the transparent encryption and disk quota facility is not mandatory. CEPH File system separates metadata and data operations by using the generating functions unlike other file systems which uses file allocation tables. With this, it is possible to distribute the complexity of data operations such as data access, replication, failure detection and recovery. As CEPH file system is characterized by its excellent Performance, Reliability and Scalability, it is suitable for the application where the Throughput, Reliability and Scalability are the necessary features. Gluster File system is a NAS file system which is also highly scalable. Gluster file system can be used in Cloud Computing, Streaming media services and content delivery networks which involves client-server models. Google file system is the proprietary file system developed by and for Google. HDFS is much similar to the Google File System but unlike Google, it is open source. It is a very large distributed file system with two key features of high throughput and fault tolerance. HDFS can be used in the applications that require the processing of large data sets by using the Map/Reduce approach [4].

References

[1] Ohad Rodeh, Avi Teperman, "ZFS - A Scalable Distributed File System Using Object Disks," 20th IEEE/11th NASA Goddard Conference on Mass Storage Systems and Technologies (MSS'03), IEEE, 2003, pp. 1-7.

[2] Diana Gudu, Marcus Hardt, Achim Streit, "Evaluating the Performance and Scalability of the CEPH Distributed Storage System," proceedings of IEEE International Conference on Big Data, IEEE, 2014, p. 1.

[3] Long, Sage A. Weil Scott A. Brandt Ethan L. Miller Darrell D. E., "CEPH: A Scalable, High-Performance Distributed File System," 7th USENIX Symposium on Operating Systems Design and Implementation, OSDI, 2006, pp. 2-4.

[4] Shweta Pandey, Dr.Vrinda Tokekar, "Prominence of MapReduce in BIG DATA Processing," Fourth International Conference on Communication Systems and Network Technologies, 2014, pp. 1-2.

[5] Mengdi Wang, Bo Li, Yongxin Zhao, Geguang Pu, "Formalizing Google File System," 20th Pacific Rim International Symposium on Dependable Computing, IEEE, 2014, p. 1.
[6] Panda, Ranjit Noronha and Dhabaleswar K., "IMCa: A High Performance Caching Front-end for GlusterFS on InfiniBand," 37th International Conference on Parallel Processing, IEEE Computer Society, 2005, p. 1.
[7] Xie Tao, Liang Alei, "Small File Access Optimization Based on GlusterFS," International Conference on Cloud Computing and Internet of Things, CCTOT, 2014, pp. 1-2.
[8] Bo Li, Mengdi Wang, Yongxin Zhao, Geguang Pu, Huibiao Zhu and Fu Song, "Modeling and Verifying Google File System," 16th International Symposium on High Assurance Systems Engineering, IEEE, p. 2, 2015.
[9] R.Vijayakumari, R.Kirankumar, K.Gangadhara Rao, "Comparative analysis of Google File System and Hadoop," International Journal of Advanced Trends in Computer Science and Engineering, vol. I, no. 1, pp. 553-558, 2014.
[10] Rahul Dutta, Annappa B, "A Scalable Cloud Platform using Matlab Distributed Computing Server Integrated with HDFS," International Symposium on Cloud and Services Computing, 2012, pp. 1-2.
[11] Kala Karun. A, Chitharanjan. K, "A Review on Hadoop – HDFS Infrastructure Extensions," Proceedings of IEEE Conference on Information and Communication Technologies, ICT, 2013, pp. 2-3.
[12] Diana Gudu, Marcus Hardt, Achim Streit, "Evaluating the Performance and Scalability of the Ceph Distributed Storage System", IEEE International Conference on Big Data, IEEE, 2014, p. 1.

Study and Analysis of Modified Aggressive Packet Combining with Consideration of the Physical Signal

Achyuth Sarkar* ,Swarnendu K Chakraborty** and C.T.Bhunia***
*Assistant Professor, Department of CSE, NIT Arunachal Pradesh
achyuthit@gmail.com
**Assistant Professor, Department of CSE, NIT Arunachal Pradesh
swarnendu.chakraborty@gmail.com
*** Director, NIT Arunachal Pradesh
ctbhunia@vsnl.com

Abstract: In a highly error prone wireless transmission medium Aggressive Packet Combining (APC) is a useful and efficient tool for receiving the correct version of the transmitted packet. In APC three copies of each packet is transmitted and the receiver performs bit wise majority voting to obtain the correct copy. However one of the major drawback of this scheme arises from the fact that it fails to correct the error when erroneous bits are present in two or more copies or at the same bit location. To address these limitations of conventional APC we proposed a new modified version of APC by considering the physical signal through which the transmitted copy can be more efficiently and coherently received by the receiver. Study and analysis made in this paper clearly reveals that the proposed scheme is indeed superior to that of conventional APC.

Keywords: Aggressive Packet Combining, Correction capability,Third bits left/right shift, circular Left shift, MSB,LSB,Physical level.

Introduction

It has been indeed a challenge to the researchers and scientists for many years to reliably and faithfully transmitdata or information from a source to a destination. Backward Error Correction (BEC) and Forward Error Correction (FEC) are the two methods that are extensively used in this regard. BEC has found its application in wired transmission and FEC in wireless transmission [1-2]. BEC technique is found to be cost effective and due to this many researchers are trying to implement BEC for wireless transmission of data. In the traditional error correction technique, an erroneous packet is dropped. But the erroneous packet will have both correct and erroneous information. Chakraborty[3] proposed PC scheme, a simple technique to explore the information present in erroneous packet. In PC scheme bit errors are located by XORing two copies of the packet. Leung [4] proposed APC. APC is a low latency error correction scheme, so it is important for wireless data network. APC is well established and well studied elsewhere performance, so it is important for wireless data network. In APC transmitter sends three copies of the packet. The receiver applies bit wise majority logic to recover a correct copy of the packet. It is well studied that APC offers better performance. But APC suffers from many flaws. The major limitation of APC is its inability to correct the errors when errors are present at same bit location of two or more copies.Receiver applies bit wise majority logic to recover a correct copy of the packet. It is well studied that APC offers better performance, so it is important for wireless data network. APC is well established and well studied elsewhere [4-8]. In APC transmitter sends three copies of the packet. The receiver applies bit wise majority logic to recover a correct copy of the packet. It is well studied that APC offers better performance. But APC suffers from many flaws. The major limitation of APC is its inability to correct errors when errors are present in same bit location in two or more copies [9-11]. The review of Packet Combining scheme is given in section II and Aggressive Packet Combining scheme in section III. The proposed modified scheme is illustrated in section IV. Finally section V concludes the paper.

Review of Packet Combining (PC) Scheme

Chakrabotry [3] suggested a very simple and elegant technique known as packet combining scheme (PC) where theReceiver will correct original data stream.

As per Chakraborty's proposal:

Let the transmiited data packet is 11100011

1st Copy: 1**0**100011 which is in error. Therefore the receiver will ask for retransmission of the packet from the transmitter and stores this erroneous copy in the buffer. The transmitter sends the second copy as
2nd Copy: **0**1100011 which is again in error.
The receiver applies bitwise XORing on the received erroneous copies to detect the error positions.

1st Copy: 1**0**100011
2nd Copy: **0**1100011
XORing: 11000000

The error positions are thus identified as 1st and 2nd bit from left. The receiver applies brute force method by bit inversion to rectify the errors and retrieve the correct packet.

Review of Aggressive Packet Combining (APC) Scheme

Aggressive Packet Combining (APC) is a modified form of Packet Combining scheme[4]. Here three copies of a packet are sent via same transmission link to the receiver. At the receiver side all the three copies are received erroneously .Receiver then applies majority logic bit by bit on the received three erroneous copies.

For Example:

Original Packet: 11000001

First Copy: 1**0**000001
Second Copy: **0**1000001
Third Copy: 110**1**0001
Majority Logic: 11000001 which is the original packet.

In case where the generated copy after applying majority logic is not correct, the receiver searches for the least reliable bits and after selection of the least reliable bits the receiver applies brute force correction to that bits followed by error detection. In case of failure, it request retransmission request to the transmitter in which the sender sends three copies for transmission.

Proposed Scheme of Modified APC

Protocol I

Suppose original packet: "11101000 ".As per proposed new protocol 1st and 3rd copy will be send as it is sender as receiver but second copy will be send as "00010111" also we are assuming that error places will be (-) error from MSB in 1st and 2nd copy and positive error at 3rd place from MSB in 3rd copy in the environment of non-repeated error syndrome.

First copy: 11001000
Second Copy 11101000
Third Copy 11101000

_____
11101000

Applying majority logic: 11101000 original copy of the packet that we have sent at sender side.
On the other hand 1st copy 10101011 and 2nd copy 01010111 and 3rd copy 10101011 here 1st and 3rd copy as it is transmitted and 2copy 1bit left shift:
So it is the part of sender copy

1st copy 10101011
2nd copy 01011011
3rd Copy 10101011

_____
10101011

The receiving side receive easily able to collect the sender copy :

The Receiving side will be

1st copy	10101011
2nd copy	10101011
3rd copy	10101011
	10101011

On the other hand with error

1st copy	10100011
2nd copy	01110111
3rd copy	10100011
	10100011

Receiver of this receive the sender information

1st copy	10100011
2nd copy	10111011
3rd copy	10100011
	10100011

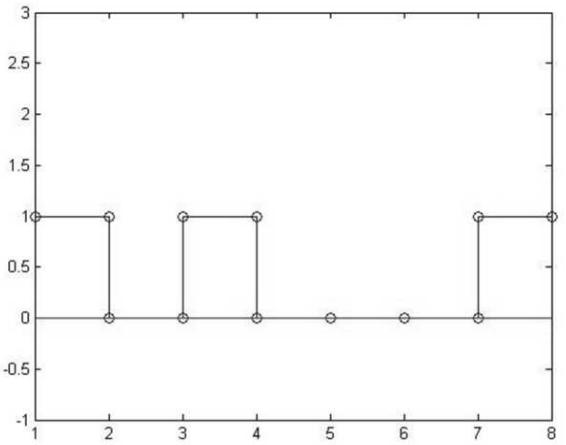


Fig 1. Physical level signal of Protocol I

Protocol II

Another example we like to describe in the environment of repeated error syndrome, suppose original data stream is 11011100.

First copy:	11101000
Second Copy	00010111
Third Copy	11101000
	11101000

Sender side 11101000 bit

Receiver side receive the information:-

First copy: 11001000

Second Copy 11101000
Third Copy 11101000

11101000

Without error of this information.

The Receiving side will be
1st copy 11101000
2nd copy 00010111
3rd copy 11101000

11101000

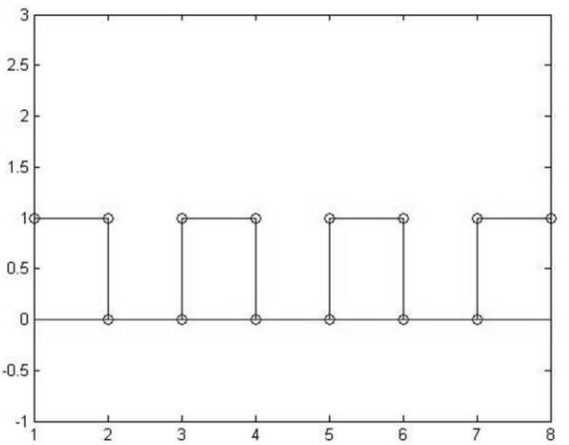


Fig 2. Physical level signal of Protocol II

Conclusion

As per the suggested protocols the correction capability and reliability of APC is improved by a considerable extent. Through these protocols the corrected copies of the packets can be recovered at the receiver with ease. Therefore the proposed scheme can be considered fruitful and beneficial in data transmission over wireless networks.

References

[1] H.Liu, H.Ma, M.E. Zarki, and S. Gupta, "Error control schemes for networks: an overview" mobile networks and applications, vol. 2, pp.167-182, 1997.
[2] O. Yuen, "Design tradeoffs in cellularl PCS systems" IEEE Commun. Mag., vo1.34, no.9, pp. 146-IS2, Sept. 1996.
[3] Shyam S. Chakraborty et aI, "An ARQ Scheme with Packet Combining," IEEE Comm Letters, vol. 2, No. 7, pp. 200-202, July'95.
[4] Yiu-Wing LEUNG, "Aggressive Packet Combining for Error Control in Wireless Networks," trans. Comm Vol. E83, No. 2, pp38-38S, Feb'2000.
[5] C T Bhunia, IT, Network & Internet, New Age International Publishers, India, 2005.
[6] C T Bhunia, "Modified Aggressive Packet Combining Scheme," Pre-print, ICTP, Italy, IC/20I0/037, pp. 1-10, June 2010 .
[7] ARKSastry,ImprovingAutomatic Repeat Request(ARQ)Performance onSatellite Channels UnderHighErrorRateConditions, IEEETransComm,April'77,pp436-439.
[8] Swarnendu K Chakraborty, Raj at S Goswami, Abhinandan Bhunia, Chandan T Bhunia, "Two new modified schemes of Aggressive Packet Combining Schemes in Achieving Better Throughput," 10th International Conference on Information Technology, IEEE, 2013.
[9] Ranita Khumukcham, Abhinandan Goswami, Yang Saring, " Four new protocols for achieving better correction capability of APC scheme", International Conference on Communication and Signal Processing, ICCSP-2015, Melmaruvathur, Tamil Nadu, India.
[10] Ranita Khumukcham, Abhinandan Goswami, Yang Saring, " Combined APC-PC scheme for random and time varying channels", International Conference on Communication and Signal Processing, ICCSP-2015, Melmaruvathur, Tamil Nadu, India.
[11] Abhinandan Goswami, Yang Saring, C.T.Bhunia ,: "Aggressive Packet Combining Scheme In Multipath Routing To Achieve Higher Throughput And Error Correction Rates", *International Journal of Electrical, Electronics and Data Communication (IJEEDC)*, Volume-4,Issue-1,pp 26-30 ,2016.

Thyroid Malignancy Detection Using K-Nearest Neighbor (KNN) and Probabilistic Neural Network (PNN)

Wrushali Mendre *, Ranjana Raut ** and Pradeep T. Karule***
*Electronics Department, YCCE, Wanadongri, Nagpur, India
vrushalimendre@gmail.com
**Electronics Department, S.G.B. Amravati University, Amravati, India
rdr24164@rediffmail.com
***Electronics Department, YCCE, Wanadongri, Nagpur, India
ptkarule@gmail.com

Abstract: The incidence of thyroid nodule malignancy is increasing. Fine Needle Aspiration Biopsy (FNAB) is used to confirm the nodule malignancy. There is a need to improve the ability of predicting thyroid malignancy prior to FNAB when performing ultrasound examination**.** This paper presents a computer- based scheme for identification of malignant thyroid nodule using textural features extracted from UltraSonoGraphy [USG] image of thyroid nodule. The study comprised 85 ultrasound patient's images that are cytological confirmed using FNAB. Thyroid nodule region is sampled from thyroid gland parenchyma to get sub-images. A set of 7 Discrete Wavelet Transform [DWT] based features has been generated. K-Nearest Neighbor [KNN] and Probabilistic Neural Network [PNN] are the two algorithms used for the differential classification of thyroid nodule as either malignant or benign. Receiver Operating Characteristic [ROC] analysis is done to compare the performances of both classifiers. The results show that, Area Under the ROC Curve [AUC] for the KNN is 0.995 whereas for PNN, it is 0.76. It is observed that KNN gives better accuracy than PNN for the given feature set.

Keywords: K-Nearest Neighbor [KNN], Probabilistic Neural Network [PNN], Thyroid Nodule Malignancy, UltraSonoGraphy [USG].

Introduction

Thyroid nodules are very common and are observed in USG of 50% of the adult population [1]. It is a discrete lesion within the thyroid gland that is sonographically distinguishable from the adjacent parenchyma. Thyroid lesion can be a malignant one in only small population (<7%). The American Cancer Society estimates for thyroid cancer in the United States for 2015 are about 62,450 new cases of thyroid cancer (47,230 in women, and 15,220 in men) and about 1,950 deaths from thyroid cancer (1,080 women and 870 men) [2,3].

Fine Needle Aspiration Biopsy [FNAB] is the gold standard for detecting malignancy in thyroid nodule [4]. The initial objective of the tests is to detect whether any nodules are present in thyroid. The next task is to check whether the detected nodule is malignant or not. Former studies have adopted FNAB as a dominant method for accurate differentiation of thyroid nodule malignancy. However it is an invasive and labor intensive test. In FNAB, only a small percent of the total thyroid nodules have the probability to be malignant. In other words, only 9 to 13% of the thyroid nodules tested using FNAB are labeled as malignant and the indeterminate results are obtained because of the low cellularity, small size and cystic nature of nodules or due to inexperience of the operator performing the FNA test.

The USG is one of the most widely used imaging technologies [5] in medicine. It is portable, free of radiation risk, and relatively inexpensive when compared with other imaging modalities, such as magnetic resonance and computed tomography. Ability of ultrasound to clearly discriminate malignant from benign nodule has been questioned by former studies [6-8]. However, the evolution of imaging technology in past years, has established USG as the major imaging modality in evaluating thyroid nodules [9]. Medical researchers did rigorous study for finding USG features which are responsible for malignancy. These features include microcalcifications, local invasion, lymph node metastases, nodule that is taller than it is wide and markedly reduced echogenicity [10]. Presence of microcalcifications is very reliable criteria on which to base increased suspicion of malignancy in thyroid nodule [11]. Microcalcifications and coarse calcifications in a predominantly solid nodule are associated with an approximately threefold and two fold increase in cancer risk, as compared with predominantly solid nodule without calcifications [12].

Accurate visual interpretation of the class of thyroid nodule images can only be done by ultrasonographers with lot of experience and training. Otherwise, it results in subjective interpretations and inter-observer variability. Such limitations have led to extensive research in developing automated efficient ultrasound image analysis techniques called Computer Aided Diagnostic [CAD] systems so as to obtain accurate, reproducible and more objective diagnosis results.

Tsantis et al. [9] generated several morphological and wavelet local maxima features, and classified the thyroid nodule as either malignant or benign nodule with 96% accuracy of classification. Two classifiers, Support Vector Machines [SVM], with number of support vector ranging from 8% to 10% of the number of training points, which is indicative of comparatively higher complexity and Probabilistic Neural Network [PNN] are used to evaluate malignancy. Microcalcifications and coarse calcifications which are two major features responsible for thyroid nodule malignancy are not considered while extracting the features of nodular thyroid region. *Keramidas et al.* [13] presented thyroid nodule tissue identification scheme based on fuzzy logic and Local Binary Patterns [FLBP] and Fuzzy Grey Level Histogram [FLGH] and could only differentiate between normal thyroid tissue and thyroid nodule tissue. Nodule malignancy is not considered in classification. Nikita Singh et al. [14] proposed segmentation and classification method for classification of thyroid nodule through USG images. Zhao et al. [15] proposed segmentation method to assist FNAB.

In all the above mentioned algorithms, there is a scope for experimentation to derive the features that suggest potential thyroid malignancy. Presence of intrinsic microcalcifications and coarse calcifications in solid and predominantly solid nodule increases the risk of malignancy. As far as correct diagnosis through USG image is concerned, accuracy of classification is also a very important parameter which needs to be addressed. Enhancement of the functionality of proposed algorithms is needed to improve the classification accuracy by considering the characteristics of malignant thyroid nodule.

The proposed research addresses the malignancy detection of thyroid nodule by considering the USG characteristics of malignant thyroid nodule with improvement in accuracy and reduction in complexity. This paper is organized as follows. Material and Methods section introduces the main components of the proposed scheme. Later, the experimental results obtained by the proposed method are discussed and finally the conclusions are presented.

Material and Methods

The proposed diagnosis system for thyroid malignancy detection consists of four components namely: pre-processing, Region of Interest [ROI] separation, feature extraction and classification. The first step involves the removal of speckle noise which is inherent in USG images. The second step involves the manual separation of ROI that includes only nodular region. The third step includes DWT based feature extraction method. In the forth step, classification algorithm is implemented to categorize the nodule as either malignant or benign. The classification pipeline used in proposed research is shown in Fig. 1.

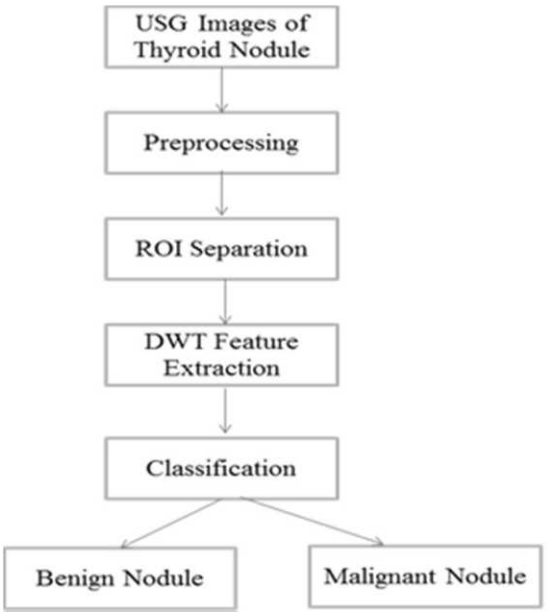


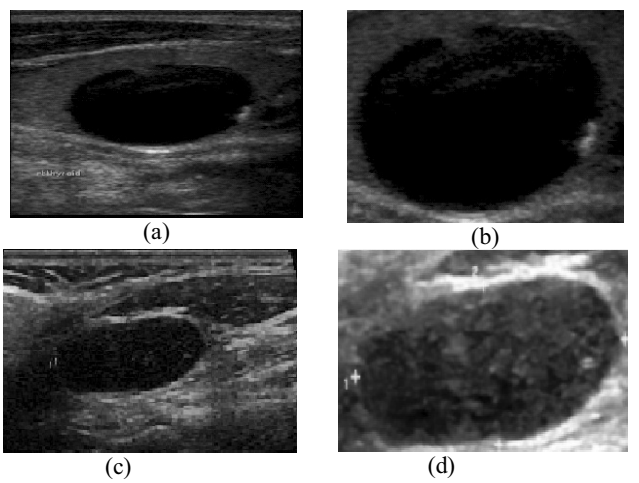
Figure 1.Classification Framework used in proposed scheme

Thyroid nodule USG images are taken from Wilmington Endocrinology PA website [16]. The images are in JPEG format with 540x410 size, 24bit depth, and true color image. This standard database comprising of 85 ultrasound images of thyroid nodule (55 benign and 30 malignant) are considered to examine the efficacy of the proposed algorithm.

Speckle Reducing Anisotropic Diffusion [SRAD], filter is used to remove the speckle noise. The performance of SRAD filter is compared with frost filter and spatiotemporal filter. It is observed that SRAD filter performs best for thyroid nodule images [17]. After preprocessing, ROI is cropped manually. Fig. 2 shows ROI separation of benign and malignant thyroid

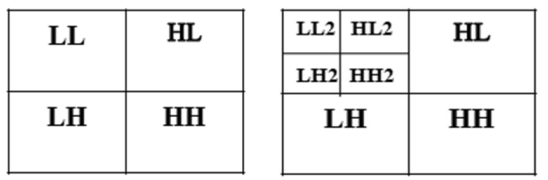
nodule. White dots can be observed in ROI of malignant nodule (Fig. 2 (d)). These (white dots) shows the microcalcifications.

DWT is a well known textural feature extraction technique. In this method, the images are sent through the series of down sampling filters which are composed up of a sequence of low pass and high pass filters. This process will decompose the input thyroid nodule image into a series of sub band images. The four sub band images consist of one approximation image and three detail images. The approximation image is actually the low frequency. Detail image contains the information of specific scale and orientation. This means that the spatial information is also retained within the sub band images. Therefore, the detail images are suitable to be used for deriving a set of wavelet features in the input image. On the other hand, the approximation image can be used for higher levels of decomposition for the input image. Down sampling operation has helped to reduce the useless and redundant samples in the decomposition process [18].



(a) (b) (c) (d)

Figure 2. Thyroid nodule image (a) benign cystic thyroid nodule (b) separation of nodular region (ROI) of benign nodule (c) malignant solid thyroid nodule (d) ROI separation of malignant thyroid nodular region



(a) (b)

Figure 3. 2D sub-band DWT (a) First Level of Decomposition (b) Second Level Decomposition

Fig. 3(a) shows the first level of DWT decomposition. Decomposition is further performed on the LL coefficient to get coarser scale of wavelet coefficients. Decomposition is stopped at second level because further decomposition didn't give significant results. Fig3 (b) shows second level of DWT decomposition. The Normalized energies A2 , V2, H2, D2, H1, V1, and D1 are obtained from matrixes LL2, HL2, LH2, HH2, LH, HL and HH respectively. Equations 1 to 7 show the formation of normalized energies from individual sub-band matrices. All the elements into the individual rows of the matrix are added and the resultant vector elements are squared and then added to form a scalar. This scalar is finally normalized by dividing it by number of rows and columns of original matrix.

$$A_2 = \frac{1}{m*n} \sum_{i=1}^{m} \left[\sum_{j=1}^{n} LL2_{ij}\right]^2 \quad (1)$$

$$V_2 = \frac{1}{m*n} \sum_{i=1}^{m} \left[\sum_{j=1}^{n} HL2_{ij}\right]^2 \quad (2)$$

$$H_2 = \frac{1}{m*n} \sum_{i=1}^{m} \left[\sum_{j=1}^{n} LH2_{ij}\right]^2 \quad (3)$$

$$D_2 = \frac{1}{m*n} \sum_{i=1}^{m} \left[\sum_{j=1}^{n} HH2_{ij}\right]^2 \quad (4)$$

$$H_1 = \frac{1}{m*n} \sum_{i=1}^{m} \left[\sum_{j=1}^{n} LH_{ij}\right]^2 \quad (5)$$

$$V_1 = \frac{1}{m*n} \sum_{i=1}^{m} \left[\sum_{j=1}^{n} HL_{ij}\right]^2 \quad (6)$$

$$D_1 = \frac{1}{m*n} \sum_{i=1}^{m} \left[\sum_{j=1}^{n} HH_{ij}\right]^2 \quad (7)$$

The feature vectors extracted from Thyroid nodule ROI are subsequently classified into a predefined set of classes. The output of, the classification phase is a class label representing either malignant or benign nodule. For the classification phase of the proposed system, two widely used classifiers have been evaluated, K-NN and PNN.

K-NN is a simple and memory based classification method in which K- nearest neighbors is used in determining the class [19]. K-NN is easy to train network with effective classification approach. It allows easy and fast incorporation of new data into an existing trained system. It requires large memory size and heavy computational load for large datasets and multidimensional feature spaces. k−NN classification has two stages; the first is the determination of the nearest neighbors and the second is the determination of the class using those neighbors. The distance between the two data points using some distance function *d (x, y)* is calculated where *x* and *y* are the data points composed of *n* features, such that x=$\{x_1, \ldots, x_n\}$ and y=$\{y_1, \ldots, y_n\}$. In the proposed K-NN model, Euclidian Distance is measured as follow:

$$d_E(x, y) = \sum_{i=1}^{n} \sqrt{{x_i}^2 - {y_i}^2} \quad (8)$$

Probabilistic neural network [PNN] belongs to the radial basis function neural networks [20]. It is based on the theory of Bayesian classification and the estimation of probability density function. It is necessary to classify the input vector into one of the two classes in Bayesian optimal manner. This theory allows for a cost function to represent the fact that it may be worse to misclassify a vector that is actually a member of benign nodule class than it is to misclassify a vector that belongs to malignant class. The Bayes rule such that the input vector belonging to benign thyroid nodule class is classified as in equation (9).

$$P_A C_A f_A(x) > P_B C_B f_B(x) \quad (9)$$

Where P_A is the priori probability of occurrence of patterns in benign thyroid nodule class, C_A is the cost associated with classifying vectors and $f_A(x)$ is the probability density function of benign thyroid nodule class. Bayes decision rule is applied mainly to estimate the Probability Density Function [PDF]. The PDF should be positive all over, should be integrable and the integral over all *x* must be 1. Equation (10) is used by the PNN neural network to estimate the PDF.

$$f_A(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} - \frac{1}{m_n} \sum_{i=1}^{m_A} exp\left[-2\frac{(x-x_A)^T (x-x_{Ai})}{\sigma^2}\right] \quad (10)$$

Where x_{Ai} is the i^{th} training pattern from benign class, *n* is the dimension of input vectors, m_A is the number of training patterns in benign class of thyroid nodule and σ is the smoothing parameter which corresponds to standard deviation of Gaussian distribution [21].

The PNN architecture for a two-class classification problem comprises 4 layers. The input layer that has a node for each feature of input data. The pattern layer in which, every pattern node corresponds to a specific training pattern. In this layer, at first the squared Euclidean distances from the input vector to the training vectors are computed. These distances are then transformed by the neuron's activation function (the exponential) in order to estimate how close this input is, to the training vector. The summation layer that comprises a summation neuron for each class in which, the outputs from the pattern nodes are summed [20]. The activation of the summation neuron for a class is equivalent to the estimated density function value of this class. The final layer is the output layer in which, the summation neurons feed their result to the output neuron. This neuron can be considered as a threshold discriminator that implements the Bayes' decision criterion. The output neuron generates two outputs: the estimated conditional probability that the test pattern x belongs to the first class, and the estimated conditional probability that the pattern belongs to the other class. The test pattern *x* is categorized to the class with the maximum probability [21].

The classification accuracy of both classifiers is evaluated by means of Receiver Operating Characteristics [ROC] curves analysis. ROC curve is a plot of the true positive rate (sensitivity) versus the false positive rate (specificity) for different thresholds over the entire range of each classifier discriminant function output value. In contrast with the classification accuracies obtained from truth tables, ROC analysis is independent of class distribution or error cost [22]. The Area under ROC Curve [AUC] can be statistically interpreted as the probability of the classifier to correctly classify malignant cases from benign cases. An additional evaluation parameter used in this study is the likelihood ratio that measures the power of each classifier for increasing certainty about a positive diagnosis [22]. Likelihood ratio can be calculated using equation (11).

$$Likelihood\ Ratio = \frac{Sensitivity}{1-Specificity} \quad (11)$$

Thyroid nodule image preprocessing, feature extraction with the design, training and testing of both classifiers are implemented in MATLAB 2010. The ROC analysis is also done in MATLAB 2010. The computer used for processing has an Intel (R), Core (TM), i3-2330M CPU running at 2.20GHz and 2 GB of RAM.

Results

The proposed system provides the quantitative analysis of the usage of DWT for feature extraction and comparative analysis of KNN and PNN for classification of thyroid nodule. All the levels of sub-bands of DWT are used for feature extraction. Feature selection is not done because it is quite evident that all sub- bands of wavelet transformation technique are important as information missing in one sub-band is always found in another sub-band. Table 1 shows the sample features for both the classes of thyroid nodule USG for DWT.

Table 1.Significant features extracted (mean±SD) using the DWT

Features	Benign	Malignant
A2	3.6E+06±1.9E+06	2.1E+06±2.19E06
V2	1.39E+04±2.0E+04	8.31E+03±6.49E03
H2	2.26E+02±2.93E+02	3.05E+02±2.79E+02
D2	4.6E+01±6.3E+01	1.69E+02±1.81+02
V1	4.52E+03±6.45E+03	2.42E+03±2.0E+03
H1	3.98E+01±5.53E+01	6.15E+01±5.51E+01
D1	6.48E+00±1.39E+00	8.57E+00±1.86E+00

Two pattern recognition models (KNN and PNN) have been developed and utilized for accessing the thyroid nodule malignancy. Table 2 reports the results of ROC analysis for both the classifiers. AUC represents the probability that a random pair of benign and malignant thyroid nodules will be correctly classified. Sensitivity depends on the detections of malignant nodules and specificity on benign nodules.

In KNN, a sample data point is assigned a class which is the most common among its K-nearest neighbors and K is the small positive number. These set of neighbors are the training samples for this classifiers for which correct classification is known. For finding the optimal value of K for which accuracy of classification will be highest, K values are varied from 1 to 5, and corresponding accuracies are analyzed. Fig. 4 shows the performance of K-NN for variation in K values. For the given feature set, AUC is highest for K=1 with 0.995 sensitivity and 0.989 specificity. The likelihood ratio is 13.9. Fig. 5 shows the ROC curve with K value of 1.

Regarding the PNN model, dataset of 85 USG images are partitioned as follows: 70% images for training and 15% images for testing and 15% images are used for cross validation. Fig. 6 displays the overall performance of PNN in terms of Mean Square Error [MSE]. It is observed that overall MSE is 0.4321 at epoch 1. Highest classification accuracy obtained is 64.1026% with AUC of 0.76. The sensitivity and specificity are 0.52 and 0.62 respectively with Likelihood ratio of 1.4.

Table 2.ROC analysis results for the KNN and PNN classifier

Model	AUC	Sensitivity	Specificity	Likelihood Ratio
K-NN with K=1	0.995	0.989	0.9286	13.9
K-NN with K=2	0.9945	0.989	0.922	12.7
K-NN with K=3	0.84	0.68	0.7143	2.4
K-NN with K=4	0.9	0.8	0.9286	11.2
K-NN with K=5	0.88	0.76	0.8571	5.3
PNN	0.76	0.52	0.62	1.4

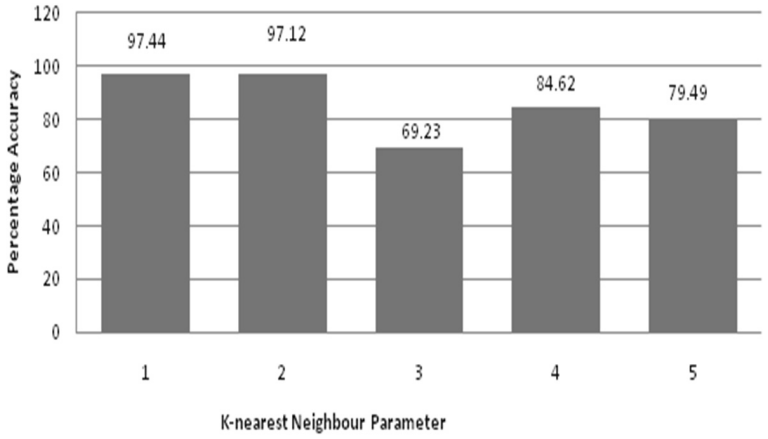


Figure 4. Accuracy variations for corresponding variations in K- values

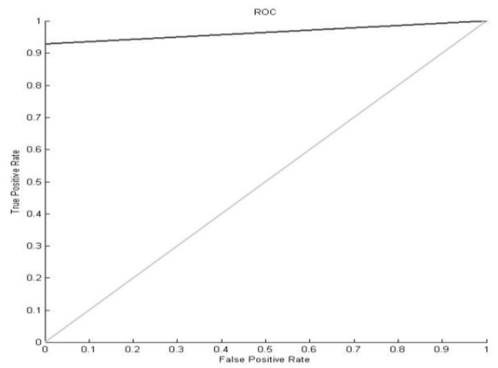


Figure 5. ROC curve for K-NN classifier with K=1

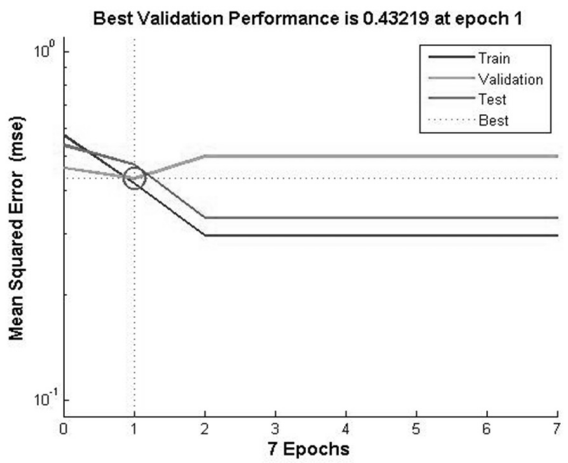


Figure 6. Overall Mean Square Error for training, testing and cross validation dataset

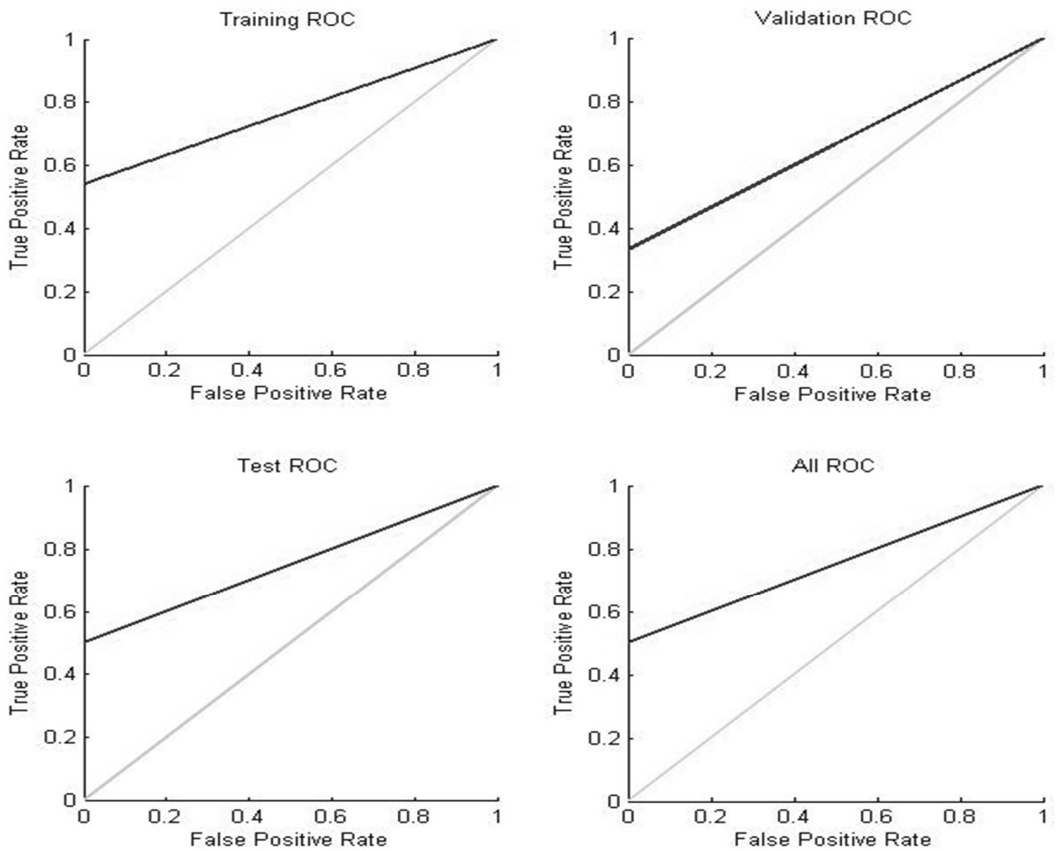


Figure 7 Train, Test, Validation and overall ROC for PNN

Fig. 7 shows the training, Test, validation and overall ROC for PNN classifier. If the performances of both classifiers are compared, then it is observed that K-NN performs best for the proposed scheme.

Discussion and Conclusion

Calcifications are small flecks of calcium within a thyroid nodule, usually seen as bright spots on USG image. These white spots, containing specific orientation, cover the whole nodule region. The presences of microcalcifications and coarse calcifications have already been reported with relatively high precision as suggestive of thyroid malignancy [10,11, 12]. In this study, the textural features that encode the information about microcalcifications and coarse calcifications have been generated using DWT. Derived DWT features extract the structure orientation, which can be in horizontal, vertical or diagonal directions. The sub-band energies (normalized energies contained in A2, V2, H2, D2, V1, H1 and D1) obtained after performing DWT contains the information about microcalcifications and coarse calcifications.

Both KNN and PNN classifiers have exploited the importance of microcalcifications and coarse calcifications towards an accurate differentiation between benign and malignant thyroid nodule. Therefore features derived in the proposed approach imbibe the expert knowledge in the system thus ensuring superior classification accuracy. The designed intelligent system thus aids the decision making of the radiologist about malignancy. Malignancy characteristics used by the radiologists are derived using the proposed system.

A comprehensive study has been made that aimed at the generation of several texture based wavelet features together with the design of two powerful classifiers (KNN and PNN) in order to evaluate the malignancy in thyroid nodule USG. These features are well defined and can be integrated in overall diagnostic procedure. Results show that the features extracted from DWT coupled with K-NN classifier with K=1, presented a classification accuracy of 97.44% with sensitivity and specificity of 0.989 and 0.9286respectively.

It is worth noting that the algorithm, presented in [13] is limited to the detection of nodule tissue from thyroid region, whereas the detection of malignancy is very important issue and is not addressed in their study. Methodologies reported in [9] used SVM for classification but the number of support vectors ranges from 8% to 10% of the number of training points. It is

indicative of the SVM complexity and less differentiation capability. The proposed method reports the accuracy of 97.44% using KNN. KNN algorithm is among the simplest of all machine learning algorithms. It must be emphasized that the SVM classifier performance reported in literature [9] is dependent on kernel function selection and number of support vectors which may increase the model complexity. It must also be emphasized, that the reported methods are all tested on independent data sets. A direct comparison may not necessarily justify their effectiveness and efficiency but a general idea of the performance might be obtained. So the summarized results are merely indicative and prove the efficacy of the approach for its clinical deployment. Our further research aims to deploy the designed system in clinics for aiding the decision making of a radiologist when applied to practical cases.

Acknowledgment

The authors would like to thank department of radiology, Smt. Kashibai Navale Medical College and Hospital, Pune for support and guidance.

References

[1] Hoang, Jenny K., et al. "US Features of Thyroid Malignancy: Pearls and Pitfalls 1." Radiographics 27.3 (2007): 847-860.
[2] NCI (National Cancer Institute) on thyroid cancer. Information available at http://seer.cancer.gov/statfacts/html/thyro.html (last accessed January 2016)
[3] ACS (American Cancer Society). Information available at http://www.cancer.org/cancer/thyroidcancer/detailedguide/thyroid-cancer-key-statistics (last accessed January 2016).
[4] Amedee, Ronald G., and Nina R. Dhurandhar. "Fine Needle Aspiration Biopsy." The Laryngoscope 111.9 (2001): 1551-1557.
[5] Bushberg, J.T., The Essential Physics of Medical Imaging, Lippincott Williams & Wilkins, 2002.
[6] Watters DA, Ahuja AT, Evans RM, ChickW, KingWW,Metreweli C, et al. Role of ultrasound in the management of thyroid nodules. American Journal of Surgery 1992;164:654–7.
[7] Gimondo P, Mirk P, Messina G, Pizzi G, Tomei A. The role of ultrasonography in thyroid disease. Minerva Medica 1993;84:671–80.
[8] Stanley F. AACE clinical practice guidelines for the diagnosis and management of thyroid nodules. Endocrine Practice 1996;2:78–84.
[9] Tsantis, Stavros, et al. "Morphological and wavelet features towards sonographic thyroid nodules evaluation." Computerized Medical Imaging and Graphics 33.2 (2009): 91-99.
[10] Hoang, Jenny K., et al. "US Features of Thyroid Malignancy: Pearls and Pitfalls 1." Radiographics 27.3 (2007): 847-860.
[11] Iannuccilli, Jason D., John J. Cronan, and Jack M. Monchik. "Risk for malignancy of thyroid nodules as assessed by sonographic criteria the need for biopsy." Journal of Ultrasound in Medicine 23.11 (2004): 1455-1464.
[12] Frates, Mary C., et al. "Management of Thyroid Nodules Detected at US: Society of Radiologists in Ultrasound Consensus Conference Statement 1."Radiology 237.3 (2005): 794-800.
[13] Keramidas, Eystratios G., Dimitris Maroulis, and Dimitris K. Iakovidis. "TND: a thyroid nodule detection system for analysis of ultrasound images and videos." Journal of medical systems 36.3 (2012): 1271-1281.
[14] Nikita Singh and Alka Jindal "A Segmentation and Comparison of Classification Method for Thyroid Ultrasound Images." International Journal of Computer Applications (0975-8887) Volume 50-No 11,.July 2011
[15] Zhao J, Zheng W, Zhang L, Tian H. Segmentation of Ultrasound Images of Thyroid Nodule for Assisting Fine Needle Aspiration Cytology. Health Information Science and Systems; 2012. 1:5.
[16] www.wilmingtonendo.com
[17] Mendre, Wrushali, R. D. Raut, and Sayali Malwadkar. "Performance Comparison of Speckle Reducing Filters for Ultrasonography USG) Images of Thyroid Glands." International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE) 4 (2015): 33.
[18] Van de Wouwer, Gert, Paul Scheunders, and Dirk Van Dyck. "Statistical texture characterization from discrete wavelet representations." Image Processing, IEEE Transactions on 8.4 (1999): 592-598.
[19] Cunningham, Padraig, and Sarah Jane Delany. "K-Nearest neighbour classifiers." Multiple Classifier Systems (2007): 1-17.
[20] Specht D. Probabilistic neural networks, Neural Networks 1990; 3:109-18.
[21] Sivanandam, S. N., and S. N. Deepa. Introduction to neural networks using Matlab 6.0. Tata McGraw-Hill Education, 2006.
[22] Lasko TA, Bhagwat JG, Zou KH, Ohno-Machado L. The use of receiver operating characteristic curves in biomedical informatics. Journal of Biomedical Informatics 2005;38:404-15.

Signature Imprintment on VLSI Core Designs

Saurabh Radheshyam Chiraniya*, Pradeep Patil** and A. B. Patki***
* ** *** E&TC Dept., College of Engineering, Pune
schiraniya@gmail.com
**pradeeppatil7748@gmail.com
***abpatki.extc@coep.ac.in

Abstract: This paper is to explore new application for the word watermark to go beyond just images and videos, but reach till complex VLSI design. Aiming to represent the identity of creator or claimer and to ensure the authenticity of the Intellectual Property Core (IP Core) and avoid the misuse of the IP core.

Keywords: VLSI, FPGA, Watermark, IP Core, ASIC, IC design.

Introduction

Reuse based IP (Intellectual Property) core design is one the most promising ways to accelerate the design of Soc products by creating the libraries of IPs of basic and frequently used modules.
There are two types of IP Cores:-

1) Soft IP Core
2) Hard IP Core

Soft IP Core are in Software forms which are yet to be manufactured but ready to be manufactured if needed. Soft IP Cores can be in the form of HDL Codes or Design form.
Hard IP Cores are in the Chip form which are realized on silicon real estate with all of its structure frozen on silicon and encapsulated in the form of a Chip or an IC Chip[1].
Now a day's dominance and dependency of human being on electronics has increased very rapidly. The increasing dominance of electronics is leading to increase in usage of ICs and this in turn has caused to boom the IC design and IC or IP Manufacturing market to new heights. All of these events are driving to make IP core designs or in simpler words IC designs as an important and valuable entity.
All these conditions have brought the need to create some mechanism or system that can protect IP cores. This arouse the need of Imprinting the identity of the creator of IP core design in design itself whether it is soft IP Core design or Hard IP Cores. There had been numerous cases of IP design theft and misuse. Most of these cases remained unsettled or as there was no specific way to identify the creator of IP from IP core itself because there was no watermark signature or any identity mark left in design by creator to represent own identity.
This paper will try to introduce the method, mechanism or system to imprint the signature of the creator in VLSI designs or Products ICs (in the case of companies). With Signature Imprinting on VLSI Designs, IP Cores and reusable modules will get safer environment. Market and R&D in new IP Core will also increase this way. SoCs will be able to use IP Designs more Aptly and these readymade IP Cores will Cut the Cost of the Design and also will increase the speed of development of the product[2].

Related Works

Current market and Industry scenario demands creating of libraries and reusable VLSI Designs in module form. Initial steps in creating mechanism for Protection of Design should intend to protect the soft IP Cores because VLSI Designs in software form are always the first step in IP Core creation. In this area there is need of defining the signature method for software which is universally accepted[3].
Software approach, once mostly uses for designing is HDL codes or circuit designs. Hard IP Cores are end results of these soft form or designs. If one embeds the signature in soft IP form itself the it is surely going to reflect in IC design[4].
Paper [2] defines a method for embedding the identity of creator in HDL (signing the HDL), in which it suggests using unused states to produce a specific predetermined result in the output.
Paper [5] suggests much simpler method of imprinting signature on VLSI design. It uses predefined lookup table as an identity carrier of mark of creator/claimer. It also states that signature can be represented using a sequence generator in IP adding an extra block for it.

Paper [6] suggests the need of IP core protection in case processors and controllers designs especially as those are amongst the most valuable VLSI Designs but some papers like a paper by Kirovski named "Intellectual Property protection by Watermarking Combinational Logic Synthesis Solutions" also put forward the need to watermark even smaller circuits too.
A paper by William H. Mengione-Smith, John Lach, Miodrag Potkojek named "FPGA Fingerprinting Techniques for protection of Intellectual Property" emphasizes on the imprinting of the signature not only on ASIC but on FPGA designs too. This paper suggests using of the unused CLBs and adding circuit overhead to FPGA circuit for such purpose.

Gaps in Research for this Topic

Most of the papers suggest creating a Lookup Tables as mark of identity or engaging the unused CLBs for identity representation for FPGA Designs and for ASIC they mostly suggest sequence generator. From the results after all embedding these methods for check, these methods are not found very reliable or stable and can be easily cracked in reverse engineering of FPGA. There is a need of time to define some guideline parameters which may serve as guideline for designing of the signature module for VLSI Design.
The signature module should fulfil the following requirements

1) Robust
2) Adding less Silicon real estate, timing and power overhead
3) Least effect on main circuit performance
4) Unrecognisable by attacker
5) Easily recognisable by creator or claimer
6) Tamperproof
7) Best fit for both ASIC and FPGA

Till date, all other signature methods or mechanisms which had been presented in papers or write-ups may fulfil couple of the above mentioned parameters but misses out most of those proving those methods ineffective.

Problem Definition

The research in signature embedding in IP core will help in creating safer and crime free environment for IP cores and VLSI designs by embedding the representation mark of creator in Design itself. This will help protect IP designs from theft and misuse and will also protect the ownership rights of design.
This paper is intend to address the audience related to field of VLSI design and IP design creation along with engineers, scientists and industry people who all are related with VLSI IP Designs and in need to protect it. The government of India has semiconductor registry act under the department of DEITY of India.[7] this department is trying to formulate the rules and policies for the VLSI designs manufacturing and for any activity relating IP Core in Semiconductor. This signature imprintment paper may lead to setup a system or mechanism for signature embedding in IP core. In future this is going to help settle many unsettled cases in design ownership conflict.

Statement

This paper is intending to formulate a way to create a module to add in VLSI design which will serve as identity mark or designer, while fulfilling most of the required parameters which are already defined in Gaps in Research section with least possible deviations.

Innovative Content

Out of all the papers mentioned in the related work section most of those are based on FPGA design protection leaving the design relating ASIC vulnerable to misused which are costlier. Out of all those only few could be applied directly on ASIC designs unlike creating lookup table and using unused CLBs. Possibility of applied the method of adding an sequence generator is the method which we can add into ASIC design, but while checked for feasibility of using sequence generator for signature module it is found that smaller length sequence are dependent of probabilities and can be decoded in reverse engineering and moreover adding larger sequences of 32bits or 64bits or more adds significantly large area overhead and also visibly caught in the HDL code and IC working

Problem Formulation

After striking out all of above possibilities, this paper tries to device out the new possibility with least deviation from parameters.

- For adding robustness in the design, the consideration of design accuracy, reliability is important. Digital circuits provide more robustness and reliability. Trying with multivalued logic and imparting in with artificially created jitter is also a possibility but this causes module circuit working less robust as in multivalued logic one has to deal with

the noise margin at all defined levels leaving it less reliable and in artificially created jitter the dealing with such high frequency mixers imparts deviation in it from robustness.

- For adding less timing overhead module had to be kept distinguished from main circuit so that is won't be able to add timing overhead in the working path of the design.
- For adding lesser power overhead the module needs a gateway before some command or signal invokes module working instead of module kept on standby and kept consuming power.
- For less area overhead, the module can't be created with complex circuits, it must be combination of simple circuits only.
- It being made out of smaller circuit it will eventually go unseen in big and complex circuitry of VLSI Design.
- One of the possibility for creating of signature module can be using and extra bit, which one uses for avoiding race condition in state machines mechanism in high speed circuit, crafting of signature, but this possible had been stroke out because it affected the main circuit working and moreover added to timing overhead.

Making way through all the visualisation of module design which may fulfil most of parameters from the list could look like a parallel multibit state machine with gateway for input so that it consumes less power and enable gateway for output too which intend to add least interference in circuit working.

Solution Methodology

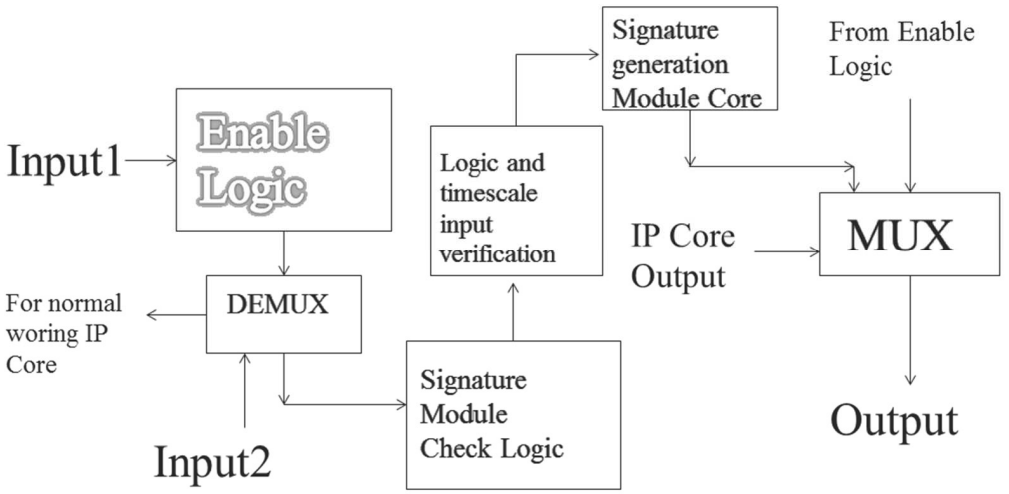


Fig1:Block Diagram of Possible Module for Signature Imprint

Currently the design for Signature module is made using the function for input to enable the module which act as a gate circuit for the module and also keep checking on for the valid input combination from bus which is represented by Input1 in block diagram. This Input1 can be any combination from all possible inputs to original circuit which only designer knows. A certain combination of signals of certain inputs paths constituting input one will enable the gateway and thereby will also enable taking in signal from input2 to signature module. Then comes the DeMux which is present in circuit to cut-off and connect the signature module from input lines of the main circuit to keep least interference with main circuit working. After enable logic found valid the DeMux enables the input 2 to signature module. The input 2 here is again same as any combination of input line out of all possible inputs to VLSI circuit which only designer knows. After inputs from input2 is redirected to signature module the signature module will keep checking for valid input combination of predefined combination which will finally enable the signature to appear in output through Mux. And this out will represent the signature.

The retrieval method is kept very specific and through all those gateways and via all multiple checks and lots of inputs to increase the diversity of all possible signals and also intend to satisfy the parameter which said the method should not be recognisable by attacker.

Algorithm of the Module

The algorithm of signature module is as follows

- Chose the combinations of inputs for making input1 bus and input 2 bus

- Assign and distinguish between input1 and input2
- Design an enable logic which will be gateway for enabling the module
- Enable logic also enable mux and DeMux in the circuit to toggle and connect module to input and output
- When mux and DeMux is applied in the circuit the power overhead in standby condition is reduced to much extent
- Then create the logic which will accept and check for validity over parallel combination of inputs in multiple time sequentially
- After all received inputs found valid then the module and on correct timing the module in the processor will output us the signature mark is predefined fashion on predefined combination of outputs

Results

For current design valid attributes are as follows:-

1) Input1- 8bit- Valid signal is 'S (83_D)' which means valid signals over 8 bit line is '01010011' (Module Enabling Gateway)

2) Input2- 8bit- Valid signal is 'V(86_D)' which means valid signals over 8 bit line is '01010110'
   5 clock unit of delay
   Valid signal is 'L(76_D)' which means valid signals over 8 bit line is '01001100'
   5 clock unit of delay
   Valid signal is 'S(83_D)' which means valid signals over 8 bit line is '01010011'
   5 clock unit of delay
   Valid signal is 'I(73_D)' which means valid signals over 8 bit line is '01001001'
3) Valid Output which will represent signature on output of the module will be as follows:-
   Output- 8bit- valid signals representing Identity is
   'C(67_D)' which means valid signals over 8 bit line is '01000011'
   5 clock unit of delay
   'O(79_D)' which means valid signals over 8 bit line is '01001111'
   5 clock unit of delay
   'E(69_D)' which means valid signals over 8 bit line is '01000101'
   5 clock unit of delay
   'P(80_D)' which means valid signals over 8 bit line is '01010000'

This design is realised and tested on Virtex 7 FPGA of Xilinx
Some of the parameters of the results are as follows:-

Table 1: Number of Cells required

Sub Module	*Number of Cells*
For enable logic	17
For DeMux	26
For Mux	24
Signature Module Check Logic	51
Logic and Timescale Verification	113
Signature generation Module	82

Time to transit from input to output:-

- If verification fails at first place 32nSec
- If verification is valid 81nsec

Power Consumed:-

- Standby Condition 19uWatt
- After module is enabled successfully 130uWatt

Conclusion

The module created shows some deviations to parameters but efforts are taken to keep those minimum and there is much scope for this module implementation to make a generalised and universally accepted design which possibly in future will be used in almost all valuable VLSI Designs. Likewise there is IUPAC nomenclature for all chemicals and compounds, same is

the need in VLSI Design policy formulation to assign the nomenclature or signature to all companies and even if some individual creates as design he may register for his own signature, which is universally accepted. From above method there is a possibility of creating billions of different possible signatures by changing the combination of valid input signals, output signals and time gap between those signals.

Future Work

Future work is to take this module till extraction level in ASIC Design and measure the effects of module on circuit parameters and try to improvise it for IC design. Next part will be minimising of the deviation from defined parameters.

Acknowledgement

Authors thank Prof. Dr. Mrs. P.P. Rege, HoD, E&TC Dept, CoEP, Pune for providing the laboratory resources for undertaking the work.

References

[1] Michael Keating,Pierre Bricaud, "REUSE METHODOLOGY MANUAL FOR SYSTEM -ON-A-CHIP DESIGNS", KLUWER ACADEMIC PUBLISHERS, Third Edition.

[2] Min-Chuan Lin, Guo-Ruey Tsai, Chun-Rong Wu, Ching-Hui Lin, "Watermarking Technique for HDL-based IP Module Protection",International conference on Advanced VLSI Technology, April 3-5, 2014.

[3] Lin Yuan, Pushkin R. Pari, and Gang Qu, "Soft IP Protection: Watermarking HDL Codes", Springer Journal, J. Fridrich (Ed.): IH 2004, LNCS 3200, pp. 224-228, 2004.

[4] E. Castillo, U. Meyer-Baese, L.Parrilla, A. Garcia, A. Lloris, "HDL- level automated Watermarking of IP Cores", 2008 SPIE Digital Library { Subscriber Archive Copy }.

[5] E. Castillo, U. Meyer-Baese, L.Parrilla, A. Garcia, A. Lloris, "Intellectual Property Protection of IP cores at HDL Design Level with Automatic Signature Spreading", IEEE, International Conference on Advances in Electronics and Microelectronics,2008

[6] Luis Parrilla, Encarnacion Castillo, Antonio Garcia, Daniel Gonzalez, Antonio Lloris, Elas Todorovich, Eduardo Boemo, "PROTECTION OF MICROPROCESSOR-BASED CORES FOR FPL DEVICES",IEEE, 2010

[7] www.deity.gov.in/content/ads-policies

Comparison of Network Codes in Distributed Storage Area Network based on Performance using NS3 Simulator

Priyanka Ramesh Nanaware* and S. K. Srivastava**
*Department of Electronics Engineering, Pillai Institute of Information Technology, Engineering, Media Studies & Research New Panvel, India, ME(Electronics Engineering) Final Year
nanawarepriyanka25@gmail.com
**Associate Professor in Department of Electronics and Telecommunication Engineering, Pillai Institute of Information Technology, Engineering, Media Studies & Research New Panvel, India
sanjeevkumar.srivastava1@gmail.com

Abstract: In this paper, it is evaluated the comparison of erasure code and regenerating code in distributed storage area network based on the performance parameters such as throughput, latency, bandwidth consumption, packet delivery ratio, recovery speed etc. Distributed storage area network is formed using ceph cluster. The results of this comparison show that the regenerating code provides the better throughput and fast recovery as compared to erasure code. At the same time regenerating code occupies less bandwidth, low latency and high packet delivery ratio. Therefore it concludes that the performance of regenerating code over erasure code and is better in this scenario.

Keywords: Ceph Cluster, Distributed Storage Area Network, Network coding, NS3 simulator, Performance parameters.

Introduction

In a recent scenario a wide variety of applications, relies on distributed environments to process and analyse large amounts of data hence it is necessary to improve the performance of such Distributed storage network. The development in information technology from last few years has led to big increases in data storage and it results data traffic. In particular the applications forces the networks to adapt and become more flexible and to increased need of network capacity and performance.

One way to improve the network performance and capacity is by introducing network coding in a distributed storage network. With the data spread out in a distributed storage network, the availability for the end user will increase. Network coding can also increase the reliability of a distributed storage network because it introduces redundancy to the system.

Network coding is a method of optimizing the flow of digital data in a network by transmitting digital evidence about messages. The "digital evidence" is itself, a composite of two or more messages. Network coding can be deployed with different coding schemes. There are different types of network coding such as Replication code, Erasure Code, Regenerative Code, Self-repairing code etc.[1][2]

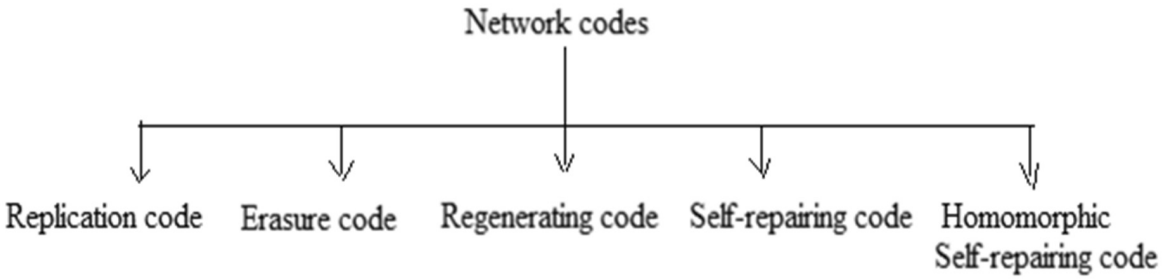


Figure 1. Types of Network codes

"Fig. 1", shows the types of network coding in distributed storage area network. In network coding, the data does not depend only on one transmitted message but also on the contents of other messages that happen to be sharing the route at the time of transmission. For this reason, network coding is more resistant to hacking, eavesdropping and other forms of attack than traditional data transmission. The extent of throughput improvement that network coding can provide.[2]

Distributed Storage Area Network

Distributed storage area network play a very important role in the field of Networking and in Advanced Computing. Such a distributed storage area network is formed using cluster. There are number of Types and ways to form a cluster such as Hadoop Cluster, High performance cluster, High availability cluster, Ceph cluster etc. In a recent scenario a wide variety of applications, relies on distributed environments to process and analyse large amounts of data hence it is necessary to improve the performance of such Distributed storage network.

In the traditional system data is stored on a single disk, suppose that disk will be failed then there is completely loss of data and data recovery is also not possible because data is stored on single disk. After that the RAID is used, In RAID mirroring technique is used. In RAID due to mirrored technique firstly original data is stored on a single disk and the same copy of original data is stored on another disk. So here even original disk get failed we can recover data with the help of another disk. But the disadvantage of RAID due to mirrored technique is its bandwidth consumption is more. Suppose traditional system requires bandwidth W, then for RAID requires 2W. So to overcome this disadvantage distributed storage area network is used.

In distributed storage area network the data is distributed in number of disks. Here if one or two disks get failed then also we can recover that using the data present in the other disks. So the main advantage of distributed storage area is the data is stored is secured form and there is no any kind of loss of data.[3] Here Distributed storage area network is formed by ceph cluster. “Fig. 2” shows the Ceph cluster.

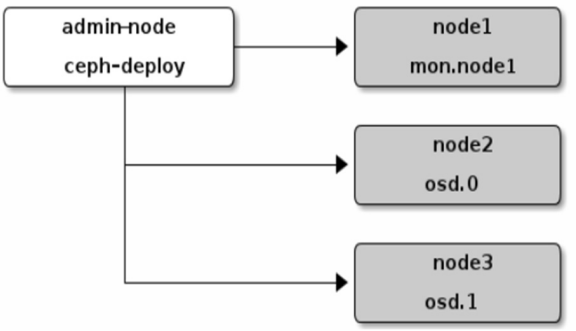


Figure 2. Ceph Cluster

Three main components of ceph cluster

1) Admin node: Ceph admin node responsible for in which way data should distribute.
2) Monitor node: It is responsible to keep track of active and failed cluster nodes.
3) OSD (Object Storage Devices): That actually store the content of files. Ideally, OSDs stores the data which is distributed by the admin node of the ceph cluster.

Ceph’s CRUSH algorithm liberates storage clusters from the scalability and performance limitations imposed by centralized data table mapping. It replicates and re-balances data within the cluster dynamically eliminating this tedious task for administrators while delivering high performance and infinite scalability. Ceph cluster supports object storage which has following advantages over block storage:

1) The data can be anything we want to store i.e. Text, images etc.
2) With block storage, files are split into evenly sized blocks of data, while object storage have flexibility in data storage
3) object that contains the data, metadata, and the unique identifier therefore they are more secure than block storage.

NS3 Simulator

On the basis of comparison of various simulation tools, it is observed that NS-3 is best suited simulation tool. NS3 is implemented using C++ With modern hardware capabilities and compilation time was not an issue like for NS2. NS3 can be developed with C++ entirely. A simulation script can be written as a C++ program, which is not possible in NS2. NS3 demonstrates the best performance among all the network simulations tools.[4][5]

Erasure Code

The first one is an optimal erasure code, called Maximum Distance Separable (MDS) code. This code improves the storage efficiency. Here the original data is divided into M fragments which are stored in n nodes. Each node stores M/k data fragments and k nodes are required to recreate the original data. MDS code achieves the optimum in the redundancy-

reliability trade-off.[6][7][8]. In Erasure code when node failure will occurs firstly the node failure is detected by the monitor node. To recover the failure node data, here original data is recreated and failure node data is recovered in BN (Brick Node). For the recovery need to recreate original data first so here bandwidth required and latency is more.

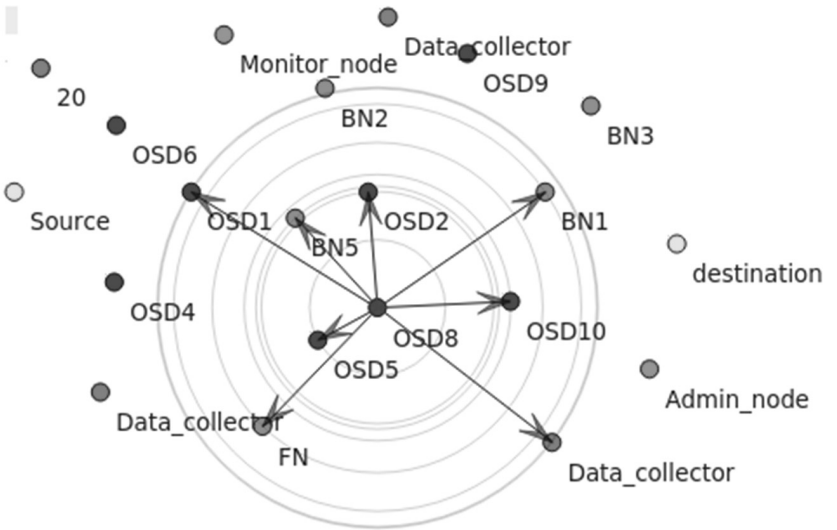


Figure 3. Erasure code data distribution and recovery

"Fig. 3" shows the nodes of ceph cluster, data distribution and recovery in erasure code. Here FN indicates the failure node, BN indicates the brick node, OSD stands for object storage devices. Recovered data of failure node is stored in brick node and then send to respective destination.

Regenerating Code

Regenerating code is another variation of the erasure code. The regenerating code provides an improved repair function. This is made possible because of the fact that the regenerating code allows a new node to connect all remaining storage nodes after a node failure. This means $k \leq d \leq n - 1$ for the regenerating code, compared to $d = k$ for the MDS code.[6]

There is a storage/bandwidth trade-off that is of great interest. It describes the relationship between the amount of stored data and the amount of data that necessary to transfer for a repair. There are two extreme points on this curve called Minimum Storage Regenerating (MSR) and Minimum Bandwidth Regenerating (MBR). They are what they sound like; MSR stores the smallest amount of data possible in each node and MBR requires the least amount of data to be transferred to make a repair.[9][10] In regenerating code when node failure occurs, monitor node detects the node failure. For recovery of failure node data here specific location of failure node is detected and depending of specific location of failure node how many nodes are required to recover the failure node data are decided.

"Fig. 4" shows the nodes of ceph cluster, data distribution and recovery in regenerating code. Here FN indicates the failure node, NC indicates the new comer node, OSD stands for object storage devices. Recovered data of failure node is stored in provider and then send to respective destination

Results

Parameter Assumptions

1. Operating System : Red Hat
2. Simulation Tool : NS3
3. Graph evaluation : Xgraph
4. Area of Experiment : 5000x5000
5. No of Nodes : 20
6. Packet Size : 400 bytes

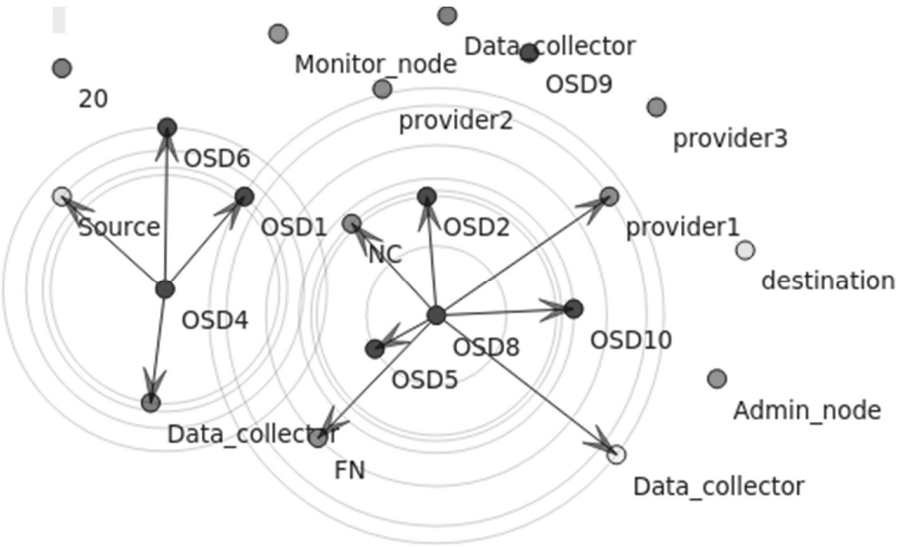


Figure 4. Regenerating code data distribution and recovery

Table 1 Comparison of Erasure code and Regenerating code for simulation time=100 msec

Performance Parameters	Network Codes	
	Erasure Code	Regenerative Code
Bandwidth Consumption(Hz)	93.4933	74.7946
Throughput(Mbps)	1914.74	15812.1
Packet Delivery Ratio	0.449333	0.549339
Latency(msec)	0.800182	0.359548
Recovery Speed(Mbps)	1.44933	1.50456

"Table 1" explains the performance parameters of Erasure code and Regenerative code for simulation time 100 msec

Table 2 Comparison of Erasure code and Regenerating code for simulation time=150 msec

Performance Parameters	Network Codes	
	Erasure Code	Regenerative Code
Bandwidth Consumption(Hz)	93.4917	74.7954
Throughput(Mbps)	2872.07	23717.7
Packet Delivery Ratio	0.449635	0.549641
Latency(msec)	0.800622	0.359988
Recovery Speed(Mbps)	1.44964	1.50487

"Table 2" explains the performance parameters of Erasure code and Regenerative code for simulation time 150 msec

Table 3 Comparison of Erasure code and Regenerating code for simulation time=200 msec

Performance Parameters	Network Codes	
	Erasure Code	Regenerative Code
Bandwidth Consumption(Hz)	128.261	102.609
Throughput(Mbps)	5253.57	43384.3
Packet Delivery Ratio	0.450003	0.550009
Latency(msec)	0.801754	0.360792
Recovery Speed(Mbps)	1.45000	1.50523

"Table 3" explains the performance parameters of Erasure code and Regenerative code for simulation time 200msec

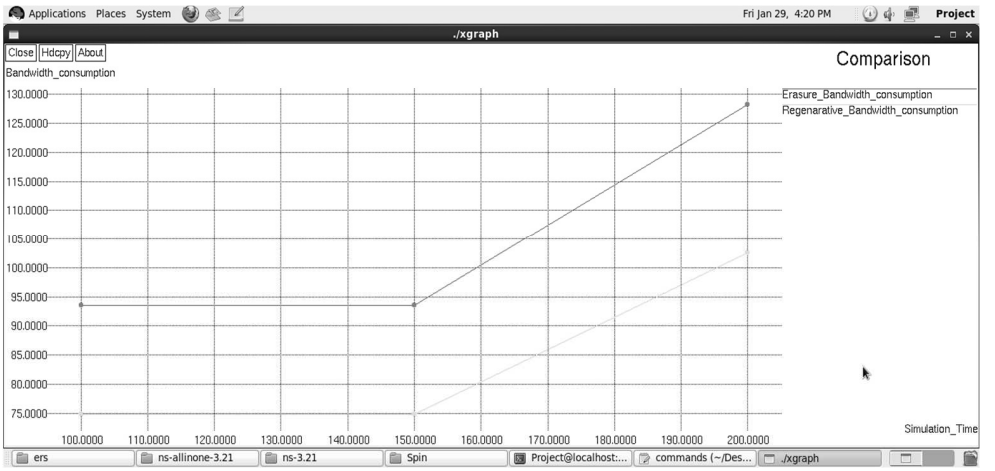


Figure 5. Comparison of Bandwidth Consumption of Erasure Code and Regenerative code

"Fig. 5" shows that Regenerative code requires less bandwidth than Erasure code. For the recovery erasure code needs to recreate the original data therefor bandwidth consumption is more in erasure code compared to regenerative code. As the simulation time increases the bandwidth consumption is more.

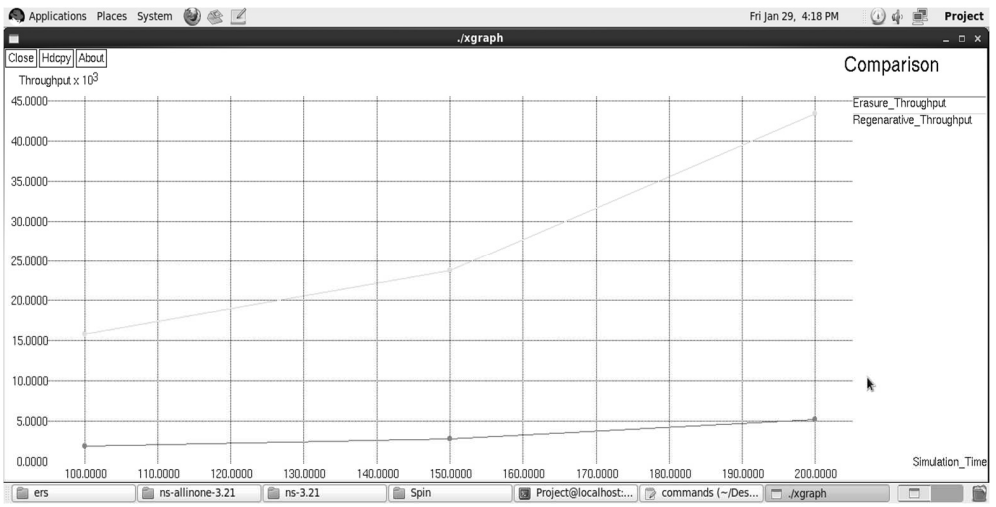


Figure 6. Comparison of Throughput of Erasure Code and Regenerative code

"Fig. 6" shows that Actual rate that information is transferred is more in Regenerative code and very lower throughput in erasure code. In erasure code as simulation time increases there is slight change in throughput. But in regenerating code as simulation time increases the throughput increase at great extent.

"Fig. 7" shows, In Regenerative code, data packets received by the destinations to those generated by the sources (i.e no of received bytes/no of transmitted bytes) are more than Erasure code. Packet delivery ratio of erasure code is very lower than regenerating code because erasure code does not support high density area.

"Fig. 8" shows, Erasure coding have a high delay. As the simulation time increases latency in Regenerative code decreases. At the time of recovery erasure code needs to recreate original data first therefore latency is more in erasure code as compared to regenerating code.

"Fig. 9" shows, Recovery speed of Regenerative code is more than Erasure code. At the time of recovery erasure code needs to recreate original data first therefor recovery speed of erasure code is less than regenerating code. In regenerating code depending on which specific block missing is detected and according to that recovery is done and here no need to recreate original data therefor recovery speed is faster than erasure code.

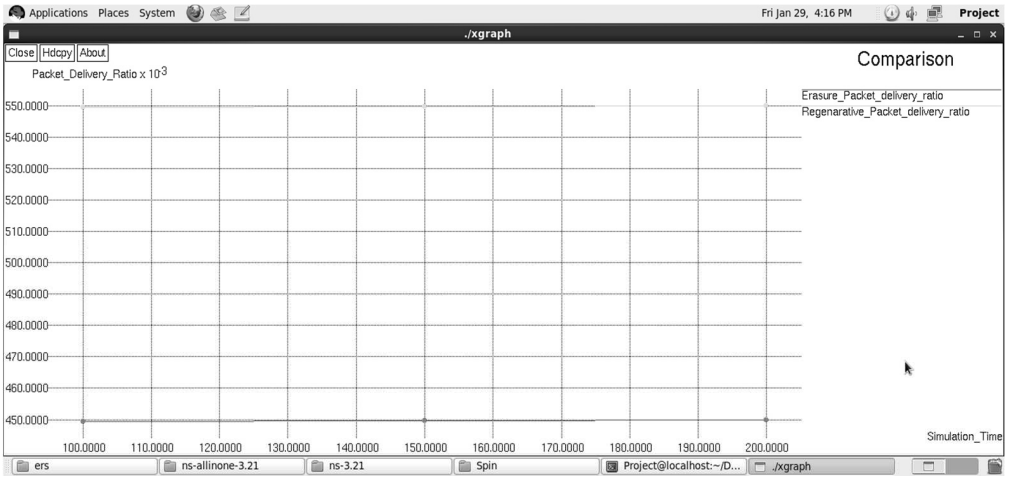


Figure 7. Comparison of Packet Delivery ratio of Erasure Code and Regenerative code

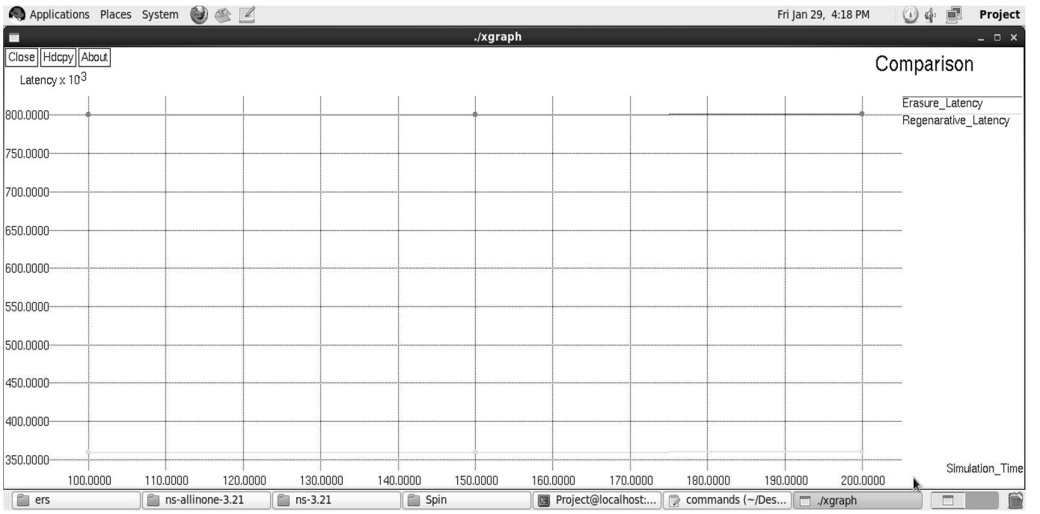


Figure 8. Comparison of Latency of Erasure Code and Regenerative code

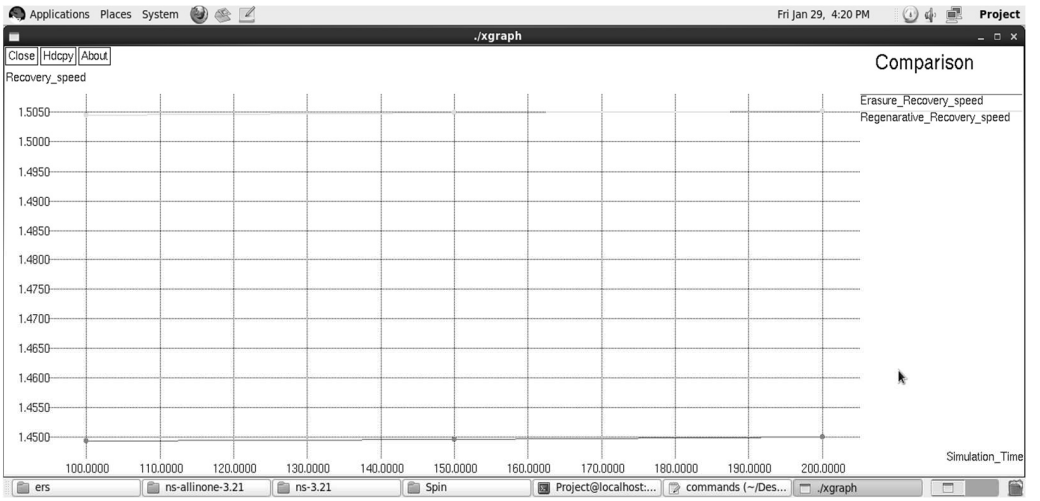


Figure 9. Comparison of Recovery Speed of Erasure Code and Regenerative code

Conclusion

The results of this comparison show that the regenerating code provides the better throughput and fast recovery as compared to erasure code. At the same time regenerating code occupies less bandwidth, low latency and high packet delivery ratio. Therefore it concludes that the performance of regenerating code over erasure code and is better in Distributed Storage Network.

Acknowledgement

The authors wish to thank Shreya Bokare, CNIE department at CDAC (Centre for Development of Advanced Computing) Kharghar, for her support.

References

[1] Alexandros G. Dimakis, Kannan Ramchandran, Yunnan Wu, Changho Suh, "A Survey on Network Codes for Distributed Storage" Proceedings of the IEEE, Vol.99,No.3,March2011.

[2] Alexandros G. Dimakis, P. Brighten Godfrey, Yunnan Wu, Martin J. Wainwrigh, Kannan Ramchandran, "Network Coding for Distributed Storage Systems", IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 56, NO. 9, SEPTEMBER 2010

[3] Diana Gudu, Marcus Hardt, Achim Streit "Evaluating the Performance and Scalability of the Ceph Distributed Storage System" 2014 IEEE International Conference on Big Data, Page No. 177 – 182

[4] Atta Rehman Khan, Sardar M. Bilal, Mazliza Othman "A Performance Comparison of Network Simulators for Wireless Networks", 30 E-28911 Leganés

[5] Suraj G. Gupta, Mangesh M. Ghonge, Parag D. Thakare, Dr. P. M. Jawandhiya, "Open-Source Network Simulation Tools: An Overview" *International Journal of Advanced Research in Computer Engineering & Technology* (IJARCET) Volume 2, Issue 4, April 2013, ISSN: 2278 – 1323

[6] RICKARD BROMAN, "A practical study of Network Coding in Distributed Storage Systems", Stockholm, Sweden 2013

[7] Hsiao-Ying Lin, Li-Ping Tung, Bao-Shuh P. Lin, "An Empirical Study on Data Retrievability in Decentralized Erasure Code Based Distributed Storage Systems" 2013 7th International Conference on Software Security and Reliability, Page No 30 – 39.

[8] Ming-Zhu Deng, Zhi-Guang Chen, Yi-Mo Du, Nong Xiao, and Fang Liu , "Erasure Codes in Big Data Era", The 2014 International Conference on Control, Automation and Information Sciences (ICCAIS 2014) December 2-5, 2014, Gwangju, South Korea, Page No. 218 – 223.

[9] Jian Li Tongtong Li Jian Ren , "Secure Regenerating Code", Globecom 2014 - Communication and Information System Security Symposium, Page No. 770 – 774.

[10] Onur Ozan Koyluoglu, Ankit Singh Rawat, and Sriram Vishwanath "Secure Cooperative Regenerating Codes for Distributed Storage Systems" IEEE Transactions On Information Theory, VOL. 60, NO. 9, SEPTEMBER 2014, Page No. 5228 – 5244.

Optimal PI Tuning with GA and PSO for Shunt Active Power Filter

N Gowtham* and Shobha Shankar**
*Research Scholar, Department of Electrical and Electronics Engineering, VVCE, Mysuru
gowtham.n@vvce.ac.in
**Professor and Head, Department of Electrical and Electronics Engineering, VVCE, Mysuru
shobha.shankar@vvce.ac.in

Abstract: The PI controller plays important role in all engineering controls. The tuning of PI controller is generally done with institutive methods, which are less effective. This paper discusses about the PI controller tuning which affects the results of the total harmonic distortion (THD) of the active power filter. The active power filter is designed for reducing the THD in non-linear load connected power system. It is identified that the variation of proportional constant (Kp) and Integral constant (Ki) affects the THD value. So a problem is formulated to minimize the THD value of the system to find the optimal Kp, Ki parameters. The well-known genetic algorithm (GA) and Particle swarm optimization (PSO) techniques are used to find the optimal value of parameters and minimum THD value. MATLAB Simulink, Sim power system and optimization tool boxes are used.

Keywords: Active Power Filter (APF), Genetic Algorithm (GA), PI tuning and Particle Swarm Optimization (PSO).

Introduction

The power quality is very important nowadays as the need of energy is increasing. The active power filter is used for improving the power quality in the non-linear load connected to the power system. In single phase and three phase system rectifiers or front end converters produces the problem of lower order harmonic injection [1]. Active filter system reduces this problem by injecting the current [2]. There are two types of active power filters, they are shunt and series compensation. The shunt takes care of the current injection and series takes care of voltage injection. Mainly shunt active filters are given priority as it handles the current because current is load dependent [3]. The inverter is used to inject the current to the supply system. The hysteresis control is used for reducing the complexity of PWM generation and current which is to be injected is derived and used as the control technique [4]. A notch based filter is useful in selecting a particular harmonic and removing it. The phase lock loop is eliminated to synchronize with the control part [5].

The Genetic algorithms are also used in the switching of inverters in active power to optimally initialize the switching of inverter, which increases the switching time faster in the initial condition [6]. The control technique used for controlling the pulses is taken from the reference generation of current. Genetic Algorithm is used to produce the reference current, which makes the control faster compared to conventional methods [7]. The hybrid passive filters are introduced to compensate the reactive power as well as harmonics, which has lesser size of inductance [8]. Adaptive filter design is made for synchronization detection and harmonic extraction, which can be used in the Flexible AC Transmission Systems (FACTS) [9]. A robust control strategy with boost converter is presented by the authors in [10] for shunt active filter to improve the performances of the filter.

The upgrade of inverter control strategies also improves the performance of the active filters. A new vector control based hysteresis control is made for improving the transient performance of the system [11]. A non-linear optimal predictive control to improve the stability and robustness of the shunt active filter [12]. The control algorithm is improved to reduce the switching nodes are presented in [13]. The power system is not dealing with three wires (RYB). It also uses four wires (RYBG). A non- linear control strategy is used in [14] to control the power factor with 4-wire system to improve power factor in asymmetrical and non-linear loading condition. A new control strategy with dq-axis is presented in [15] to improve the performance of the shunt active filter. Proportional integral (PI) and Proportional Integral Derivative (PID) based Active power line conditioner is used to improve the power factor is presented in [16, 17]. The analytical analysis of shunt active filter is done for improving the stability and power factor is presented in [19]. The pq theory is meant for identifying the power component. A literature [20] describes that the instantaneous nature of pq is not considered and while considering it gives better results. It is known as Instantaneous reactive power theory. And it is described in analytical method for improving the stability [21]. The unified power quality conditioner is nothing but the combination of shunt and series compensator. It provides the sag and swell compensation of voltage and harmonic compensation in current [22].

Optimization is very important in engineering field to improve the efficiency of the existing system. The PI tuning problem of shunt active filters is not considered in any literature. This problem is formulated and the solution is made with genetic algorithm [18, 23] and Particle swarm optimization[23] are used to find the optimum values of K_p and K_i with minimization of THD.

Problem Definition

The DC link voltage is controlled with PI controller. This PI controller is responsible for total harmonic distortion.

$$I_c = K_p * (V_{dc\ ref} - V_{dc*}) + K_i * \int_0^t (V_{dc\ ref} - V_{dc*})dt \quad (1)$$

Here,

I_c = Compensation current
V_{dc*} = DC measured voltage at DC link
$V_{dc\ ref}$ = DC reference voltage
K_p = Proportional constant
K_i = Integral constant

Objective function is

Minimize $\sum_{i=0}^{n}$ mean (THD) (2)

Subjected to

$P1 \leq K_p \leq P2$ (3)

$I1 \leq K_i \leq I2$ (4)

Here,

n = Total number of samples
P1 = lower limit of K_p value
P2 = Upper limit of K_p value
I1 = Lower limit of K_i value
I2 = Upper limit of K_i value

Solution Methods

There are many solution algorithms available in meta-heuristic optimization techniques. GA and PSO algorithms are used due its simplicity.

Genetic Algorithm

Genetic algorithms are based on natural genetics and natural selection. The good properties of parents will produce the child. The natural genetic operations are reproduction, crossover and mutation [23]. Reproduction is used to select the good strings with a probability constant. It is known as probability of reproduction. The equation can be written as

$$P_i = \frac{F_i}{\sum_{j=1}^{n} F_j} \ ; i = 1,2,\ldots,n \ ; j = n+1 \quad (5)$$

$P_i - Mating\ pool$

$F_i - Fitteness\ function$

Crossover is used to create a new strings by exchanging the information among the strings. Mutation is the better fitness values for new generations. The mutation probability changes between 1 and 0. The solution steps are as follows

1. Read the input data like population size, string length, probability of cross over (p_c), probability of mutation (p_m). Populate random kp and ki values (size m).
2. Evaluate the fitness equation (2).
3. Carry out the reproduction process
4. Carry out the crossover operation using the crossover probability p_c.
5. Carry out mutation using probability p_m to find new generation (m).
6. Do the step 2.
7. Check for end of iteration.
8. Stop and show the result.

Particle Swarm Optimization

The PSO algorithm is based on the food searching behavior of the birds or fishes, which is taken as particles. The particle which can get the food faster is made as a mathematical algorithm. The steps of the algorithm is given below,

1. Assume the swarm number is N.
2. Generate the Kp and Ki values. This is called as initial population (X).
3. Evaluate the objective function (2).
4. Find the velocities of the particles. Using the below equation.

$$V_j(i) = V_j(i-1) + c_1 r_1 \left[P_{best,j} - X_j(i-1)\right] + c_2 r_2 \left[G_{best} - X_j(i-1)\right]; \quad (6)$$

j = 1, 2, ….N.

where,

V – velocity of the particle
i – iteration number
j – particle number
P_{best} – Particle best
G_{best} – Global best
c_1, c_2 – cognitive and social learning rates (chosen as 2)
r_1, r_2 – uniformly distributed random numbers in range 0 and 1.

5. Update the X values using the following equation,

$$X_j(i) = X_j(i-1) + V_j(i); j = 1,2,\ldots.,N \quad (7)$$

6. Repeat the step 3 till the end of iteration count.
7. Check for convergence.

Shunt Active Power Filter with Tunning Algorithm

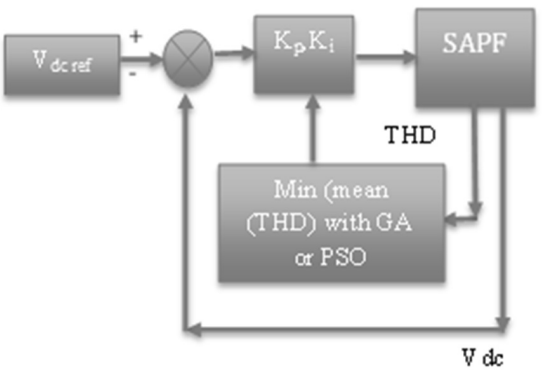


Figure 1. Block diagram of proposed system

The above figure shows the control diagram of the proposed model. The reference dc voltage is given to the comparator. The other side of the comparator is given from the measured V_{dc}. The output of the comparator is given to the PI controller where the K_p and K_i parameters are used. The output of the PI controller is given to the control loop of current in the shunt active power filter (SAPF). The total harmonic distortion (THD) is taken out and given to the objective function. The algorithms generate initial random population of the K_p & K_i parameters. So till the optimum results are reached the algorithm run and produce the optimum value of K_p and K_i value with lesser THD.

Results and Discussions

The simulation is carried out using the parameters given in table 1. The non-linear load taken here is rectifier with RL load. Due to the non-linear load the waveforms are taken as shown in the figure 3.

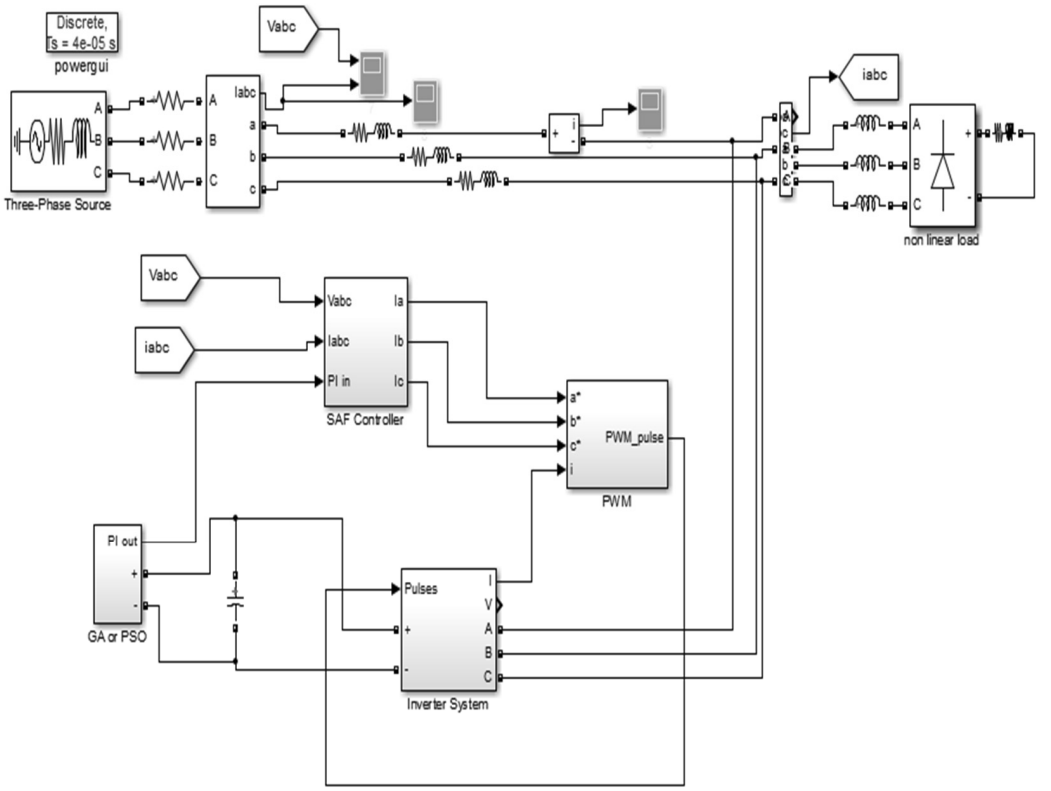


Figure 2. Simulation of the test system

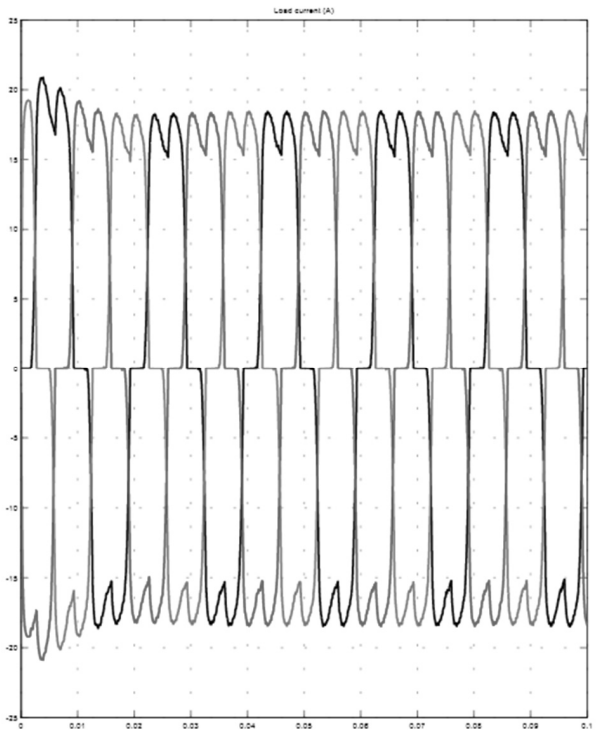


Figure 3. Three phase current waveform before compensation

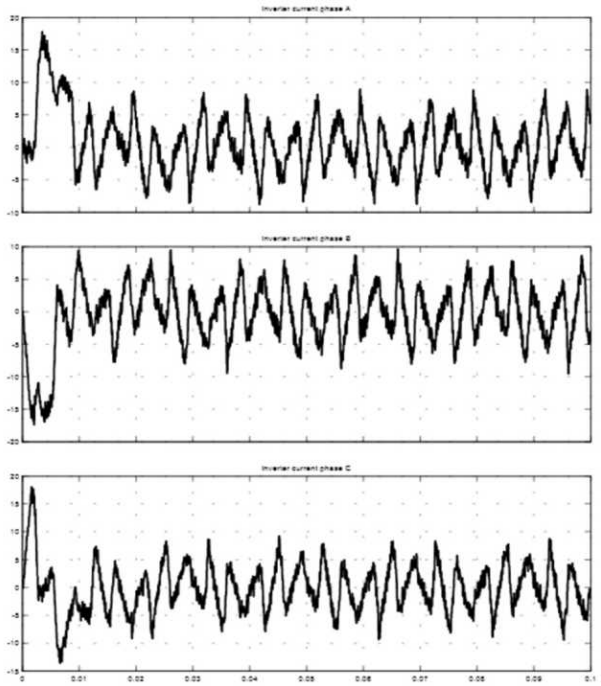


Figure 4. Compensation current produced by the controller

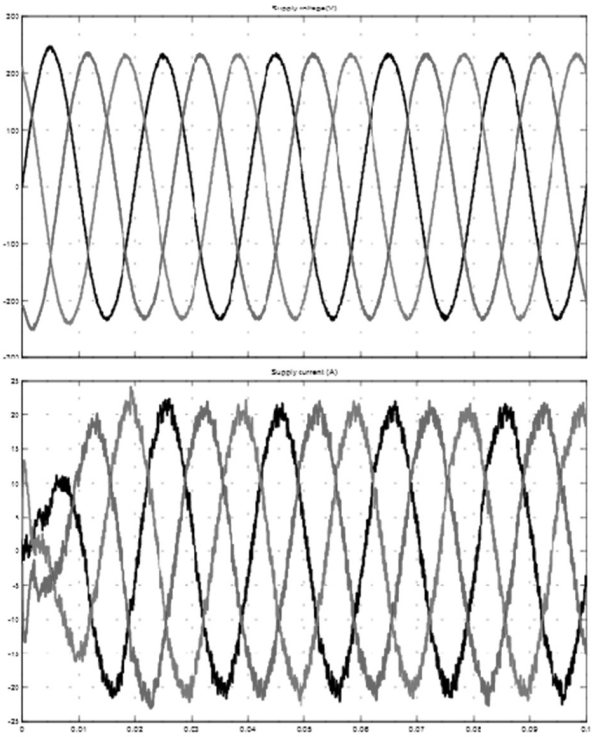


Figure 5. Three phase voltage and current after compensation

The figure 3 shows the current before compensation or load current. Figure 4 shows the compensation current produced by the controller. After injection of the compensation current, the voltage and current waveforms are shown in figure 5.

Table 1. Parameters of SAF

Phase voltage & frequency	310V,50 Hz
Line impedance	R_s =1 Ω, L_s=0.1mH
R-C load	R_L=32Ω,C_L=500μF
R-L load	R_L=26Ω,L_L=10mH
DC bus voltage & capacitance	V_{dc}=500V, C_{dc}= 2000 μF
Shunt active filter ratings	1.81 KVA

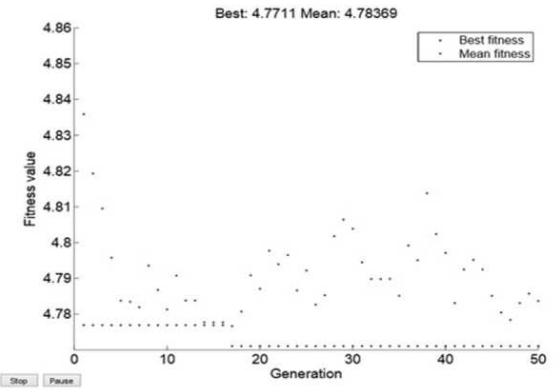


Figure 6. Convergence graph of Genetic Algorithm (GA)

The figure 3 shows the convergence graph of the GA algorithm. The graph shows the THD value nearly 4.77 % which satisfies the IEEE standard for THD.
The parameters used are given below (for both GA and PSO),
Population size : 20
Number of Variables: 2(K_p, K_i)
Range of Variables : 0-0.002
Maximum iteration : 100

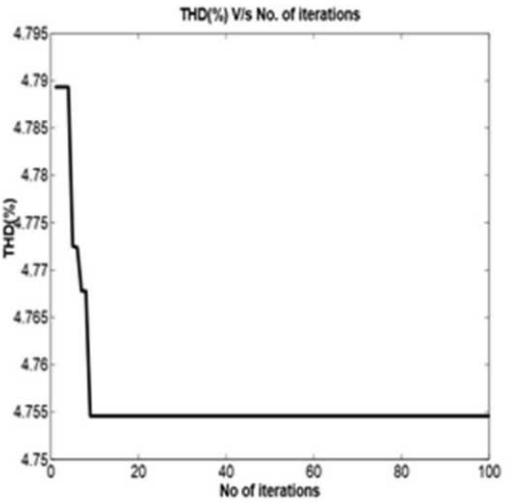


Figure 7. Convergence graph of PSO algorithm

The above figure shows the results of PSO algorithm after solving the objective function. The solution converges in 4.755 % of THD. And it is less compared to the GA method.

Conclusion

The shunt active filter is designed with PI controller and pq theory. The parameter optimization of the shunt active filter is done with two algorithms GA and PSO. The THD value minimization is taken as objective and it is minimized. The optimal values of kp and ki values varies with algorithm and the THD value is also varied. Finally PSO algorithm gives better results compared to GA. The optimal value of kp and ki values are identified.

References

[1] Ned Mohan 2002, 'Power Electronics: Converters, Applications, and Design' 3rd Edition', Wiley publications

[2] Bhattacharya,S Frank,TM, Divan,DM & Banerjee, B 1998, 'Active filter system implementation', IEEE Ind. Appl. Mag., 4(5): 47–63

[3] Bhattacharya, C. Chakraborty, & Bhattacharya,S 2009, 'Current compensation in shunt type active power filters', IEEE Ind. Electron. Mag., 3(3): 38– 49.

[4] Bhim Singh, Kamal Al-Haddad & Ambrish Chandra 1999, 'A New Control Approach to 3-phase Active Filter for Harmonics and Reactive Power Compensation', IEEE Trans. on Power Systems, 46(5):133– 138.

[5] Davood yazdani, Alireza Bakhshai & Praveen K.Jain 2010, 'A Three- phase Adaptive Notch filter based Approach to Harmonic /Reactive current extraction and harmonic decomposition', IEEE Transactions on Power electronics, 25(4): 914-922.

[6] El-Habrouk,M & Darwish,M,K 2002, 'A new control technique for active power filters using a combined genetic algorithm conventional analysis', IEEE Transactions on Industrial Electronics, 49(1):58-66.

[7] Gonzalez, SA, Garcia-Retegui,R, & Benedetti,M 2007, 'Harmonic computation technique suitable for active power filters', IEEE Trans. Ind. Electron., 54(5): 2791–2796.

[8] Hamadi,A , Rahmani,S & Al-Haddad, K 2010, 'A hybrid passive filter configuration for VAR control and harmonic compensation', IEEE Trans. Ind. Electron., 57(7): 2419–2434.

[9] Karimi, H, Karimi-Ghartemani,M, & Iravani,MR 2003, 'An adaptive filter for synchronous extraction of harmonics and distortions', IEEE Trans. Power Del., 18(4): 1350- 1356.

[10] Marconi,L, Ronchi,F & Tilli,A 2004, 'Robust perfect compensation of load harmonics in shunt active filters', in Proc. 43rd IEEE Conf. Decision Contr., 3: 2978–2983.

[11] Mansour Mohseni,Syed.M.Islam 2010, ' A New Vector based Hysteresis Current Control Scheme for three phase PWM Voltage Source Inverters', IEEE Trans.Power Elec.,25(9):2299-2309

[12] Mendalek,N,Fnaiech, F, Al-Haddad,K &Dessaint, LA 2002, 'A non- linear optimal predictive control of a shunt active power filter', in Conf. Rec. IEEE 37th Ind. Appl. Soc. Annu. Meeting, 1: 70–77.

[13] Singh, BN ,Singh,B, Chandra,A, Rastgoufard,P & Al-Haddad, K 2007, 'An improved control algorithm for active filters', IEEE Trans. Power Del.,22(2): 1009–1020.

[14] Sommer,JR, Espinoza,JR &Moran,LA 2006, 'A non-linear control strategy for instantaneous power factor correction in 3- 4-wire electrical systems under asymmetrical and non-linear loads', in Proc. Int. Symp. Ind. Electron. 2: 1206–1211.

[15] Salem Rahmani, Abdelhamid Hamadi, 2009, 'A New Control Technique for Three-Phase Shunt Hybrid Power Filter', IEEE Transactions On Industrial Electronics, Vol. 56, No. 8.

[16] Karuppanan, P & Kamal Kanta Mahapatra 2010, 'PI with Instantaneous Power theory based shunt APLC for power quality', National conference on Emerging technological Trends (NCETT-2010).

[17] Karuppanan, P & Kamala kanta Mahapatra 2010, 'PID with PLL Synchronization controlled Shunt APLC under Non-sinusoidal and Unbalanced conditions', National Power Electronics Conference (NPEC) Proceedings, IIT-Roorkee.

[18] Kennedy, J, Eberhart, RC, 1995 'Particle Swarm Optimization' 1995 IEEE International conference on neural network, Vol.4 Piscataway, NJ Institute of Electrical and Electronics Engineers pp.1942-48.

[19] Kuo,H, Yeh,SN & Hwang,JC 2001, 'Novel analytical model for design and implementation of three-phase active power filter controller', Proc. Inst. Elect. Eng.—Electr. Power Appl., vol. 148, no. 4, pp. 369–383.

[20] Leszek,S, Czarnecki 2006, 'Instantaneous Reactive Power p-q Theory and Power Properties of Three-Phase Systems', IEEE Trans on Power Vol.21, No. 1, pp 362-367.

[21] Liao,JC & Yeh,SN 2000, 'A novel instantaneous power control strategy and analytical model for integrated rectifier inverter systems', IEEE Trans. Power Electron., vol. 15, no. 6, pp. 996–1006.

[22] Luis,FC, Monteiro, Jose,C, Costa,C, Mauricio Aredes, & JoaoLAfonso 2005, 'A Control Strategy for Unified Power Quality Conditioner' Brazilan power electronics conference.

[23] Singiresu S. Rao, 2009, " Engineering Optimizatoin: Theory and Practice", wiley publications.

Optimal Control of DC Motor using Linear Quadratic Regulator

Meetty Tomy* and A.G Thosar**
*M.E student, Government Engineering College, Aurangabad, India
meettymeetsaim@gmail.com
**Head of the Electrical Department, Government Engineering College, Aurangabad, India
aprevankar@gmail.com

Abstract: This paper provides the implementation of optimal control for an armature-controlled DC motor. The selection of error weighted Matrix and control weighted matrix in order to implement optimal control theory for improving the dynamic behavior of DC motor is presented. The closed loop performance of Armature controlled DC motor with derived linear optimal controller is then evaluated for the transient operating condition (starting). The result obtained from MATLAB is compared with that of PID controller and simple closed loop response of the motor.

Keywords: Optimal Control, Dc motor, Performance Index, MATLAB.

Introduction

With exploration of space research and missile technology, the complexity and demand of high accuracy for space probe and missiles introduced the world optimal control theory during the 1970s. At that time the challenge to scientists and engineers was to minimise the weight of satellites and missiles and control them accurately. Ref. [1] The linear quadratic control problem has its origins in the celebrated work of N.Wiener on mean –square filtering for weapon fire control during World war II (1940-45) . Ref. [2] Wiener solved the problem of designing filters that minimize a mean square error criterion of the form

$$J = E\ \{e^2(T)\}$$

Where e(t) is the error, and E{x} represents the expected value of the random variable x . Ref. [3] For a deterministic case, the above error criterion is generalized as an integral quadratic term as

$$J = \int e'(t)\ Q\ e(t)\ dt$$

Where, Q is some positive definite matrix. R. Bellman in 1957 introduced the technique of dynamic programming to solve discrete-time optimal control problems. Ref. [4] But, the most important contribution to optimal control systems was made in 1956 by L S Pontryagin and his associates, in development of his celebrated maximum principle Ref. [5]. In United States, R.E Kalman in 1960 provided linear quadratic regulator (LQR) and Linear quadratic Gaussian (LQG) theory to design optimal feedback controls. Ref. [6][7] He went on to present optimal filtering and estimation theory leading to his famous discrete Kalman filter and the continuous Kalman filter with Bucy Ref. [8]. Kalman had a profound effect on optimal control theory and the Kalman filter is one of the most widely used technique in applications of control theory to real world problems in variety of fields.

This paper presents optimal control of a DC motor such that it improves the transient response of the DC motor and reduces the steady state error in the motor speed. Section 2 details the optimal control theory. Section 3gives the state space model of DC motor derived from machine equations and fundamental laws of science. Section 4 compares the closed loop performance of motor without compensation, with PID Control action and with linear quadratic regulator. Section 5 summarizes all the significant conclusions drawn from the paper.

Optimal Control Theory

The main objective is to determine control signals that will cause a process(plant) to satisfy some physical constraints and at the same time extremize (maximize or minimize) a chosen performance criterion (performance index or cost function).Ref. [8] A designer aims at finding the optimal control vector that will drive the plant P from initial state to final state with some constraints on controls and states and at the same time extremizing the given performance index J .

The formulation of optimal control problem requires

1. A mathematical description (or model) of the process to be controlled (generally in state variable form),
2. A specification of performance index, and

3. A statement of boundary conditions and the physical constraints on the states and /or controls.

Performance Index

The typical performance criteria in classical control design are system time response to step or ramp input characterized by rise time, settling time, peak overshoot, and steady state accuracy; and the frequency response of the system characterized by gain and phase margins, and bandwidth.

Performance Index (PI) can be calculated or measured and used to evaluate system performance. Ref. [8] System performance is needed to be observed because of the need for – parameter optimization, design of optimum system and operation of modern adaptive control system.

The optimal control problem is to find a control which causes the dynamical system to reach a target or follow a state variable or trajectory and at the same time extremize a performance index(PI) which may take several forms described as :

1. PI for Time-optimal Control system: In minimum time system to be transferred from an arbitrary initial state $x(t_0)$ to a specified final state $x(t_f)$.

$$J = \int_{t0}^{t} dt$$

2. PI for fuel-optimal Control system: For minimization of the total expenditure of fuel. Assume magnitude $|u(t)|$ is proportional to the rate of fuel consumption, where $|u(t)|$ is the thrust of rocket engine.

$$J = \int_{t0}^{t} |u(t)| \; dt$$

3. Performance Index for Minimum-Energy Control system: For minimization of the total expended energy. Consider a resistive circuit where R_i is the resistance of the i_{th} loop and u_i is the current in that loop.

$$J = \int_{t0}^{t} \sum_{i=1}^{m} R_i \, |u_i(t)| \; dt$$

4. PI for terminal Control system: For minimization of error between the desired target position $x_d(t_f)$ and the actual target $x_a(t)$ at the final time t_f.
5. PI for general optimal control system

$$J = x'(t_f) F x(t_f) + \int_{t_0}^{t_f} [\, x'(t) Q x(t) + u'(t) R u(t)] \, dt$$

Where R is a positive definite matrix, Q and F are positive semi-definite matrix respectively. Q and R may be time varying. This particular form is called quadratic (in terms of states and controls)form.

Constraints

The control u(t) and state x(t) vectors are either unconstrained or constrained depending upon the physical situation. The unconstrained problem is less involved and gives rise to some results. From the physical considerations, we have controls and states, such as currents and voltages in an electrical circuit, speed of a motor, thrust of a rocket, constrained as

$$U_+ \leq u(t) \leq U_- \text{ and } X_- \leq x(t) \leq X_+$$

Where + and – indicate the maximum and minimum values the variables can attain.

Formal Statement Of Optimal Control System

The optimal control problem is to find the optimal control $u^*(t)$ (* indicates extremal or optimal value) which causes the linear time-invariant plant (system)

$$\dot{x}(t) = A\, x(t) + B\, u(t)$$

to give the trajectory $\dot{x}(t)$ that optimizes or extremizes(minimizes or maximizes) a performance index

$$J = x'(t_f)\, F\, x(t_f) + \int [\, x'(t)\, Q\, x(t) + u'(t)\, R\, u(t)\,] dt$$

Or which causes the non linear system $\dot{x}(t) = f(x(t),u(t),t)$ to give the state $x^*(t)$ that optimizes the general performance index

$$J = S(x(t_f),t_f) + \int V(x(t),u(t),t) dt$$

With some constraints on the control variable u(t) and/or the state variables x(t) . The final time t_f may be fixed or free, and the final (target) state may be fully or partially fixed or free.

Ref. [8] The optimal control systems are studied in three stages:

1. In the first stage, we consider the performance index and use the well known theory of calculus of variations to obtain optimal functions.
2. In second stage, we bring in the plant and try to address the problem of finding optimal control $u^*(t)$ which will drive the plant and at the same time optimize the performance index.
3. Finally, the topic of constraints on the controls and states is considered along with the plant and performance index to obtain optimal control.

Different Type of Systems

Different cases depending on the statement of problem regarding the final time t_f and the final state $x(t_f)$

1. Type (a) Fixed- final time and fixed- final state system: Here, since t_f and $x_(t_f)$ are fixed or specified and there is no extra boundary condition to be used other than those given in the problem formulation.
2. Type (b) Free-final Time and fixed-final state system: t_f is free or not specified in advance, δt_f is arbitrary and $x(t_f)$ is fixed or specified.
3. Type (c) Fixed-final time and free-final state system: Here t_f is specified and $x(t_f)$ is free.
4. Type (d) Free-final time and dependent free-final state system : t_f and $x(t_f)$ are related such that $x(t_f)$ lies on a moving curve with respect to t_f.
5. Type (e) Free-final time and Independent Free-final state: t_f and $x(t_f)$ are not related.

Open-loop Optimal Control

One has to construct or realize an open-loop optimal controller and in most cases it is very tedious. Also, changes in plant parameters are not taken into account by open-loop optimal controller which prompts us to think in terms of closed loop optimal controller i.e. to obtain optimal u*(t) in terms of state x*(t). Ref. [8] This closed loop optimal controller will have advantages such as sensitive to plant parameter variations and simplified construction of the controller.

Matrices in Performance Indices

1. The error weighted matrix Q(t) : In order to keep the error e(t) small an error squared non-negative, the integral of the expression 0.5 e'(t)Q(t)e(t) should be non-negative and small. Thus, the matrix Q(t) must be positive semi-definite. Due to the quadratic nature of the Weightage, large errors require more attention than small errors.
2. The control weighted matrix R(t): The quadratic nature of the control cost expression 0.5u'(t)R(t)u(t) indicates that larger control effect requires higher cost. Since the cost of the control has to be a positive quantity, the matrix R(t) should be positive definite.
3. The control signal u(t): The assumption that there are no constraints on the control u(t) is very important in obtaining the closed loop optimal configuration.
4. The terminal cost weighted matrix $F(t_f)$: The main purpose of this term is to ensure that the error e(t) at the final time t_f is small as possible. To guarantee this, the corresponding matrix $F(t_f)$ should be positive semi-definite.
   Further, without loss of generality, we assume that the weighted matrices Q(t), R(t) and F(t) are symmetric. The quadratic performance index has attractive features:
   - It provides an elegant procedure for the design of closed-loop optimal controller.
   - It results in the optimal feedback control that is linear in state function.
5. Infinite Final time : When the final time t_f is infinity, the terminal cost term involving $F(t_f)$ must be zero.

Closed-loop Optimal Control

To formulate a closed-loop optimal control i.e. to obtain the optimal control u*(t)

$$u^*(t) = - R^{-1}(t)\, B'(t)\, P(t)\, x^*(t)$$

which is now a negative feedback of the state x*(t). This negative feedback resulted from 'theoretical development' or 'mathematics' of optimal control procedure and not introduced intentionally .Fig. 1 shows a closed loop optimal control system with infinite final time which means that the system needs to be follow the desired signal for the whole processing time.

Matrix Differential Riccati Equation

P(t) is not dependent on the initial state which satisfies the matrix differential equation

$$\dot{P}(t) + P(t)\, A(t) + A'(t)\, P(t) + Q(t) - P(t)\, B(t)\, R^{-1}(t)\, B'(t)\, P(t) = 0$$

This is the matrix differential equation of the Riccati type and often called the matrix differential Riccati Equation (DRE). Also, P(t) is called the Riccati coefficient matrix or simply Riccati matrix or Riccati coefficient and the optimal control (feedback) law can be given as

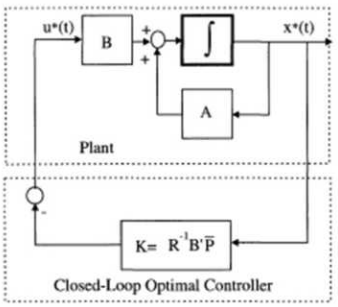


Figure 1: Implementation of closed-loop optimal control: Infinite final time

$$u^*(t) = - R^{-1}(t)\, B'(t)\, P(t)\, x^*(t)$$

The matrix DRE can also be written in a compact form as

$$\dot{P}(t) = - P(t)\, A(t) - A'(t)\, P(t) - Q(t) + P(t)\, E(t)\, P(t)$$

Where $E(t) = B(t)\, R^{-1}(t)\, B'(t)$

Symmetric Property of the Riccati Coefficient Matrix: An important property of Riccati matrix P(t) is that n x n matrix P(t) is symmetric for all $[t_0,t_f]$ i.e. P(t)= P'(t). Ref. [8] The matrices $F(t_f)$, Q(t) and R(t) are symmetric and therefore, the matrix $B(t)\, R^{-1}(t)\, B'(t)$ is also symmetric.

Salient features of state regulator system and matrix differential Riccati equation

1. Riccati Coefficient P(t) : P(t) is a time-varying matrix which depends upon the system matrices A(t) and B(t), the performance index (design) matrices Q(t), R(t) and $F(t_f)$ and the terminal time t_f but P(t) does not depend upon the initial state $x(t_0)$ of the system.
2. P(t) is symmetric and hence it follows that the n x n order matrix DRE represents a system of n(n+1)/2 first order non-linear, time-varying, ordinary differential equations.
3. Optimal Control: The optimal control u*(t) is minimum (maximum) if the control weighted matrix R(t0 is positive definite (negative definite).
4. Optimal state: Using the optimal control ,
   $u^*(t) = - R^{-1}(t)\, B'(t)\, P(t)\, x^*(t)$
   in state equation $\dot{x}(t) = A\, x(t) + B\, u(t)$
   we have,
   $\dot{x}(t) = [A(t) - B(t)\, R^{-1}(t)\, B'(t)\, P(t)]\, x^*(t) = G(t)\, x^*(t)$
   where
   $G(t) = A(t) - B(t)\, R^{-1}(t)\, B'(t)\, P(t)$
   The solution of this state differential equation along with the initial condition $x(t_0)$ give sthe optimal state x*(t). There is no condition on the closed-loop matrix G(t) regarding stability as long as the finite final time (t_f) system is considered.
5. Optimal Cost: The minimum cost J* is given by
   $J^* = (1/2)\, x^{*\prime}(t)\, P(t)\, x^*(t)$ for $t= [t_0,t_f]$
   Where, P(t) is the solution of matrix DRE and x*(t) is the solution of the closed-loop optimal system.
6. Definiteness of the Matrix P(t) : Since $F(t_f)$ is positive semi-definite, and $P(t_f) = F(t_f)$ this reflects that $P(t_f)$ is positive semi-definite, symmetric matrix.
7. Independence of Riccati Coefficient Matrix : The matrix P(t) is independent of the optimal state x*(t), so that once the system and the cost are specified, i.e. once the system/plant matrices A(t) and B(t) are given, and the performance index matrices $F(t_f)$, Q(t) and R(t), the matrix P(t) can be independently computed before optimal system operates in the forward direction from its initial condition.
8. Implementation of the Optimal Control: Fig.2 represents the block diagram implementing the closed-loop optimal controller as shown below.
   The figure shows clearly that the closed loop optimal controller gets its values of P(t) externally, after solving the matrix DRE backward in time from $t= t_f$ to $t = t_0$ and hence there is no way that teh closed loop optimal control configuration can be implemented.
9. Linear Optimal Control: The optimal feedback control u*(t) given as
   $u^*(t) = - K(t)\, x^*(t)$ where, the Kalman gain $K(t) = R^{-1}(t)\, B'(t)\, P(t)$

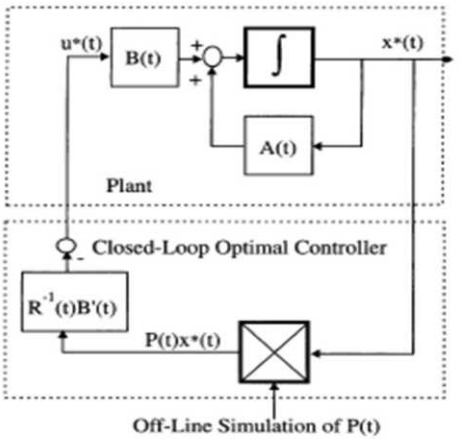


Figure 2: Closed-Loop Optimal Control Implementation

10. Controllability: As long as it is a finite time system, the need of controllability condition on the system for implementing the optimal feedback control is not there because the contribution of those uncontrollable states to the cost function is still a finite quantity only.

State Space Model of DC Motor

A physical plant can be described by a set of linear or non-linear differential equations. Fig.3 depicts the equivalent circuit of a DC motor armature is based on the fact that the armature winding has a resistance R_a, a self inductance L_a, and an induced emf V_b. Ref. [9] In case of a motor , the input energy is electrical energy and the output is the mechanical energy, with an air gap torque T_m at rotational speed ω_m. Armature controlled dc motor uses armature Voltage V_a as the control variable.

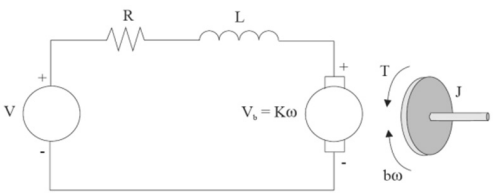


Figure 3: Model of a DC motor

The voltage equation of DC Motor is given as:

$$V_a(s) = R_a I_a(s) + L_a s I_a(s) + V_b(s)$$

In steady state , the armature current is constant and the rate of change of armature current is zero. The voltage equation of DC Motor is given as:

$$V_a(s) = R_a I_a(s) + V_b(s)$$

The air gap power is expressed in terms of the electromagnetic torque and speed as

$$P_a = \omega_m T_m = V_b I_a$$

Motor torque is given as

$$T_m(s) = K_m I_a(s)$$

In case of permanent magnet dc motor Φ is constant because the field winding here is replaced by permanent magnet. K_m is function of permeability of magnetic material and also known as the torque constant.

From Faraday's law,the induced emf if the armature conductors are divided into 'A' parallel paths (neglecting the sign) is

$$V_b = Z \Phi_f P N_r / (60 A)$$

There two possible arrangements of conductors in the armature, wave winding s and lap windings. A= 2 for wave windings and A= P for lap winding (P represents the number of poles of the motor). In compact form we have

$$V_b = K \Phi_f \omega_m$$

Where $\omega_m = 2\pi N_r/60$ rad/sec and $K = (P/A)Z(1/2\pi)$

If field flux is constant, then the induced emf is proportional to the rotor speed and the constant of proportionality is known as the induced emf constant or back emf constant. Then the induced emf is represented as

$$V_b = K_b \omega_m$$

Where K_b is the induced emf constant , given by $K_b = K \Phi_f$ Volt/(rad/sec)

For simplicity, the load is modelled as a moment of inertia,J,in Kg-m^2/sec^2, with a viscous friction coefficient B in N-m/(rad/sec). Then the acceleration torque Ta in N-m drives the load and is given by

$$J(d\,\omega_m/dt) + B\,\omega_m = T_m - T_l = T_a$$

Where T_l is the load torque.
The dynamic equations are cast in state space form and are given by
The state equation,

$$\begin{bmatrix} pI_a \\ p\omega_m \end{bmatrix} = \begin{bmatrix} \frac{-R_a}{L_a} & \frac{-K_m}{L_a} \\ \frac{K_m}{J} & \frac{-B}{J} \end{bmatrix} \begin{bmatrix} I_a \\ \omega_m \end{bmatrix} + \begin{bmatrix} \frac{1}{L_a} \\ 0 \end{bmatrix} V$$

Where p is the differential operator with respect to time. The above equation is expressed compactly in the form given by $\dot{x} = A\,x + B\,u$ where $x = [I_a\ \omega_m]'$, $u = [V]$, x is the state variable vector and u is the input vector[9].
The output equation is given by,
$\omega_m = [1\ \ 0]\ \omega_m$

Results

The LQR design minimizes a weighted squared state error and control effort to achieve a state feedback controller design. The optimal feedback state regulation, minimizes the quadratic cost function (performance index)
$J = \int (x^T(t)\,Q\,x(t) + u^T(t)\,R\,u(t))$ for [limits tending from 0 to infinity]
The specifications of separately excited DC Motor are shown in Table 1. The transfer function of the motor in closed loop is obtained from MATLAB script file given as

$$\theta(S)/Va(S) = \frac{0.023}{0.005\ s\text{^}2 + 0.01002\ s + 0.02356}$$

To control the speed of DC Motor, a traditional PID Controller is designed using trial and error method first. After several iterations, the values of Ki, Kp and Kd are chosen as Kp=10,Ki=10,Kd=5. Speed response of traditional PID controller is shown here in section B. The characteristic equation with PID controller is
0.005 s^3 + 0.01002 s^2 + 0.000559 s = 0
In state space representation the system matrices obtained from the MATLAB Script file and the motor specifications which are as follows
A= [-0.003 2.3 ; -0.046 -2]
B= [0;2]
C= [1 0]
D= [0]
Weighting matrices Q and R are selected as Q = [0.1 0,0 0.0001]' and R[0.1]. The step response of closed loop optimal control DC motor is shown in section C. Table 2 lists the performance of PID controller and LQR along with open loop system MATLAB Script file is used to describe the system performance. In order to compare the results of closed loop optimal controller with closed loop system and PID controller with MATLAB environment is used.
The analysis from section D and table 2 shows that the motor performance with optimal control is best suited because the performance measures such as settling time is the least for optimal control, no overshoot and zero steady state error. The Riccati coefficient, P obtained is, P = [0.0772 0.0487; 0.0487 0.0400], which is positive semi-definite, symmetric matrix. This Riccati coefficient is further used to obtain the optimal state x*(t) (here Ia i.e. Armature current and ω_m i.e. angular speed of the armature). State feedback gain matrix, K= [0.9749 0.8009]. The feedback gain matrix is then used to obtain the optimal control u*(t) i.e. V (Armature voltage). Desired pole location at which system is optimal is, E = [-1.8024 + 1.1630i;-1.8024 - 1.1630i]. The optimal performance index can then be obtained from the optimal state x*(t) and Riccati coefficient P. It is clear from the results that the motor speed tracks the desired signal with good transient response. Here the results of the control system are response for step change in desired signal.

Table 1. Separately excited DC Motor Specifications

DC motor specifications		
Particulars	Value	Unit
Moment Of Inertia, J	0.01	Kg-m^2
Back emf constant & Torque constant, Ka	0.023	Nm/A
Viscous friction constant, b	0.00003	Nms
Armature resistance, R	1	ohm
Armature Inductance, L	0.5	Henry

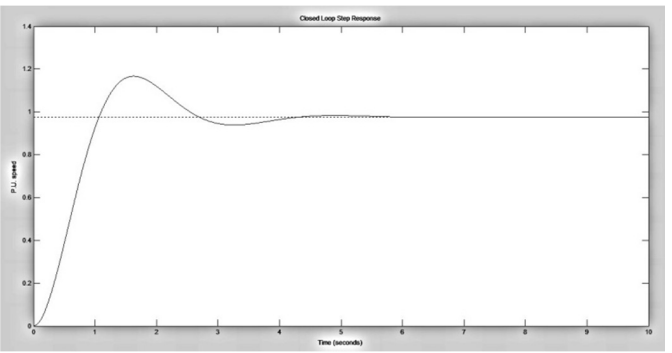


Figure 4. Step response of closed loop system

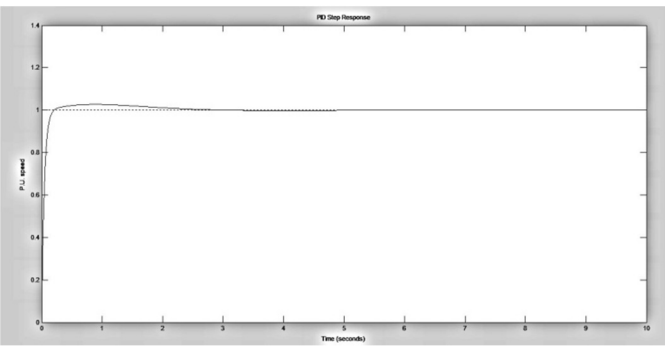


Figure 5. Step response of closed loop system with PID control action

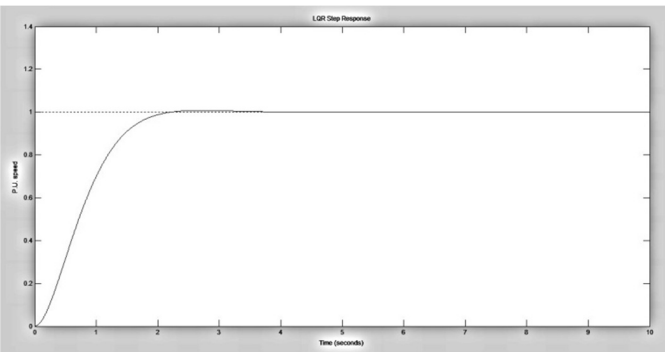


Figure 6. Step response of closed loop optimal control system

Table 2. Comparison of Results

System Performance	Comparison		
	Closed loop system	Closed loop system with PID controller	Closed loop optimal controller
Settling time, Ts	5.77 sec	3.19 sec	2.24 sec
Maximum Overshoot, Mp	Amplitude-1.17 At 1.62 sec	Amplitude – 1.03 At 0.748 sec	0
Steady state error,e_{ss}	0.023	0	0

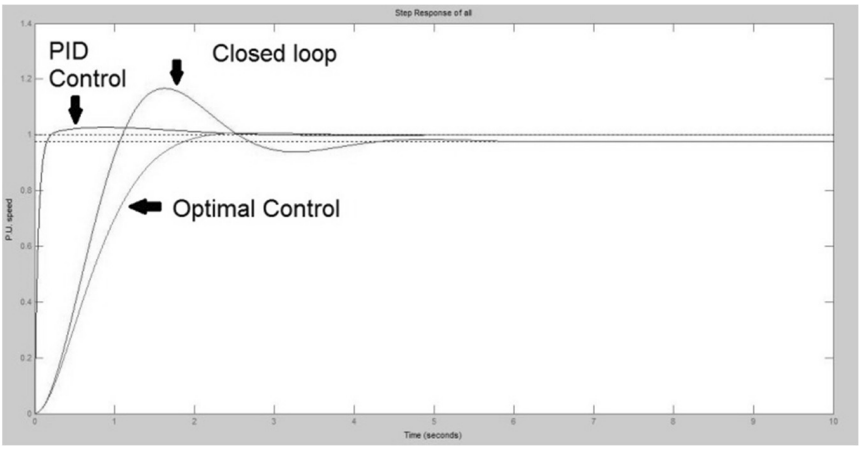


Figure 7. Step response of closed loop, motor with PID control and optimal control

Conclusion

This paper proposes an approach to control design of a DC motor based on LQR control design. The mechanical and electrical parameters of DC motor are used to obtain the response for the system. The LQR design provides an optimal state feedback control that minimizes the quadratic error and control effort. On comparison between the simple closed loop system, closed loop system with PID control action and closed loop optimal controller, the transient response and steady state error in the response due to closed-loop optimal controller yields the best result.

References

[1] N. Wiener. Extrapolation, Interpolation, and Smoothing of Stationary Time Series. Technology Press, Cambridge, MA, 1949.
[2] N. Wiener. Cybernetics. Wiley, New York, NY, 1948.
[3] R.E Bellman ,Dynamic Programming, Princeton University Press. Princeton, NJ,1957
[4] V.G Boltyanskii, R V Gamkrelidze and L.S Pontryagn.On the theory of optimal processes.Dokl.Akad.Nauk SSSR, 110:7-10,1956
[5] R.E Kalman, Anew approach to linear filtering in prediction problem,ASME journal of asic engineering,82:34-45,March 1960
[6] R.E Kalman and R.S Bucy. New results in linear filtering and prediction theory. Transaction ASME J.Base Engineering,83:95-107,1961.
[7] L S Pontryagin, V G Boltyanskii, R V Gamkredlize and E F Mishchenko.The mathematical theory of optimal processes,Wiley Interscience,NewYork,NY162.
[8] Desineni Subbaram Naidu, Optimal Control Systems,CRC Press LLC, 2000 N.W. Florida 33231,TJ213.N2655 2002,ISBN 0-8493-0892-5.
[9] R. Krishnan,Electic Drives- Modeling, Analysis and control, ISBN-978-81-203-2168-7,PHI Leaning Pvt Ltd

A Review on Implementation of UART with Optimum Noise Immunity

Pallavi R. Upase* and P.S. Choudhari**
* M.E.Student, Department of EEE,
Prof. Ram Meghe College of Engineering Administration and Management, Amravati, Maharashtra, India
p.anil.deshpande@gmail.com
** HOD, Department of EXTC, Prof. Ram Meghe College of Engineering Administration and Management, Amravati, Maharashtra, India
pschoudhari@gmail.com

Abstract: For asynchronous transmission of serial data between a DTE (data terminal equipment) and a DCE (data communication equipment) a UART (universal asynchronous receiver/transmitter) is used. It is a serial communication protocol widely used for short-distance, low speed, low-cost full duplex data exchange between computer and an output device. In working environments employing multiprocessor communication such as industries, the received data may be erroneous due to noise. The consequences may be hazardous due to improper control. The present day UART are designed with an error detection logic which requests retransmission of data frame from the transmitter if any error is detected. This task demands extra time. In the proposed paper we have reviewed the various approaches made so far for detecting and correcting errors, and improving the noise immunity of the system in order to obtain an optimized error free reception of data.

Keywords: UART (Universal Asynchronous Receiver/ Transmitter), BIST (Built in Self Test), FEC (Forward Error Correction), Hamming Codes, SEC-DED (Single error correction- Double error detection).

Introduction

The UART simultaneously perform transmit and receive functions as the transmitter and receiver have separate clock signals and control signals. Also they share a bidirectional data bus, which allow them to operate virtually independently of one another. The primary functions performed by a UART are [1]:

- Parallel to serial data conversion in the transmitter and serial to parallel data conversion in the receiver
- Error detection by inserting parity bits in the transmitter and checking parity bits in the receiver
- Insert start and stop bits in the transmitter and detect and remove start and stop bits in the receiver
- Formatting data in the transmitter and receiver
- Provide transmit and receive status on CPU
- Voltage level conversion between the DTE and the serial interface and vice-versa
- Provide a means of achieving bit and character synchronism.

During the data communication, transmission errors may occur due to electrical interference by nearby sources such as fluorescent lights, motors, generators, power lines etc or by some system discrepancies. In this paper we have presented an overview of different methods to control these errors proposed by various authors so far.

Data frame format of UART

UART has a transmitter, a line control register (LCR), a baud rate generator (BRG), and a receiver. The nature of the data is to be specified before transmitting the data in either direction through an UART. This is done by programming an eight-bit control word into the UART line control register (Fig.1.) at the transmitter end.

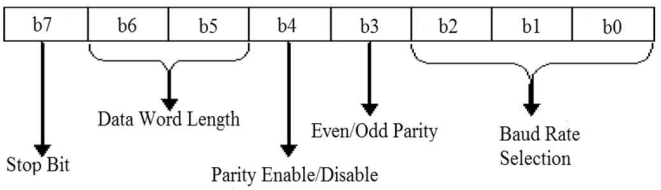


Fig.1. Line control register (control word register) bit description

Table 1 specifies exact logic levels of selection lines for selecting different baud rates, odd/even parity and data word length.

Table 1 : LCR Bit Description

BIT	Value			Description
0,1,2	b2	b1	b0	Baud rate selection
	0	0	0	57600
	0	0	1	38400
	0	1	0	19200
	0	1	1	9600
	1	0	0	4800
	1	0	1	2400
	1	1	0	1200
	1	1	1	600
3	0			Even Parity
	1			Odd Parity
4	0			Parity Disable
	1			Parity Enable
5,6	b6	b5		Data Word length
	0	0		5
	0	1		6
	1	0		7
	1	1		8
7	0			1 Stop bit
	1			2 Stop bits

An 11 bit frame format of UART can be as seen in Fig. 2. The signal line is assigned with two states- using logic high and logic low to distinguish a '1' and '0' respectively.

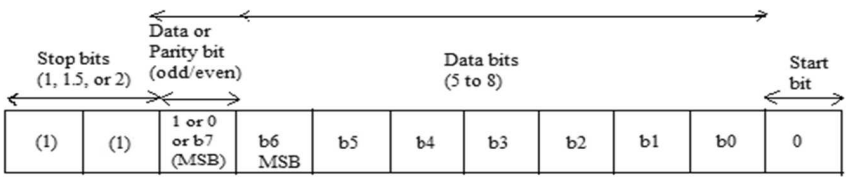


Fig. 2. UART data frame format [1]

UART frame format consist of a start bit, data bits, parity bit and stop bit. After the Start Bit, the Data Bits are sent, with the Least Significant Bit (LSB) sent first. The start bit is always low and the stop bit is always high. When the complete data word has been sent, it adds a Parity Bit .This parity bit may be used by the receiver to perform error checking. Then 1 or 1.5 or 2 Stop Bits is/are sent by the transmitter. Because asynchronous data are "self-synchronizing", if there is no data to transmit, the transmission line will be idle.

Sources of errors in UART

With asynchronous data, the transmit and receive clock may not be continuously synchronized. UART bit synchronization is achieved by establishing a timing reference at the centre of each start bit. Therefore, it is imperative that a UART detect the occurrence of a valid start bit early in the bit and establish a timing reference before it begins to accept data. The difference in time between when a sample is taken (i.e., when a data bit is clocked into the receive shift register) and the actual centre of data bit, is called the sampling error as shown in Fig. 3.

The difference in time between the beginning of the start bit and when it is detected is called the detection error. The higher the receive clock rate, the earlier a start bit would be detected. Because of the detection error, successive samples occur slightly off from the centre of the data bit. With asynchronous clocks, the magnitude of the sampling error for each successive sample would increase (the clock would slip over or slip under the data), eventually causing a data bit to be sampled twice or not sampled at all, depending on whether the receive clock is higher or lower than the transmit clock. [1] Hence the received data may be erroneous.

Errors can also occur if the transmission medium is not an ideal one. However internal defects, interference, channel noise, etc. keeps the transmission media far from ideal.

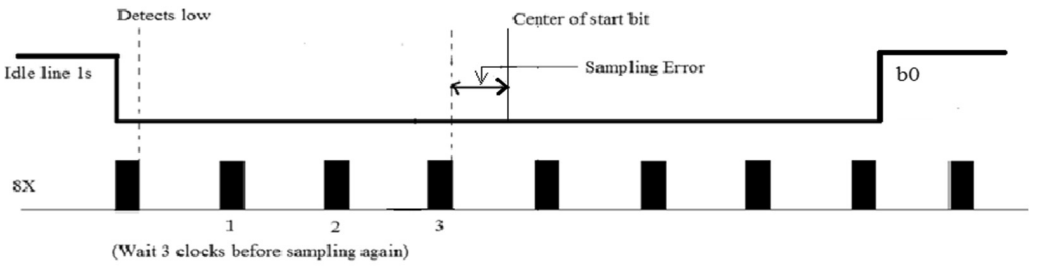


Fig.3. Sampling error

Data communication errors can be classified as single bit, multiple bits or burst. A single bit is when only one bit within the data string is erroneous. A multiple bit error occurs when two or more non-consecutive bits within the data string are in error. Whereas when two or more consecutive bits within the string are erroneous, it is a burst error. The length of burst errors is measured from the first corrupted bit to the last corrupted bit in the string. Some errors in between may not be corrupted ones. So far designers have developed two basic strategies for handling transmission errors; error detecting codes and error correcting codes. Parity bits, block and frame check characters and cyclic redundancy characters are examples of error detecting codes which include enough redundant information with each transmitted message to enable the receiver to determine when an error has occurred. Error correcting codes include sufficient extraneous information along with each message so that the receiver can determine in which bit the error has occurred and hence correct it.

Previous Work Done for Dealing with Errors

Mohd Yamani Idna Idris, Mashkuri Yaacob and Zaidi Razak [9] have implemented the BIST (Built IN Self Test) technique in the UART design which is a method of self testing on a system on chip. For testing sequential networks, we need to observe the state of flip-flops instead of the outputs only. The network output needs to be verified for each input combination and each state of flip-flop before going to the next correct state. This can be done by arranging the flip-flops to form a shift register. As seen in Fig.4, “Register A” and Register B” are configured by mode control bits B1 and B2. The state of the flip-flop is shifted out bit-by-bit using a single serial-output pin ’Q’ on the IC. This is called scan path testing.

A built-in-logic-block-analyser (BILBO) architecture is a scan register that can be modified to serve as a state register, a pattern generator, a signature register, or a shift register by setting the bits B1B2 of the BILBO scan register as shown in Table 2.

This technique checks if any logic errors present in the UART. Both transmitter as well as the receiver of the UART is checked by this method. The LFSR replaces the function of the external tester features such as a test pattern generator by automatically generating pseudo random patterns to give fault coverage to the UART module. The MISR acts as a compression tool, compressing the output result when automatic pseudo random pattern is fed to the UART. The shift register minimizes the input/output overhead by shifting the parallel signature produced by MISR into serial signature. Although time consuming (receiver has to wait for the signal from the transmitter) this method takes advantage of 100% fault coverage [10].

Table 2: BILBO Operating Modes

B1B2	Operating Mode
00	Shift Register
01	Linear feedback shift register(LFSR)/ Pseudo Random Pattern Generator (PRPG)
10	Normal
11	Multiple Input Signature Register (MISR)

Himanshu Patel, Sanjay Trivedi, R. Neelkanthan and V. R. Gujraty [3] have utilized recursive running sum filter to remove noisy samples from data samples at the receiver. Recursive Running Sum (RRS) is simple low pass filter. Input data signal is directly sampled with system clock and samples are accumulated over a window size. Serial receive data signal is directly sampled with system clock and samples are fed to RRS filter. The window size of the filter is user programmable and it decides baud rate. For correct data decoding, the selection of window size (down sampling factor-M) is important.

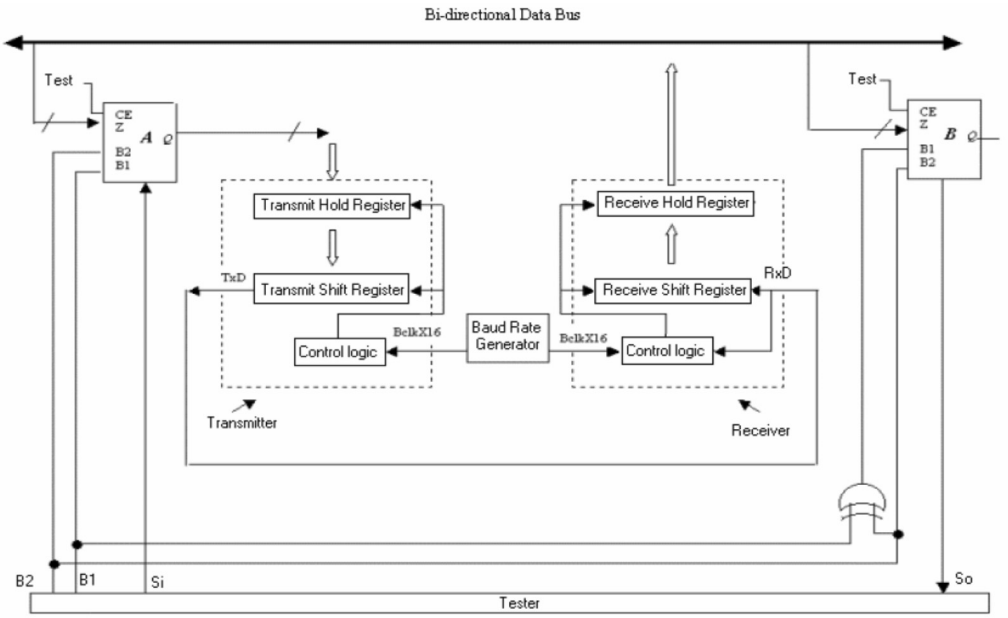


Fig.4. UART with BILBO register and Tester [9]

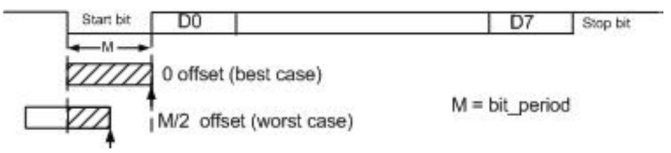


Fig 5.a Window offset when M = bit_period

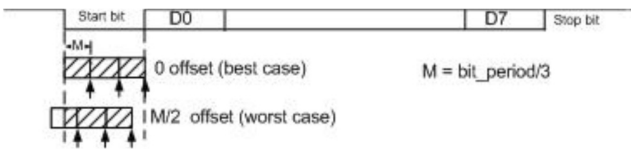


Fig 5.b Window offset when M = bit_period/3

For application of this filter in the asynchronous communication where data bits are transmitted asynchronously, the offset between window and start bit is important. When the window size (down sampling factor- M) is equal to bit period, different situations of offset between window and start bit is shown in the Fig-5.a. The window size is user programmable and it is set to one third of required bit period as shown in Fig.5.b. The baud rate equation is as follows.

$$Baud\_rate = \frac{clock\_freq}{M*3} \qquad (1)$$

Where, *clock_freq* is system clock frequency, M is window size, (1 to 2^{16}).
As the baud rate is decided by the window size so there is no need of any external "timer module" which is normally required for standard UARTs. The intermediate data bit is decoded using magnitude comparator as seen in Fig. 6.

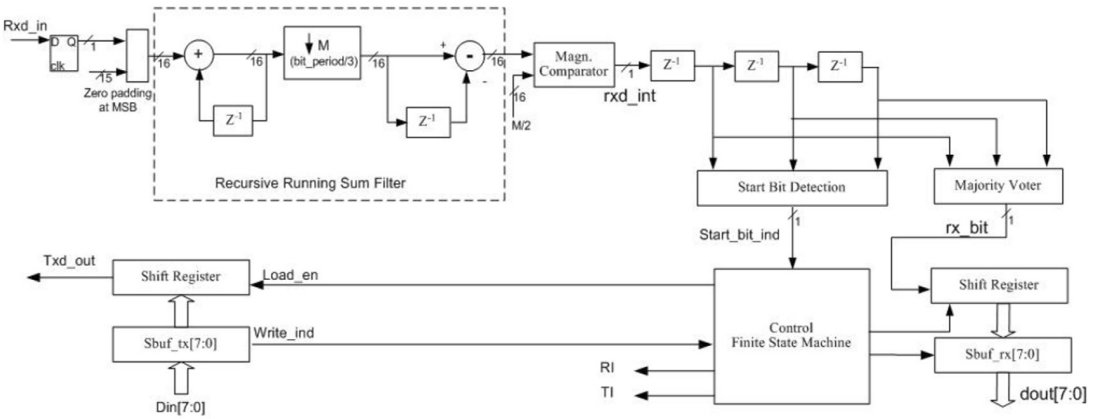


Fig.6. Robust UART architecture using running sum filter.[3]

A majority voter is used to decode actual data bit from three intermediate data bits. As three samples are available in a bit period, in the worst case too, two full windows are available in one bit period. Therefore, two out of three samples are correct. Hence the bit will be correctly decoded by the majority voter.
Comparison of simulation results at different noise level obtained by the authors, show that the robust UART described had far better performance than standard UART at higher noise levels as seen n Fig. 7. Percentage Ratio of correctly received bytes and total transmitted bytes is represented on the x-axis of graph. The y-axis of the graph indicates the percentage of average corrupted samples in a bit period. It should be noted that it is impossible to recover data bits if more than 50% of data samples in a bit period are corrupted by noise. The graph clearly indicate that the performance of standard UART deteriorates if more than 6% of data samples are corrupted in a bit period, while for the proposed robust UART the performance deteriorates only after 37%.

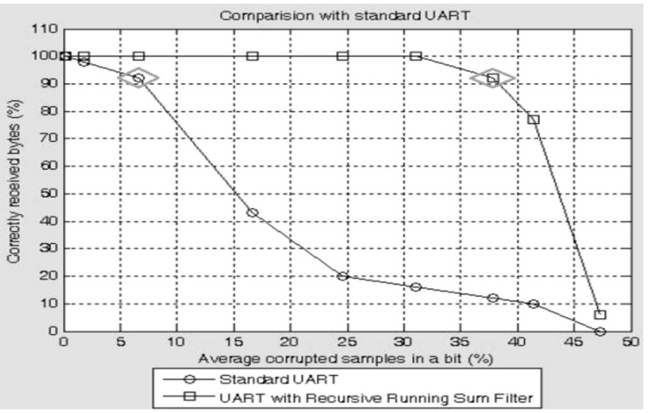


Fig. 7 Comparison with standard UART at different noise levels

Naresh Patel, Vatsalkumar Patel and Vikaskumar Patel [2] have proposed architecture of UART using the Status register which indicates parity error, framing error, overrun error and break error during the data reception.
In Fig.8, the error logic block handles four types of errors. PL bit will be set if the received parity does not match with the parity generated from data bits which indicates that parity error occurred. SL bit is set if frame error occurs when receiver fails to detect correct stop bit or when 4 samples do not match. OL bit is set when overrun error occurs if the receiver FIFO register is full and other data arrives at the RHR (receiver hold register). BL bit is set when there is a break in received data and break error occurs i.e. if the RXIN pin is held low for longer time than the frame time [2].

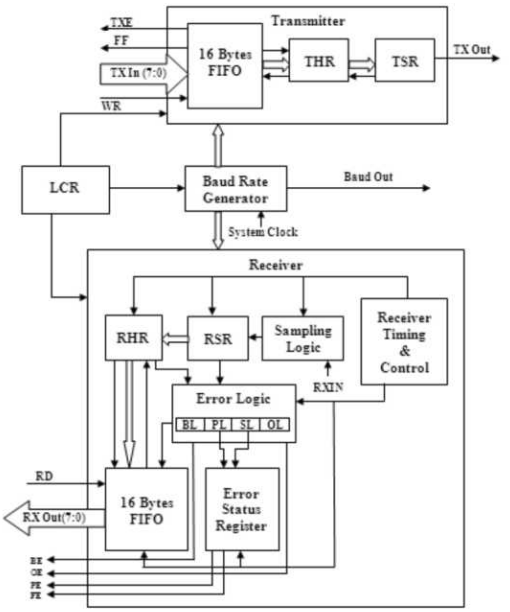


Fig.8. UART architecture with status register [2]

In the methods discussed so far, the major drawback of the logic used is, if an error is detected in received data, retransmission of corresponding data frames is required which take additional time for automatic repeat request (ARQ) and retransmission of data.

Sindhuaja Muppalla and Koteswara Rao Vaddempudi [5] have proposed a novel technique which improves the noise immunity of the UART system. They have implemented the extended hamming codes which have forward error correction (FEC) capability. These are also called Single Error Correction and Double Error Detection (SEC-DED) codes.

D3	D2	D1	D0	P0	P1	P2	P3

Fig. 9. a.-, Data format of (8,4) Hamming code

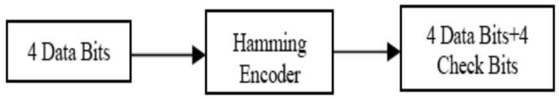


Fig. 9. b. Extended(8,4) Hamming code generation

In Fig.9.a- D0 to D3 are the data bits whereas P0 to P3 are the redundancy bits also called the Parity check bits formed by following equations [5]:

$$P3 = D0 \oplus D1 \oplus D2 \quad (2)$$
$$P2 = D0 \oplus D1 \oplus D3 \quad (3)$$
$$P1 = D0 \oplus D2 \oplus D3 \quad (4)$$
$$P0 = D1 \oplus D2 \oplus D3 \quad (5)$$

Hamming codes are used for error detection and correction. Hamming encoder (Fig. 9.b) generates the redundancy bits using a formula:

$$2^r \geq D+r+1 \quad (6)$$

In equation (6), r = number of redundancy bits and D = number of information data bits.

Redundancy bits are those extra bits which are required to detect and correct errors. The redundancy bits are added to the information bit at the transmitter and removed at the receiver. The receiver is able to detect the error and correct it because of these redundancy bits. [6]

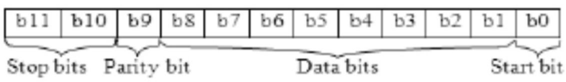


Fig. 10.a- Frame format for error detection mode

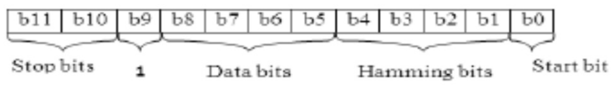


Fig. 10.b- Frame format for error correction mode

The frame formats used in this hamming code base architecture is as shown in Fig.10.a and Fig.10.b for the error detection and error correction modes respectively. The status register in the two modes are as shown in Fig.11.a and Fig.11.b. [6].

BE	PE	FE	OE

Fig.11.a. Error status register in error detection mode

BE	NE	SEC	DED	OE

Fig.11.b. Error status register in error correction mode

The different flags are set as follows:

BE : Break error – set when break in the received data string occurs.

PE: Parity Error- set when the received parity does not match the generated parity at the transmitter end.

FE: Frame error –set when the receiver fails to detect the correct stop bit

OE: Overrun error- set when data arrives at the receiver Hold register even though the FIFO register in receiver is full.

NE: No error- set when no errors occur (desired)

DED: Double error detected- set when two bit in the data string are erroneous

SEC- Single error corrected- set when only one erroneous bit is detected and consequently corrected by changing it.

At the receiver end, in the Hamming decoder, the syndrome is decoded to identify the error position if any, and the error is corrected if it is a single error and the DE flag is set if in case a double error is detected.

The decoding algorithm is as shown by following 4 equations:

$$S0 = D3 \oplus D2 \oplus D1 \oplus D0 \oplus P0 \oplus P1 \oplus P2 \oplus P3 \quad (7)$$
$$S1 = D3 \oplus D2 \oplus D1 \oplus P0 \quad (8)$$
$$S2 = D3 \oplus D2 \oplus D0 \oplus P1 \quad (9)$$
$$S3 = D3 \oplus D1 \oplus D0 \oplus P2 \quad (10)$$

All possible values of S0, S1, S2, S3 and their corresponding meanings are represented in Table 3.

Table3. Error Logic Register

S0	S1S2 S3	NE/SE/DE	Error Position	Flag set
0	000	No Error	-	NE
1	011	SE	D0	SEC
1	101	SE	D1	SEC
1	110	SE	D2	SEC
1	111	SE	D3	SEC
0	001	DE	Not predictable	DED
0	010	DE	Not predictable	DED
0	100	DE	Not predictable	DED
0	011	DE	Not predictable	DED
0	101	DE	Not predictable	DED
0	110	DE	Not predictable	DED
0	111	DE	Not predictable	DED

The test bench waveforms observed by the authors for the transmitter are as shown in Fig. 12.a . The waveforms for the RSR (receive shift register) and RHR (receive hold register) of the receiver part are given in Fig. 12.b and Fig.12.c.

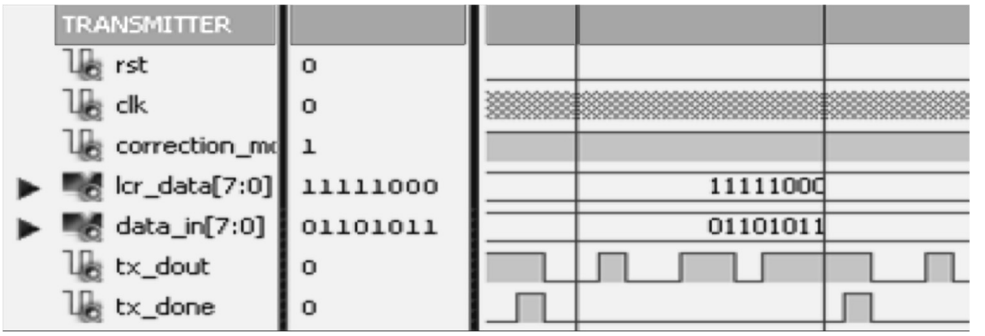


Fig 12.a Test bench waveforms for transmitter

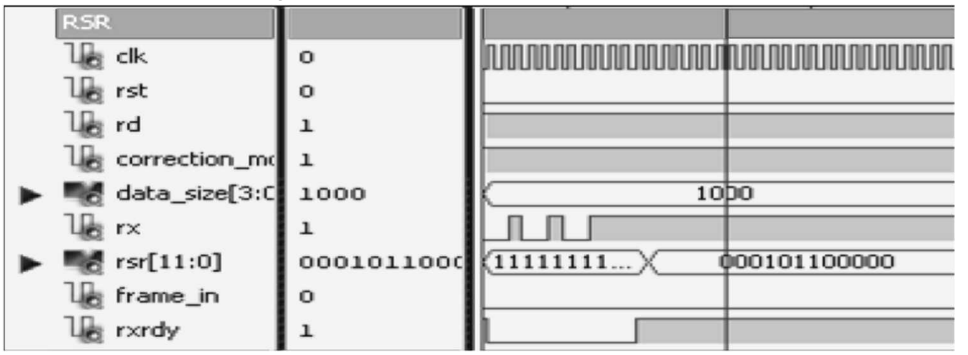


Fig. 12.b Test bench waveforms for Receive shift register

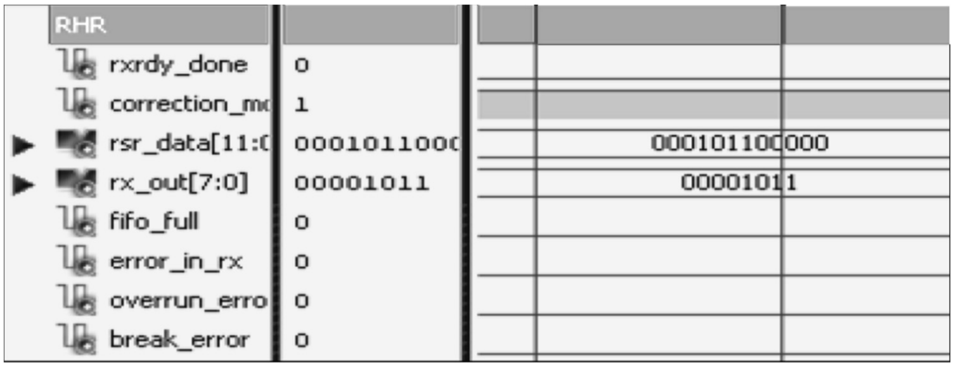


Fig 12.c Test bench waveforms for the receive hold register

As seen in Fig 12.a, only the 4 LSBs out of the 8 bit data "1011" (D3 to D0) are transmitted. The frame is formed starting with start bit '0', hamming bits "0100" (P0 to P3) and a '1' followed by two stop bits '11'. The RSR (receive shift register) received above data with an error in bit position D1. This error is corrected and the data output from RSR is given with all bits, except the data bits in the frame, replaced with zeros as "000101100000" shown in Fig12.b.The RHR (receive hold register) extracts the four data bits "1011" concatenated with "0000" in MSB position giving out a data byte as indicated in Fig.12.c.

Wilfried Elmenreich and Martin Delvai [7] have examined applicability of common UARTs in time-triggered systems and developed mathematical means to guarantee that a UART frame stays within its given time slot so that the communication is reliable.

A typical time-triggered protocol (TTP) defines a network cycle during which each node is given a dedicated time window during which it may broadcast a message. Each cycle begins with a reference message that contains the schedule for the upcoming cycle. Time-triggered messages are scheduled into exclusive windows, to ensure that there are no collisions. With no collisions occurring, the time-triggered approach is able to guarantee delivery time, and hence be deterministic. Today most microcontrollers already include a unit for serial transmission. TTP use a time triggered communication schedule to achieve predictable timing behavior and provide clock synchronization for imprecise cheap on-chip oscillators [11]. To achieve the intended goal of predictable communication, it must be ensured, that the UART communication of a microcontroller with an on-chip oscillator can hold the timing requirements imposed by the communication protocol, although UARTs were not originally designed for this application class.

The authors have studied the timing properties of a UART communication. The baud rate of a UART is usually configured by integer values - the arithmetic rounding error leads to baud rate deviations. The architecture of common UARTs furthermore leads to intrinsic delays at the sending UART. Moreover one has to regard the qualities of the clocks of the communication partners. Five timing deviations were observed [7]:

i. **Arithmetic error in baud rate setting**: If we want to set the UART to a given (ideal) baud rate, for every clock frequency *fclk* we can determine a UART Baud Rate Setting (UBRS) that best approximates the desired baud rate. Thus we can give an upper bound for the ratio between the fastest and the slowest approximated baud rate.

ii. **Send jitter problem**: At initialization of the UART, the baud rate generator is started. Running at a frequency corresponding to the configured baud rate, the baud rate generator periodically generates ticks which are possible start events of a message bit. When the UART receives a transmit signal, it starts the transmission at the next tick from the baud rate generator. Thus, depending on the internal state of the baud rate generator the transmission of a message may be delayed up to the interval time of two subsequent baud rate generator ticks. Usually the state of the baud rate generator cannot be read, leading to an indeterministic send delay jitter.

iii. **Clock drift**: The components performing a UART communication are set apart from each other and are clocked by different clock sources. To synchronize these clocks periodically, clock synchronization algorithms are used. Clocks can drift apart from each other after synchronization, depending on their drift rate. The drift of a physical clock is the ratio between the actual clock frequency *fclk* and the reference frequency *fref.* Good clocks have drifts that are very close to 1, so the drift rate ρ is introduced as in equation 11:

$$\rho = \frac{f_{clk}}{f_{ref}} - 1 \quad \text{.. (11)}$$

A perfect clock will have a drift rate of 0. Real clocks have varying drift rates depending on temperature, voltage variations, or aging effects. The baud rate depends on the frequency of the UART clocking and is therefore influenced by the clock drift.

iv. **Clock offset**: This problem refers to the offset between the particular clocks of the communicating nodes. The offset of the local clock affects the instant when the communication partners send or expect to receive a message.

v. **Signal runtime**: Electric signals in a cable travel at approximately 2/3 rd of the speed of light. This results in a delay as given in equation 12:

$$t_{Signal} \approx \frac{l}{2 \times 10^8 \frac{m}{sec}} \quad \text{.....(12)}$$

Where, *l* is the length of cable between two transceivers.

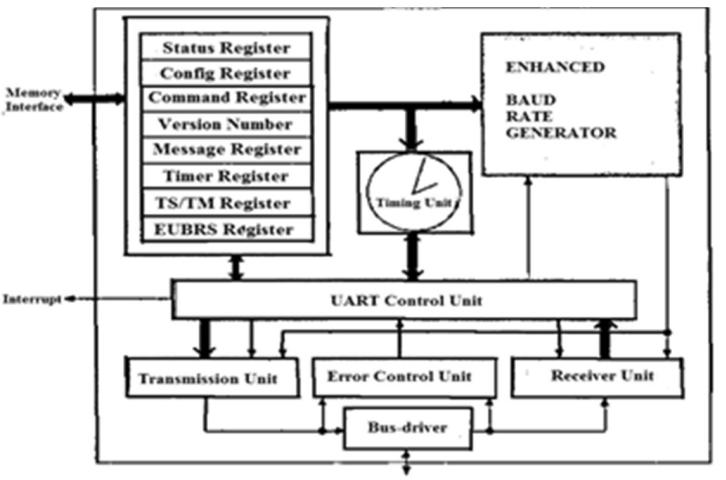


Fig. 13 Block diagram of UART I/O Module for Time Triggered Systems

When the send instant is a priori known to the sender and collisions between subsequent UART frames are obviated, the signal delay can be compensated by sending the time of minimal delay earlier. The condition for a successful and correct communication is given in Equation 13, where *n* is the number of bits in a frame including the start, stop and parity bits and BR is the Baud Rate:

$$\frac{BR_{fast}}{BR_{slow}} < \frac{n}{n - 1/2} \quad \text{..... (13)}$$

Fig.13 illustrates the block diagram of the UART I/O module proposed by Wilfried and Martin. This UART component overcomes the disadvantages of send jitter and arithmetic error in baud rate setting. It starts immediately after receiving the transmit signal with the message transmission. In this way the send jitter is completely eliminated.

Conclusion

We have analyzed that the received data may be erroneous due to three basic reasons:

(i) The noise induced by external nearby electric machines as dealt in [2] with the help of status register, in [3]with recursive running sum filter and in [5] using encoder and decoder;

(ii) The faults in the system on chip (gate level) as handled in [4] by implementing built-in self test technique; and

(iii) Timing problems due to imprecise on chip oscillators as tackled in [7] by adjusting the baud rate of transmitter and receiver.

We conclude that keeping the difference between the baud rate of the transmitter and receiver to minimum; keeping the receiver clock rate slightly higher than the transmitter; and, employing the extended Hamming code encoder and decoder at transmitter and the receiver end respectively can optimize error free reception of data, hence making the UART more immune to noise. Further rearranging the data bits as N columns and sending first bit of each followed by second bit of each, before implementing Hamming scheme may enable us to deal with burst errors; or modifying the Hamming matrices may facilitate to tackle adjacent errors, making the data more impervious to noise.

References

[1] Wayne Tomasi – "Electronic Communications systems - fundamentals through advanced"- fifth edition, Pearson Prentice Hall

[2] Naresh Patel, Vatsalkumar Patel, Vikaskumar Patel, "VHDL Implementation of UART with Status Register", 2012 IEEE International Conference on Communication Systems and Network Technologies, 2012.

[3] Himanshu Patel, Sanjay Trivedi, R. Neelkanthan, V. R. Gujraty , "A Robust UART Architecture Based on Recursive Running Sum Filter for Better Noise Performance", Held jointly with 6th International Conference on Embedded Systems, 20th International Conference on VLSI Design, pp. 819-823, Jan. 2007.
[4] Idris, M.Y.I, Yaacob M, "A VHDL implementation of BIST technique in UART design", 2003 Conference on Convergent technologies for Asia-Pacific Region (TENCON), pp. 1450-1454, 2003.
[5] Sindhuaja Muppalla, Koteswara Rao Vaddempudi," A Novel VHDL Implementation of UART with Single Error Correction and Double Error Detection Capability", SPACES-2015, Dept of ECE, KL UNIVERSITY
[6] Nutan Shep, Mrs. P.H. Bhagat," Implementation of Hamming code using VLSI", International Journal of Engineering Trends and Technology- Volume4Issue2- 2013
[7] Wilfried Elmenreich, Martin Delvai, "Time-Triggered Communication with UARTs", 4th IEEE International Workshop on Factory Communication Systems, Vasteris, Sweden, August 28-30, 2002
[8] Dipanjan Bhadra, Vikas S. Vij, Kenneth S. Stevens," A Low Power UART Design Based on Asynchronous Techniques", IEEE 978-1-4799-0066-4/2013
[9] Mohd Yamani Idna Idris, Mashkuri Yaacob, Zaidi Razak, "A VHDL Implementation of UART Design With BIST Capability", Malaysian Journal of Computer Science, Vol. 19 (1), 2006, pp 73-86
[10] Dr. T.V.S.P.Gupta,Y. Kumari, M. Ashok Kumar," UART realization with BIST architecture using VHDL " International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 3, Issue 1, January -February 2013, pp.636-640
[11] H. Kopetz, W. Elmenreich, and C. Mack, "A comparison of LIN and TTP/A", Proceedings of the 3rd IEEE International Workshop on Factory Communication Systems, Porto, Portugal, pages 99-107, September 2000.

Introduction to Energy Efficient Motors

Himanshu Shekhar* and C R Bharath**
* Exe. Engineer, L&T Power
himanshu.shekhar@lntpower.com
** Asst. Engineering Manager, L&T Power
crbharath@lntpower.com

Abstract: Selection, Design, Manufacturing, Testing & Installation of Motor is becoming critical day by day because of huge capital cost & energy saving potential involved into it. Motors manufactured prior to 1975 were designed and constructed to meet minimum performance levels as a trade-off for a lower purchase price. Efficiency was maintained only at levels high enough to meet the temperature rise restrictions of the particular motor. In 1977, National Electrical Manufacturers Association (NEMA) recommended a procedure for labeling standard three-phase motors with an average nominal efficiency.

Motor efficiency depends on mechanical and electrical imperfections of the motor. Study found that Resistance ($I^2 R$) losses in the stator windings and rotor bars can constitute up to a 15 percent loss in efficiency in three-phase motors. Magnetization losses in the stator and rotor cores cause about a 1 percent to 7 percent efficiency loss. Friction losses in the bearings and inefficiency in the cooling fans result in 0.5 percent to 1.5 percent loss in motor efficiency. Friction and magnetization losses are independent of motor load and relate solely to motor size and design. The remaining losses are referred to as stray load losses.

To minimize the above losses **Energy Efficient Motors of Efficiency Class IE1, IE2 (High), IE3 (Premium), IE4 (Super Premium)** have been introduced in the Industry. For an example, a 4 –Pole 375 kW IE3 Energy Efficient Induction Motor shall have 96 percent Nominal Efficiency according to IS 12615:2011. Thus, a motor that is 96 Percent efficient converts 96 percent of the electrical energy input into mechanical energy. The remaining 4 Percent of the electrical energy is dissipated as heat, evidenced by a rise in motor temperature. The higher efficiency is attained by improved motor design and using high quality materials to reduce motor losses.

When we consider energy efficient motors, two factors will affect the payback period: power cost and operating hours per year. Where electricity is inexpensive or operating time is low, it may take several years for the savings from installation of high efficiency motors to outweigh the difference in initial cost. On the other hand, where power costs and the operating hours per year are high, it may be possible to replace an existing standard efficiency motor with an energy efficient motor and realize a payback in less than one year.

The present trend of Energy Efficient Motors being adopted in Indian Industry and promoted by Policy Makers is really commendable. It shows the trend in terms of huge energy & cost saving drive, which is very much in need of the hour for Indian Economy.

Keywords: Motor, Energy Saving, Efficiency, NEMA, Losses, Power Cost, Operating Hour, Payback Period.

Introduction

Global warming is a reality and world over people are working towards reduction in carbon foot print. Electric motor applications, in Indian industry, consume about seventy percent of the generated electrical energy. Improving efficiency of the motor is therefore a major concern in energy-efficiency efforts. Electric motors with improved efficiency, in combination with frequency converters can save about 7% of the total worldwide electrical energy. Roughly one quarter to one third of these savings come from the improved efficiency of the motor [2].

Table 1. Estimated Energy Saving potential of Motors in various Sectors

Sector	Total Conservation Potential (%)	Energy Handled By Motors (%)
Industrial Sector	Up to 25	70-75
Agriculture Sector	Up to 25	20-25
Domestic Sector	Up to 20	2-3
Commercial Sector	Up to 30	4-5

Challenges in Implementation

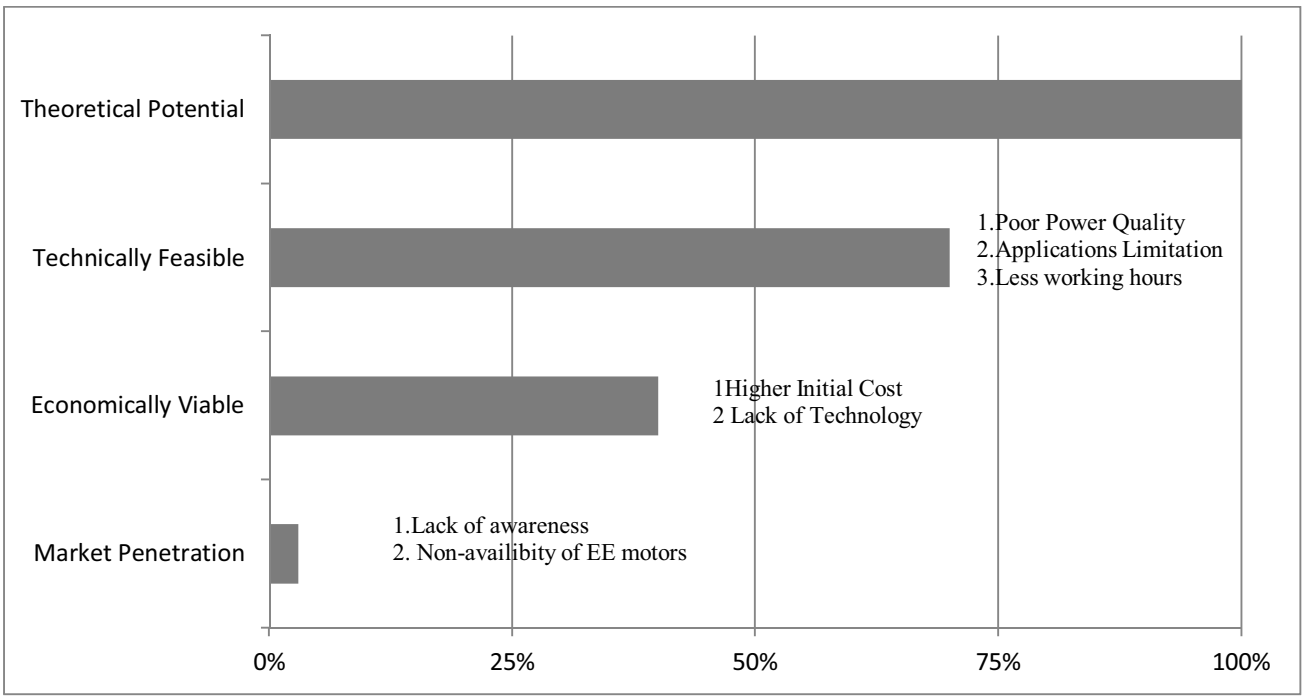
There is lack of awareness among the motor purchasers, regarding the potential for energy and cost savings by using more energy-efficient Electrical Motor Driven system (EMDS). In the Company organizational structure, Procurement budget is made independent of operations and maintenance budgets.

Motor is often integrated as a part of main equipment produced by OEMs before sale to the final end user.

The major barrier is that SMEs have to change their mindset toward the life cycle cost of the motor. Every purchase in industry units needs to weigh the savings that can be achieved through energy-efficient motors. There is always a tendency in the Indian market to go for the lowest initial cost instead of analyzing the life-cycle cost. However, as this cost is hidden, people tend to take the decision to purchase a motor based on the initial cost. The initial purchasing cost contributes only 4 per cent compared to the 88 percent of the total lifetime cost incurred due to consumed energy or the running cost of the motor.

The unavailability of complete Technical Standard is another grey area, which need to be explored before full-fledged introduction of Energy-efficient motors into the Indian Market. Even though standards are published by Bureau of Indian Standard (BIS) & International Electrotechnical Commission (IEC) but several critical information such as practical testing of efficiency, starting time & Starting current are missing.

Hence every purchase and technical specification for these EMD systems need to be revised and updated with new standard, i.e. by including IE2 and IE3 motors in the specifications.



Graph-1 Barriers in penetration of high-efficiency motors in Indian market

Global Scenario of Energy-efficient Motors

In recent years, market share of more efficient motors has been increasing in many regions and countries. This is particularly the case for the United States, China and other countries and to a certain extent for Europe. The United States and Canada are leaders in terms of setting motor energy-efficiency standards. Four standardized efficiency classes (IE1, IE2, IE3 and IE4) are currently recognized.

The United States and Canada are leaders in terms of setting motor energy-efficiency standards, as they introduced regulations for motors in the late 1990s. As early as 2002, China defined Minimum Energy Performance Standard (MEPS) for electric motors. The European Union passed MEPS legislation for electric motors in 2009 as an implementing measure under the eco-design directive; these will replace the previous industrial voluntary agreement. Australia, Korea, Brazil, Mexico, Taiwan and some other countries with large electricity consumption from motors have already adopted MEPS. However, some large motor using economies such as India, Japan and Russia are yet to adopt MEPS.

Energy efficient motor can save 3% of global annual electricity, which has a savings potential of 132 TWh. Annual global electricity consumption is 18,400 TWh. 40% of total Annual global electricity consumption of 7,360 TWh is by industrial

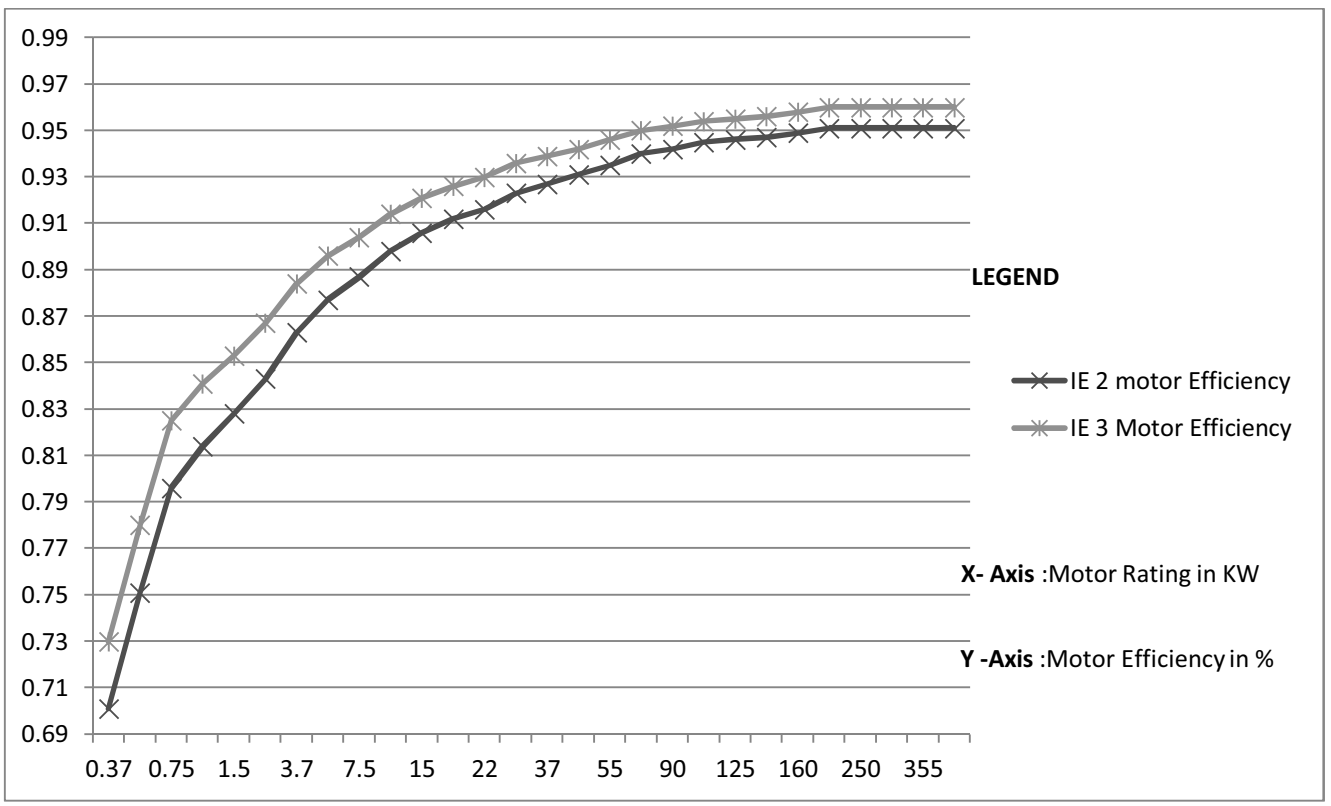
units and 60% of the industrial consumption of 4,400 TWh is used by motor systems. A savings potential of 3% on motor system amounts to 132 TWh [6].

Scope of Energy Saving

There is huge scope in power saving specially in the Indian Power Industries. For an example ,only LT motor being used in following application are considered for our case study and outcome of the possible energy saving is tabulated below.

Table 2. Application of Motor in a Typical 660 MW Thermal Power Plant

Application (for LT Motor)	Approx. Power Consumption (in KW)	Power Saving (in KW) Using IE3 Motor
Air Conditioning	530	9.01
Ventilation	1675	30.24
Boiler & Auxlliaries	554	8.08
Turbine & Auxlliaries	335	4.05
Cooling Tower	4560	48
Balance Of Plant	3086	37.419
Coal Handling Plant	4379	61
Ash Handling Plant	3322	44.9



Graph-2 Efficiency Comparison for IE2 & IE3 Motors

The above graph shows the difference in efficiency of IE2 & IE3 Motors for the ranging starting from 0.37 KW to 355 KW.

Conclusion

With the growing energy cost, Energy Efficiency is no longer a luxury but an urgent necessity [7]. Typical useful life of motors is 15 to 20 years .Therefore investment for replacing all old motors less efficient than IE2 or IE3 can be paid back in

less than 10-16 months. When entire investment can be returned in such a short period of time, the cumulative gains from life cycle cost can be much more than the initial cost of IE2/IE3 motors.
The initial purchasing cost contributes only 4 per cent compared to 88 percent of the total lifetime cost incurred due to consumed energy or the running cost of the motor.
There are other advantages for switching over to Energy Efficient Motor such as:

- Almost constant efficiency between 60 to 100 % load
- Ability to operate at Higher Temperature
- Reduction of Greenhouse Gases

For an example, When IE3 motors are used continuously for one year the power savings can be derived as mentioned in Table 3.

Table 3. Annual Power Saving by replacing IE2 with IE3 motors

Motor Rating (in KW)	Power Savings Per Year (in KW)	Motor Rating (in KW)	Power Savings Per Year (in KW)	Motor Rating (in KW)	Power Savings Per Year (in KW)
0.37	94	15	131400	125	1095000
0.55	140	18.5	162060	132	1156320
0.75	191	22	192720	160	1401600
1.1	260	30	262800	200	1752000
1.5	329	37	324120	250	2190000
2.2	463	45	394200	315	2759400
3.7	681	55	481800	355	3109800
5.5	915	75	657000	375	3285000
7.5	1117	90	788400	-	-
11	1542	110	963600	-	-

The global energy shortage has a significant impact on the design and application of electric motors. The opportunity is now available to select motors which will have substantial energy savings. Prudent decision, selection and use of energy-efficient motors not only support our national energy conservation goal but also provides an attractive financial reward to the motor end user.
The era of the energy-efficient motor has arrived. Hence energy consumption is a significant parameter, which needs to be keenly considered while selecting a motor irrespective of KW rating & area of application.

Acknowledgement

The authors greatly acknowledge the support from the Management of the L&T POWER. The work has been carried out as a part of our ongoing Power Project Engineering work. We would also like to thank our seniors and our colleagues for their help and kind support rendered to us. This paper would not be possible without their support.
We specially thank our mentors Mr. Pankaj Malhotra (Unit Head), Mr. G.V.Ashok Kumar (Dept. Head) & Mr. Shrikant Jagdish Keole (Sr. DGM), Mr. Arun Rao Mandokar (AGM) for their valuable guidance & incessant encouragement for improvising the existing Engineering Practices.
We would also like to acknowledge our Learning & Development Team (Personnel & HR Dept.) for their unconditional support in this journey.

References

[1] EE Motor Standards in India, Bureau of Energy Efficiency (BEE), Government of India
[2] Energy Efficient Motor by Bharat Bijlee
[3] K.N. Hemanth Kumar “India driving fast toward energy-efficient motors”, EPR, 19 May 2014

[4] A giant step towards an energy-efficient future, Technical Write-up by SIEMENS
[5] Green Riverside “Energy Efficient Motors”, Library-ENERCOM
[6] International Copper Promotion Council India “Adoption of High Energy Efficient Motors”
[7] Power House- Expert in High efficiency motors and motor system “Cost competitiveness in manufacturing through Energy Efficiency in Motor systems”.
[8] Howard E. Jordan “Energy-Efficient Electric Motors and Their Applications”, Springer Science+ Business Media, LLC

Computation of Steady State Temperature Distribution in the Stator of a 75 kW Slip-Ring Induction Motor by Finite Element Method

Pradyumna Ranjan Ghosh*, Nirmal Kr. Bhattacharya** and Debasis Sarkar***
* ME Student, EE Department, IIEST, Shibpur
prghosh2007@gmail.com
**Jadavpur University
nirmal_k_bhattacharya@yahoo.co.in
***Professor and Ex-HOD, EE Department, IIEST Shibpur
debasissrkr@yahoo.co.in

Abstract: For development of electric machines, particularly induction machines, temperature limit is a key factor affecting the efficiency of the overall design. Since conventional loading of induction motors is often expensive, the estimation of temperature rise by tools of mathematical modeling becomes increasingly important and as a result of which computational methods are widely used for estimation of temperature rise in electrical machines. This paper describes the problem of two dimensional steady state heat flow in the stator of induction motor. The stator being static is prone to high temperature and the study of thermal behavior in the stator is useful to identify the causes of failure in induction machines. The temperature distribution is obtained using finite element formulation and employing arch shaped elements in the r-θ plane of the cylindrical co-ordinate system. This model is applied to one 3-phase slip ring induction motor of 75 kW rating.

Keywords: FEM, Induction Motor, Thermal Analysis, Design Performance, Transients.

Introduction

Considering the extended use of squirrel cage induction machine in industrial or domestic applications both as motor and generator, the improvement of the energy efficiency of this electromechanical energy converter represents a continuous challenge for the design engineers, any achievements in this area meaning important energy savings for the world economy. Thus to design a reliable and economical motor, accurate prediction of temperature distribution within the motor and effective use of the coolant for carrying away the heat generated in the iron and copper are important to designers[18].

Traditionally, thermal studies of electrical machines have been carried out by analytical techniques or by thermal network method [8], [9]. Numerical techniques based on either finite difference method [10], [11] or finite element method [1]-[8], [14]-[20] are more suitable for analysis of complex system. Most of the earlier designers and researchers have traditionally adopted analytical methods such as separation of variables, conformal mapping, and resistance analog networks for prediction of temperatures. The analytical work is largely limited and that too with many major assumptions. Even though the resistance analog method predicts average temperatures quite accurately, the method failed in predicting hot spot temperatures. Finite difference is one of the popular numerical methods widely used. The estimation of core iron and copper winding temperatures in electrical machines, the finite difference method had been normally employed. Even though this method predicts hot spot temperatures, the method is not as flexible as finite element method in handling complex boundary condition and geometry. Rajagopal [5] and Sarkar [7] have carried out extensive two-dimensional steady state and transient thermal analysis of TEFC machines using Finite Element Method (FEM). Recent advances in Computational Electromagnetics, encouraged by continuing increase of power and speed of computers, make finite elements and related techniques an attractive alternative to well established semi-analytical and empirical design methods, as well as to the still popular 'trial and error' approach

Temperature rise forms a basic factor in the design of electrical machines and the heating is due to the losses occurring in the machine. As heat is developed in a machine its temperature rises, and if it is unable to dissipate heat to the surrounding environment, its temperature will increase to an extremely high value. However, the rates of heat transfer from the outer surface of the machine to the surrounding medium increases with the rise in temperature until after a certain time when the temperature ceases to rise. This occurs when the amount of heat dissipated into the surrounding atmosphere per unit time equals the heat dissipated in that time.

In this paper, the finite element method is used for predicting the temperature distribution in the stator of an induction motor under steady state conditions, using arch-shaped finite elements with explicitly derived solution matrices. A general derivation of the finite element equations by the method of weighted residuals (Galerkin's method) is introduced. A 39-

element two-dimensional slice of armature iron, together with copper winding bounded by planes at mid-slot, mid-tooth and mid-package, are used for solution to a heating problem, and this defines the scope of this technique. The requirements of computer storage for a large number of elements have been reduced by the use of half band-width of the symmetric matrices so that the computing costs can be greatly reduced. The model is applied to one slip-ring induction motor of 75 KW rating and the temperatures obtained are found to be within the permissible limit in terms of overall temperature rise computed from the resulting loss density distribution.

Steady Heat Conduction

The general form of the heat conduction equation can be described by the following relations

$$q = -V \nabla T \tag{1}$$

$$\nabla \cdot q = Q \tag{2}$$

Where, T is the potential function (temperature)
V is the medium permeability (thermal conductivity)
q is the flux (heat flux)
Q is the forcing function (heat source)

Combining equations (1) and (2), one obtains the general partial differential equation describing the two dimensional heat conduction problem.

$$\nabla \cdot \left(V \nabla T \right) = -Q \tag{3}$$

In cylindrical polar co-ordinates, equation (3) can be expanded as

$$\frac{1}{r}\frac{\delta}{\delta r}\left(V_r r \frac{\delta T}{\delta r} \right) + \frac{V_\theta}{r^2}\frac{\delta^2 T}{\delta \theta^2} + Q = 0 \tag{4}$$

Where, V_r, V_θ are thermal conductivities in the radial and circumferential directions respectively.

Finite Element Analysis (Galerkin's Method)

The Galerkin's criterion for heat conduction equations are used for furnishing thermal problems. When only the governing differential equations and their boundary conditions are available, Galerkin's methods are convenient in a way that this approach surfaces the variational method in generality and further broadens the range of applicability of the finite element method. Through the element equations derived for those problems were explicitly evaluated only for the simplest type in each case, the equations are general and apply for many element shapes and displacement models. The popularity of the method stems mainly from the case with which irregular geometries and implicit natural boundary conditions can be handled. Another important advantage is that the method allows development of general computer program that can solve variety of thermal problems simply by accepting different input data. The computer program illustrates how a real problem is actually solved by the finite element methods. It is envisaged that such programs would be useful for future studies of more complicated problems.

The solution of equation (4) can be obtained by assuming the general functional behavior of the dependent field variable in some way so as to approximately satisfy the given differential equation and boundary conditions. Substitution of this approximation into the original differential equation and boundary condition then results in some error called a residual. This residual is required to vanish in some average sense over the entire solution domain.

The approximate behavior of the potential function within each element is prescribed in terms of their nodal values and some weighting functions N_1,N_2 ... such that

$$T = \Sigma N_i T_i \tag{5}$$

i=1,2,….,m

The weighting functions are strictly functions of the geometry and are termed interpolation functions. These interpolation functions determine the order of the approximating polynomials for the heat conduction problem.

The method of weighted residuals determine the 'm' unknowns T_i in such a way that the error over the entire solution domain is small. This is accomplished by forming a weighted average of the error and specifying that this weighted average vanishes over the solution domain.

The required equations governing the behaviour of an element is given by the expression

$$\iint_{D^{(e)}} N_i \left[\frac{\delta}{\delta r}\left(V_r \frac{\delta T^{(e)}}{\delta r} \right) + \frac{\delta}{\delta \theta}\left(\frac{V_\theta}{r^2} \frac{\delta T^{(e)}}{\delta \theta} \right) + Q \right] r dr d\theta = 0 \qquad (6)$$

Equation (6) expresses the desired averaging to the error or residual within the element boundaries, but it does not admit the influence of the boundary. Since we have made no attempt to choose the N_i so as to satisfy the boundary conditions, we must use integration by parts to introduce the influence of the natural boundary conditions.

Arch-Element Shape Functions

Consider the arch-shaped element of Fig. 1 formed by circle arcs radii a, b, radii inclined at an angle 2α.

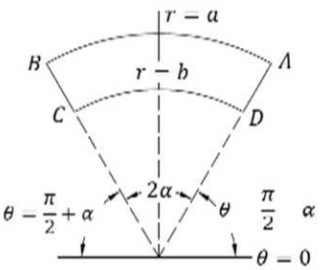


Fig. 1. 2D Arched shaped element suitable for discretization of induction motor stator

The shape functions can now be defined in terms of a set of non-dimensional coordinates with the help of cylindrical polar coordinates r, θ using the formula given below

$$\rho = \frac{r}{a}, \gamma = \frac{\theta - \pi/2}{\alpha}$$

The arch element with non-dimensional co-ordinates is shown in Fig. 2.

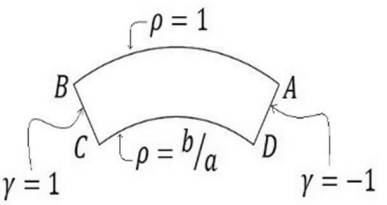


Fig. 2. The non-dimensional arch shaped element

The temperature at any point within the element be given in terms of its nodal temperatures by

$$T = T_A N_A + T_B N_B + T_C N_C + T_D N_D \qquad (7)$$

Where the N's are shape functions chosen as follows:

$$N_A = \frac{(\rho - \frac{b}{a})(\gamma - 1)}{-2(1 - \frac{b}{a})} \quad N_B = \frac{(\rho - \frac{b}{a})(\gamma + 1)}{2(1 - \frac{b}{a})}$$

$$N_C = \frac{(\rho - 1)(\gamma + 1)}{-2(1 - \frac{b}{a})} \quad N_D = \frac{(\rho - 1)(\gamma - 1)}{2(1 - \frac{b}{a})}$$

It is seen that the shape functions satisfy the following conditions:

(a) That at any given vertex 'A' the corresponding shape function N_A has a value of unity, and the other shape functions N_B , $N_{C, \ldots\ldots}$ have a zero value at this vertex. Thus at node j, $N_j = 1$ but $N_i = 0$, $i \neq j$.

(b) The value of the potential varies linearly between any two adjacent nodes on the element edges.

(c) The value of the potential function in each element is determined by the order of the finite element. The order of the element is the order of polynomial of the spatial co-ordinates that describes the potential within the element. The potential varies as a cubic function of the spatial co-ordinates on the faces and within the element.

Boundary Condition

In this analysis the two dimensional domain of core iron and winding chosen for modelling the problem is shown in Fig. 3 and the geometry is bounded by planes passing through the mid-tooth and the mid-slot. The temperature distribution is assumed symmetrical across two planes, with the heat flux normal to the two surfaces being zero. From the other two boundary surfaces, heat is transferred by convection to the surrounding gas. It is convected to the air gap gas from the teeth, to the back of core gas from the yoke iron. The boundary conditions may be written in terms of $\frac{\delta T}{\delta n}$, the temperature gradient normal to the surface

1) Mid-slot surface $\frac{\delta T}{\delta n_s} = 0$

2) Mid-tooth surface $\frac{\delta T}{\delta n_t} = 0$

3) Air-gap surface $h\left(T - T_{AG}\right) = -V_r \frac{\delta T}{\delta n_{AG}}$

Where T=Surface temperature, T_{AG} =Air gap gas temperature

4) Back of core surface $h\left(T - T_{BC}\right) = -V_r \frac{\delta T}{\delta n_{BC}}$

Where T_{BC} = Back of core gas temperature

Approximate Numeric Form

The heat flow equation may be formulated in Galerkin's form, the solution being obtained by specializing the general functional form to a particular function, which then becomes the approximate solution sought. Focusing our attention on equation (6), we obtain through integration by parts.

$$\iint_{D^{(e)}} N_i \frac{\delta}{\delta r}\left(V_r \frac{\delta T^{(e)}}{\delta r}\right) r d\theta dr = \int_{S_2^{(e)}} V_r \frac{\delta T^{(e)}}{\delta r} N_i r d\theta - \iint_{r,\theta} V_r \frac{\delta T^{(e)}}{\delta r}\frac{\delta N_i}{\delta r} r d\theta dr$$

$$= \int_{S_2^{(e)}} V_r \frac{\delta T^{(e)}}{\delta r} n_r N_i d\Sigma - \iint_{r,\theta} V_r \frac{\delta T^{(e)}}{\delta r}\frac{\delta N_i}{\delta r} r d\theta dr \qquad (8)$$

Where n_r is the r component of the unit normal to the boundary, and d$\sum$ is a differential arc length along the boundary. Equation (8) takes the form

$$-\iint_{D^{(e)}} \left(V_r \frac{\delta T^{(e)}}{\delta r}\frac{\delta N_i}{\delta r} + \frac{V_\theta}{r^2}\frac{\delta T^{(e)}}{\delta \theta}\frac{\delta N_i}{\delta \theta}\right) r d\theta dr - \iint_{D^{(e)}} \left(N_i Q\right) r d\theta dr$$

$$+ \int_{S_2^{(e)}} \left(V_r \frac{\delta T^{(e)}}{\delta r} n_r + \frac{V_\theta}{r^2}\frac{\delta T^{(e)}}{\delta \theta}\right) N_i d\Sigma^{(e)} = 0 \qquad (9)$$

For i=A, B, C, D

The surface integral (boundary residual) in equation (9) now enables us to introduce the natural boundary conditions. Equation (9) can be written with respect to the nodal temperatures as

$$\iint_{r,\theta}\left[V_r \frac{\delta T^{(e)}}{\delta r}\frac{\delta}{\delta T_i}\left(\frac{\delta T^{(e)}}{\delta r}\right) + \frac{V_\theta}{r^2}\frac{\delta T^{(e)}}{\delta \theta}\frac{\delta}{\delta T_i}\left(\frac{\delta T^{(e)}}{\delta \theta}\right) - Q N_i\right] r d\theta dr$$

$$+ \int_{S_2^{(e)}} \left(q N_i + h\left[N\right]\left\{T\right\}^{(e)} N_i - h T_\infty N_i\right) d\Sigma^{(e)} = 0 \qquad (10)$$

For i=A, B, C, D

There are four such equations as (10) for the four vertices of the element. These equations, when evaluated lead to the matrix equation

$$[[S_R] + [S_\theta] + [S_H]][T] = [R] + [S_C] \qquad (11)$$

$[S_R]$, $[S_\theta]$ are symmetric coefficient matrices (thermal stiffness matrices).

$[S_H]$ is the heat convection matrix.

[T] is the column vector of unknown temperatures.

[R] is the forcing function (heat source vector).

$[S_C]$ is the column vector of heat convection.

Method of Solution [12]

The system of global equations as determined by equation (11) has to be solved in order to determine the nodal temperatures. The solution of this set of linear simultaneous equations is determined by Gauss method which takes the advantage of banded nature of the matrices. To save computer memory the symmetric matrix of half band width, efficiently stored by Gauss routine, by which extremely large problems can be effectively solved.

Application to Slip-Ring Induction Motor

Optimal design of large induction motor is a process that involves electrical and mechanical skills as well as thermal and fluid dynamic skills. It is necessary to improve the ventilation structure to decrease stator windings temperature below the permissible insulation material temperature. It has been observed that with proper optimization of machine parameters the convection heat transfer can be improved and as a consequence the energy efficiency of the motor will also improve.

The problem concerns heat flow through core iron and winding in the stator of an induction motor. The stator being static is prone to high temperature and the temperature distribution of the stator only is computed here. The hottest spot is generally in the copper coils. The heat from the outer surfaces, i.e. the back of core surface and the air-gap surface is lost through convective mode of heat transfer. Thermal conductivity of copper and insulation are taken together for calculation.

As the temperature is maximum at the central plane, the temperature distribution in the plane can be determined approximately by taking this as a two-dimensional r-θ problem with the following assumptions:-

1) The temperature in the strip of unit thickness on the central axis is assumed to be fixed axially i.e., no axial flow of heat is assumed in the central plane. This assumption is permissible because in the central plane where the temperature distribution is maximum, the temperature gradient in the axial direction is zero.
2) The convection is taken care of only at the cylindrical surfaces neglecting the convection at the end surfaces.

Because of this assumption, the temperatures calculated in the central plane will be slightly higher than the actual.

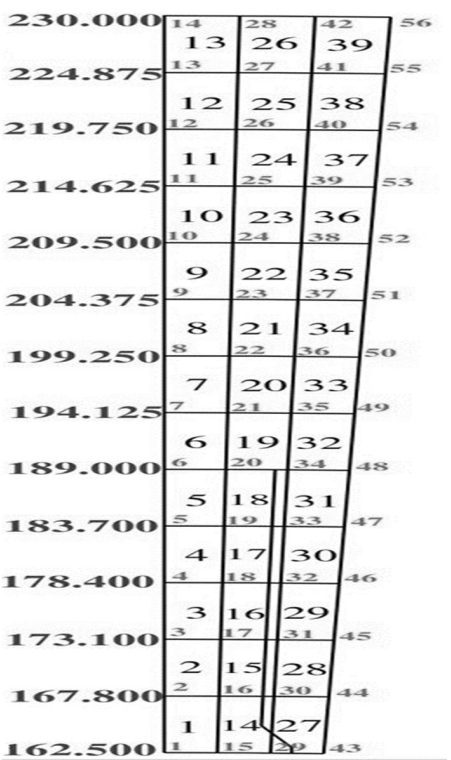


Fig. 3. Slice of core iron and winding of stator of slip ring induction motor bounded by planes at mid-slot and mid-tooth divided into arch shaped elements

In this analysis, because of symmetry the two dimensional domain in cylindrical polar co-ordinate of core iron and winding chosen for modeling the problem and the geometry bounded by planes passing through the mid-tooth and the mid-slot are divided into finite elements as shown in Fig. 3. Arch shaped elements are used throughout the solution region.

Convective Heat Transfer Coefficient

Two separate values of the convective heat transfer coefficient have been taken for the cylinder curved surface over the stator frame and the cylindrical air gap surface.

[1] Natural Convection:- The natural convection heat transfer coefficient on cylindrical curved surface over the stator frame are dependent on Grashof number and Prandtl number and calculated as h=4.795 watt/m^2 °C.

[2] Forced Convection:- The heat transfer co-efficient on forced convection for turbulent flow in cylindrical air gap surface are dependent on Reynolds number and Prandtl number and calculated as h= 76.84 Watt/m^2 °C.

Thermal Conductivity

For the steady state problem in two dimensions under considerations, the following properties are taken for each different material.

1) Thermal conductivity in radial direction, V_r Watt/m°C

2) Thermal conductivity in circumferential direction, V_θ Watt/m°C

Table 1. Typical Set of Material Properties for Induction Motor Stator

Material Property	Magnetic Steel Wedge	Copper and Insulation
V_r	33.07	2.007
V_θ	0.826	1.062

Calculation of Heat Losses [13]

Heat losses in the teeth and yoke of the core are based on calculated magnetic flux densities (1.56 wb/m^2 and 1.26 wb/m^2 respectively). In these regions, tooth flux lines are predominantly radial and yoke flux lines predominantly circumferential [13]. The grain orientation of the core punching's in these two directions and therefore influences the heating for a given flux density.

Table 2. Heat Losses Generated From Different Portions of Stator Slice

Portion of Stator Slice	Losses in Watt/mm^3
Stator Core	0.000032597
Stator Tooth	0.000063033
Copper Windings in Stator Slot	0.000304952

Data Analysis

In this report the two-dimensional finite element method was applied to analyze the temperature rise in stator slice portion of 75 KW slip ring induction machine. The heat flux is assumed to be travelling along the stator of induction motor in order to obtain an optimum temperature distribution. It is observed that the stator copper loss has the highest value. So, it is evident that that the copper losses in the slot will have the highest operating temperature. In the stator slice, node numbers 44 to 48 are present in the slot portion where the copper conductors are present and have maximum temperatures. It is further observed that the value of temperature increases from air gap boundary surface towards the back of core surface. Temperature distribution profile in various parts of the machine of stator slot, stator tooth and stator core have been depicted in Figs 4, 5 & 6 respectively.

Conclusion

The two-dimensional steady state finite element procedure for the thermal analysis of large induction motor stator provides the opportunity for the in-depth studies of stator heating problems. By virtue of the new, explicitly derived arch element, together with an efficient bandwidth and Gauss routine, extremely large problems can be efficiently solved. Though the results are approximate, the method is fast, inexpensive and leads itself to immediate visual pictures of the temperature pattern in a two-dimensional slice of iron core & winding in the stator of an induction motor. A number of other analysis can also be carried out, involving mechanical and electromagnetic fields, particularly where more than one of these fields are present.

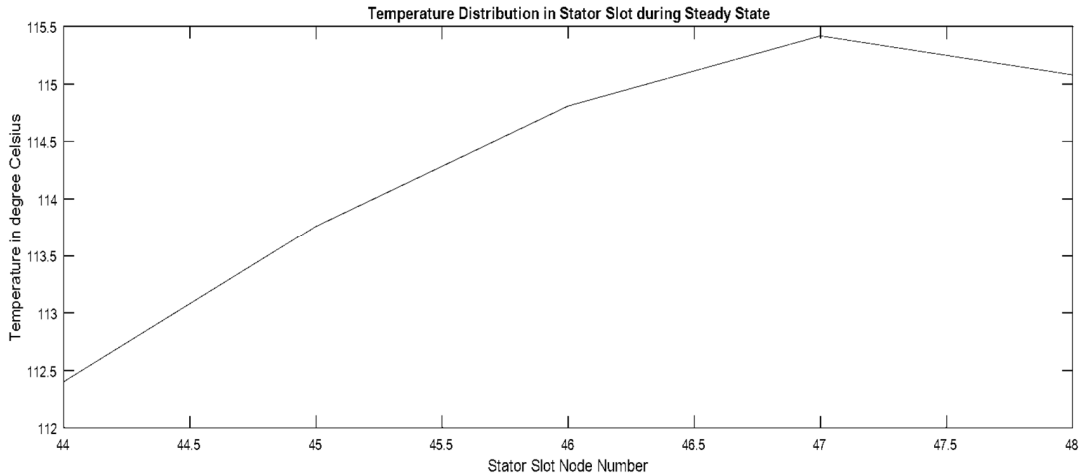


Fig 4:-Temperature Distribution in Stator Slot

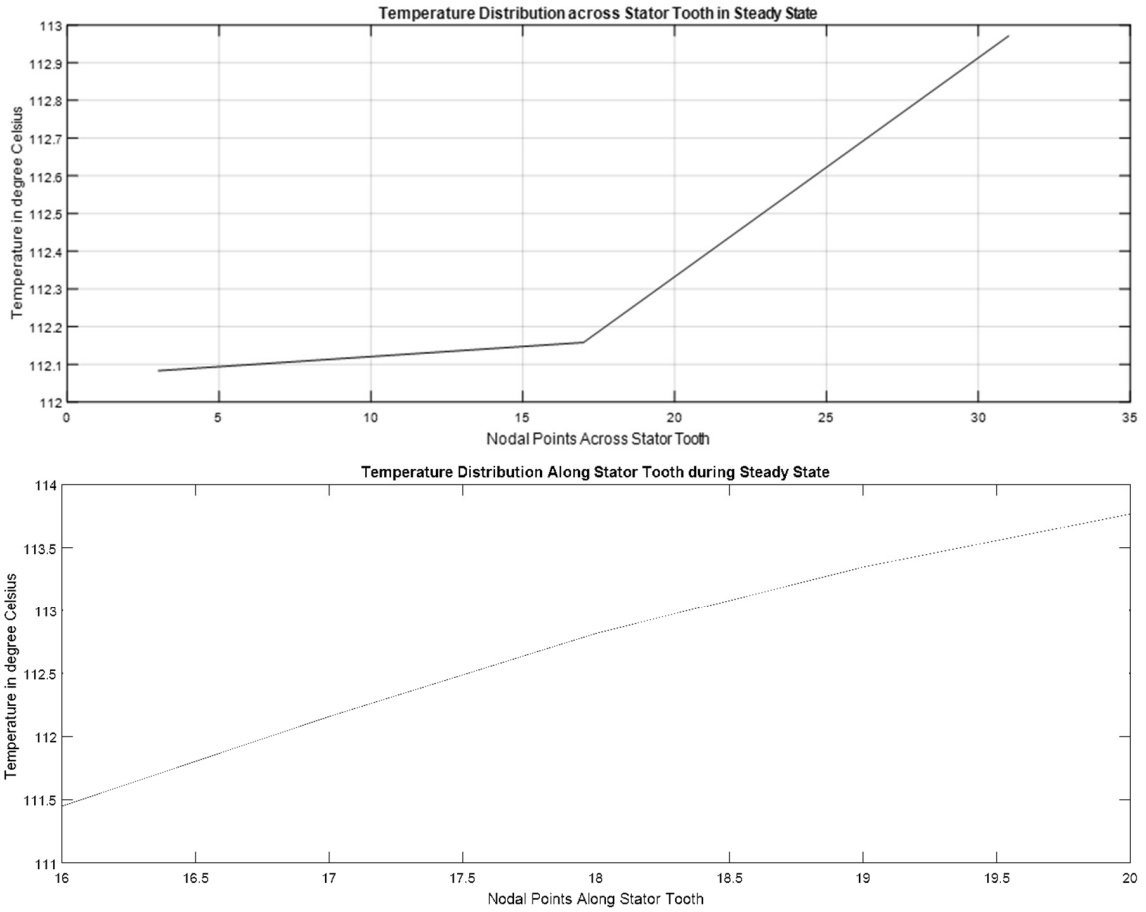


Fig 5:-Temperature Distribution across and along Stator Tooth

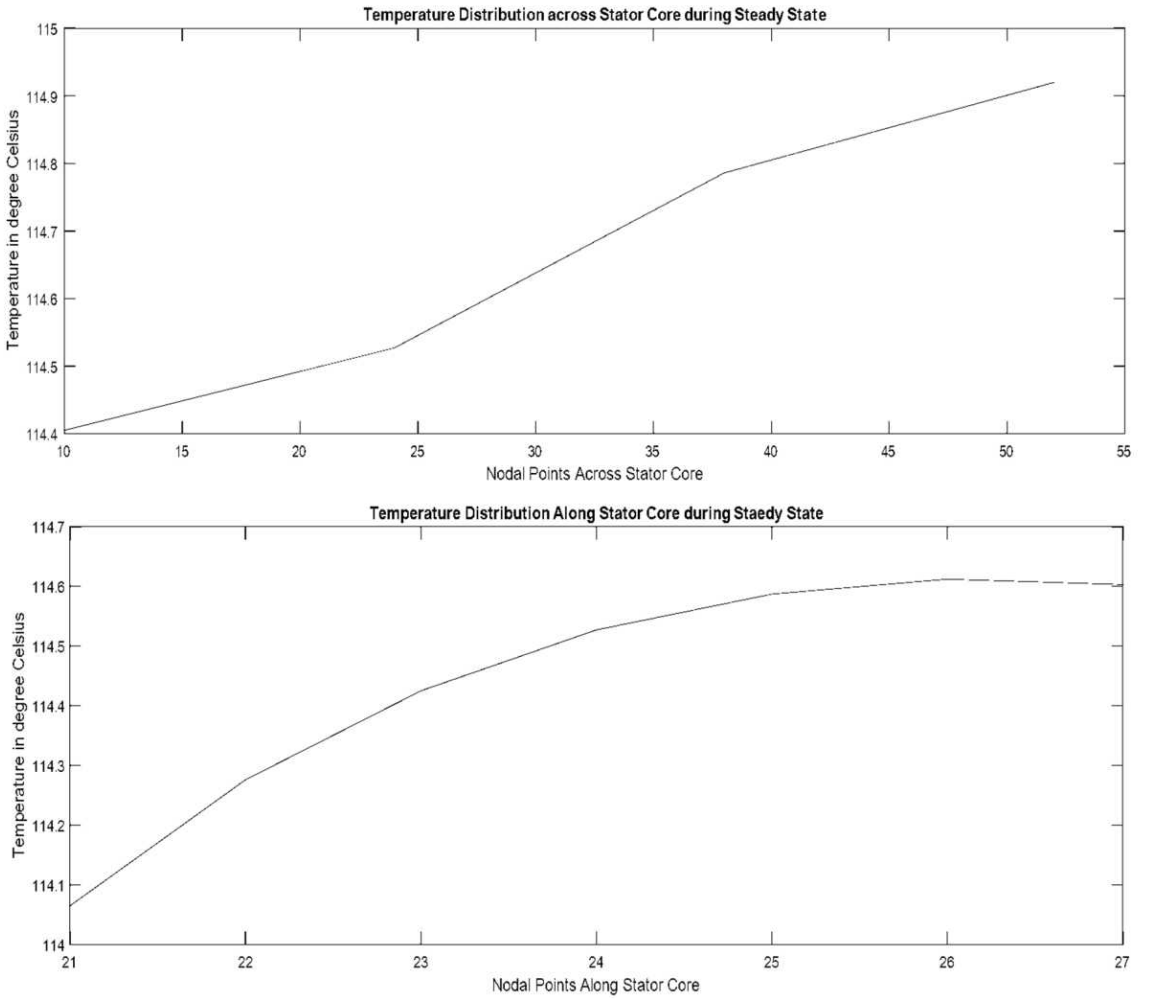


Fig 6:-Temperature Distribution across and along Stator Core

Appendix

Symmetric Coefficient Matrices

$$[S_R] = \frac{V_r \alpha}{6q^2}\left(1 - \frac{b^2}{a^2}\right)\begin{bmatrix} 2 & 1 & -1 & -2 \\ & 2 & -2 & -1 \\ & & 2 & 1 \\ SYM & & & 2 \end{bmatrix}, \quad [S_\theta] = \frac{V_\theta}{4\alpha q^2}\begin{bmatrix} A & -A & B & -B \\ & A & -B & B \\ & & C & -C \\ SYM & & & C \end{bmatrix}$$

Where,

$$A = 1 - 4\frac{b}{a} + 3\frac{b^2}{a^2} - 2\frac{b^2}{a^2}\log_e\frac{b}{a} \qquad B = -1 + \frac{b^2}{a^2} - 2\frac{b}{a}\log_e\frac{b}{a} \qquad C = -3 + 4\frac{b}{a} - \frac{b^2}{a^2} - 2\log_e\frac{b}{a}$$

Forcing Function Vector

$$[R] = \frac{Q a^2 \alpha}{6\left(1 - \frac{b}{a}\right)}\begin{bmatrix} G \\ G \\ H \\ H \end{bmatrix}$$

Where,

$$G = 2 - 3\frac{b}{a} + \frac{b^3}{a^3}, \qquad H = 1 - 3\frac{b^2}{a^2} + 2\frac{b^3}{a^3}$$

Formulation of the Heat Convection Matrix on Cylindrical Curved Surface

$$[S_H] = ha\rho\alpha \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ & \frac{2}{3} & 0 & 0 \\ & & 0 & 0 \\ SYM & & & 0 \end{bmatrix}$$

Formulation of the Heat Convection Vector on Cylindrical Curved Surface

$$[S_C] = hT_\infty a\rho\alpha \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

References

[1] A.F.Armor., and M. V. K. Chari., "Heat flow in the stator core of large turbine generators by the method of three-dimensional finite elements, Part-I:Analysis by Scalar potential formulation. Part - II: Temperature distribution in the stator iron," IEEE Trans,Vol. PAS-95, No. 5, pp.1648-1668, September 1976.

[2] A.F. Armor, "Transient, three-dimensional, finite element analysis of heat flow in turbine-generator rotors," *IEEE Trans. PAS-99*, no. 3, pp. 934–946, May/June 1980.

[3] Boglietti, A. Cavagnino, D. Staton, Martin Shanel, M.Mueller, and C.Mejuto,"Evolution and modern approaches for thermal analysis of electrical machines", IEEE Transactions on Industrial Electronics", 56(3):871– 882, March 2009.

[4] D. Staton, A. Boglietti, and A. Cavagnino., "Solving the more difficult aspects of electric motor thermal analysis in small and medium size industrial induction motors." Energy Conversion, IEEE Transactions on, 20(3):620–628, Sept 2005.

[5] M.S., Rajagopal,D.B.,Kulkarni,K.N., Seetharamu,and P.A. Ashwathnarayana., "Axi-symmetric steady state thermal analysis of totally enclosed fan cooled induction motors using FEM", *2nd Nat Conf.on CAD/CAM*,19-20 Aug,1994.

[6] D.W. Pepper and J.C. Heinrich. "The Finite Element Method: Basic Concepts and Applications. Series in Computational and Physical Processes in Mechanics and Thermal Sciences." Taylor & Francis, 2005.

[7] D. Sarkar, P. K. Mukherjee, and S. K. Sen, "Use of 3-dimensional finite elements for computation of temperature distribution in the stator of an induction motor," *IEE Proceedings*, pt. B, vol. 138, no. 2, pp. 75–86,1991.

[8] G. M. Rosenberry, Jr., "The transient stalled temperature rise of cast aluminium squirrel-cage rotors for induction motors," *AIEE Trans.*, vol. PAS-74, Oct. 1955.

[9] T. J. Roberts, "The solution of the heat flow equations in large electrical machines," *Proc. Instn. Mech. Engrs.*, pt.3E, vol.184, pp. 70–83, 1969–1970.

[10] K.Reichert, "The calculation of the temperature distribution in electrical machines with the aid of the finite difference method," Elektro Tech Z-A, Bd 90, H6, 1969.

[11] C. E. Tindall and S. Brankin, "Loss-at-source thermal modelling in salient pole alternators using 3-dimensional finite difference techniques," *IEEE Trans. Magnetics*, vol. 24, no. 1, Jan. 1988.

[12] L.L.,Bhirud,"Matrix operations on the computer", Bombay, Oxford & IBH Publishing Co., 1975.

[13] M.V.Deshpande, "Design and Testing of Electrical Machines", Wheeler Publishing, 1983.

[14] E. Dlala, "Comparison of models for estimating magnetic core losses in electrical machines using the finite-element method", IEEE Transactions on Magnetics, 45(2):716-725, Feb. 2009.

[15] Ruoho S., Santa-Nokki T., Kolehmainen J., Arkkio A., "Modeling magnet length within 2d finite-element analysis of electric machines", IEEE Transactions on Magnetics 45 (2009) 8, pp. 3114-3120.

[16] Pippuri J.E., Belahcen A., Dlala E., Arkkio A., "Inclusion of eddy currents in laminations in two-dimensional finite element analysis", IEEE Transactions on Magnetics 46 (2010) 8, pp. 2915–2918.

[17] Nannan Zhao, Zhu, Z.Q., Weiguo Liu., "Thermal analysis and comparison of permanent magnet motor and generator", International Conference on Electrical Machines and Systems (ICEMS), pp. 1 – 5. 2011.

[18] Hai Xia Xia; Lin Li; Jing Jing Du; Lei Liu.,"Analysis and calculation of the 3D rotor temperature field of a generator-motor" International Conference on Electrical Machines and Systems (ICEMS), 2011, pp.1 – 4.

[19] Ganesh B. Kumbhar, Satish M. Mahajan., "Analysis of short circuit and inrush transients in a current transformer using a field-circuit coupled FE formulation", IJEPES- Volume 33, Issue 8, Pages 1361–1367, October 2011.
[20] B.L. Rajalakshmi Samaga, K.P. Vittal.,"Comprehensive study of mixed eccentricity fault diagnosis in induction motors using signature analysis", IJEPES- Volume 35, Pages *180-185,* 2012**.**

Parametric Study of Drag on Orientation of High Rise Building

Amol .S. Pote * and Smrutirekha Mishra **
* Department of Civil Engineering, G. H. Raisoni College of Engineering and Management, Pune 412207, India
amol.pote@raisoni.net
** Research Scholar, Department of Civil Engineering, G. H. Raisoni College of Engineering and Management, Pune 412207, India
smruticeb@yahoo.co.in

Abstract: In India high rise buildings are becoming a common feature of urban landscape. Limited information is available regarding wind flow around such buildings. Effect of wind is predominant on structure depending on location, orientation of building. The flow is affected by architectural features. In this work a rectangular building is studied under the effect of wind force and how orientation affects the wind load on tall buildings shall be studied based on wind tunnel tests with rigid models, Wind pressure distributions on building with different aspect ratio will be measured and analyzed. The investigations can offer some basic understanding for estimating wind load distribution on building & facilitate wind resistant design of cladding components & their connections considering wind load.
It is also proposed to study the interference when two tall buildings are built side by side. The study will involve different orientations & separating distance between the two tall buildings. Models will be tested in wind tunnel for different angles of orientation & different wind speeds.

Keywords: Damping, Rectangular building, structural dynamics, wind force, wind tunnel.

Introduction

Wind is a phenomenon of great complexity because of the many flow situations arising from the interaction of wind with structures. Wind is composed of a multitude of eddies of varying sizes and rotational characteristics carried along in a general stream of air moving relative to the earth's surface. The emergence of modern materials & construction techniques resulted in structures that are often to a degree unknown in the past, remarkably low in damping,& light in weight .Generally such structures are more affected by the action of wind. The structural should ensure that the structure should be safe & serviceable during its anticipated life even if it is subjected to wind loads. Wind forms the predominant source of loads in tall free standing structures .The effect of wind on tall structure can be divided into two components they are

- Along wind effect
- Across wind effect

Along wind loads are caused by the drag components of the wind force whereas across wind loads are caused by the corresponding lift components.
An important problem associated with wind induced motion of buildings is concerned with human response to vibration and perception of motion. Therefore, for most tall buildings serviceability considerations govern the design and not strength issues.

Wind Speed

At great heights above the surface of the earth, where frictional effects are negligible, air movements are driven by pressure gradients in the atmosphere, which in turn are the thermodynamic consequences of variable solar heating of the earth. This upper level wind speed is known as the gradient wind velocity. In practice, it has been found useful to start with a reference wind speed based on statistical analysis of wind speed records obtained at meteorological stations throughout the country.

Design Wind Loads

The characteristics of wind pressures on a structure are a function of the characteristics of the approaching wind, the geometry of the structure under consideration, and the geometry and proximity of the structures upwind. The pressures are not steady, but highly fluctuating, partly as a result of the gustiness of the wind, but also because of local vortex shedding at the edges of the structures themselves.

The complexities of wind loading, should be kept in mind when applying a design document. Because of the many uncertainties involved, the maximum wind loads experienced by a structure during its lifetime, may vary widely from those assumed in design. Wind loading governs the design of some types of structures such as tall buildings and slender towers. It often becomes attractive to make use of experimental wind tunnel data in place of the coefficients given in the Wind Loading Code for these structures.

Types of Wind Design

Typically for wind sensitive structures three basic wind effects need to be considered.

1. Environmental wind studies
2. Wind loads for façade
3. Wind loads for structure
4. Design Criteria

In terms of designing a structure for lateral wind loads the following basic design criteria need to be satisfied.

a. Stability
b. Strength
c. Serviceability

Static Analysis

This method assumes the quasi-steady approximation. It approximates the peak pressures on building surfaces by the product of the gust dynamic wind pressure and the mean pressure coefficients. The mean pressure coefficients are measured in the wind-tunnel or by full-scale tests.

The quasisteady model has been found to be fairly reliable for wind loading on small structures.

In static analysis, gust wind speed (Vz) is used to calculate the forces, pressures and moments on the structure.

Along and Cross-Wind Loading

The wind approaching a building a complex phenomenon, but the flow pattern generated around a building is equally complicated by the distortion of the mean flow, flow separation, the formation of vortices, and development of the wake. Large wind pressure fluctuations due to these effects can occur on the surface of a building.

Under the collective influence of these fluctuating forces, a building tends to vibrate in rectilinear and torsional modes, as illustrated in Fig. 1. The amplitude of such oscillations is dependent on the nature of the aerodynamic forces and the dynamic characteristics of the building.

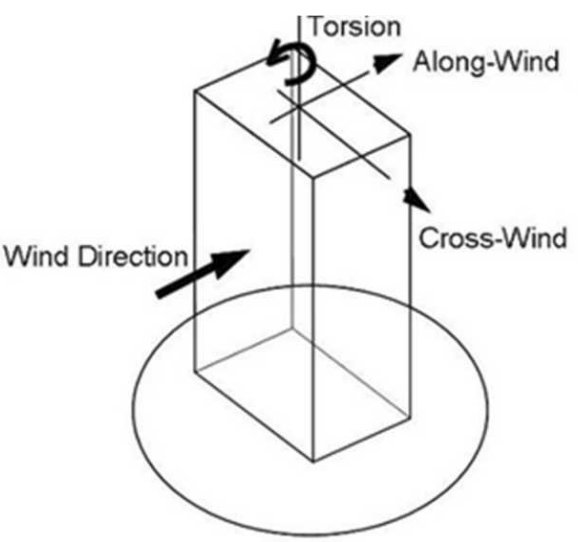


Figure 1. Wind Response Directions

Along-Wind Loading

The along-wind loading or response of a building due to buffeting by wind can be assumed to consist of a mean component due to the action of the mean wind speed (eg, the mean-hourly wind speed) and a fluctuating component due to wind speed variations from the mean.

Cross-Wind Loading

There are many examples of slender structures

That are susceptible to dynamic motion perpendicular to the direction of the wind. Crosswind excitation of modern tall buildings and structures can be divided into three mechanisms (AS/NZ1170.2, 2002) and their higher time derivatives, which are described as follows:

a) Votex Shedding.
b) The incident turbulence mechanism.
c) Higher derivatives of crosswind displacement.

Wind Tunnel

There are many situations where analytical methods cannot be used to estimate certain types of wind loads and associated structural response. Wind tunnel testing is now common practice for design of most tall buildings. It is a tool which is used in aerodynamic research to study various effects of air moving on solid objects. It consists of a tubular passage with an object under test which is mounted in the middle position of tunnel .Air is made to move over the object by a powerful suction fan system .The test object is also called a wind tunnel model consists of suitable sensors which is used to measure aerodynamic forces &outlet velocity & a manometer to measure pressure distribution over the object surface. The wind tunnel used is of suction type with an axial flow fan powered by a suitable motor. Wind tunnel testing is used in design of most major tall buildings to identify the wind induced structural loads and responses for which the superstructure has to be designed. The first wind engineering task for the engineer is to determine whether to design according to local conditions i.e. local wind loading code or standard, or to employ wind tunnel testing results, so it is advised to just analysis the result of wind tunnel for a structure and implement them with local condition codes using coefficients.

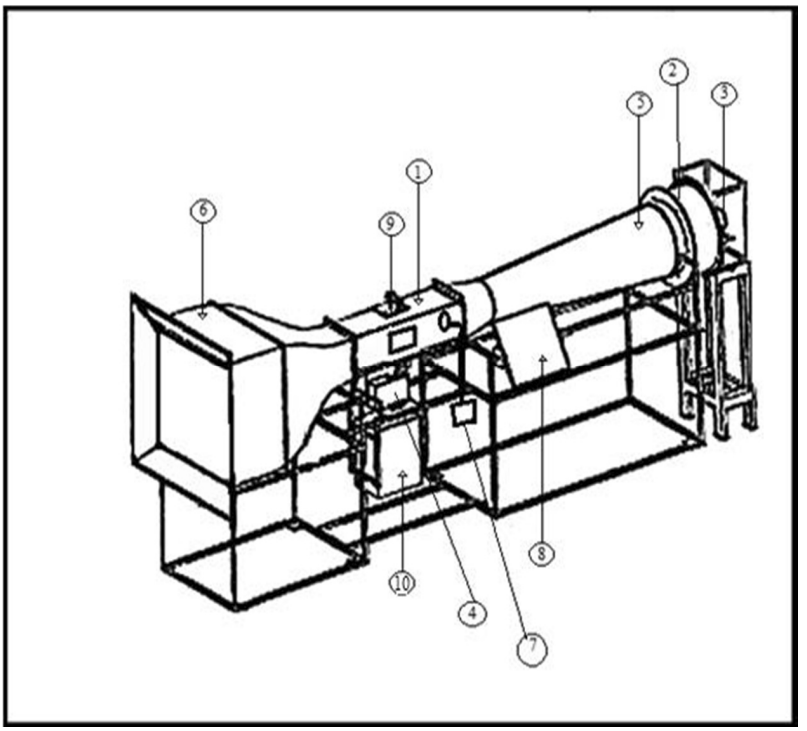


Fig 2: Wind Tunnel

Table 1. Specification of Wind Tunnel

Sr no.	Type	Open type wind tunnel
1	Test section	-
	Main Duct	Ms with powder coating
	Side glass material	Acrylic Sheet- 8mm Thick.
	Size	300mm x 300mm X1000mm long
2	Blower Fan	5 Blades- Aluminum Die cast
3	A.C. Motor	3 H.P. -2880 RPM
	Speed Variation	10% to 100% by frequency Drive Controller
4	Strain Gauge Balance	Two Channel, Digital display
		Lift force 0-20 Kg
		Drag force 0-20 Kg
	Lift/Drag force sensor	0 to 20 Kg beam type load cell
5	Diffuser section	-

Wind Tunnel Tests

Wind tunnel tests Wind tunnel testing is a powerful tool that allows engineers to determine the nature and intensity of wind forces acting on complex structures.

Wind tunnel testing is particularly useful when the complexity of the structure and the surrounding terrain, resulting in complex wind flows, does not allow the determination of wind forces using simplified code provisions.

The facility is a closed circuit, single return tunnel with a 7 ft wide by 5 ft high by 25 ft long test section. With a contraction ratio of 15:1 and the use of multiple diffusion screens, very low turbulence is achievable. The tunnel is capable of continuous operation at velocities of over 150 mph. large windows on the side and top of the test section allow for excellent model viewing. Force and moment data are obtained from a below-floor balance system capable of supporting models having gross weights of over 300 pounds. It resolves all six force and moment components to an accuracy of one part in 3,000, with less than 0.1 percent interaction error.

The wind tunnel facility also has instrumentation for the measurement and recording of operating pressures, the angle of attack or angle of yaw, tunnel static pressure, dynamic pressure and temperatures. A computer/data acquisition system is used to average and record the results. Boundary layer velocity profiles on the tunnel walls or on the model can be obtained via Pitot tubes or hot wire probes.

The laboratory has extensive facilities for flow visualization. The introduction of helium-filled bubbles or smoke into the flow, small vanes mounted on a given surface to show flow direction, coating the model with oil and the use of tufts are some of those methods. Photographic systems from still to motion (5 frames/sec up to 1 million frames/sec) are available. Video cameras and recorders are also available.

Wind Tunnel Instrumentation

Pitot tube Pitot tube is the basic instrument used for measuring wind speed in a wind tunnel

Hot-Wire Anemometer used to measure both mean velocity and turbulence. They can measure rapid changes of velocities with frequency response higher than I kHz

Bell Mouth and Entry the entry is shaped to guide the air smoothly into tunnel.

Test Section Working section consists of 300mm x 300mm x 1000mm test section with two windows to insert the models or probes

Diffuser Section Diffuser reduces dynamic pressure which, leads to reduction in power losses at the exit.

Axial Fan and Motor

Which gives smooth variation of air velocity in the test section which can be seen on the anemometer or Pitot tube and one can set the velocity of air to desired level.

Smoke generator

A smoke generator and probe that allows students to see air flow in subsonic wind tunnels and other low flow rate air flow products. It is a control unit that pumps oil to the tip of a probe.

Results and Discussion

Building Geometry & Experimental details

A 50 storey building H=150 m tall prototype building has a rectangular cross section with constant dimension 45.72m*30.48 m over the building height.

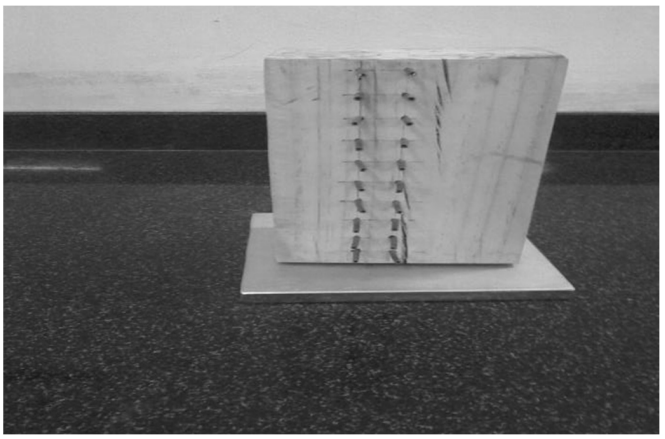


Figure 3 : Building Model with pressure tappings

The x & y are normal to long &short sides of the rectangle respectively .The direction of the wind speed is defined by counter clockwise angle with respect to y axis .the experiments will be conducted in wind tunnel.

Codes and Standards

The AISC Specifications (AISC 1978) contains a provision in Section 1.3.5, Wind, requiring that "proper provision shall be made for stresses caused by wind, both during erection and after completion of the building.

Risks

The tacit attitude seems to prevail in the design/construction industry that structural integrity during construction is less important than in permanent structures, therefore greater risks are acceptable.

There are various means, during the planning-design- construction process, of controlling the wind created risks:

1. During the planning phase Establishment
2. During the design phase
3. During the construction phase

Wind-Structure Interaction

Wind-structure interaction depends on many factors such as the size and shape of the structure, on the exposure and shading of the components and on the dynamic characteristics of the structure. All these factors are very different during construction than after completion of the project. Some elements, such as temporary support and access structures, do not even exist after completion. Not only are these factors different but they continually change during the day-to-day progress of the work.

Projected areas are generally less during construction than after completion: for example, a structure skeleton versus the enclosed building. But these wind areas in their temporary configurations generate different force coefficients and give rise to different gust response factors than the enclosed building does. The wind-structure interaction terms of force coefficients and gust response factors in ANSI and other codes were developed for complete structures of regular shapes.

At present, the best an engineer can do is calculate and sum up in a conservative manner the wind forces on all individual components exposed to the wind, or, if the funds and time are available, perform wind tunnel tests of the most vulnerable configurations anticipated during the construction.

Conclusion

After the observation it has been concluded that In long wall wind pressure remains constant with alternate pressure tappings i.e. 0 & 60 degree and vice versa. At the highest point pressure may be changed with angle 30, 45 &60 degree that depends on aspect ratio. In short wall wind pressure increases towards increasing the height of building with angle of orientation .At bottom pressure changes with constant angle of orientation 45 degree.

References

[1] Wind Desig N Problem S With Building Structure S Durin G Construction By Robert T. Ratay,1 Fellow, Asce
[2] Ratay, R. T., And Suprenant, B. A. (1987). "Assessment And Control Of Wind Created Risk During Construction." Proc. Of Risk Assessment Of Constructed Facilities Un› Der Wind, Dynamics Of Structures, Asce Structures Congress '87, Orlando, Fla., August.
[3] Mehta, K. C. (1983). "Use Of Ansi A58.1 For Solar Collector Wind Loads." Asce Preprint 83-507, Houston, Tex., Oct. 17-19.
[4] Wind Loading On Tall Buildings P. Mendis, T. Ngo, N. Haritos, A. Hira The University Of Melbourne, Australia , B. Samali University Of Technology Sydney, Australia J. Cheung Monash University, Australia
[5] As/Nz1170.2 (2002) Australian/New Zealand Standard, Structural Design Actions, Part 2: Wind Actions, Standards Australia & Standards New Zealand.
[6] Tamura Y., Ohkuma, T., Okada, H., Kanda, J. (1999) Wind Loading Standards And Design Criteria In Japan, J. Wind Engineering And Industrial Aerodynamics , 83, 1, 555-566.
[7] Cooney R.C. And King, A.B. (1988) Serviceability Criteria Of Buildings
[8] Ansi Code (1982). Ansi A58.1-1982. Minimum Design Loads For Buildings And Other Structures. American National Standards Institute
[9] Soong T.T. And Costantinou, M.C. (1994). Passive And Active Structural Vibration Control In Civil Engineering. New York: Springer

An Overview of Developments in Adsorption Refrigeration Systems: Principle and Optimization Techniques

Jaimin K. Patel*, N.S.Mehta** and J.B.Dabhi***
* ** ***Department of Mechanical engineering, L.D.R.P-ITR College, Gandhinagar, Gujarat
Email id: jaiminmech012@gmail.com nirveshmehta11@gmail.com

Abstract: The Negative environmental impacts and limited sources of fossil fuels are the major reason to focus on renewable energy sources, predominantly solar energy. Refrigeration and air-Conditioning processes consumes fifteen percent of the electricity produced in the whole world. Environment friendly adsorption cooling systems are possible alternatives to conventional electricity-driven vapour compression refrigeration systems. Comparatively bigger sizes of adsorption based cooling units, due to their low specific cooling power & manufacturing cost, are major roots that preventing successful commercialization of the technology. Efforts are on, to enhance the performance of adsorption systems through improvements in working pair's properties, improved heat and mas transfer by efficient design of adsorbent bed. This paper traces the basic principle, improvement techniques & analyses the barriers to wide spread use of adsorption refrigerators.

Keywords: Adsorbate, Adsorbent, Renewable energy, Working pairs, Solar energy.

Introduction

The standard of living of people are increasing drastically, and proportionally day by day, demand of cooling ,comfort air conditioning and refrigeration are increasing. Orthodox cooling technologies are generally electricity driven customary vapor compression refrigeration systems which usually operated with synthetic refrigerants [1], such as CFCs, HCFCs or HFCs, they causes the ozone layer depletion and/or cause greenhouse effect. As a result, several protocols, like the Montreal Protocol (1987) or the Kyoto Protocol (1997), were established in order to considerably reduce, the emissions of these refrigerants [3, 4]. However ,the situation continue demanding for the development of alternative technologies operating with environmentally friendly substances, especially due to the increasing emissions of HFCs, although the emission of CFCs and HCFs have been decreasing since the late 1980s [5,6]. Furthermore, the ever increasing energy consumption worldwide makes it urgent to find new ways to use the energy resources in a more efficient and rational way. It is estimated that the global energy consumption will increase by 71% between 2003 and 2030 [7]. In addition, currently 80% of the energy on Earth comes from fossil fuel resources [8].

Usual vapor compression refrigeration cycles are electrically powered, which leads to fossil energy consumption; and also largely contributing to the greenhouse effect. In another perspective, a considerable percentage of the world population is in remote areas, where the electricity supply is rare deficient or even non-existent, The need for refrigeration systems in these locations is of extreme importance since ,due to the electricity shortage, conventional refrigeration equipment cannot be used ,for instance ,in food and medicines storage ,Ice making or even for air-conditioning. Therefore ,the awareness on issues such as the decrease of fossil fuel resources, the severe environmental problems or even the location challenges (e.g., remote areas) require the development of new technologies and led the human kind to look with greater interest for ecological and renewable energy sources. These include wind, solar, hydropower biomass and geothermal energies, or even thermal waste from various processes.

Solar energy currently a subject of great interest, and refrigeration is a particularly attractive application due to the coincidence between the peak of cooling demand and the solar radiation availability. Recently, adsorption refrigeration processes have been investigated (theoretically and experimentally)and proposed as an alternative to vapor compression refrigeration systems ,attempting to preserve the production and efficiency level of traditional systems, and becoming one of the most promising solar refrigeration methods [4]. Adsorption systems are not cost-competitive, and have some technical drawbacks, such as low coefficients of performance (COP),low specific cooling powers(SCP),and poor heat and mass transfer on the adsorbent beds ,which makes the systems more bulky and expensive [10,11]. Moreover, in the case of solar refrigerators, the energy source is intermittent and can also be highly irregular. However, these systems promote significant primary energy savings in comparison with common mechanical vapor compression refrigeration systems, and have simpler control, no vibration, no noise, lower operation and maintenance costs, lower environmental impact, and are simpler and more robust [11]. Compared with the absorption systems, adsorption systems can be powered over a large range of heat source temperatures, are more robust and less sensitive to physical impacts ,do not present corrosion problems ,and are less complex because they contain fewer moving parts[11,12].Therefore, the adsorption refrigeration systems appear as a good

alternative to replace(or integrate)the traditional refrigeration systems by more environmentally friendly systems ,which can be powered by renewable energy sources. Researchers worldwide are working to improve the performance of adsorption cooling systems in order to overcome its current technical and economic issues.

The process

Principles of adsorption

The adsorption is a surface phenomenon, which result from the interaction between a solid and a fluid (refrigerant) based on a physical or chemical reaction. Physical adsorption occurs when the molecules of refrigerant (adsorbate) fix themselves at the surface of a porous solid element (adsorbent), which is due to Vander Waals forces and electrostatic forces. By applying heat, this process can be reversed in which adsorbate molecules can be released (Which is called desorption process).In turn, the chemical adsorption results from the ionic or covalent bonds formed between the adsorbate molecules and the solid substance. The bonding forces are much greater than that of physical adsorption, releasing more heat. However ,the process cannot be easily reversed .Besides, this type of bonding promotes the chemical alteration of the adsorbed substance ,thus the adsorbate and adsorbent molecules never keep their original state after adsorption .Therefore, most of the adsorption refrigeration systems mainly involve physical adsorption [12,14].These cycle leads to intermittent operation ,with the adsorbent bed alternating between the adsorption and desorption stages. Thus, when continuous cooling effect is required, two or more adsorbent beds must be operating out of phase, which require that heat source is always available, which is not the case of solar radiation.

Applications of the Adsorption process

- Solid adsorbents, in combination with suitable adsorbate can be used in air separation systems to separate gases,
- The principle of adsorption is used in refrigeration cycles to provide air-conditioning or for ice-making purposes. Several companies have successfully commercialized adsorption chillers,
- Desiccants such as silica gel and zeolite are also used in many systems to extract moisture from the air and prevent damage to products such as medicines, shoes, etc.
- Chemical Plant Uses;CO2 removal from ammonia synthesis gas,CO2 removal from H2/CO mixture,NOx removal from Gases,
- CO2 and water removal from air in air separation plants.

Basic solar adsorption refrigeration cycle

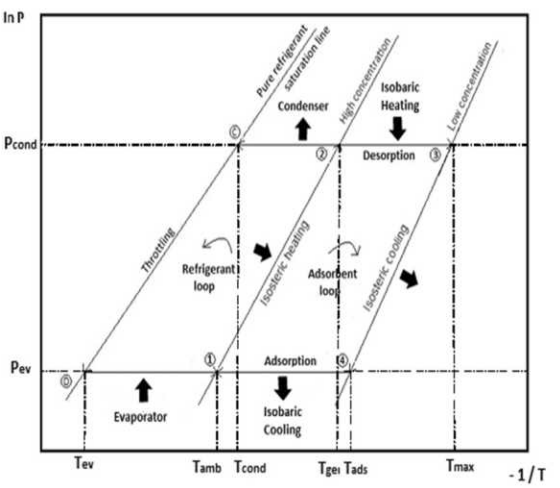


Fig.1. Theoretical adsorption cooling cycle indicating,
Phase 1: Isosteric heating (Process 1-2)
Phase 2: Isobaric desorption (2-3)
and condensation (2-C)
Phase 3: Isosteric cooling (3-4)
and throttling (C-D)
Phase 4: Isobaric cooling (4-1)
and Evaporation (D-1)

Selection of the working pair

Since the performance of the system greatly depends upon chosen working pair, choice should be made carefully The proper selection depends on the temperature of the heat source ,the desired characteristics of the refrigeration system, the properties of the working pair constituents and the affinity between them(which depend on the chemical , physical and thermodynamic properties of the substances),and also on their cost ,availability and environmental impact.

Choice of adsorbent

The most important features for choosing a suitable adsorbent are [13, 17]:

-Ability to adsorb a large amount of adsorbate when cooled to ambient temperature,

-Desorption of most (ideally all) of the adsorbate when heated by the available heat source,

-Higher apparent density; high pore volume; high surface area,

-Low specific heat; Good thermal conductivity, to shorten the cycle time; High porosity (in the order of 600 m^2/g)

-Chemically and physically compatible with the chosen refrigerant; Non-toxic and noncorrosive; Low cost and wide availability.

While selection there must be compromise between the high porosity required for rapid vapor diffusion and the high density required for good thermal conductivity [20]. The most commonly used adsorbents are activated carbon, zeolite and silica-gel .Activated carbon offers a good compromise between high adsorption and desorption capacities. Natural zeolites need to be present in large quantities since only a small amount of adsorb ate is desorbed during the temperature increase. However, the adsorption isotherms of zeolites have extremely non-linear pressure dependence, which is relevant for solar refrigeration applications .Contrarily, activated carbon and silica-gel present almost linear pressure-dependent isotherms. Silica-gel satisfies most of the criteria listed above but it is expensive and may not be available in most countries. Besides, the deterioration phenomenon of the adsorption capacity and aging of silica-gel is another current issue [13, 18, and 21].

Improvement in adsorbent

Basic adsorbent can we doped to present better performance when applied in adsorption refrigeration.Gordeeva et al. [31] presented new family of methanol sorbents ”salts in mesoporous silica” in adsorptive air conditioning system. They concluded that these synthesized composite sorbent has a higher adsorption capacity. Composite Licl/sio2 shows the highest sorption capacity Wads = 0.8 g/g.Tso et al [32] carried out study on composite adsorbent that were synthesized from activated carbon, silica gel and cacl2.Their result showed that the maximum adsorption capacity of 0.23 kg water/kg adsorbent, as recorded at 27˚ C and a water vapor pressure of 900 Pa .The maximum adsorption capacity of the raw activated carbon was 0.02 kg water /kg adsorbent at the same conditions. Wang et al [33] studied a specially treated activated carbon fiber and concluded that it might be a good substitute as refrigeration capacity and adsorption time are 3 times more and 1/5 to 1/10 of those normal AC, respectively .Other technique is to use consolidated adsorbents to increase the thermal conductivity. Which increase the density and thermal conductivity by 20% and 145-209 % respectively.

Choice of adsorbate

The adsorbate or refrigerant must fulfill the following requirements [9, 17]:

-Evaporation temperature below 0 degree C (for refrigeration purposes; it can be higher in the case of air conditioning applications); Small molecular size so as to facilitate the adsorption effect;

-High latent heat of vaporization and low specific volume when in the liquid state; High thermal conductivity; Low viscosity.

-Thermally stable with the adsorbent in the operating temperature range; chemically stable in the operating temperature range;

-Non-toxic, non-corrosive and non-flammable

-Low saturation pressures (between 1 to 5 atm) at normal operating temperature; Absence of ecological issues, unlike common refrigerants.

The most commonly used refrigerants are ammonia, methanol and water, which have relatively high latent heat values (1368, 1160 and 2258kJ/kg, respectively) and low specific volumes (of the order of 0.001 m3/kg). Water and methanol operate at sub-atmospheric saturation pressures at the operating temperatures. In the case of ammonia, it operates at higher pressures so small leakages can be tolerated in some cases; Ammonia is an example of positive refrigerant. Ammonia is toxic and corrosive, while water and methanol are not, but the second is flammable .Water is the most thermally stable adsorbate, closely followed by methanol and ammonia, in that order .However, water cannot be used for cooling purposes below 0 ˚C [18].

Review of working pairs

There are some observations regarding adsorbent–adsorbate working pairs for adsorption refrigeration system. The most commonly used working pairs are: zeolite-water, silica gel– water, activated carbon-methanol and activated carbon-ammonia [9]. Silica gel–water is ideal due to its low regeneration temperature; require low grade heat sources, commonly below 85

degree C. Moreover, water has the advantage of having a greater latent heat than other conventional refrigerants. It is suitable for air- conditioning application .(It is widely used in adsorption chillers).However, this pair has a low adsorption capacity as well as low vapor pressure ,which can hinder mass transfer .Furthermore ,this working pair requires vacuum conditions in the system ,where by any non-condensable gas will cause a significant reduction in the systems performance [5,25,27,28]. Activated carbon-methanol is one of the most common working pairs in adsorption refrigeration systems. It also operates at low regeneration temperatures (care must be taken since regeneration temperatures above 120˚ C promote the decomposition of methanol), while its adsorption-evaporation temperature lift is limited to 40 degree C. This pair is also characterized by its large cyclic adsorption capacity, low adsorption heat, low freezing point and high evaporation latent heat of methanol .However, activated carbon has a low thermal conductivity, acting like a thermal insulator and causing a decrease in the system's COP, and activated carbon–methanol also operates at vacuum conditions. Besides, methanol must be used with caution due to its high toxicity and flammability [15, 27, 29]. Sequentially, the activated carbon–ammonia pair requires regeneration temperatures that can exceed 150 ˚ C. Its adsorption heat is similar to that of the pair activated carbon–methanol, but it requires higher operating pressures(about 1600 kPa),which enhances the heat and mass transfer performance and reduces the cycle time, also preventing the infiltration of air into the system. All these factors help to increase the specific cooling capacity of the system. Moreover, this pair suitability to high temperature energy sources and the high cooling capacity of ammonia.However,the activated carbon has a lower adsorption capacity with ammonia than with methanol; furthermore, care must be taken due to the ammonia toxicity, irritating odor(even at low concentrations)and corrosive nature [4,26]. For the zeolite–water pair, the regeneration temperatures can go beyond 200˚ C, with an adsorption-evaporation temperature lift up to 70˚ C or more. This pair remains stable at high temperatures, and the water latent heat is much higher than that of methanol or other traditional refrigerants. However, a system operating with the zeolite-water pair is more fitted for air-conditioning applications due to the solidification temperature of water, which restrains the freezing process. Other disadvantages of this pair are the low adsorption quantity, which is about 0.2kg/kg. Anyanwu and Ogueke [29] evaluated the thermodynamic performance of different working pairs when designing a solar adsorption refrigerator .It was concluded that the activated carbon–ammonia pair presents the best results for ice making, deep freezing and food conservation applications. In turn, the zeolite–water pair is better suited for air-conditioning applications. Because the lowest evaporating temperature of water is 0 degree, and due to its high latent heat of vaporization, suitable for producing chilled water, it is a proper choice for air-conditioning purposes. The activated carbon–methanol pair is also suitable for ice production and freezing applications [3].

Improvement in conventional pair

Habib et al. [33] investigated three pairs of adsorbent/adsorbate according to Malaysia climate conditions. The selected pairs were activated carbon-methanol; A.C fiber ethanol, silica gel water, among them amount of adsorbate, adsorbed/desorbed was highest for A.C fiber ethanol and then for A.C/methanol and the lowest was for the silica gel/water. Simulations for the six working pairs [ACF (A-15)/ethanol, ACF (A-20)/ethanol, silica gel/water, Chemviron/R134a, Fluka/R134a and Maxsorb II/R134a] are carried out at partial vacuum and pressurized conditions. By loh et al. Among these working pairs, Maxsorb II/R134a has the highest uptake capacity about 0.36 kg/kg which is followed, respectively by ACF (A-20)/ethanol, ACF (A-15)/ethanol, Fluka/R134a, silica gel/water and Chemviron/R134a pairs. [34].El-sharkawy et al. [35] studied that Maxsorb III/methanol pair has superiority among other carbonaceous adsorbent/methanol pair for both A.C & Ice making. They concluded that adsorption capacity 1.76 times than that of activated charcoal/methanol. Allouhi et al [36] observed the optimal performance for 7 pairs of adsorbent/adsorbate (A.C fiber/methanol, A.C/methanol, A.C/ethanol, silica gel/water, zeolite/ethanol, and zeolite/water) according to Moroccan city fez, morocco. The maximum uptake was obtained by A.C fiber/methanol (0.3406 kg/kg) whereas the maximum SCOP was about 0.384 for silica gel/water.

Optimization of structure

To achieve the high performance of adsorbent bed, one of the method is to expedite the pass of the heat absorbed by bed to adsorbent for desorption of the adsorbate by improving the heat transfer structure of bed .second one is to enlarge the mass transfer channels in the adsorbent bed to make the desorbed adsorbate get into condenser quickly, for that bed pressure should remain steady in favorable to more adsorbate desorption. Another method is to reduce heat loss by improving thermal insulation. Xu ji et al. [37] has developed an enhanced heat and mass transfer finned tube casing. Comparing with the metal casing with the same dimension, finned tube has 51.4 % more heat transfer per unit length. The aluminum alloy was used to build the casing due to its high thermal conductivity, low specific heat. Also, some experiments corresponding to the adsorption/desorption process with and without valve control were implemented. The cooling efficiency with control valve is higher than that without valve control.

A majority of analysis are developed by using evacuated tube or flat plate collector, whereas less attention has been given to PTCs. Abu-Hamdeh et al. [38] has developed model which uses parabolic trough solar collector to improve its overall performance and productivity .It uses olive waste (as an adsorbent) with an methanol (as an adsorbate).They obtained optimum adsorbent mass, tank volume, collector area .but it require sun tracking system.

Table 1. Some typical research group on adsorption refrigeration system

Research Group	**Application**	**Pair**	**Source Temperature Or Solar radiation**	**Arrangement**	**Ice mass per day**	**COP**	**Remarks**
Li et al.[40]	Ice making	A.C-Methanol	100˚ to 130˚ C	Flat plate collector	4-5 kg	0.12-0.14	Temp. of ice reached to -1˚ C
N.M.Khattab [41]	Ice making	Charcoal-Methanol	120˚ to 130˚ C	FPC with side reflector	6.9-9.4 kg/m^2	0.136 -0.159	Granular dia. 5-7 mm Porosity-46.45%
Z.Taminot-Telto[42]	Ice making	A.C-NH3	90˚ to 120˚ C	Heating coil and boiler	SCP : 60 W/kg	0.12	Evaporation T up to -20˚C
Boubakri et al.[43]	Ice making	A.C-Methanol	19-29 MJ/m^2	FPC	5-11.5 kg/m^2	0.33 - 0.36	Collector Condenser mechanism
Abu-Hamdeh et al.[44]	Chilled water/A.C	Olive waste-methanol	95 to 120˚ C	Parabolic trough solar collector	-	0.75	Best cooling adsorbent mass: 30-40 kg, Collector area : 3.5-5 m^2
Xu ji et al.[45]	Ice making	A.C-Methanol	11-20 MJ/m^2	FPC with finned tube casing	6.5 Kg	0.039 - 0.122	Under a typical weather condition
Saha et al.[46]	Chilled Water	Silicagel-Water	55-75˚ C	Hot water supply	3.2 KW	0.36	Two stage adsorption chiller, Properties are given
Wang et al.[47]	Chilled Water	A.C-Methanol	80 -110˚ C	Waste heat driven	2.6 kg/kg &SCP 150 W/kg	0.4	Dual refrigeration & A.C
W.Wang et al.[48]	Review paper	Activated carbon & ACF	-	-	-	-	ACF provide better surface area and better mass transfer.

The adsorption refrigeration tube (ART) has drawn increasing attention because of its advantages of no moving parts, compact structure. Zhao et al. [39] has developed a new design of ART, in which activated carbon–methanol was selected as the working pair for either refrigeration or air-conditioning purposes. They employed condenser, evaporator and generator in a single tube. A concept of transient boundary i.e. transient pressure (and transient vapor density) was introduced for the first time into the model.

Conclusion

Low specific cooling power of the system leading to bigger sizes of the chillers and comparatively higher investment cost are major causes to prevent successful commercialization of the technology. It still needs a lot of research on adsorbent materials, improved heat and mass transfer, advanced cycles, etc. to make this technology a competitive one.

This work will help to understand basic phenomenon of ARS as well as selection criteria to select suitable pair according to application. In our opinion, the adsorption technology would direct towards development of Ecofriendly green technology for refrigeration purpose in remote areas.

Acknowledgment

It is with immense pride and pleasure to express my sincere gratitude to my guides (Prof.) Dr. N.S.Mehta & Jaspal Dabhi, for his encouragement and constant help throughout this research work right from its inception. Authors also thankful to the principal and Head of the mechanical engineering department of L.D.R.P-ITR, for their kind support and necessary guidance.

References

[1] S.C.Kaushik, A.Mahesh,"Solar adsorption cooling system: some materials and collectors aspects", Centre for Energy Studies, IIT Delhi.

[2] Edmunds JA, Wuebles DL, Scott MJ,"Energy and radiative precursor emissions", in: Proceedings of the 8th Miami International Conference on alternative energy sources; 1987 December p.14–16.

[3] Choudhury B,Chatterjee PK, Sarkar JP,"Review paper on solar powered air conditioning through adsorption route", Renew Sustain Energy Rev 2010;14(8),2189–95.http://dx.doi.org/10.1016/j.rser.2010.03.025.

[4] Fan Y, Luo L, Souyri B,"Review of solar sorption refrigeration technologies: development and applications", Renew Sustain Energy Rev 2007; 11 (8):1758–75, http://dx.doi.org/10.1016/j.rser.2006.01.007.

[5] Ullah KR, Saidur R, Ping HW, Akikur RK, Shuvo NH,"A review of solar thermal Refrigeration and cooling methods", Renew Sustain Energy Rev 2013;24:499–513,http://dx.doi.org/10.1016/j.rser.2013.03.024.

[6] Scientific Assessment of Ozone Depletion: 2010.Executive Summary. Global Ozone Research and Monitoring Project Report No.52.World Meteorological Organization Geneva, Switzerland; 2011. Available from: http://www.esrl.noaa.gov/csd/assessments/ozone/2010/executive summary/.Last accessed 22 May2014.

[7] Sarbu I, Adam M,"Applications of solar energy for domestic hot water and buildings heating/cooling", IntJ Energy2011; 5(2):34–42 Available From: http://www.naun.org/multimedia/NAUN/energy/20-553.pdf.Last accessed 10.10.2013.

[8] Desideri U, Proietti S, Sdringola P,"Solar-powered cooling systems: Technical and Economic analysis on industrial refrigeration and air-conditioning applications", Appl Energy2009; 86(9):1376–86.

[9] Choudhury B,Saha BB, Chatterjee PK ,Sarkar JP,"An over view of developments In adsorption refrigeration systems towards a sustainable way of cooling",ApplEnergy2013;104:554–67,http://dx.doi.org/10.1016/j.apenergy.2012.11.042

[10] Wang RZ, Oliveira RG. Adsorption refrigeration – An efficient way to make good Use of waste heat and solar energy.ProgEnergyCombustSci2006; 32(4):424–58, http://dx.doi.org/10.1016/j.pecs.2006.01.002.

[11] Wang LW, Wang RZ, Oliveira RG,"A review on adsorption working pairs for refrigeration", Renew Sustain Energy Rev2009;13(3):518-34, http://dx.doi.org/10.1016/j.rser.2007.12.002.

[12] Alghoul MA,Sulaiman MY, Azmi BZ, Wahab MA, "Advances on multi-purpose solar Adsorption systems for domestic refrigeration and water heating", Appl Therm Eng. 2007;27(5-6):813–22,http://dx.doi.org/10.1016/j.applthermaleng.

[13] Sumathy K, Yeung KH, Yong L,"Technology development in the solar adsorption refrigeration systems", Prog Energy Combust Sci 2003; 29(4):301–27, http://dx.doi.org/10.1016/S03601285 (03)000285.

[14] Anyanwu EE,"Review of solid adsorption solar refrigeration II: an over view of the principles and theory", Energy Convers Manag 2004; 45(7-8):1279–95, http://dx.doi.org/10.1016/j.enconman.2003.08.003

[15] Vasta S,Maggio G,Santori G,Freni A,Polonara F,Restuccia G,"An adsorptive solar ice-maker dynamic simulation for north Mediterranean climate",EnergyConversManag2008;49(11):3025–35;http://dx.doi.org/10.1016/j.enconman.2008.06.020.

[16] Riffel DB, Belo FA., Leite APF,"Ar condicion ado solar por adsorção: Fundamentos E estado da arte.In: I CBENS – I Congresso Brasileiro de Energia Solar", Fortaleza, Brazil; 2007.8–11April.Portuguese.

[17] Odesola IF, Adebayo J," Solar adsorption technologies for ice-making and recent Developments in solar technologies: a review", Int J Adv. Eng. Technol 2010; 1(3):284–303.

[18] Ruthven DM,"Principles of adsorption and adsorption processes", NewYork: John Wiley&Sons; 1984.

[19] Tather M ,Tantekin-Ersolmaz B, Erdem-Senatalar A,"A novel approach to enhance Heat and mass transfer in adsorption heat pumps using the zeolite-water pair", Micropor MesoporMater1999;27(1):1–10, http://dx.doi.org/ 10.1016/S1387-1811(98)00174-7.

[20] Wang D,Zhang J,Xia Y,Han Y,Wang S,"Investigation of adsorption performance Deterioration in silica gel–water adsorption refrigeration",EnergyConversManag2012;58:157-62,http://dx.doi.org/10.1016/j.enconman.2012.01.013.

[21] Gordeeva LG ,Freni A ,Restuccia G ,Aristov YI,"Influence of characteristics of methanol sorbents "salts in mesoporous silica" on the performance of adsorptive air conditioning cycle",IndEngChemRes2007;46(9):2747–52,http://dx.doi.org/10.1021/ie060666n.

[22] Vasiliev LL ,Gulko NV ,Khaustov VM,"Solid adsorption refrigerators with active carbon-Acetone and carbon ethanol pairs", In: Proceedings of the International Sorption Heat Pump Conference,vol.1,Montreal,Quebec,Canada; 1996.p.3–6.

[23] Wang RZ ,Jia JP, Zhu YH,TengY,Wu JY,Cheng J," Study on a new solid adsorption Refrigeration pair: active carbon fiber-methanol", J Sol Energy Eng. 1997;119(3):214–8,http://dx.doi.org/10.1115/1.2888021.
[24] Wang D, Zhang J,Tian X,Liu D,Sumathy K,"Progress in silica gel–water adsorption Refrigeration technology", Renew Sustain Energy Rev2014;30:85–104,http://dx.doi.org/10.1016/j.rser.2013.09.023.
[25] Al-Alili A,Hwang Y,Radermacher R,"Review of solar thermal air conditioning technologies",Int J Refrig 2014;39:4–22, http://dx.doi.org/10.1016/j.ijrefrig. 2013.11.028.
[26] Wang DC ,Li YH ,Li D ,XiaYZ ,Zhang JP,"A review on adsorption refrigeration technology and adsorption deterioration in physical adsorption systems", Renew Sustain Energy Rev2010;14(1):344–53, http://dx.doi.org/10.1016/j.rser.2009.08.001.
[27] Wang D, Z hang J, Yang Q, Li N, Sumathy K."Study of adsorption characteristics in silica gel-Water adsorption refrigeration", ApplEnergy2014; 113:734–41, http://dx.doi.org/10.1016/j.apenergy.2013.08.011.
[28] Mahesh A, Kaushik SC, "Solar adsorption cooling system: An overview", J Renew Sustain Energy2012; 4(2):022701, http://dx.doi.org/10.1063/1. 3691610.
[29] Anyanwu EE, Ogueke NV, "Thermodynamic design procedure for solid adsorption solar refrigerator", Renew Energy2005; 30(1):81–96, http://dx.doi.org/10.1016/j.renene.2004.05.005.
[30] Lewis JS, Chaer I, Tassou SA,"Fostering the development of technologies and practices to reduce the energy inputs into the refrigeration of food -reviews of Alternative refrigeration technologies", Centre for energy and built environment Research school of engineering and design brunel university;2007.
[31] Larisa G. Gordeeva ,Angelo Freni ,Giovanni Restuccia, and Yuri I Aristov,"Influence of Characteristics of Methanol Sorbents "Salts in Mesoporous Silica "on the Performance of Adsorptive Air Conditioning Cycle", Ind. Eng. Chem. Res., 2007, 46 (9), pp. 2747–2752.
[32] Tso C.Y. and Chao C. Y.H,"Activated carbon, silica-gel and calcium chloride composite adsorbents for energy efficient solar adsorption cooling and dehumidification systems", International Journal of Refrigeration 35 (2012) 1626 – 1638
[33] Koyama Khairul Habib,Bidyut Baran Saha Shigerub,"Study of various adsorbent refrigerant pairs for the application of solar driven adsorption cooling in tropical climates", Applied Thermal Engineering 72 (2014) 266e274.
[34] W.S. Loh ,I.I. El-Sharkawy , K.C. Ng , B.B. Saha,"Adsorption cooling cycles for alternative adsorbent/adsorbate pairs working at partial vacuum and pressurized conditions", Applied thermal Engineering 29 (2009) 793-798.
[35] I.I. El-Sharkawy, M. Hassan, B.B. Saha, S.Koyama, M.M. Nasr, "Study on adsorption of methanol onto carbon based adsorbents", international Journal of Refrigeration 32 7 (2009), 1579-86.
[36] A. Allouhia, T. Kousksoub ,A. Jamila, T. El Rhafikic, Y. Mourada, Y. Zeraoulib," Optimal working pairs for solar adsorption cooling applications", Energy 79 (2015)235-247.
[37] Xu Ji ,Ming Li ,Jieqing Fan ,Peng Zhang , Bin Luo ,Liuling Wang, "Structure optimization and performance experiments of a solar-powered finned-tube adsorption refrigeration system", Applied Energy 113(2014),1293-1300.
[38] Nodal H. Abu-Hamdeh, Khaled A. Alnefaie, Khalid H. Almitani,"Design and performance characteristics of solar adsorption refrigeration System using parabolic trough collector: Experimental and statistical optimization technique", Energy Conversion and Management 74(2013) 162-170.
[39] Y.L. Zhao, Eric Hu, Antoni Blazewicz,"A non-uniform pressure and transient boundary condition based dynamic modeling of the adsorption process of an adsorption refrigeration tube", Applied Thermal Engineering 29 (2009) 793–798.
[40] M. Li , R.Z. Wang, Y.X. Xu, J.Y. Wu, A.O. Dieng,"Experimental study on dynamic performance analysis of a flat-plate solar solid-adsorption refrigeration for ice maker", Renewable Energy 27 (2002) 211–221
[41] N.M. Khattab,"A novel solar-powered adsorption refrigeration module", Applied Thermal Engineering 24 (2004) 2747–2760.
[42] Z.Tamainot-Telto, R.E critoph,"Adsorption refrigerator using monolithic carbon-ammonia pair", Int J. Refri. Vol. 20, No. 2, pp. 146-155, 1997.
[43] A. Boubakri, J. J. Guilleminot , F. meunier,"adsorptive solar powered ice maker: experiments and model", Solar Energy Vol. 69, No. 3, pp. 249–263, 2000.
[44] Nidal H.Abu-Hamdeh , Khaled A. Alnefaie , Khalid H. Almitani,"Design and performance characteristics of solar adsorption refrigeration system using parabolic trough collector: Experimental and statistical optimization technique"; Energy Conversion and Management 74 (2013) 162–170.
[45] Xu Ji, Ming Li, Jieqing Fan, Peng Zhang, Bin Luo a, Liuling Wang, "Structure optimization and performance experiments of a solar-powered finned-tube adsorption refrigeration system", Applied Energy 113 (2014) 1293–1300
[46] B.B. Saha, A. Akisawa, T. Kashiwagi, "Solar/waste heat driven two-stage adsorption chiller: the prototype", Renewable Energy 23 (2001) 93–101.
[47] R.Z. Wang, J.Y. Wu, Y.X. Xu, W. Wang, "Performance researches and improvements on heat regenerative adsorption refrigerator and heat pump", Energy Conversion & Management 42 (2001) 233-249
[48] L.W. Wang, R.Z. Wang, R.G. Oliveira,"A review on adsorption working pairs for refrigeration", Renewable and Sustainable Energy Reviews 13 (2009) 518–534.

Pedestrian and Vehicle Detection for Advanced Driver Assistance Systems

P Lavanya*, G Harshith**, Chiraag*** and S S Shylaja****
* Student, PESIT
lava281995@gmail.com
**Student, PESIT
harshit.gowda78@gmail.com
***Student, PESIT
chiraag7@gmail.com
****Professor, PESIT
shylaja.sharath@pes.edu

Abstract: Application of image processing to real-time systems is growing at a rapid speed and has great potential for exploitation. Human errors are inevitable and a system designed with predictable response time and a deadline, aid in overcoming these shortcomings. This paper implements a pedestrian and vehicle detection system that combines Histogram of Gradients (HOG) and Support Vector Machine (SVM) with a Cascade Classifier and Haar. The implementation described processes every 10th frame of the video and is tested on standard datasets. Error percentage was calculated using Weak law of large numbers for the algorithms and the improvised result after preprocessing the image.
Our approach extends the functionality of the inbuilt people detector and trains a classifier to detect 4 wheelers, it minimizes processing delay and errors by tweaking the region of interest as per the requirement of a real-time application.

Keywords: HOG, SVM, Cascade Classifier, HAAR, Weak law.

Introduction

Urbanization has led to the rapid increase in the number of vehicles on the road, thereby increasing the need to control congestion due to vehicular traffic and improvise safety measures for pedestrians. Amongst the top causes for road accidents the most prominent ones are due to distracted driving, recklessness and weather adversities. Traffic monitoring and surveillance are important for road safety and management. Advanced driver system assistance has played an integral role in development of Intelligent Transportation System. Looking at a statistical projection of traffic fatalities, the death toll due to road accidents is about 1 million per annum, urging the need to improvise safety measures overcoming the ignorance and carelessness of humans. Our implementation of an advanced driver system aims to do so by integrating a pedestrian and vehicle detection system under adverse weather conditions like fog.

Related Work

There is an extensive literature on object detection and here we mention a few relevant examples that helped us decide implementation specifics for vehicle and pedestrian detection.
Corner detection was tested; initially the drawback was low accuracy and false detection. Sift and Surf was not accurate enough for pedestrians. The trade-off between accuracy and output latency time was solved by choosing HOG and SVM which were comparatively slow but accurate for pedestrians and Haar Feature-based Cascade Classifiers for vehicles which would show much lesser variations in comparison to pedestrians.
Histograms of Oriented Gradients for Human Detection [1] combined with ideas from Chris McCormick [3] tutorial implements a Histogram normalization and block normalization technique that make it invariant to illumination and to an extent posture changes as well. Hence OpenCV, comprising of a class to run the HOG person detector was used for implementing the pedestrian detection. Haar Feature-based Cascade Classifiers [2] implemented for face detection was extended and applied using a machine learning technique where a cascade function is trained from a set of positives and negative images, which is then used to detect cars in test images/frames. Our implementation combines both techniques to extract the best features in each and tries to minimize existing flaws and response time.

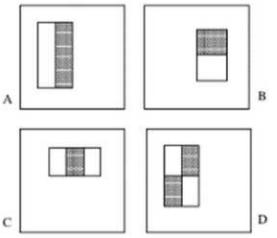


Figure 1. Haar rectangle features [4]

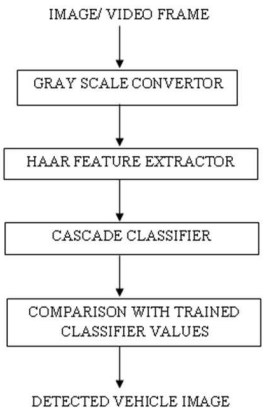


Figure 2. Data flow diagram for vehicle detection

Constraints on a Real Time System

Continuous and interactive processing: Must process high frame rate video input to prevent missing detail under fixed time constraints.
Fixed Deadlines: Missing a deadline implies a system is malfunctioning at its peak.
Predictable response time: In order to alert a driver, the response time should be within a specific range.
Efficient functioning under adverse weather conditions: Detection must also be successful under fog which produces images with high noise.

Proposed Method

Vehicle Detection

To identify vehicles, the cascade classifier (Haar) was trained to recognize objects within the image; this required a set of positive images that contain the object we are trying to detect and negative images. This completed the phase of asset preparation. [4]. A metadata file is then generated, which contains the file name, number of objects, position and dimensions of objects. This metadata file is then used to create a vector file that will be used by the classifier containing all positive images within a single file. The script available in OpenCV library is then invoked to train the classifier. A Python script using the classifier to detect vehicles and to highlight each detected vehicle was implemented.

Pedestrian Detection

HOG was chosen as a feature descriptor [3] as it provides the functionality of reproducing the same or similar feature descriptor when an object is viewed under noise conditions. The HOG person detector is used that implements a sliding window detector technique which traverses the entire image that in turn influences the processing delay. For every position of the window, a HOG feature descriptor is generated which is then fed into a trained SVM, which classifies it as either “pedestrian” or “not a pedestrian”. The implementation of the HOG person detector used here takes reference from Chris Mccormick’s work [3] that uses a detection window of 64X128 pixels confining it to a 9-bit histogram ranging from 0 to 180 degrees. To make the process invariant to changes in illumination or noise, the cells are first grouped into blocks prior to normalization (Fig. 3). This implementation of HOG was used from the OpenCV documentation to detect pedestrians.

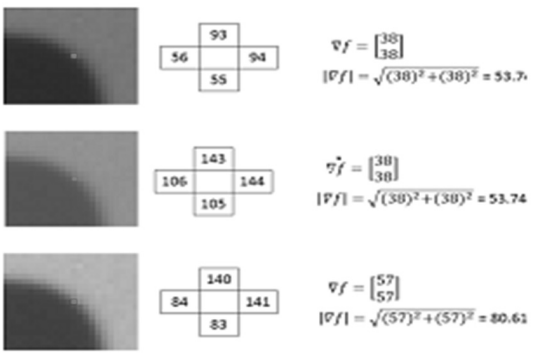


Figure 3. Illumination and pixel invariance [3]

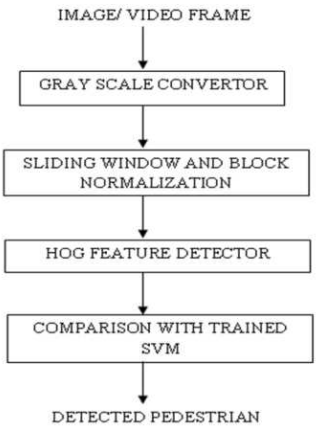


Figure 4 .Data flow diagram for pedestrian detection

Video Processing

The trade-off between output latency and accuracy of detection was solved by processing every 10^{th} frame of the video. The region of interest is chosen so that the unnecessary parts of the image are not considered for processing. Both pedestrian and vehicle detection are integrated to get a combined result.

Error Calculation

A mathematical model for the determination of error in the inbuilt HOG and HAAR was built using Weak Law of Large Numbers.

Error Calculation for pedestrians

To provide a mathematical model for percentage of error in the feature descriptor provided by HOG for person detection, we used weak law of large numbers [8] to state the error percentage with an accuracy meter and confidence meter. Various symbols used are explained in Table1.

Applying Pollsters algorithm,

$$P(|\mu - f| \geq \delta) \leq C$$

Applying Chebyshev Inequality states,

$$P(|\mu - f| \geq C) \leq \frac{\sigma 2}{c2}$$

Combining the two algorithms, we establish a result for n=504, the probability of detection of the pedestrian was found to be 58± 9% with a confidence of 95%.

Table 1. Symbols and their meaning

Symbol	Quantity
F	Probability of detection
μ	Sample probability of the sample space
X	1 for successful and 0 for unsuccessful detection
Δ	Error
C	1-{confidence of μ lying in (f ± δ) }
σ^2	Variance
σ_{mn}	σ^2/total number of objects

Figure 5. Processing stages for pedestrian detection

Figure 6. Detection under Fog condition

Figure 7. Processing stages for vehicle detection

Error calculation for vehicles

To minimize the detection of extra objects, we used a preprocessing technique that changed the region of interest on which the algorithm is implemented. The upper 33% of the image is cropped to remove extra objects detected in that horizon and above, also significantly reducing the processing delay. The region of interest is confined to the road and eliminates detection on the side pavements by mathematical manipulations, increasing the accuracy factor by 73%.

Result

We have implemented a detection system for vehicles using a Haar Feature-Based Cascade classifier, combined with HOG and SVM for pedestrian detection that provides low latency outputs using preprocessing techniques.

The algorithm was tested on the following datasets:

Frames of video [5]

A mathematically modeled proof for improvement in accuracy rate was determined using weak law of large numbers, pedestrian detection was found to have a success rate of 58%±9% (error rate) with a confidence of 95% in the video frames. The modification of region of interest improved accuracy of detection of vehicles in the video by 73%.

Pedestrian dataset from MIT [6]

Using weak law of large numbers, pedestrian detection was found to have a success rate of 84%±5% (error rate) with a confidence of 90% in the training data.

The algorithm successfully detects pedestrians under fog conditions (Fig. 6).

Conclusion

A detection system using a feature descriptor was implemented that detects both pedestrians and vehicles, decreasing latency time of the current algorithm. It currently employs a preprocessing technique to minimize false positives, although there is room for optimization using algorithmic techniques. The code detects pedestrians when full body is visible, expansion of this to two wheelers by increasing the size of dataset has potential scope.

Acknowledgement

We wish to express our sincere gratitude to our college PES Institute of Technology, Bengaluru for providing us the opportunity to implement this project and for their support.

References

[1] NavneetDalal and Bill Triggs, “Histograms of Oriented Gradients for Human Detection”, CVPR Conference.
[2] http://docs.opencv.org/master/d7/d8b/tutorial_py_face_detection.html#gsc.tab=0
[3] https://chrisjmccormick.wordpress.com/2013/05/09/hog-person-detector-tutorial/
[4] P Viola and M Jones. “Rapid Object detection using a boosted cascade of simple features”, CVPR Conference.
[5] https://www.youtube.com/watch?v=98t-yh4MC6Y
[6] http://cbcl.mit.edu/software-datasets/PedestrianData.html
[7] http://cogcomp.cs.illinois.edu/Data/Car/
[8] https://www.youtube.com/watch?v=j9WZyLZCBzs&list=PLu14u3cNGP61 MdtwGTqZA0MreSaDybji8

Artificial Neural Network for Detection and Location of Faults in Mixed Underground Cable and Overhead Transmission Line

Ankita Nag * and Anamika Yadav **
* **Department of Electrical Engineering
National Institute of Technology, Raipur, C.G., India
Email id: nag2408ankita@gmail.com, ayadav.ele@nitrr.ac.in

Abstract: Power grid in India is laid almost with the overhead lines (OH). With the capacity of XLPE cables to transmit high voltages has led power system engineers to take keen interest in mixing OH lines with underground (UG) cables where the environmental effects, increasing population or the areas where right of way is a constraint for connecting the grid line. Faults in mixed system require broader aspects of consideration and analysis as the UG cable and OH lines both exhibits different characteristics. Conventional tracer and terminal methods are time consuming therefore computer based methods are emerging as a solution to remove their drawbacks and to provide more accurate results for fault detection, classification and location faster. This paper deals with use of ANN for protection of mixed line taking into account the parameters at each point for cables and sequence components for OH line. In MATLAB/SIMULINK software simulation is carried out for verification of results.

Keywords: Artificial Neural Network, Discrete Fourier Transform, Cable Parameters.

Introduction

In present scenario when the demand for power is increasing along with the increase in the population the hybrid transmission will be the solution when there is constraint of right of way in densely populated area. This combination can also be implemented to connect the existing grids with the off-shore wind farms with better reliability. To make the system operate successfully the response to the fault for the detection and location purpose should be minimal. This helps to reduce the time required to restore the system back to normal condition. Thus with the advancement in the technologies future power grids can be implemented with relays employing the knowledge based techniques.

For protective relaying of mixed systems generally two methods are used for finding fault location in UG cable and OH line [1].1. Phasor based method also called as impedance based method which uses fundamental component of signals [2, 3, 4]. 2. Travelling based method [5]. Travelling based method are sensitive to reflected waves when fault occurs in nearby lines and even to noise. Combination of these two methods has been used in [6, 7]. Unequal impedance is the major problem of the mixed system. Adaptive Network based Fuzzy Inference System is also been used for fault location of ground faults and line faults in mixed lines [8]. Discrete wavelet transformation and Support Vector Machine for hybrid transmission line where DWT is used to extract transient information from measured voltages and SVM to identify faulty section has also been proposed [9]. A new single-ended travelling wave fault location algorithm in which samples just from voltage transients generated by fault clearing action of circuit breaker are taken from the sending end of cable line has been proposed for fault location in mixed line [10].

Researches based on ANN for the protection of power system have taken up the interest of researchers. ANN which works on the learning ability of human from its surroundings, adapting itself to it and responding accordingly is utilized for the protection purpose by training the network with past records, measurements, available data and observations [2]. ANN imitates the biological capability of solving linear [11] and non-linear problems. ANN is been implemented for complete protection scheme for fault detection, classification and location in Transmission lines [12]. ANN is also applied for fault analysis in distribution networks [13]. However there are few researches done considering shunt faults in hybrid underground cable and overhead transmission line using ANN. This paper presents an Artificial Neural Network approach to simulate the hybrid model for distance protection. Simulation is done to detect and locate the faults. Fig. 1 shows the single line diagram of a three-phase mixed transmission line connected with source and Fig. 2 shows the process flow chart of the proposed model.

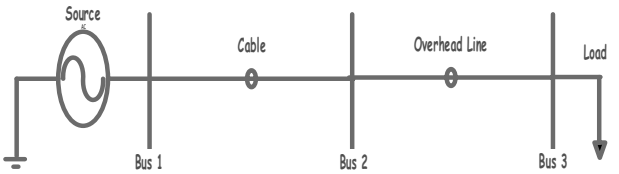


Figure 1. Single line diagram of a three-phase mixed transmission line connected with source

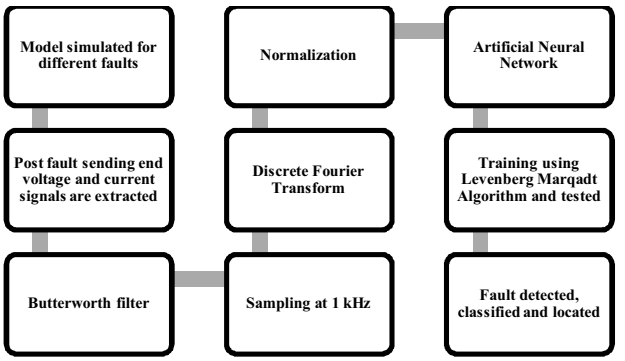


Figure 2. Process flow chart of the proposed model

Mixed Underground Cable and OH Transmission Line Single Line Diagram

A single line diagram of mixed underground cable and overhead transmission line is shown in Fig. 3 and during fault condition at cable in Fig 4, at junction in Fig 5 and at overhead line in Fig. 6.

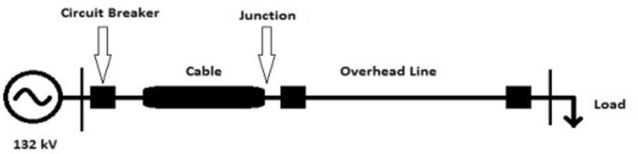


Figure 3. No fault

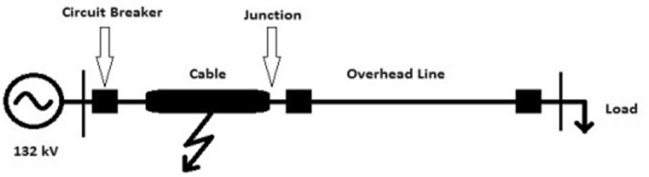


Figure 4. Fault at UG cable section

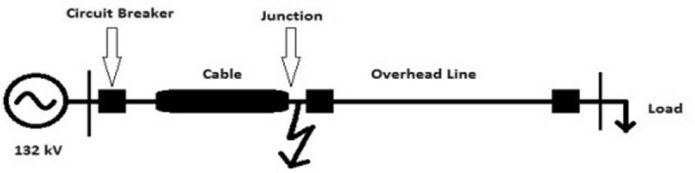


Figure 5. Fault at junction

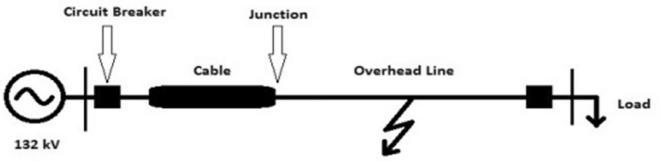


Figure 6. Fault at OH line

System under Study

The system simulated for single line to ground fault is composed of 132 KV, 50 Hz, 15 km mixed transmission lines with 3 km UG cable and 12 km OH line section, connected to a source at one end and to three-phase series RLC load, as shown in Fig. 7. All components are modeled by the MATLAB R2013a Simulink and SimPowerSystem toolbox. Short circuit capacity of the equivalent Thevenin source of the line is considered to be 1250 MVA. *X*/*R* ratio is 10. The cable is simulated using 3 core cable with screen (pi-model) where cable parameters are taken into consideration and can be calculated using power_cableparam command in MATLAB. This command open cable geometry dialog box where according to requirement we can give various specifications. This data is then fed as preloadFcn in callbacks of model properties of the simulation model. Then in initFcn the R, L and C matrix in Ohm, Henry and Farad per km respectively is defined. The overhead line is simulated using three phase pi-section line model. Sequence components are used for line pi-section . The cable specifications used is shown in Fig. 8. Various transmission line parameters are shown in Table 1 [10].

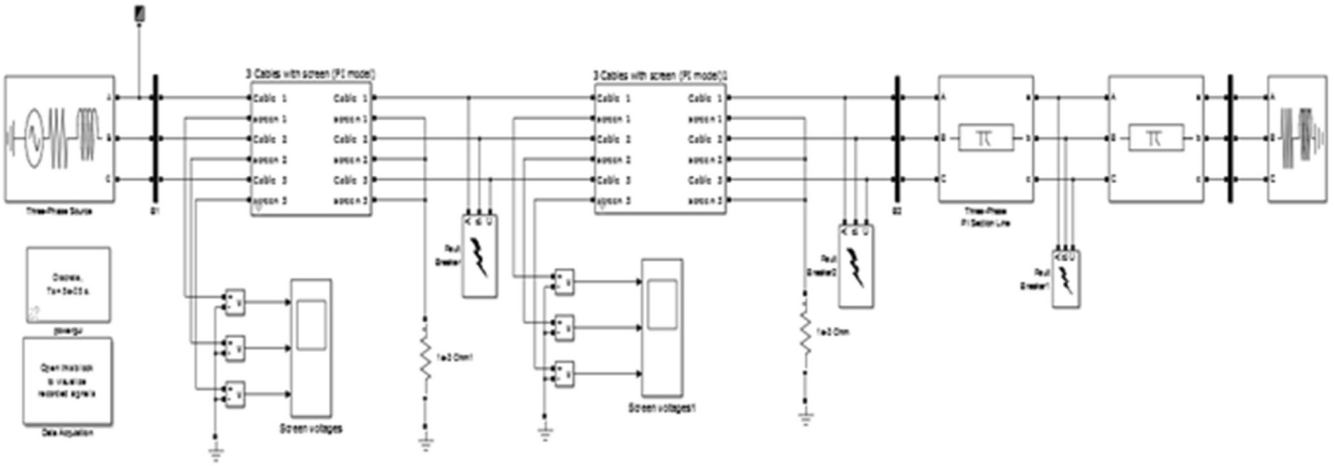


Figure 7. Power system model simulated in MATLAB Simulink software

Figure 8. Cable Specifications

Table 1. Line Parameters

Positive sequence resistance R1, Ω/km	0.3317
Zero sequence resistance R0, Ω/km	0.4817
Positive sequence inductance L1, H/km	1.326e-3
Zero sequence inductance L0, H/km	4.595e-3
Positive sequence capacitance C1, F/km	0.008688e-6
Zero sequence capacitance C0, F/km	0.00476e-6
Line length, km	12

Data Generation

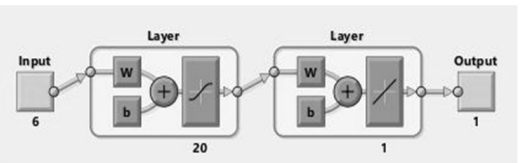
The sending end voltage and current values for single line to ground fault are passed through filter having cut-off frequency of 400 Hz. Then by using sampling frequency of 1 kHz it is passed to Discrete Fourier Transform block to extract the fundamental values. The obtained values are then separated into training inputs six in number (V_{af}, V_{bf}, V_{cf}, I_{af}, I_{bf}, I_{cf}) and testing inputs and fed to the Artificial Neural Network for training purpose. Being knowledge based algorithm target is provided for learning the various features. The algorithm used by the Artificial Neural Network is Levenberg Marquardt.

Sequence of Simulation and generation of input data

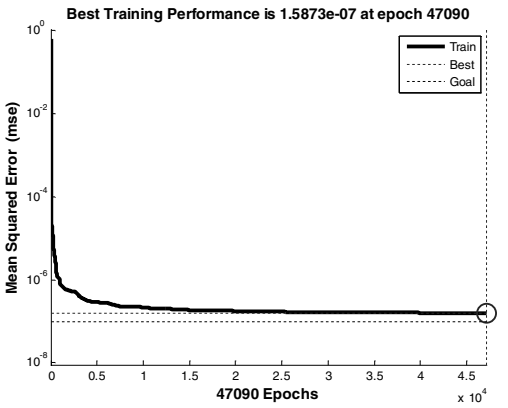
- Single line to ground fault for fault location and all types of fault for fault detection is applied for various fault resistances (FR) and fault inception time (FIT) at different section of cable keeping the line parameters and length of line fixed. Fundamental values are obtained.
- Single line to ground fault for fault location and all types of fault for fault detection is applied for various fault resistances and fault inception angle is then applied at different section of overhead line keeping the cable parameters and length of cable fixed. Obtain the fundamental values.
- Similarly keeping cable and line parameters and length constant fault is applied at junction.
- To create input set merge the cable data, junction data and overhead line data.
- Set the target values for fault detection and fault location.

Fault Detection

Fault detection is done by giving target for faulty data as input 1 and non-faulty data as input 0 and this logic is shown in table 2. The Artificial Neural Network is then trained using one hidden layer having 20 neurons with tansig as activation function and one output layer showing presence or absence of fault with purelin activation function. The ANN architecture and performance plot is shown in Fig. 9(a) and (b) respectively.



(a)



(b)

Figure 9. (a) ANN architecture, (b) Performance index and

Training and testing pattern for fault detection used in the proposed work

The training has been done with FIT= 40 ms, FR= 0, 50, 100 Ω for all types of ground fault. The testing has been done for all types of fault. Results of simulation for AG fault at 2.5 km, FR=20 Ω, FIT=43.333 ms in UG cable, ABG fault at junction, FR=40 Ω, FIT=45 ms and ABC fault at 11 km, FR=60 Ω, FIT=50 ms in OH line is shown in simulation results.

Table 2. Logic for fault detection

Fault Detection	Logic Output
Healthy Condition	0
Faulty Condition	1

Fault Location

Fault location is done by giving input data set (V_{af}, V_{bf}, V_{cf}, I_{af}, I_{bf}, and I_{cf}) and trained and tested with target as location of fault with 2 hidden layers with activation function as tansig and one output layer with purelin and giving output as location to the fault. The ANN architecture and performance plot is shown in Fig. 10(a) and (b) respectively.

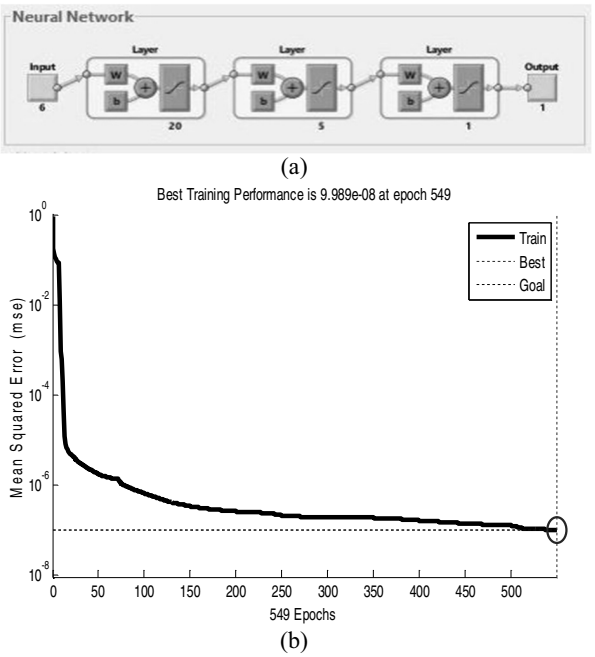


Figure 10. (a) ANN architecture, (b) Performance index

Training and testing pattern for fault location used in the proposed work

Training and testing pattern for fault resistance variation is shown in Table 3 and Table 4 respectively and inception angle variation is given in Table 5 and Table 6. Voltage and Current waveforms for phase A to ground fault is shown in Fig. 11. Voltage and Current Fundamental Values for phase A to ground Fault is shown in Fig. 12(a) and (b) respectively.

Case I. Variation in fault resistances

- Training
  The training is done keeping the fault inception time 40 ms and the fault resistance values as 0,50 and 100 for various lengths.
- Testing
  The testing is done keeping the fault inception time constant and changing the fault resistance values and lengths which are not used for training purpose.

Case II. Variation in Fault Inception Angle

- Training
  The training is done keeping the fault inception angle 0, 60,120,180 degree 40 ms and the fault resistance values as 0,50 and 100 for various lengths.

- Testing
  The testing is done keeping the fault inception 30, 90,150 and the fault resistance values as 2,48,52,98 and lengths which are not used for training purpose.

The relative error for the location purpose is given in equation (1) and the absolute error w.r.t. total line length is given in equation (2):

$$\text{Relative Error for Location} = E = \frac{\text{Actual Distance} - \text{Calculated distance}}{\text{Actual Distance}} \quad (1)$$

$$\text{Absolute Error w. r. t total length} = \frac{\text{Actual Distance} - \text{Calculated distance}}{\text{Total length of the mixed transmission line(15 km)}} \quad (2)$$

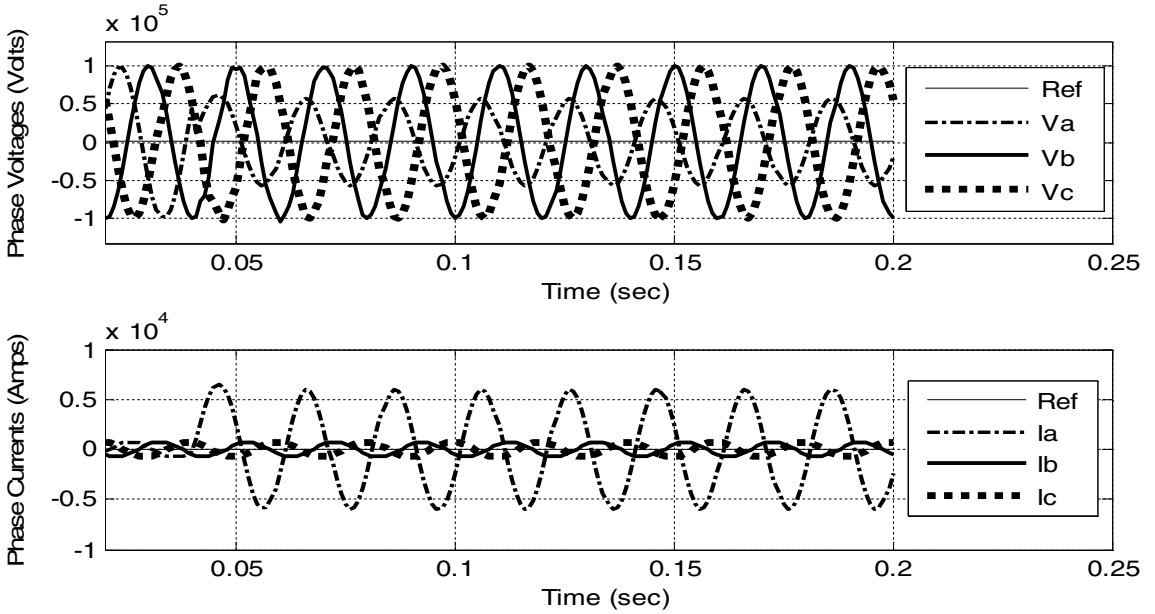


Figure 11. Phase voltages and currents for fault at phase A and Ground

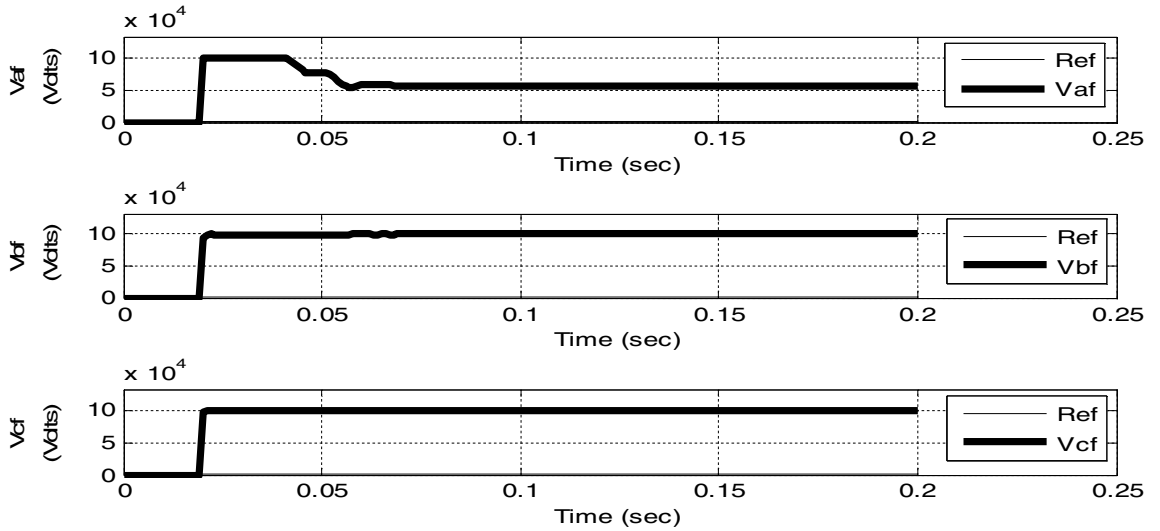


Figure 12. (a) Voltages Fundamental Values for phase A to ground fault

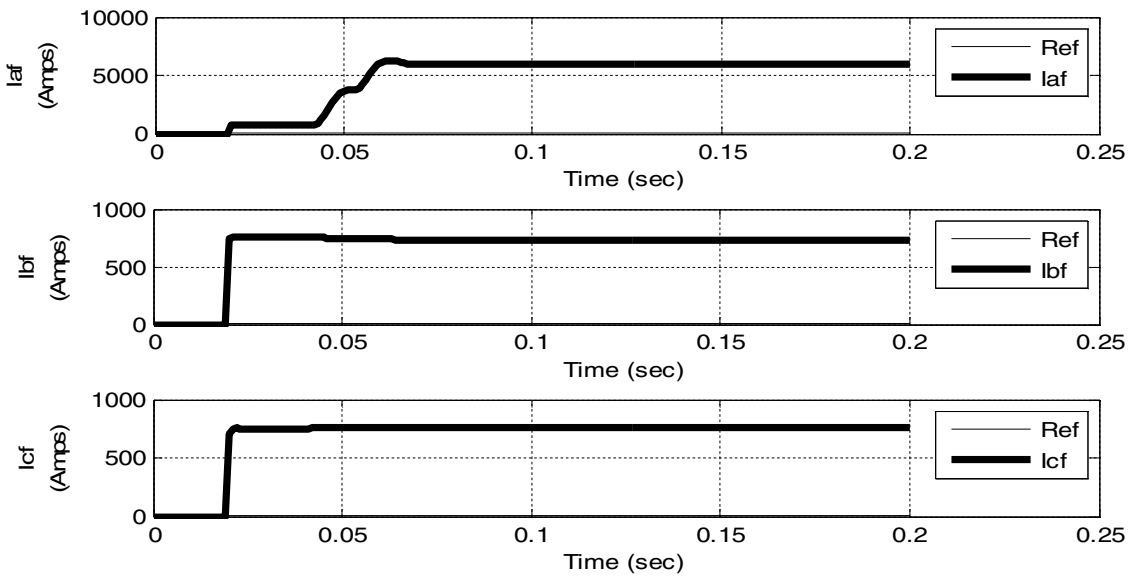


Figure 12. (b) Current Fundamental Values for phase A to ground Fault

Table 3.Variation in fault resistance (train pattern)

1.	Fault Inception Angle		40 ms (0 degree fault inception angle)
2.	Fault Resistance		0,50,100 (in Ohms)
3.	Fault Location	Underground cable	0.05, 0.15, 0.25,… 2.95 (in km)
		Overhead line	0.5, 1.5, 2.5,… 11.5 (in km)

Table 4.Variation in fault resistance (test pattern)

1.	Fault Inception Angle		0 degree fault inception angle
2.	Fault Resistance		2,48,52,98 (in Ohms)
3.	Fault Location	Underground cable	0.1,0.2,0.3,… 2.9 (in km)
		Junction	3.01 (in km)
		Overhead line	1,2,3,4,5,6,7,8,9,10,11 (in km)

Table 5.Variation in fault inception angle (train pattern)

1.	Fault Inception Angle		0,60,120,180 (in degrees)
2.	Fault Resistance		0, 50, 100 (in Ohms)
3.	Fault Location	Underground cable	0.05,0.15,0.25,... 2.95 (in km)
		Overhead line	0.5,1.5,2.5,… 11.5(in km)

Table 6.Variation in fault inception angle (test pattern)

1.	Fault Inception Angle		30,90,150 (in degree)
2.	Fault Resistance		2,48,52,98 (in Ohms)
3.	Fault Location	Underground cable	0.1,0.2,0.3,… 2.9 (in km)
		Junction	3.01 (in km)
		Overhead line	1,2,3,4,5,6,7,8,9,10,11 (in km)

Simulation Result

The proposed scheme for fault detection and location has been designed using ANN and tested with wide variation in fault parameters which are discussed in detail as below:

Fault detection

In Fig. 13 AG fault at 2.5 km, FR=20 Ω, FIT=43.333 ms in UG cable has been detected in 1.67 ms, in Fig. 14 ABG fault at junction, FR=40 Ω, FIT=45 ms has been detected in 2.00 ms and in Fig. 15 ABC fault at 11 km, FR=60 Ω, FIT= 50 ms in OH line has been detected in 3 ms.

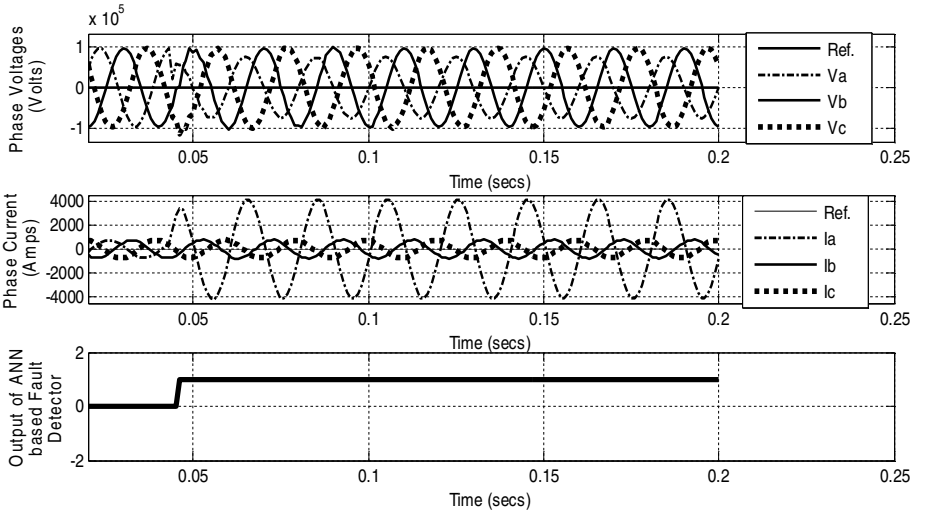


Figure 13. Test Result for AG fault at 2.5 km, FR=20 Ω, FIT=43.333 ms in UG cable

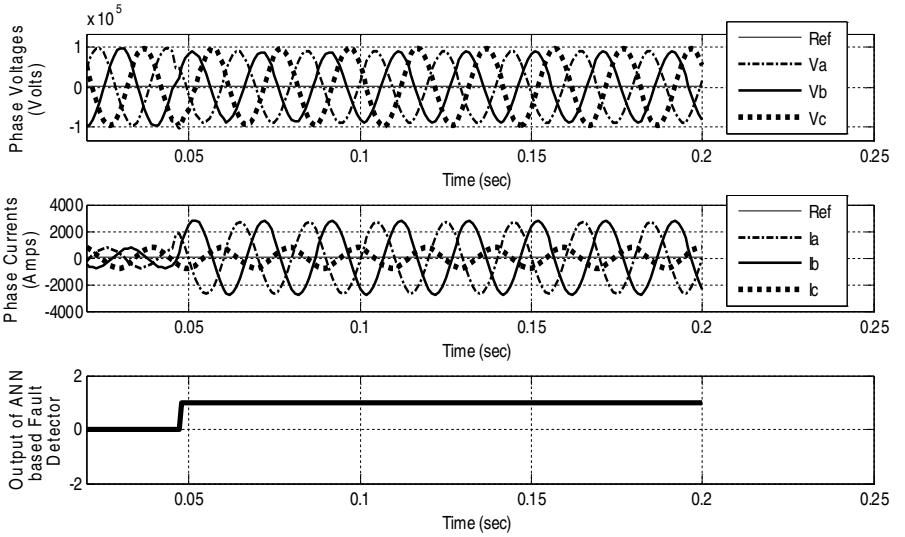


Figure 14. Test Result for ABG fault at junction, FR=40 Ω, FIT=45 ms

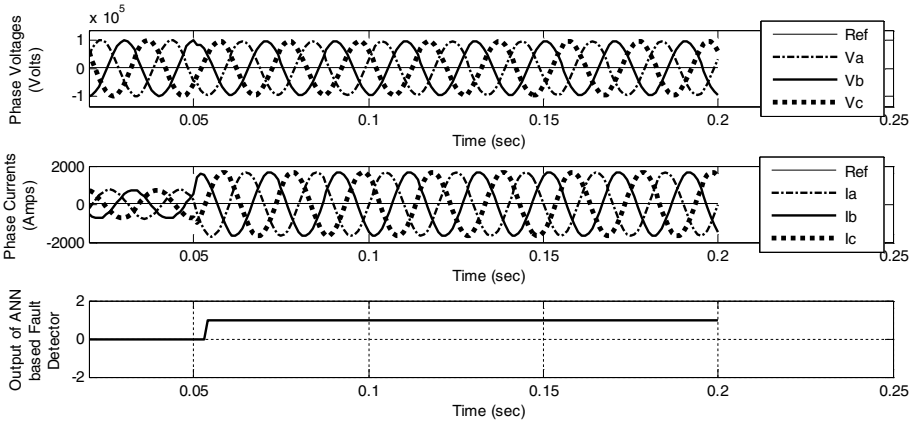


Figure 15. Test Result for ABC fault at 11 km, FR=60 Ω, FIT=50 ms in OH line

Fault location

Test results of simulation for variation in fault resistance and fault inception angle are given in Table 7 and Table 8.

Table 7. Fault Resistance Variation

S. No.	Fault Resistance (in Ohms)	Section of fault	Actual distance (in km)	Calculated distance (in km)	Error w.r.t. total length
1.	2	Cable	1.8	1.789	0.00073
		Junction	3.01	3.009	0.000066
		Line	11	10.87	0.0086
2.	48	Cable	1.8	1.810	-0.00066
		Junction	3.01	2.91	0.0066
		Line	11	10.99	0.00066
3.	52	Cable	1.8	1.799	0.000066
		Junction	3.01	2.981	0.00193
		Line	11	10.88	0.008
4.	98	Cable	1.8	1.67	0.0086
		Junction	3.01	3.35	-0.0226
		Line	11	10.37	0.041

Table 8. Fault Inception Angle Variation

S. No.	Fault Inception angle (in degree)	Section of fault	Actual distance (in km)	Calculated distance (in km)	Error w.r.t. total length
1.	30	Cable	1.8	1.76	0.00266
		Junction	3.01	2.99	0.00133
		Line	11	11.06	-0.00400
2.	90	Cable	1.8	1.78	0.1192
		Junction	3.01	2.94	0.00466
		Line	11	11.138	-0.0092
3.	150	Cable	1.8	1.828	-0.00186
		Junction	3.01	3.038	-0.00186
		Line	11	11.14	-0.00955

Conclusion

In this paper, Artificial Neural Network has been used for fault detection and distance location in Mixed Underground Cable and Overhead Transmission Line. For fault detection and distance location, the Fundamental component of voltages and currents measured at only one end of the line has been used. The Supervised algorithm for training of the ANN used is Levenberg Marquardt algorithm. The performance of the proposed scheme has been also investigated for variation in fault resistance and fault inception angle. Detection and location of fault is done correctly with higher rate of accuracy as shown in simulation results.

References

[1] IEEE Guide for Determining Fault Location on AC Transmission and Distribution Lines, IEEE Std C37.114™-2004.
[2] Eisa Bashier M. Tayeb Omer A/Aziz A/Rhim, "Transmission Line Faults Detection, Classification and Location using Artificial Neural Network", IEEE (978-1/11) 2012.
[3] J. B. Lee, C. W. Ha and C. H. Jung, "Development of Digital Distance Relaying Algorithm in Combined transmission Lines with Underground Power Cables" 0-7803-7173-9/01, 2001 IEEE.
[4] E. S. T. Eldin, M. M. A. Aziz, D. K. Ibrahim, and M. Gilany, "Fault Location Scheme for Combined Overhead Line with Underground Power Cable," Elect. Power Syst. Res., vol. 76, no. 11, pp. 928–935, Jul. 2006.
[5] Junyu Han, Peter A Crossley, "Travelling Wave Fault Locator for Mixed, Overhead, and Underground Teed Transmission Feeders", IEEJ Transactions on Electrical and electronics Engineering, IEEJ Trans 2015; 10: 383-389.
[6] Junyu Han, Peter A Crossley," Fault Location on Mixed Overhead Line and Cable Transmission Networks"
[7] Junyu Han, Peter A Crossley, "Fault Location on a Mixed Overhead and Underground Transmission Feeder Using a Multiple-Zone quadrilateral Impedance Relay and a double-ended Travelling Wave fault Locator".
[8] Javad Sadeh, Hamid Afradi, "A New and Accurate Fault Location Algorithm for Combined Transmission Line using Adaptive Network based Fuzzy Inference System."

[9] Hanif Livani,, and C. Yaman Evrenosoglu, “A Machine Learning and Wavelet based Fault Location Method for Hybrid Transmission Line “IEEE Transaction on Smart Grid, Vol. 5, No. 1, January 2014
[10] Ismail Niazy, Javad Sadeh, ”A new single-ended fault location algorithm for combined transmission lines considering fault clearing transients without using line parameters”, Electrical Power and Energy Systems 44 (2013) 816-823
[11] Roy Batruni, “A Multilayer Neural Network with Piecewise-Linear Structure and Back Propagation Learning”, IEEE Transaction on Neural Networks, Vol. 2. No. 3 MAY 1991.
[12] M. Oleskovicz, D V Coury, R K Aggrawal, “A Complete Scheme for Fault Detection, Classification and Location in Transmission Lines using Neural Networks”, Development in Power System Protection, Conference Publication No. 479, IEEE 2001.
[13] K.L. Butler and J. A. Momoh, “A Neural Net Based approach for Fault Diagnosis in Distribution Networks”, 0-7803-0/98 IEEE 1998.

Delineating Data Security and Privacy Boundary in Bioinformatics

Mukti Routray
Silicon Institute of Technology, Bhubaneswar, Odisha, India
mukti@silicon.ac.in

Abstract: This work draws a visible boundary for data privacy with respect to Big Data in bioinformatics. Defining Big Data boundary is a growing field where researchers have found an enormous repository of data which has huge business attraction leading to inappropriate usage of such data. This work provides a visibility to that thin line of ethical frontier which needs to be identified by data contributors and preservers so that the data fetchers are limited to data analysis or pattern analysis and the source remains encrypted and protected from invasion. Genomic data as well as bio-medical data have a true potential in their exploitation by commercial sectors which should be clogged with more privacy and security.

Keywords: big data; bioinformatics; security; privacy; ethics.

Introduction

Many areas of big data security have been observed among which underutilization of data, attacks with malefic intensions, erroneous data reporting, and passive interpretation of the scope of big data are the potential and significant threats. The security feature not only is multi-dimensional by itself but also generate an immediate requirement for policies [1] on every other feature of data stated above.

In bioinformatics the growth of data size is enormous in the recent years. The European Bioinformatics Institute (EBI) has almost 40 PB of data about genes, proteins, and small molecules in 2014, in comparison to 18 petabytes in 2013 [2]. Now this is a huge task as the data itself is not confined to general perceivable boundaries. As in case of underutilization of data, somewhere, the retention period [3] is questionable. Voices have been raised by privacy organizations that have monitored massive data retained for five years at the National ANPR (Automatic Number Plate Recognition) Data Center in Hendon, North London. Also a major question is unanswered when the data itself finds no single user right from its time of origination. Such zero-utility data in full or fragments occupy significant storage units. Multimedia data from CCTV cameras produce data for years and stored for another four to five years out of which the data may not find a single user. The data is stale from the very beginning. The data remain there with a negligible probability of utilization.

Big data is indeed enormous. It is a well-known fact now that Google has more than three billion searches stored daily. [4] From way back time in April 1994, the World Wide Web Worm [5] received an average of almost 1500 queries daily which increased to almost roughly 20 million queries per day with Altavista in 1997. Google from 9800 requests daily in 1999 to 60 million in 2000 to 200 million in 2004 and to 4.7 billion in 2011 [6] . Twitter generates over 340 million tweets daily [7].

Medical bioinformatics is concerned with sensitive and expensive data mostly contributing projects for drug design or in environments like hospitals. The distribution of data increases the complexity and involves data transfer through many network devices. Thus, data loss or corruption can occur. In absence of well-defined policies to classify the usage of such bursts of data, they tend to be misused or overused among unintended groups who bestow interest in such data.

An aspect of security of big data is in its significance in real time applications. Such applications expect the arrival of streams within strict time constraints [8]. A bunch of security attacks over transmission and related security in multimedia big data in absence of guided rules or policies may result in its unethical usage in regular online activities.

Functional genomics information is very much separated with genome-sequencing data for privacy aspect. Specifically, release of low level data sequencing leads to extraction of variation in genomes as in genome sequencing. However proteins annotated may be recovered in very less amount, at most with 5% of typed human genome. [9]

Open source multi-lingual intelligence has gained milestones with predicting criminal behavior and has tracked criminal transits in countries with malicious intentions. However this intelligence itself is exposed if the criminal faction gains superiority in locating the same data and generates a different outcome in pattern. Not only policies are to be framed, they need to be guided by several layers to address the importance of big data security. Protecting critical cyberspace infrastructure is a form of defense against catastrophic terrorism in security informatics [10] especially bioinformatics. There are several other areas of importance as global economy rests on the heap of big data exploration. The need of the hour is a policy directed definition of cyber security over big data usage, justifying the privacy of the data donor and building a strong

economy as a whole. Financial systems are the nerves of global economy. On a long run the resilience of the global financial system nurtured on big data will also depend on cyber security [11].

Privacy in Big Data and Application Level Security

Generally genomic types are accessed over web-servers through the inter-network. Most applications having a GUI based user interactive module which responds to queries from the users or sometimes perform a method of identifying types and sequences based on some machine learning or deep learning approaches.

This kind of data submissions must be related to generic use of web infrastructure in data gathering.

Facebook did an experiment [12] with human data through its newsfeed content sampled across its billions of users and later publishing the observations from the big data in PNAS.

Susan T. Fiske and Robert M. Hauser have voiced for regulation of human participation [13] in such experiments. They had expressed concern about the differences between academic and commercial research with respect to such experiments.

An ethical perspective allowing the forward movement of social-computing research [14] needs a policy framework for upholding the ethics.

The application layer is found to be more vulnerable in big data mostly where the computing takes place in the cloud. Cloud provides Software as a Service (SaaS) where the deployment is either over the internet or at a convenient location of the user. SaaS throughout the current years have become vulnerable to security threats and mostly the threats results with the application layer breakdown. Security breaches were almost reported with a higher level of 39 % [15]. Attacks were targeted through a popular SQL injection [16] which exploited 18% known and 5% unknown vulnerabilities [17].

Other most common application layer security threats faced by web application were OS and LDAP injections. Scripts are uploaded through left open doors in the web and users unknowingly execute the automated script without any knowledge of exposing their vital data to the attackers.

Access to the cloud environment exposes all data of users of the space and the cloud mostly becomes a high valued target [18][19].

These security threats are tightly coupled with privacy concerns in the application layer with unethical human tendencies to use data maliciously.

As the employees of the SaaS providers have access to enormous data, inherent policy needs to be framed for an infrequent detailed check on their behavior. A concern similar to this has been given importance by the US government which initiated the ADAMS [20] analysis for detailed understanding of human behavior. A project is floated to detect and prevent insider threats where the potential threat is more harmful as in case of government employee may abuse access privileges and share classified information.

Erroneous data reporting with the increase use of mobile applications generating real time data may be checked with higher precision in the application layer. As for crisis data management [21] multiple communities are likely to interact using the mobile applications where decisions are taken very quickly in response to vital information. Community interactions are highly flawed due to inaccurate reporting by the applications.

Scalable data analysis and management also stresses appropriate systems [22] for specific set of application requirements.

Data privacy problems can be solved by several restrictions that are provided by Data as a Service (DaaS) platform [23] where the processing of critical data may be limited to private cloud infrastructures. Public cloud resources should utilize the publicly available data.

Role based access control [24] in the application layer is needed by business organizations in order to secure its data being threatened by several external attacks.

A new multilevel [25] security model is proposed which controls data flowing in multiapplicative smart cards. The model is also effective in detecting illegal data sharing from the smart cards.

A data center infrastructure dealing with cloud platform for computing needs both outside and inside securities. Deep packet inspections and enhanced security is assured through content aware networking paradigm [26]. A growth of new type of consumers in this big data era could be seen as people might tend to buy personalized services and may lend not only their “likes” but also their profile data which may create more and more organizations to come up with near to exact, need based solutions to a margin of society. Privacy alone can be bargained for gaining an extra edge. Therefore anonymity in big data still remains a challenge in the future evolution of business related to big data and ever evolving human needs catering for more unified solutions.

Privacy and security in Big Data Usage in Biomedical Data and Bioinformatics

Schadt had pointed out [27] that it is very difficult to protect medical and genomics data when data have grown tremendously where privacy is the biggest challenge. Commercial Bio-medical sector have shown a lot of interest in widespread adoption of cloud computing which is there to handle this big volume. Behind this interest lies the perceived exploitation of the security of handling such sets. Clinical sequencing meets several regulations, mostly by Health Insurance Portability and Accounting Act of 1996 (HIPAA) in US, though presently there are no specific regulations pertaining to molecular/ genetic

diagnostics in India and even though cloud computing is being cautiously considered in such cases in US, in the developing countries like India where such regulatory mechanism is missing might be a favorable ground for illegal practices which might lead to severe exploitation of Big Data through cloud computing. In US HIPAA standards must be complied by several services as well as platform and infrastructure layers. Amazon released a paper that complies with a bunch of the standard regulatory body. [28]. Hybrid Clouds undertaken by Dell along with TGen to support first personalized medicine trial for pediatric cancer [29]. Key issues in sequencing should have detailed encryption mechanisms mostly key based. Issues such as incorrect data deletion and access to deleted data of a customer by another should be addressed.

Degree of Recognition of Actual Threats

Data security has been tougher with ever increasing dimensions of big data.

Genomic and bio-medical Big Data should include encryption for related information. Drug development should involve queries from regulatory authorities to the users of the cloud service which should be mandatory in drug development. Moreover when a person gives his or her consent to use data in a particular way, the researchers have no right to deviate. [30]

This work finds the significant areas of concern for big data security which may be traced with the composition of the data source. Generally a voluminous data source is polluted with data underutilization and data with errors. This work proposes the following categories of data underutilization and erroneous data submissions more of a generic kind however well applied in bio-medicine and genomics:

Underutilization of Data

1) Careless Submissions:
   Data is carelessly submitted through web forms or scripts overcrowding the database.
2) Data Littering:
   Less knowledge of prevailing security threats leading to unacknowledged data donations.
3) Left-over:
   Stripping of data during extraction and cleansing leads to such underutilizations.
4) Less resources available for processing:
   Utilization drastically reduces with reduced resources for processing big data.
5) Zero withdrawals:
   No available procedure for data withdrawals for voluntary donated data.
6) Data redundancy:
   There is typical growth of redundant data across and within distinct servers.
7) Data archives:
   Archives leads to huge data storages with significant reduction of users over time.
8) Careless AI:
   Mechanical interpretation often leads to chaos with a useful data.

Erroneous Data

1) Inaccurate readings from sensors/devices:
   Data generated from inaccurate readings from machines are accessed in parallel in absence of proper identification.
2) Faulty machines:
   Machines or devices with fault generating inaccurate readings.
3) Human data reporting:
   In absence of mechanized data reports transfer, human resource is highly engaged which contributes in error prone reporting.

And there are many other similar causes in overall data source pollution. This leads to a security trap for big data where the data underutilization invites malefic attacks and erroneous data leads to misrepresentation of facts. Additionally data underutilized often leads to a metadata of errors in representation. The complexity typically grows with the growth of voluminous data. There is a set intersection represented for misrepresentation of facts and malefic attacks. As attacks with intentional misrepresentation often becomes indistinguishable.

Thus the security threats can now be broadly classified as actual security threats and misrepresentation of facts. Security threats needs to evaluated in actual as most of the time the attackers disguises behind misrepresentation of facts. Such misrepresentation is widely fuelled by both data under-utilization and error in the data. There are two broad types of misrepresentation of data, actual representation of erroneous data and forced representation of unutilized data.

This work also proposes two broad types of misrepresentation of data, actual representation of erroneous data and forced representation of unutilized data.

Here degree of recognition of actual threats is directly proportional to the probability of countering underutilized data given they are error free.

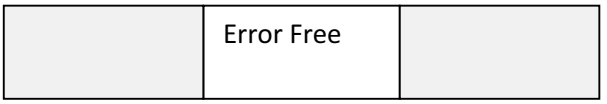


Figure 1. The whole region indicates data source with both underutilization and data with errors. Typically a portion may be error free

$$degree(R) = P(U/E_f)$$

Where R is recognition of actual threats, U is underutilized data and E_f is error free data.
The error-prone underutilized data is ignored as mostly harmless though there is a need for a separate study of such false positives where such data may be processed for any critical system.

Conclusion and Future Work

On demand cloud infrastructure which is a must in the near future in bio-medical data access may be promoted to ensure better privacy and security in the future. All of the above security breaches or threats in the application layer affecting privacy may be reduced with developing security based architecture with advanced cryptography. It is all about the knowledge of distributing the key to access private data.

Acknowledgment

The author wish to thank Dr. Swati Vipsita. Assistant Professor, IIIT, Bhubaneswar, for her support in acquiring information related to this area.

References

[1] US survey 2015-Survey WhiteHouse FAQ-Security Release-Whitepaper.
[2] EMBL-European Bioinformatics Institute, “EMBL-EBI annual scientific report 2013,” 2014.
[3] Reported by Paul Lewis on Monday, http://www.theguardian.com/uk/2008/sep/15/civilliberties.police, 15th September, 2008,
[4] Genta, Robert M., and Amnon Sonnenberg. "Big data in gastroenterology research." Nature Reviews Gastroenterology & Hepatology 11, no. 6 (2014): 386-390.
[5] Brin, Sergey, and Lawrence Page. "The anatomy of a large-scale hypertextual Web search engine." Computer networks and ISDN systems 30, no. 1 (1998): 107-117.
[6] Coté, Mark. "Data motility: The materiality of big social data." Cultural Studies Review 20.1 (2014): 121.
[7] Das, T. K., D. P. Acharjya, and M. R. Patra. "Opinion mining about a product by analyzing public tweets in Twitter." In Computer Communication and Informatics (ICCCI), 2014 International Conference on, pp. 1-4. IEEE, 2014.
[8] Singh, Dilpreet, and Chandan K. Reddy. "A survey on platforms for big data analytics." Journal of Big Data 2.1 (2014): 1-20.
[9] Dov Greenbaum1,2,3,4,5, Andrea Sboner1,2¤, Xinmeng Jasmine Mu1, Mark Gerstein. PLoS Computational Biology December 2011, Volume 7, Issue 12, e1002278
[10] Chen, Hsinchun, Roger HL Chiang, and Veda C. Storey. "Business Intelligence and Analytics: From Big Data to Big Impact." MIS quarterly 36.4 (2012): 1165-1188.
[11] http://www.business-standard.com/article/news-ians/at-g20-modi-pushes-for-coordination-to-tackle-black-money-114111600146_1.html, 2014.
[12] Kramer, Adam D. I. and Guillory, Jamie E. and Hancock, Jeffrey T. "Experimental evidence of massive-scale emotional contagion through social networks." Proceedings of the National Academy of Sciences 111, no. 24 (2014), 8788-8790.
[13] Fiske, Susan T. and Hauser, Robert M.. "Protecting human research participants in the age of big data." Proceedings of the National Academy of Sciences 111,no. 38 (2014), 13675-13676.
[14] Kahn, Jeffrey P. and Vayena, Effy and Mastroianni, Anna C."Opinion: Learning as we go: Lessons from the publication of Facebook’s social-computing research." Proceedings of the National Academy of Sciences 111, no. 38 (2014), 13677-13679.
[15] Wade HB, David Hylender C, Andrew Valentine J. Verizon Business 2008 data breach investigation report, 2008 http://www.verizonbusiness.com/resources/security/databreachreport.pdf [accessed on 23 October 2014].
[16] Auger R. SQL Injection, 2009 /http://projects.webappsec.org/SQL-Injection [accessed on: 12 July 2014].
[17] Subashini, Subashini, and V. Kavitha. "A survey on security issues in service delivery models of cloud computing." Journal of Network and Computer Applications 34, no. 1 (2011): 1-11
[18] Bernard Golden. Defining private clouds, 2009 /http://www.cio.com/article/492695/Defining_Private_Clouds_Part_One [accessed on: 10 August 2014].
[19] Kaufman LM. Data security in the world of cloud computing, security and privacy. IEEE 2009;7(4):61–4.
[20] Fact sheet: Big data across the federal government (2012). http:// www.whitehouse.gov/sites/default/files/microsites/ostp/big data fact sheet 3 29 2012.pdf

[21] Herbert, Katherine G., Emily Hill, Jerry Alan Fails, Joseph O. Ajala, Richard T. Boniface, and Paul W. Cushman. "Scientific Data Infrastructure for Sustainability Science Mobile Applications." In Big Data (BigData Congress), 2014 IEEE International Congress on, pp. 804-805. IEEE, 2014.
[22] Agrawal, Divyakant, Sudipto Das, and Amr El Abbadi. "Big data and cloud computing: current state and future opportunities." In Proceedings of the 14th International Conference on Extending Database Technology, pp. 530-533. ACM, 2011.
[23] Cheptsov, Alexey, Axel Tenschert, Paul Schmidt, Birte Glimm, Mauricio Matthesius, and Thorsten Liebig. "Introducing a New Scalable Data-as-a-Service Cloud Platform for Enriching Traditional Text Mining Techniques by Integrating Ontology Modelling and Natural Language Processing." In Web Information Systems Engineering–WISE 2013 Workshops, pp. 62-74. Springer Berlin Heidelberg, 2014.
[24] Schumacher, Markus, Eduardo Fernandez-Buglioni, Duane Hybertson, Frank Buschmann, and Peter Sommerlad. "Security patterns." Integrating Security and Systems Engineering. Hoboken (2006).
[25] Girard, Pierre. "Which security policy for multiapplication smart cards." InUSENIX workshop on smartcard technology. 1999.
[26] Bolsens, Ivo, Georges GE Gielen, Kaushik Roy, and Ulf Schneider. "" All Programmable SOC FPGA for networking and computing in big data infrastructure"." In ASP-DAC, pp. 1-3. 2014.
[27] Schadt EE. The changing privacy landscape in the era of big data. Mol Syst Biol 2012;8:612.
[28] Creating HIPAA-Compliant Medical Data Applications With AWS. <http://aws.amazon.com/about-aws/whats-new/2009/04/06/whitepaper-hipaa/>.
[29] Managing data in the Cloud Age. <http://www.dddmag.com/articles/2012/10/managing-data-cloud-age>.
[30] Vivien Marx. "Biology: The big challenges of big data", In Nature 498, 255–260 (13 June 2013)

Sentiment Analysis of Transliterated Hindi and Marathi Script

Mohammed Arshad Ansari* and Sharvari Govilkar**
*Department of Information Technology, PIIT
mansari@student.mes.ac.in
** Department of Information Technology, PIIT
sgovilkar@mes.ac.in

Abstract: There is a growing research on sentiment analysis of various languages, which is being supplanted heavily by those same techniques and methods being applied on the mix code or transliterated text for the same purpose. This growing research is a result of necessity created through the advent of social media as well as textual analysis of the data being collected online. This paper, rather than being a pioneer, is about extending that research for further improvement. Herein, we assess the existing status, standards and achievements of the researchers in the given field and supplant it without proposed methodology to increase precision. Although, the current work is a proposal with improvements over established techniques, it is also however going to be quite comparative when it comes to the existing findings. The idea is to not just improve what has already been built or shown to be true, but also check if the simplest approach is still the best way to proceed or not. By this we mean the existing direct supervised learning for sentiment analysis, without much NLP or language specific work. Since we shall be testing our approach against the existing state of the art as well as entering the area previously not under coverage (Marathi transliterated text), this work is bound to make great strides in the field of sentiment analysis.

Keywords: Mixed code script, English, Hindi, Marathi, Transliteration, Sentiment Analysis.

Introduction

Sentiment analysis is a process of analyzing natural language and figuring out the sentiments involved or expressed through the source material, with respect to the topic. The basic idea behind sentiment analysis is that each textual sentence may or may not contain some kind of polarity, expressing a degree of emotions along with the information. It is much easier to read in to those polarities when the text is spoken and not written due to the tone of the speaker; whereas, in case of written text, it is the context that is useful while determining the polarities in the statements. Sentiment analysis has grown to be one of the most important research areas when it comes to textual analysis on the web. Reason being, obviously, is to be able to make sense of the data as well as to understand the tone of information being provided. There are numerous applications, ranging from product/customer support review to improve quality of service (QOS) by corporations to understanding geo-political motivations when certain news breaks. People react on social media, especially when they are charged emotionally and when emotions take the form of textual content to vent, it has been observed that it does in a manner which is more close to a person's mother tongue.

Hindi is spoken by more than 500 million people around the world, making it one of the most spoken language in the world. Besides, English has turned out to be an international language, a lot of people speak English on the internet, however; as described above, there are instances when people use English language to phonetize and express in a foreign language. This is seen far more in India subcontinent, where people prefer to write using English alphabets, but most often, use the words from the mother tongue. If we only look at all the YouTube comments (especially if they are about some controversial issues), we would see a lot of usage of such transliterated messages or mix-script writing. Another behavior worth noting is related to vocabulary. People from subcontinent use words such as 'Bye', 'Thank you', 'Good night', 'Please', 'Sorry' and intermix them with their native tongues. This mixture of language has been observed profoundly at varying levels of society. Therefore, it would not be very far-fetched to say that the languages are evolving by mixing language themselves. This forms the necessary reason for why there needs to be analysis of mixed-languages and it starts with analyzing that which is mostly available, the mixed-script. Here, we are not going to invent something new, nor are we going to do something entirely differently. However, the purpose behind this work is to stand on the shoulders of giants and take the research of what has already been done to what it can be. This we strive to do, by improving the performances by innovatively applying techniques which have worked better in other cases. Therefore, as it will be seen, our proposed approach as well is a mixture of disparate attempts in varying domains (even slightly) to come together for better whole.

Sentiment analysis is a lot tougher for languages that are outside Eurozone, due to their lexical syntax being very different from European languages as well as due to majorly, less amount of work being done on it. Semantic analysis requires

annotated text corpus to train classifiers, which is most of the time a very huge manual task. It has been undertaken for English and for many other European languages, while at the same time, work from one supplementing work for another language, due to the similarities existent in those languages. When it comes to languages such as Hindi and Marathi, such resources are very less compared to the above mentioned languages. More so for Marathi, since a lot of work has been done and progress made in case of Hindi. Most of the reason for the under development of the research for these languages are (1) Not much annotated textual corpus needed for training, (2) Lack of basic language tools like taggers and parsers. These problems will be solved in time and this work is a part of all the woks which will finally solve this problem. Having expressed the problem, in this work, we also laude the work that has already gone in to this respective field, without which this would not have been possible. It is really interesting to note, that a great amount of effort has just started pouring in for this particular part of sentiment analysis. It is naught with great anticipation that this work is being progressed. Besides, as the sentiment analysis of the textual data being to shape more and more, the greater the benefit will be to the field of general AI. When it comes to human capacity, not representing emotions would be the biggest gap in the domain, which exactly is sentiment analysis has started to fill.

Motivations

We can discuss two kinds of motivation herein, one being in general about sentiment analysis and the other a bit specific about this very work itself. Sentiment analysis has tons of applications, especially in the current era of social media. The entire planet's population is connecting with one another and learning about each other's cultures and assimilating ideas from one another and then followed by sharing ideas, concerns and assaults on the social media. Whatever may be the case, all the information being transferred is textual in nature. It becomes of vital import that the sentimental exchanges happening on such channel is being monitored. For example, twitter and Facebook led to the entire Arab world to be engulfed in flames. The previous example was just to elaborate how the social media and in extension written media, with the emotional content, can literally change the world. Having understood the important, let's look at many possible applications for sentiment analysis.

- Product / Service Review; Product here can extend from movies, daily usage products to books, etc. They are usually reviewed either on social media or sites dedicated to such reviews. Examples being Amazon, Google/IPhone App store, Good Reads, etc. These reviews allow other users to decide whether they want to buy reviewed product or not. Similarly, service providers like ISP, Telecom providers, etc. are interested in review of the service they are providing, including customer service reviews. These reviews help the providers to improve the service by addressing the negative aspects and focusing more on positive ones.
- Discourse Analysis on topics that range from philosophical to wars between countries are also candidates for such analysis. It becomes really important when taking decisions whether the debate pertaining to such decisions are emanating from emotions or logically grounded arguments.
- Feedback Analysis for teachers from students or about government from population. They all have one thing in common, that is, they are all textual and by extension is candidate for sentiment analysis.
- Other areas such as emails analysis, twitter/Facebook feed analysis or blog analysis, help in understanding the emotions of authors regarding the topics, which help focus on problems, which can be the cause of negative emotions.

Above being the motivation for sentiment analysis in general, let's consider the motivation for this specific work. As explained in introduction as well as in the above listed areas of interest; where the source is always textual data, this data usually consists of mixed script in terms of language. That being the what, let's look at why. And more specifically why Hindi and Marathi. Marathi by itself is playing catch with Hindi, where Hindi language is making strides in this research. Although there aren't many Marathi speakers in comparison to Hindi itself, but it is still spoken by millions of people in the region of Maharashtra, which by the way has a lot of literature that has yet to be digitized and benefit the world with it. The real reason is the aspect of completion. Many terms like 'Layi Bhari' and many other slangs have crept from Marathi to Hindi and then taken to the country as whole through Bollywood movies. Many inside jokes in many Bollywood movies have their roots in Marathi language and regional aspects. These will never be easily covered if Marathi itself doesn't become partial focus of the research itself. Although, there are very few input sources to consider for Marathi, there still exists some, in form of YouTube comments, etc., which can be part of this research. Going back to Hindi itself, a lot of textual resource considers mixed script statements as noise, which definitely contains gold from sentiment analysis perspective. And, therefore, we have decided to augment the existing research by improving the precision where research is being performed and pave the way where the research is still lagging behind.

Related Work

Code Mixing, Language Identification and Transliteration

Code mixing has been done for more than a couple decades and was investigated during initial period by Gold [1] for the purpose of language identification. The same phenomenon for Indian languages was worked upon by Annamalai [2], pioneering the research field for the subcontinent languages. Recently, it was investigated by Elfarti [3] and was termed as linguistic code switching by the research group. Karimi [4] made the case for machine translation for the purpose of transliteration in the survey and suggested transliteration based on phoneme based approach and transliteration generation using bilingual corpus, while presenting the key issues that arise during the transliteration process. Dewaele [5] pointed out the strong emotional presence as being the main marker for the existence of code switch that happens in textual corpus. Gupta et. al. [6] mined the transliteration pairs between Hindi and English from the music lyrics of Bollywood songs for Fire'14 shared task, which is quite handy for training in language sentiments. Deepti Bhalla et.al. [7] have given a rule based approach to perform transliteration from English to Punjabi language through machine transliteration. Although, the focus is primarily on named entity.

The issue of identification of language of the code - mix script is another challenge that has been answered by the research community. A statistical approach was proposed by Kundu and Chandra et. al. [8] for the automatic detection of English words in Bengali + English (Benglish) text. A conditional random field model for weakly supervised learning model was used for word/token labelling by King and Abney [9] with a good result of > 90\%. Barman [10] used Facebook user data for identifying the language in mixed script and concluded that the supervised learning outperforms the dictionary based approaches. POS Tagging and transliteration efforts for Hindi + English data on social media was experimented upon by Vyas et. al. [11] and came to the conclusion that any operation on transliteration text will largely benefit from POS tagging. Another approach was given by Kulkarni et al. [12] to use genetic algorithm for identification of Marathi and Sanskrit words. In another work by Jhamtami [13], it was shown that the technique which use POS tags of adjoining words along with char sequence to achieve F1 score of 98%.

Sentiment Analysis

Although sentiment analysis is being worked upon for quite some time now and it has already entered the mainstream application. There are works being done for the transliteration of Indian languages, out of which some have been covered in this section. A survey of sentiment classification was performed by Pandey [14] covering the techniques being used for Indian languages. Joshi et.al. [15] performed experiment to compare three approaches for the sentiment analysis of Hindi text and found that HSWN performs better than Machine Translation approach but under performs in language training of sentiment corpus in Hindi. This was, however; performed in 2010 and the HSWN has been continually improving past these experiments. The same result was reiterated with by Balamurali et. al. [16]. Kashyap [17] found a way to perform Hindi Word Sense Disambiguation using WordNet with encouraging results for nouns. Subjective lexical resource was developed by Bakliwal et. al. [18] by using only WordNet and graph traversal algorithm for adverbs and adjectives.

Balamurali A R et. al. [19] performed experiment to figure out in language supervised training of sentiments against the machine translated source for sentiment analysis. They found that the MT based approach under performs much worse compared to in language training of sentiment. Fuzzy Logic membership function was used to determine the degree of polarity of the sentiment for a given POS tagged preposition by Rana [20]. Hindi SentiWordNet was developed by Balamurali A. R. et. al. [19] using the SentiWordNet by using linked WordNet. HSWN along with negation discourse was applied by Pandey [21] and Mittal et. al. [22] or sentiment analysis of Hindi language text corpora, with the accuracy of 80.21 achieved. Work on multi language sentiment analysis on twitter feeds of Sweden politicians was undertaken by Lucas Bronnimann [23], which heavily depended on the use of emoticons for emotional polarity detection. S Tembhurnikar [24] used LDA for sentiment classification on mixed tweets, however; filtered out the non-English words from the tweets before the analysis. Jagmeet Singh [25] used twitter feeds API for performing sentiment analysis in mixed languages with techniques such as dictionaries look up, taxonomy and ontology based analysis. There is only one work done on the sentiment analysis of Hindi transliteration by Srinivas [26], [27] and the approach taken was to tag words with identified language and then run against respective POS tagger for languages and sentiment analysis done on the output. The approach yielded 85\% precision.

Proposed Approach

The current word considers Hindi/Marathi text in Romanized script as input which may contain phonetic words, sounds; however, it is not considering social language like gr8, rt, f9, etc. as input source. For now, it is considered as noise for the result of this work. Although, we are not performing sentiment analysis on English text as part of this text, which is simply because it has been under taken in many works preceding this one. Therefore, concentration will solely be on the text which is transliterated Hindi or Marathi. Also, the architecture of proposed approach has integration in mind and hence, it will be

able to plug social sentiment analysis or twitter sentiment analysis or plain English analysis and will work only for the transliterated text, while taking inputs from the mentioned analyzers for their established polarity. The results can be merged and shown to have improved the overall accuracy.

System Architecture

There are going to me multiple approach for testing to be implemented as part of this work. The purpose of those work will be to ensure that the proposed system performs better than what has been accomplished by other researchers. Although, here we will only go into the actual proposed system to understand its working and predict the possible improvements.

The proposed approach we are to take in this work comprises of extending the work of Srinivas [26] with multiple improvement points at multiple level of the process. Each step is listed below with the improvement suggested from this work. All the work done on that paper has been uploaded to the website [28], which we shall be using in this paper extensively and building on top of it.

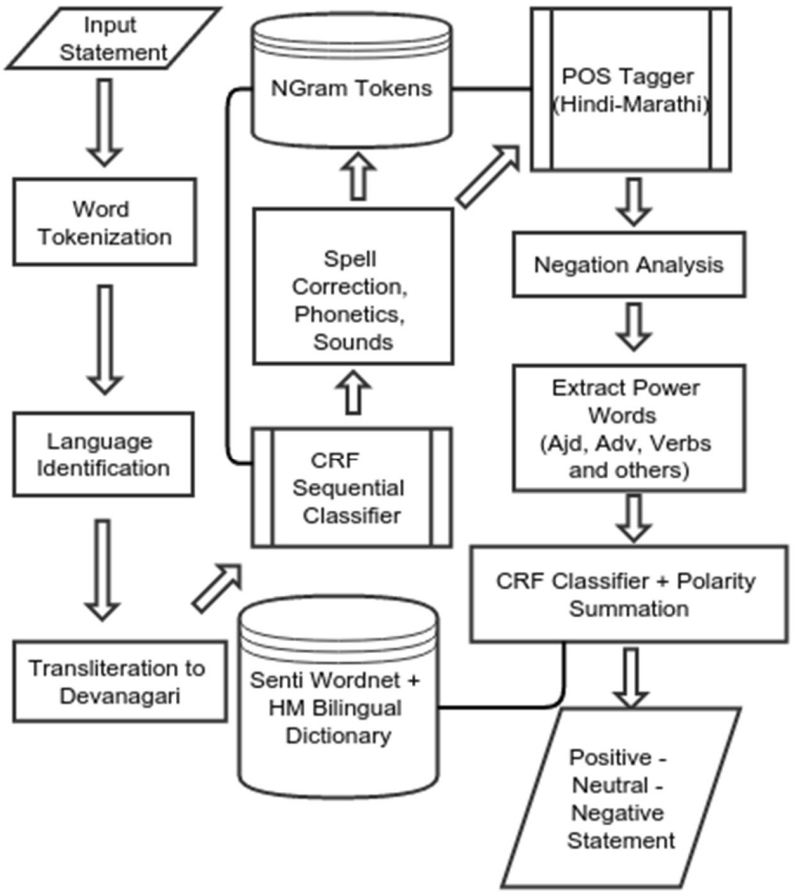


Figure 1. Process flow diagram

Text Normalization

Text normalization step has been covered heavily in work of Srinivas [27], which has following steps:

- Language Identification; Tagging of the words as <word>|<tag>, where <tag> can be E or English, H for Hindi and M or Marathi,
- Spelling Corrections; There are multiple ways to write Mujhe in Hindi such as muze, muje, etc. To come to common and widely used spelling becomes very important. [29]
- Ambiguous words; Words such as 'me' means same thing in English and Marathi, where as in Hindi it is sometimes used to say inside with another spelling being 'mein'. [17] [13]
- Sounds; Words such as aww, oohh, ouch, ewww, etc. They do contain rich information when it comes to sentiments.

- Phonetic words; Words such as plees usually is misspelling of word please, spoken in some areas of the subcontinent. It gets written too in the similar manner.[30]
- Transliteration; Conversion of Hindi/Marathi words written in English to appropriate Devanagari script.

All then above enumerated steps have been covered by Srinivas [27] and doesn't require us to go in the details of those, however; we will look at some of those steps to get a proper grip on the subject. At this point, this work will simply reuse those steps for Hindi and try to closely perform them for Marathi as well. [31]

Work-Token Normalization

The process here is really simple to explain, but quite interesting to develop. This is a required step as a sort of preprocessor, to enable the words to be converted appropriately to their respective languages. Words like "aww", "reaaaly", etc. needs to be normalized and the techniques to do are covered by Srinivas [26], [27] both. These methods are already tested with some accuracy in the above mentioned papers themselves.

Language Identification Tagging

The first step is to tag the language identifier for every word-token, using the techniques described in Kundu and Chandra et. al. [8] and King and Abney [9]. The output of this step will be more or less like in example given below:

"Yeh acchha din hai. Let's go now"
"Yeh|H acchha|H din|H hai|H. Let's|E go|E now|E".

Table 1. Language Identification Algorithm

Algorithm: Language Identification Description: Works only with bilingual settings. Here other language is could either be Hindi or Marathi
Input: Sentence Output: Tagged Sentence Given: English Dictionary Slang Dictionary [Besides slang words, it also contains emotional expressions] Other Dictionary [Hindi / Marathi]
Steps: For every word [u] in Sentence: 1. If word in slang: a. Y = give S[u] probability of 0.7 b. N = give S[u] probability of 0 2. If word [u] passes the phonetic emotions expression test, then convert it to normal spelling and give S[u] probability of 0.9 3. if word exists in Transliterated Other Language Dictionary a. Y = give word O[u] probability of 0.7, E[u] probability of 0.3, S[i] probability of S[u] * 0.3 b. N = give word O[u] probability of 0.5, E[u] probability of 0.5 4. Transliterated the word and check if word exists in Other Language Dictionary a. Yes = give word O[u] probability of 0.7 * O[u] and E[u] probability of 0.3 * E[u], S[u] probability of S[u] * 0.3 b. No = give word O[u] probability of 0.3 * O[u] and E[u] probability of 0.7 * E[u] 5. If word exists in English Dictionary: a. Yes = give word O[u] probability of 0.7 * O[u] and E[u] probability of 0.3 * E[u], S[u] probability of S[u] * 0.3 b. No = give word O[u] probability of 0.3 * O[u] and E[u] probability of 0.7 * E[u] 6. If S[u] > O[u] and S[u] > E[u]: a. Identify Word[u] as Slang 7. Else If O[u] < E[u]: a. Identify Word [u] as English 8. Else: a. Identify Word [u] as Other [English / Marathi] Return Sentence updated with identified language

To this effect, same approach can be accomplished for Marathi text. Once we have tagged all the word-tokens with their corresponding language identifier, we can move to next step. However, there are going to be ambiguous words "ho", "me",

etc., which shall use semi supervised learning for handing based on context. The plan is to also test again HMM models to discover the accuracy difference.
Language identification will be based on the approach of using hidden markov models trained on the n-gram generated from corpus to be able to produce probabilities for each work-token when considered with its neighboring words to identify the language to which the token belongs.
In all the related work, there was a need for transliteration mechanism in play on the fly. The reason for it being the de facto method of choice is because it allows the usage of POS tagging to work with the text, which would only work on respective script for a given language. This database will be trained using Hindi - English transliteration pairs collected from Fire 2013 found at [28] as well as result of another previous work by Gupta et. al.[6]. This trained model will then be used to convert all the words in Hindi WordNet to ensure greater coverage of incoming input word tokens. In case of Marathi, a similar thing will be done and associated with Hindi WordNet through Hindi -Marathi bilingual dictionary.

POS Tagging, Discourse analysis, SentiWordNet
Before sentiment analysis can be performed, it is necessary to deal with few important things. We are striving to extend and improve upon earlier work such as Srinivas [27] and therefore following much in the same footsteps. Both the steps are explained further below. Once we have the document in English or Hindi, the next step is to run it through POS tagger based on respective language. The approach will be straight forward as detailed here [11]. The POS Tagged prepositions then shall be run through the negation discourse analysis to invert the POS tagged adjectives and adverbs in case of negative discourse as explained by Pandey [21] and Mittal et. al. [22].

Table 2. Polarity Identification Algorithm

Algorithm: Polarity Identification
Description: Works only with bilingual settings. Here other language is either be Hindi or Marathi

Input: Sentence [Language Tagged]
Output: Sentence Polarity P = {} for each index j to hold the polarity of each word
Given:
POS Tagger

Steps:

1. Convert Sentence into language based phrases for each word collection belonging to same language as a phrase
2. For every phrase [u] in Sentence:
    a. If the root in phrase is negation then mark Negation = True, otherwise False
    b. POS Tag the phrase [u] using the corresponding language POS Tagger
    c. Identify Named Entity in the phrase [u]
    d. For every word [j] in Phrase [u]:
        1. if word [j] is named entity:
           continue
        2. if word [j] is tagged as slang or English:
           P [j] is polarity of slang in English SentiWordNet
        3. If word is tagged Hindi:
           P [j] is polarity of the word [j] in HSWN
        4. If word is tagged Marathi
            a. Get approximate meaning* of word from Hindi – Marathi bilingual dictionary as jnew
            b. P [j] is polarity of word [jnew] in HSWN
        5. If Negation is true, then inverse the polarity of all words in the phrase by multiplying it with -1
3. Return P containing each word with its corresponding polarity

Return Sentence updated with identified language

The output of POS Tagger shall be used to look up senti-word identifier for the word groups using SentiWordNet or HSWN, for English and Hindi, respectively. HSWN has been improved by Pandey [21] by making additions to it and that will be used in this work. Here, there are three major improvements we are considering. Since, it was established [26] that the basis for sentiment analysis being POS tagged adjectives and adverbs gives much better result that depending directly on lexicon or WordNet look up for each work, we would be going that route. Secondly, addition of discourse analysis would further enhance on the existing work [27].

Sentiment classification using classifier

Once we get sentiword identifier for each token - word, next step is to put it through the classifier which will give the polarity of the statement provided. This polarity checking decision can be as simple as simple summation of all word-token sentiment polarities or further analysis can be performed to figure out what really is the polarity of word-token and its membership with negative, positive or neutral. This step being the vital one can be accomplished using most trust classifier like SVM, Random Forests, however; impetus shall be given on naive Bayes classifier for brevity's sake.

Example Flow

'Kitni der se ticket cancel nahi ho rahi hai' becomes kitni|H= कितनी der|H= देर se= से|H ticket|E cancel|E nahi|H= नहीं ho|H= हो rahi|H= रही hai|H= है

Table 3. Word POS Tagged

कितनी	Adjective	QF
देर	Adverb	NN
से	Verb	PSP
Ticket	Unk	JJ
Cancel	Unk	NN
नहीं	Adverb	NEG
हो	Verb	VM
रही	Verb	VAUX
है	Verb	VAUX

Negation discourse analysis and polarity extraction example

नहीं and हो are closely associated with one another. Negation discourse analysis [21] works on the subtree level, which in this case is post the word हो. So all the words following 'nahi' will be part of its subtree. Hence, all the polarity from that point onwards will be reverse.

polarity = कितनी|adj=INC + देर|adv=NEG + से|v=NEU + ticket|NN=Neu + cancel|=NEG + नहीं|adv=NEG + हो|v=NEU रही|v=NEU + है|v=NEU

polarity = (INC * NEG) + NEU + NEG + NEG REVERSE (NEU + NEU + NEU) = (2 * -1) + 0 + -1 + (-1 * (0 + 0 + 0) = -3

Quite clearly, we have input as POS tagged statements with greater emphasis on adjectives and adverbs that are inverted in case of negation present in the preposition. Once we have this tagged information, we would like to test on both the process of simple polarity count summation of the given input and training the classifier, in order to come up with the best possible result in terms of accuracy.

Conclusion

The most important aspect of this work i.e. the results are what is coming next. We will show that the approach proposed in this work performs better than all the work presented here in literature, when considered independently. It is the synergy, which the approach presented there, promises. The implementation will happen for all ways that differ from the approach too, so that comparisons can be made and conclusions drawn without the strawman arguments.

There is a lot of work to be performed before any concrete conclusion can be expressed, however; There is a great possibility that the approach suggested in the given work will result in improvement in the field of sentiment analysis, that can again be extended for greater language coverage in as well as out of Indian languages. These strides towards such improvements will result in machine's being able to understand human sentiments better, which is one of the greatest challenge being faced by the research in general AI. Ours is but a small step towards that goal. It will not be too farfetched to believe that the improvements will range from 5 to 10 percent improvement where we will see the accuracy reach 95 percent.

The authors wish to thank all the peers who helped enable this work to come to fruition.

References

[1] E. M. Gold and T. R. Corporation, “Language identification in the limit,” *Inf. Control*, vol. 10, no. 5, pp. 447–474, May 1967.

[2] E. Annamalai, “The anglicized Indian languages: A case of code mixing,” *Int. J. Dravidian Linguist.*, vol. 7, no. 2, pp. 239–247, 1978.

[3] H. Elfardy and M. T. Diab, "Token Level Identification of Linguistic Code Switching.," in *COLING (Posters)*, 2012, pp. 287–296.
[4] S. Karimi, F. Scholer, and A. Turpin, "Machine transliteration survey," *ACM Comput. Surv.*, vol. 43, no. 3, pp. 1–46, Apr. 2011.
[5] J.-M. Dewaele, "Emotions in Multiple Languages." 2010.
[6] K. Gupta, M. Choudhury, and K. Bali, "Mining Hindi-English Transliteration Pairs from Online Hindi Lyrics," *Proc. Eighth Int. Conf. Lang. Resour. Eval.*, pp. 2459–2465, 2012.
[7] D. Bhalla, N. Joshi, and I. Mathur, "Rule Based Transliteration Scheme For English To Punjabi," *Int. J. Nat. Lang. Comput.*, vol. 2, no. 2, pp. 67–73, 2013.
[8] B. Kundu and S. Chandra, "Automatic detection of English words in Benglish text: A statistical approach," in *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, 2012, pp. 1–4.
[9] B. King and S. Abney, "Labeling the Languages of Words in Mixed-Language Documents using Weakly Supervised Methods," *Proc. NAACL-HLT*, no. June, pp. 1110–1119, 2013.
[10] U. Barman, A. Das, J. Wagner, and J. Foster, "Code Mixing: A Challenge for Language Identification in the Language of Social Media," in *First Workshop on Computational Approaches to Code Switching*, 2014, pp. 21–31.
[11] Y. Vyas, S. Gella, J. Sharma, K. Bali, and M. Choudhury, "POS Tagging of English-Hindi Code-Mixed Social Media Content," pp. 974–979.
[12] P. P. Kulkarni, S. Patil, and G. Dhanokar, "Marathi And Sanskrit Word Identification By Using Genetic Algorithm," vol. 2, no. 12, pp. 4588–4598, 2015.
[13] H. Jhamtani, "Word-level Language Identification in Bi-lingual Code-switched Texts," pp. 348–357, 2014.
[14] P. Pandey and S. Govilkar, "A Survey of Sentiment Classification Techniques," *Ijser*, vol. 3, no. 3, pp. 1–6, 2015.
[15] A. Joshi, B. A. R, and P. Bhattacharyya, "A Fall-back Strategy for Sentiment Analysis in Hindi: a Case Study," no. October 2015, 2010.
[16] a R. Balamurali, A. Joshi, and P. Bhattacharyya, "Harnessing WordNet Senses for Supervised Sentiment Classification," *Proc. Conf. Empir. Methods Nat. Lang. Process.*, no. 2002, pp. 1081– 1091, 2011.
[17] M. Sinha, M. K. R. R, P. Bhattacharyya, P. Pandey, and L. Kashyap, "Hindi Word Sense Disambiguation."
[18] P. A. V. V. Akshat Bakliwal, "Hindi subjective lexicon: A lexical resource for hindi adjective polarity classification," *In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), Chair*. .
[19] A. Balamurali, "Cross-lingual sentiment analysis for Indian languages using linked wordnets," *Proc. COLING 2012*, vol. 1, no. December 2012, pp. 73–82, 2012.
[20] S. Rana, "Sentiment Analysis for Hindi Text using Fuzzy Logic," *Indian J. Appl. Res.*, vol. 4, no. 8, p. 16, 2014.
[21] P. Pandey and S. Govilkar, "A Framework for Sentiment Analysis in Hindi using HSWN," vol. 119, no. 19, pp. 23–26, 2015.
[22] N. Mittal and B. Agarwal, "Sentiment Analysis of Hindi Review based on Negation and Discourse Relation," in *Sixth International Joint Conference on Natural Language Processing*, 2013, pp. 57–62.
[23] L. Brönnimann, "Multilanguage sentiment - analysis of Twitter data on the example of Swiss politicians," 2013.
[24] S. D. Tembhurnikar, "Sentiment Analysis using LDA on Product Reviews : A Survey," no. Ncacc, pp. 22–24, 2015.
[25] J. Singh and K. Mahajan, "SENTIMENTAL ANALYSIS IN MASHUP LANGUAGES," vol. 8, no. 4, pp. 663–667, 2015.
[26] S. Sharma, P. Srinivas, and R. C. Balabantaray, "Sentiment analysis of code - mix script," in *2015 International Conference on Computing and Network Communications (CoCoNet)*, 2015, pp. 530–534.
[27] S. Sharma, P. Srinivas, and R. C. Balabantaray, "Text normalization of code mix and sentiment analysis," in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2015, pp. 1468–1473.
[28] S. Sharma and P. Y. K. L. Srinivas, "Linguistic Resource Website," 2015. [Online]. Available: http://linguisticresources.weebly.com/downloads.html.
[29] E. Clark and K. Araki, "Text normalization in social media: Progress, problems and applications for a pre-processing system of casual English," *Procedia - Soc. Behav. Sci.*, vol. 27, no. Pacling, pp. 2–11, 2011.
[30] D. S. Rawat, "Survey on Machine Translation Approaches used in India," no. 6, pp. 36–39, 2015.
[31] D. Pisharoty, P. Sidhaye, H. Utpat, S. Wandkar, and R. Sugandhi, "Extending Capabilities of English to Marathi Machine Translator."

An EOQ Model Dealing with Weibull Deterioration with Shortages, Ramp Type Demand Rate and Unit Production Cost Incorporating the Effect of Inflation

Jayshree* and Shalini Jain**
* Department of Mathematics, C.C.S. University, Meerut, 250004
Jayashri000@gmail.com
** Centre for Mathematical Sciences, Banasthali University, Rajasthan, 304022
Shalini.shalini2706@gmail.com

Abstract: As the long arm of the grinding, deep financial crisis continues to haunt the global economy and the assumption that the goods in inventory always preserve their physical characteristic are not true in general, the effects of inflation and deterioration cannot be oblivious to an inventory system. In today's market structure, partial backlogged shortages are a more practical important assumption for better business performance. Therefore, bearing in mind these facts, we have developed an order level inventory system for deterioration with shortages and demand rate as a ramp type function of time by exploring two different cases when the demand rate is stabilized after and before the production stopping time end unit production cost is inversely proportional to the demand rate. Optimal costs are determined by two different cases incorporating the effect of inflation.

Keywords: Ramp Type Demand, Weibull Deterioration, Unit Production Cost, With Shortage,inflation.

Introduction

Traditionally items kept as inventory are tacitly assumed to have an infinite lifespan or presumed to be perfect throughout the business cycle. In today's competitive business world, it is of the unrealistic assumption that all produced items are of good quality. But for managing inventory in a realistic scenario, many products impair in quality due to changing technological trends, it would be unethical if the phenomenon deterioration are not considered. Deterioration of items is a frequent and natural phenomenon which cannot be ignored. In realistic scenario the life cycle of seasonal product, fruits, food items, electric component, volatile liquid, food etc are short and finite usually can undergo deterioration. Thus the item may not serve the purpose after a period of time and will have to be discarded as it cannot be used to satisfy the future demand of customers. The term deterioration means spoilage, vaporization and obsolescence, loss of character or value in a product along time. At first Wagner and Whitin (1958) dealt with an inventory model for deteriorating items at the end of the shortage period. Researchers have been progressively modifying the existing models by using the deterioration function of various types; it can be constant type or dependent on time. .In our proposed model we have considered the Weibull distribution as the function of deterioration. In past few decades it is observed by Berrotoni (1962) that both the leakage failure for the dry batteries and the life expectancy of ethical drugs could be expressed in Weibull distribution. Covert and Philip (1973) has influenced by the work of Berrotoni (1962) to develop an inventory model for deteriorating items with variable rate of deterioration .They have explained two parameter Weibull distribution to contemplate deterioration as distribution of time. Misra (1975) proposed an inventory model with two parameter Weibull distribution and finite rate of replenishment. The research has been summarized in different survey papers Goyal and Giri (2001), Raafat (1991), Ruxian et al. (2010), Goyal et al. (2013).

For long, inventory models have discussed with the case where demand is either a constant or a monotonic function. Almost the time varying demand patterns considered in most of the papers where assum es demand rate to be either increasing or decreasing continuously. But in practice, it is not possible. In real world, it stabilizes at the mature stage of the product life cycle once the product has been accepted in the market. The kind of stabilization has been termed as "ramp type". The ramp type demand is commonly seen when some fresh fruits are brought to the market. Hill (1995) first time considered increases linearly at the beginning and then after maturation it becomes a constant, a stable stage till the end of the inventory cycle. Deng at al. (2007) developed a note on the inventory models the deteriorating items with ramp-type demand rate by exploring two cases where the time point occurs before and after the point where the demand is stabilized. Skouri et al. (2011) studied with ramp type demand rate and time dependent deterioration rate with unit production cost and shortages. This type of demand patterns examined by Hariga (1995), Wu and Ouyang (2000), Yang et al. (2001), Girl et al. (2003), Manna and Chaudhuri (2006), Panda et al. (2008), Chen et al.(2006) etc.

Apart from the above mentioned facts due to the globalization of market inflation cannot be shrugged off while doing the inventory models as the market fluctuates from time to time. So inflation is a crucial attribute of today's esoteric economy. However the most common economic meaning of inflation is: reduction in the value of money. Buzacott (1975) developed the model by considering inflationary effect assuming constant inflation rate. Thereafter a great deal of research efforts have been developed or Liao et al. (2000), Yang et al. (2001), Sarkar et al. (2011). Roy and Chaudhuri (2011) dealt with an EOQ model with ramp type demand under finite time horizon and inflation.

Moreover in today's customer are often fickle and increasingly less loyal, some customers would like to wait during the shortage period, but others would not. The customer's demand for the item is lost in the lost sales case and an assuredly filled by a competitor, which can be manifestly as the loss of profit on the sales. Resultantly the opportunity cost resulted from lost sales should be deliberated in the model. Some studies have mutated inventory policies by considered the partial backlogging rate and some assumed to be completely backlogged. Two types of backlogging accumulated such as constant type and time dependent partial backlogging rate dependent on the wailing time up to the next replenishment have been studied extensively by many researches such as Abad (1996), Chang and Dye (1999), Wang (2002), Wu et al. (2006), Singh and Singh (2007), Dye et al.(2007), Singh et al.(2008), Sicilia et al. (2009). However in market structure another important factor is shortages which no retailer would prefer, and in practice are partially backlogged and partially lost.

Present work is the extension of Manna et al(in press) without shortage model where (a) the demand rate is stabilized into two different cases i.e. after the production stopping time and before the production stopping time and (b) Deterioration rate is constant .In the proposed model we at first have the demand rate which is realistic as any new brands product launch in the market the demand rate linearly depends on time and is stabilized after the production stopping the time and before the time when inventory level reaches zero. Secondly we can ignore the effect of deterioration over time as it is a natural phenomenon so we have taken the deterioration as Weibull distribution. In our model we have taken costs are considered under the effects of learning and inflation.

Further, the paper is validated with the help of a numerical example. The model has explored the effects of deterioration, inflation and backlogging parameter and finds the optimal costs in two different cases with the effect of learning and inflation.

Assumptions and Notations

The following notations and assumptions are considered to develop the inventory model

Notations

K- Unit Production cost (units /unit time)
c_1 – Holding cost per unit per unit of time
c_3 – Deterioration cost per unit per unit of time
c_4 – Shortage cost per unit per unit of time
c_5 – Lost sale cost per unit per unit of time
X – Total average cost for a production cycle
r- Inflationary rate
δ – Backlogging rate

Assumptions

(1) Demand rate in ramp type function of time, i.e. demand rate R= f (t) is assumed to be a ramp type function of time f(t)= D_0[t-(t-μ) H(t-μ)] , $D_0 > 0$ and H(t) is a Heaviside's function:
H (t-μ) = $\begin{cases} 1 & if\ t \geq \mu \\ 0 & if\ t < \mu \end{cases}$
(2) Deterioration varies unit time and it is function of two parameter Weibull distribution of the time, i.e. $\alpha\beta t^{\beta-1}$, $0 < \alpha < 1, \beta \geq 1$, where t denote time of deterioration .
(3) Lead time is zero.
(4) Inflation is considered.
(5) Shortage are Allowed and partially backlogged.
(6) K= γ f (t) is the production rate where γ (> 1) is a constant.

The unit production cost $v = \alpha_1 R^{-s}$ where $\alpha_1 > 0, s > 0$ and s ≠ 2.
α_1 is obviously positive since v and R are both non-negative. Also higher demands result in lower unit cost of production. This implies that v and R are inversely related and hence, must be non-negative i.e. positive.
Now,

$\frac{dV}{dR} = -\alpha_1 s R^{-(s+1)} < 0.$

$\frac{d^2V}{dR^2} = \alpha_1 s(s+1) R^{-(s+2)} > 0.$

Thus, marginal unit cost of production is an increasing function of R. These results imply that, as the demand rate increases, the unit cost of production decreases at an increasing rate. Due to this reason, the manufacture is encouraged to produce more as the demand for the item increases. The necessity of restriction s≠ 2 arises from the nature of the solution of the problem.

Mathematical Formulation of the Model

Case 1 ($\mu \leq t_1 \leq t_2$)

The stock level initially is zero. Production starts just after t=0. When the stock attains a level q at time t= t_1, then the production stops at that time. The time point μ occurs before the point t=t_1, where demand is stabilized after that the inventory level diminishes due to both demand and deterioration ultimately falls to zero at time t = t_2. After time t_2 shortages occurs at t=T, which are partially backlogged and partially lost. Then, the cycle repeats.

Let Q(t) be the inventory level of the system at any time t($0 \leq t \leq t_2$). The differential equations governing the system in the interval [0,t_2]are given by

$\frac{dQ(t)}{dt} + \alpha\beta t^{\beta-1} Q(t) = K - F(t)$ $\quad 0 \leq t \leq \mu$ (1)

with the condition Q(0)=0

$\frac{dQ(t)}{dt} + \alpha\beta t^{\beta-1} Q(t) = K - F(t)$ $\quad \mu \leq t \leq t_1$ (2)

with the condition $Q(t_1) = q$

$\frac{dQ(t)}{dt} + \propto \beta t^{\beta-1} Q(t) = -F(t)$ $\quad t_1 \leq t \leq t_2$ (3)

with the condition $Q(t_1) = q, Q(t_2)$=0

$\frac{dQ(t)}{dt} = -e^{-\delta(T-t_2)} F(t)$ $\quad t_2 \leq t \leq T$ (4)

with the condition $Q(t_2)$=0

Using ramp type function F(t), equation (1),(2),(3),(4) become respectively

$\frac{dQ(t)}{dt} + \propto \beta t^{\beta-1} Q(t) = (\gamma - 1) D_0 t$ $\quad 0 \leq t \leq \mu$ (5)

with the condition Q(0) = 0

$\frac{dQ(t)}{dt} + \propto \beta t^{\beta-1} Q(t) = (\gamma - 1) D_0 \mu$ $\quad \mu \leq t \leq t_1$ (6)

with the condition $Q(t_1) = q$

$\frac{dQ(t)}{dt} + \propto \beta t^{\beta-1} Q(t) = D_0 \mu$ $\quad t_1 \leq t \leq t_2$ (7)

With the conditions $Q(t_1)$= q, $Q(t_2)$= 0,

$\frac{dQ(t)}{dt} = -e^{-\delta(T-t_2)} D_0 \mu$ $\quad t_2 \leq t \leq T$ (8)

with the condition Q(t_2)=0

(5),(6),(7),(8) are first order linear differential equations

For the solution of equation (5) we get

$Q(t)e^{\alpha t^\beta} = (\gamma - 1) \int D_0 t e^{\alpha t^\beta} + C$

$= (\gamma - 1) D_0 [\frac{t^2}{2} + \frac{\alpha t^{\beta+2}}{\beta+2} + \frac{\alpha^2 t^{2\beta+2}}{2(2\beta+2)} + --] + C$ (9)

By using the condition Q (0) = 0 (10)

$Q(t) = (\gamma - 1) D_0 e^{-\alpha t^\beta} [\frac{t^2}{2} + \frac{\alpha t^{\beta+2}}{(\beta+2)} + \frac{\alpha^2 t^{2\beta+2}}{2(2\beta+2)} + --], 0 \leq t \leq \mu$ (11)

for the solution of equation (6) we have

$\int_\mu^t d[e^{\alpha t^\beta} Q(t)] = (\gamma - 1)\, D_0\, \mu \int_\mu^t e^{\alpha t^\beta} dt$

$e^{\alpha t^\beta} Q(t) - e^{\alpha \mu^\beta} Q(\mu) = (\gamma - 1)\, D_0\, \mu \int_\mu^t e^{\alpha t^\beta} dt = (\gamma - 1) D_0\, \mu \int_\mu^t 1 + \alpha t^\beta +$

$e^{\alpha t^\beta} Q(t) - e^{\alpha \mu^\beta} Q(\mu) = (\gamma - 1)\, D_0\, \mu [t + \frac{\alpha t^{\beta+1}}{\beta+1} + \frac{\alpha^2 t^{2\beta+1}}{2(2\beta+1)} + ----]$

$e^{\alpha t^\beta} Q(t) = (\gamma - 1)\, D_0\, \mu \left[\frac{\mu}{2} + \frac{\alpha \mu^{\beta+1}}{\beta+2} + \frac{\alpha^2 \mu^{2\beta+1}}{2(2\beta+2)} \pm --\right] + (\gamma - 1) D_0 \mu [t - \mu + \frac{\alpha}{\beta+1}\left(\frac{t^{\beta+1}}{\mu^{\beta+1}}\right) + \frac{\alpha^2}{2(2\beta+1)} (t^{2\beta+1} - \mu^{2\beta+1}) +]$

$= (\gamma - 1)D_0\mu e^{-\alpha t^\beta}\ [t - \frac{\mu}{2} + \frac{\alpha t^{\beta+1}}{(\beta+1)} + \frac{\alpha^2 t^{2\beta+1}}{2(2\beta+1)} - \frac{\alpha\mu^{\beta+1}}{(\beta+1)(\beta+2)} - \frac{\alpha^2\mu^{2\beta+1}}{2(2\beta+1)(2\beta+2)}], \mu \le t \le t_1$ (12)

The solution of equation (7) is given by

Q(t) $e^{\alpha t^\beta} = -D_0\mu \int e^{\alpha t^\beta} dt + C$

= - $D_0\mu \int [1 + \alpha t^\beta + \frac{\alpha^2 t^{2\beta}}{2} + \quad] + C$

Q(t) $e^{\alpha t^\beta} = -Do\mu \left(t + \frac{\alpha t^{\beta+1}}{\beta+1} + \frac{\alpha^2 t^{2\beta+1}}{2(2\beta+1)} + \quad \right) + C$

Putting $Q(t)_1$ = q we get

$$qe^{\alpha t_1\beta} = -Do\mu \left(t_1 + \frac{\alpha t_1\beta + 1}{\beta + 1} + \frac{\alpha^2 t_1^{2\beta+1}}{2(2\beta + 1)} + \quad \right) + C$$

C=$qe^{\alpha t_1^\beta} + Do\mu \left(t_1 + \frac{\alpha t_1\beta+1}{\beta+1} + \frac{\alpha^2 t_1^{2\beta+1}}{2(2\beta+1)} + \quad \right)$ (13)

Using initial condition $Q(t_2)$=0 in equation (13) we have,

$qe^{\alpha t_1^\beta} = D_0\,\mu\,(t_2 + \frac{\alpha t_2{}^{\beta+1}}{\beta+1} + \frac{\alpha^2 t_2{}^{2\beta+1}}{2(2\beta+1)} + \ - - -) - D_0\mu\,(t_1 + \frac{\alpha t_1\beta+1}{\beta+1} + \frac{\alpha^2 t_1{}^{2\beta+1}}{2(2\beta+1)} + \quad)$

q= $D_0\,\mu e^{-\alpha t_1^\beta}(t_2 + \frac{\alpha t_2{}^{\beta+1}}{\beta+1} + \frac{\alpha^2 t_2{}^{2\beta+1}}{2(2\beta+1)} + \ ) - D_0\mu\, e^{-\alpha t_1^\beta}(t_1 + \frac{\alpha t_1\beta+1}{\beta+1} + \frac{\alpha^2 t_1{}^{2\beta+1}}{2(2\beta+1)} + --)$

Substitute q in equation (13) the solution of equation (7) is

$e^{\alpha t^\beta} Q(t) = -D_0\,\mu\left(t + \frac{\alpha t^{\beta+1}}{\beta+1} + \frac{\alpha^2 t^{2\beta+1}}{2(2\beta+1)} + - - -\right) + D_0\mu\left(t_2 + \frac{\alpha t_2{}^{\beta+1}}{\beta+1} + \frac{\alpha^2 t_2{}^{2\beta+1}}{2(2\beta+1)} + - -\right) - D_0\mu\left(t_1 + \frac{\alpha t_1{}^{\beta+1}}{\beta+1} + \frac{\alpha^2 t_1{}^{2\beta+1}}{2(2\beta+1)} + - -\right) + D_0\mu\,(t_1 + \frac{\alpha t_1\beta+1}{\beta+1} + \frac{\alpha^2 t_1{}^{2\beta+1}}{2(2\beta+1)} + --)$

$Q\,(t) = D_0\,\mu\, e^{-\alpha t^\beta}[(t_2 - t) + \frac{\alpha}{\beta+1}(t_2{}^{\beta+1} - t^{\beta+1}) + \frac{\alpha^2}{2(2\beta+1)}(t_2{}^{2\beta+1} - t^{2\beta+1}) + - -]$

$t_1 \le t \le t_2$ (15)

The solution of equation (8) is

$\frac{dQ(t)}{dt} = -D_0\mu\, e^{-\delta(T-t_2)} \qquad t_2 \le t \le T$

with boundary condition $Q(t_2) = 0$

$Q\,(t) = -D_0\,\mu \int [1 - \delta(T - t_2)]\ dt + c$

$Q\,(t) = -D_0\,\mu\ [t - \delta(T - t_2)t] + c$

By using $Q(t_2)$ =0, we get

$Q(t) = D_0\,\mu[(t_2 - t) - \delta(T - t_2)(t_2 - t)]$ (16)

Shortage cost over the period [0,T] is defined as

$\int_{t_2}^{t} \theta(t)dt = -\int_{t_2}^{T} D_0\,\mu[\{t_2 - t) - \delta[T - t_2](t_2 - t)]dt\, e^{-rt}$

$= -D_0\,\mu \int_{t_2}^{T}(1 - rt)[(t_2 - t) - \delta(T - t_2)(t_2 - t)]dt$

$= -D_0\,\mu\left[\left\{Tt_2 - \frac{t_2^2}{2} - \frac{T^2}{2} - \delta(T - t_2)\left(t_2T - \frac{t_2^2}{2} - \frac{T^2}{2}\right)\right\} - \left\{r\left(\frac{T^2}{2} - \frac{t_2^2}{2}\right)t_2 - r\left(\frac{T^3}{3} - \frac{t_2^3}{3}\right) - \delta r(T - t_2)\left(\frac{t_2T^2}{2} - \frac{t_2^3}{2} - \frac{T^3}{3} + \frac{t_2^3}{3}\right)\right\}\right]$

$= -D_0\,\mu\left[\left\{Tt_2 - \frac{t_2^2}{2} - \frac{T^2}{2} - \delta\left(\frac{3t_2T^2}{2} - \frac{3t_2^2T}{2} - \frac{T^2}{2} + \frac{t_2^3}{2}\right) - \left\{r\frac{T^2t_2}{2} + r\frac{t_2^3}{2} + r\frac{T^3}{3} + \delta r\,(T - t_2)\left(\frac{t_2T^2}{2} - \frac{t_2^3}{6} - \frac{T^3}{3}\right)\right\}\right]$,

(17)

Lost sale cost per cycle is

LS= $D_0\,\mu \int_{t_2}^{T}(1 - e^{-\delta(T-t_2)})\,dt$

LS= $D_0\,\mu \int_{t_2}^{T} \delta(T - t_2)(t_2 - t)\,dt$

Lost sale cost over the period [0,T] is

LS= $D_0\ \mu \int_{t_2}^{T} \delta(T - t_2)(t_2 - t)e^{-rt}\,dt$

LS= $D_0\,\mu\delta \int_{t_2}^{T}(T - t_2)(t_2 - t)\,(1 - rt)d$

LS= $D_0\,\mu\delta \int_{t_2}^{T}(Tt_2 - Tt - t_2^2 + t_2t - rTt_2t + rTt^2 + rt_2^2t - rt_2t^2)\,dt$

LS= $D_0\,\mu\delta\left[Tt_2(T - t_2) - T\left(\frac{T^2}{2} - \frac{t_2^2}{2}\right) - t_2^2(T - t_2) + t_2\left(\frac{T^2}{2} - \frac{t_2^2}{2}\right) - rTt_2\left(\frac{T^2}{2} - \frac{t_2^2}{2}\right) + rT\left(\frac{T^3}{3} - \frac{t_2^3}{3}\right) + rt_2^2\left(\frac{T^2}{2} - \frac{t_2^2}{2}\right) - rt_2\left(\frac{T^3}{3} - \frac{t_2^3}{3}\right)\right]$

LS= $D_0\,\mu\delta\left[\frac{3T^2t_2}{2}-\frac{3t_2^2T}{2}-\frac{T^3}{2}+\frac{t_2^3}{2}-\frac{5rt_2T^3}{6}-\frac{rt_2^4}{6}+\frac{rTt_2^3}{6}+\frac{rT^4}{3}+\frac{rt_2^2T^2}{2}\right]$ (18)

The total inventory over the period [0, t_2] is

$$\int_0^{t_2} Q\,(t)dte^{-rt}=\int_0^{\mu} Q\,(t)e^{-rt}dt+\int_{\mu}^{t_1} Q\,(t)e^{-rt}dt+\int_{t_1}^{0} Q\,(t)e^{-rt}dt$$

$$\int_0^{\mu} Q(t)e^{-rt}dt=\int_0^{\mu}(\gamma-1)D_0e^{-\alpha t^{\beta}}[\frac{t^2}{2}+\frac{\alpha t^{\beta+2}}{(\beta+2)}+\frac{\alpha^2t^{2\beta+2}}{2(2\beta+2)}+--]e^{-rt}dt$$

$$=(\gamma-1)D_0\int_0^{\mu}\left(\frac{t^2}{2}-\frac{\alpha\beta t^{\beta+2}}{2(\beta+2)}-\frac{\alpha^2\,(3\beta+2)t^{2\beta+2}}{2(\beta+2)(2\beta+2)}\right)(1-rt)dt$$

$=(\gamma\text{-}1)\,D_0\ \int_0^{\mu}(\frac{t^2}{2}-\frac{\alpha\beta t^{\beta+2}}{2(\beta+2)}-\frac{\alpha^2(3\beta+2)t^{2\beta+2}}{2(\beta+2)(2\beta+2)}-\frac{rt^3}{2}+\frac{\alpha r\beta t^{\beta+3}}{2(\beta+2)}-\frac{\alpha^2(3\beta+2)t^{2\beta+3}}{2(\beta+2)(2\beta+2)})$ dt

$=(\gamma\text{-}1)\,D_0\,[\frac{t^3}{6}-\frac{\alpha\beta t^{\beta+3}}{2(\beta+2)(\beta+3)}-\frac{\alpha^2(3\beta+2)t^{2\beta+3}}{2(\beta+2)(2\beta+2)(2\beta+3)}-\frac{rt^4}{8}+\frac{r\alpha\beta t^{\beta+4}}{2(\beta+2)(\beta+4)}-\frac{r\alpha^2(3\beta+2)t^{2\beta+4}}{2(\beta+2)(2\beta+2)(2\beta+4)}]_0^{\mu}$

$= (\gamma\text{-}1)\,D_0\ [\frac{\mu^3}{6}-\frac{\alpha\beta\mu^{\beta+3}}{2(\beta+2)(\beta+3)}-\frac{\alpha^2(3\beta+2)\mu^{2\beta+3}}{2(\beta+2)(2\beta+2)(2\beta+3)}-\frac{r\mu^2}{8}+\frac{r\alpha\beta\mu^{\beta+4}}{2(\beta+2)(\beta+4)}-\frac{r\alpha^2(3\beta+2)\mu^{2\beta+4}}{2(\beta+2)(2\beta+2)(2\beta+4)}]$

$$\int_{\mu}^{t_1} Q(t)e^{-rt}dt=D_0\,\mu(\gamma-1)\int_{\mu}^{t_1}[t-\frac{\mu}{2}+\frac{\alpha t^{\beta+1}}{(\beta+1)}+\frac{\alpha^2t^{2\beta+1}}{2(2\beta+1)}-\frac{\alpha\mu^{\beta+1}}{(\beta+1)(\beta+2)}-\frac{\alpha^2\mu^{2\beta+1}}{2(2\beta+1)(2\beta+2)}-\alpha t^{\beta+1}+\frac{\alpha\mu t^{\beta}}{2}-\frac{\alpha^2t^{2\beta+1}}{(\beta+1)}+\frac{\alpha^2t^{\beta}\mu^{\beta+1}}{(\beta+1)(\beta+2)}]\,e^{-rt}\,dt$$

$$=D_0\mu\,(\gamma-1)\int_{\mu}^{t_1}t-\frac{\mu}{2}+\frac{\alpha t^{\beta+1}}{(\beta+1)}+\frac{\alpha^2t^{2\beta+1}}{2(2\beta+1)}-\frac{\alpha\mu^{\beta+1}}{(\beta+1)(\beta+2)}-\frac{\alpha^2\mu^{2\beta+1}}{2(2\beta+1)(2\beta+2)}-\alpha t^{\beta+1}+\frac{\alpha\mu t^{\beta}}{2}-\frac{\alpha^2t^{2\beta+1}}{(\beta+1)}+\frac{\alpha^2t^{\beta}\mu^{\beta+1}}{(\beta+1)(\beta+2)}-rt^2+rt\frac{\mu}{2}-\frac{\alpha rt^{\beta+2}}{(\beta+1)}-\frac{r\alpha^2t^{2\beta+2}}{2(2\beta+1)}+\frac{\alpha r\mu^{\beta+1}t}{(\beta+1)(\beta+2)}+\alpha t^{\beta+2}r-\frac{r\alpha\mu t^{\beta+1}}{2}+\frac{r\alpha^2t^{2\beta+2}}{(\beta+1)}-\frac{r\alpha^2t^{\beta+1}\mu^{\beta+1}}{(\beta+1)(\beta+2)}$$

$$\int_{t_1}^{t_2} Q(t)e^{-rt}dt=D_0\mu\int_{t_1}^{t_2}[(t_2-t)+\frac{\alpha}{(\beta+1)}\left(t_2{}^{\beta+1}-t^{\beta+1}\right)+\frac{\alpha^2}{2(2\beta+1)}\,(t_2{}^{2\beta+1}-t^{2\beta+1})]e^{-\alpha t^{\beta}}\,e^{-rt}dt$$

$$=D_0\mu\int_{t_1}^{t_2}[(t_2-t)+\frac{\alpha}{(\beta+1)}\left(t_2{}^{\beta+1}-t^{\beta+1}\right)+\frac{\alpha^2}{2(2\beta+1)}\left(t_2{}^{2\beta+1}-t^{2\beta+1}\right)-\alpha t^{\beta}(t_2-t)+\frac{\alpha^2}{(\beta+1)}t^{\beta}(t_2{}^{\beta+1}-t^{\beta+1})]e^{-rt}\,dt$$

$$=D_0\,\mu\int_{t1}^{t_2}[(t_2-t)+\frac{\alpha}{(\beta+1)}\left(t_2{}^{\beta+1}-t^{\beta+1}\right)+\frac{\alpha^2}{2(2\beta+1)}\left(t_2{}^{2\beta+1}-t^{2\beta+1}\right)-\alpha\left(t_2t^{\beta}-t^{\beta+1}\right)+\frac{\alpha^2}{(\beta+1)}\left(t_2{}^{\beta+1}t^{\beta}t^{2\beta+1}\right)-r\left(t_2t-\frac{t^2}{2}\right)-\frac{\alpha r}{(\beta+1)}\left(t_2{}^{\beta+1}t-\frac{t^{\beta+2}}{2}\right)-\frac{\alpha^2r}{2(2\beta+1)}\left(t_2{}^{2\beta+1}t-t^{2\beta+2}\right)+\alpha r\left(t_2t^{\beta+1}-t^{\beta+2}\right)-\frac{\alpha^2r}{(\beta+1)}(t_2{}^{\beta+1}t^{\beta+1}-t^{2\beta+2})]$$

$$=D_0\,\mu\,[t_2(t_2-t_1)-\frac{{t_2}^2}{2}+\frac{{t_1}^2}{2}+\frac{\alpha}{(\beta+1)}t_2{}^{\beta+1}(t_2-t_1)-\frac{\alpha}{(\beta+1)}\left(\frac{t_2{}^{\beta+2}-t_1{}^{\beta+2}}{(\beta+2)}\right)+\frac{\alpha^2}{2(2\beta+1)}t_2{}^{2\beta+1}+(t_2-t_1)-\frac{\alpha^2}{2(2\beta+1)}-\left(\frac{t_2{}^{2\beta+2}-t1^{2\beta+2}}{2(\beta+2)}\right)-\alpha t_2\left(\frac{t_2{}^{\beta+1}-t_1{}^{\beta+1}}{(\beta+1)}\right)+\alpha\left(\frac{t_2{}^{\beta+2}-t_1{}^{\beta+2}}{(\beta+2)}\right)+\frac{\alpha^2}{(\beta+1)}t_2{}^{\beta+1}\left(\frac{t_2{}^{\beta+2}-t_1{}^{\beta+1}}{(\beta+1)}\right)-\frac{\alpha^2}{(\beta+1)}\left(\frac{t_2{}^{2\beta+2}-t_1{}^{2\beta+2}}{(2\beta+2)}\right)-rt_2(t_2-t_1)+r\left(\frac{{t_2}^2-{t_1}^2}{2}\right)-\frac{r\alpha}{(\beta+1)}t_2{}^{\beta+1}(t_2-t_1)+\frac{r\alpha}{(\beta+1)}\left(\frac{t_2{}^{\beta+3}-t_1{}^{\beta+3}}{(\beta+3)}\right)-\frac{\alpha^2r}{2(2\beta+1)}t_2{}^{2\beta+1}($$

$$(*\ ition\ \ 27/11/15\ Hundred\ Eighteen\ Only\ ven\ Only$$

$$t_2-t_1)+\frac{\alpha^2r}{2(2\beta+1)}(\frac{t_2{}^{2\beta+3}-t_1{}^{2\beta+3}}{(2\beta+3)})+\alpha rt_2(\frac{t_2{}^{\beta+2}-t_1{}^{\beta+2}}{(\beta+2)}-\alpha r\left(\frac{t_2{}^{\beta+3}-t_1{}^{\beta+3}}{(\beta+3)}\right)-\frac{\alpha^2r}{(\beta+1)}t_2{}^{\beta+1}\left(\frac{t_2{}^{\beta+2}-t_1{}^{\beta+2}}{(\beta+2)}\right)+\frac{\alpha^2r}{(\beta+1)}(\frac{t_2{}^{2\beta+2}-t_1{}^{2\beta+2}}{(2\beta+3)})]$$

$$=D_0\,\mu[\frac{t_2^2}{2}-t_2t_1+\frac{t_1^2}{2}+\frac{(\alpha)t_2^{\beta+2}\beta}{(\beta+1)(\beta+2)}-\frac{\alpha t_1t_2^{\beta+1}}{(\beta+1)}-\frac{\alpha\beta t_1^{\beta+2}}{(\beta+1)(\beta+2)}-\frac{\alpha^2t_2^{2\beta+1}t_1}{2(2\beta+1)}+\frac{\alpha^2t_2^{2\beta+2}(\beta+3)}{2(2\beta+2)(\beta+1)}+\frac{\alpha^2t_1^{2\beta+2}(5\beta+3)}{2(2\beta+1)(2\beta+2)(\beta+1)}+\frac{\alpha t_2t_1^{\beta+1}}{(\beta+1)}-\frac{\alpha^2t_2^{\beta+1}t_1^{\beta+1}}{(\beta+1)^2}-\frac{rt_2^2}{2}+rt_2t_1-\frac{rt_1^2}{2}-\frac{r\alpha t_2^{\beta+2}}{(\beta+1)}+\frac{r\alpha t_2^{\beta+1}t_1}{(\beta+1)}+\frac{r\alpha t_1^{\beta+3}\beta}{(\beta+3)(\beta+1)}-\frac{\alpha^2t_2^{2\beta+2}r}{2(2\beta+1)}+\frac{r\alpha t_2^{\beta+3}(2\beta+3)}{(\beta+1)(\beta+2)(\beta+3)}-\frac{3\alpha^2\beta t_2^{2\beta+3}r}{2(2\beta+1)\,(\beta+2)(2\beta+3)}-\frac{\alpha^2(5\beta+3)\,t_1^{2\beta+3}r}{2(2\beta+1)\,(\beta+1)(2\beta+3)}-\frac{\alpha t_2t_1^{\beta+2}}{(\beta+2)}+\frac{\alpha^2t_2^{\beta+1}t_1^{\beta+2}}{(\beta+1)(\beta+2)}]$$

Therefore, the total inventory in [0, t_2] is given by

$$\int_0^{t_2} Q(t)e^{-rt}dt=(\gamma-1)\,D_0[\frac{\mu^3}{6}-\frac{\alpha\beta\mu^{\beta+3}}{2(\beta+2)(\beta+3)}-\frac{\alpha^2(3\beta+2)\mu^{2\beta+3}}{2(\beta+2)(2\beta+2)(2\beta+3)}-\frac{r\mu^2}{8}+\frac{r\alpha\beta\mu^{\beta+4}}{2(\beta+2)(\beta+4)}-\frac{r\alpha^2(3\beta+2)\mu^{2\beta+4}}{2(\beta+2)(2\beta+2)(2\beta+4)}]+$$

$$D_0\,\mu\,(\gamma-1)\left[\frac{{t_1}^2}{2}-\frac{\mu t_1}{2}-\frac{\alpha\beta{t_1}^{\beta+2}}{(\beta+1)(\beta+2)}+\frac{\alpha\beta\mu^{\beta+2}}{2(\beta+1)(\beta+2)}-\frac{\alpha^2(3\beta+1){t_1}^{2\beta+2}}{2(2\beta+1)(2\beta+2)(\beta+1)}+\frac{\alpha^2\beta\mu^{2\beta+2}}{2(2\beta+1)(2\beta+2)(\beta+2)}-\frac{\alpha\mu^{\beta+1}t_1}{(\beta+1)(\beta+2)}-\frac{\alpha^2\mu^{2\beta+1}t_1}{2(2\beta+1)(2\beta+1)}+\right.$$

$$\frac{\alpha^2 {t_1}^{\beta+1}\mu^{\beta+1}}{(\beta+1)^2(\beta+2)}-\frac{r{t_1}^3}{3}+\frac{r\mu^3}{12}+\frac{r{t_1}^2\mu}{4}+\frac{\alpha r{t_1}^{\beta+3}\beta}{(\beta+1)(\beta+3)}+\frac{r\alpha\beta\mu^{\beta+3}}{2(\beta+1)(\beta+3)(\beta+2)}-\frac{r\alpha^2{t_1}^{2\beta+3}}{2(2\beta+1)(2\beta+3)}+\frac{r\alpha\mu^{\beta+1}{t_1}^2}{2(\beta+1)(\beta+2)}+\frac{r\alpha^2\mu^{2\beta+1}{t_1}^2}{4(2\beta+1)(2\beta+2)}+\frac{r\alpha{t_1}^{(\beta+3)}}{(\beta+3)}-$$
$$\frac{\alpha\mu r{t_1}^{\beta+2}}{2(\beta+2)}+\frac{\alpha^2 r{t_1}^{2\beta+3}}{(\beta+1)(2\beta+3)}-\frac{r\alpha^2{t_1}^{\beta+2}\mu^{\beta+1}}{(\beta+1)(\beta+2)^2}+\frac{r\alpha^2\mu^{2\beta+3}(-7\beta^2-12\beta-4)}{8(\beta+1)(\beta+2)^2(2\beta+3)}+\frac{\alpha\mu{t_1}^{\beta+1}}{2(\beta+1)}\Big]+D_0\,\mu[\frac{t_2^2}{2}-t_2t_1+\frac{t_1^2}{2}+\frac{(\alpha)t_2^{\beta+2}\beta}{(\beta+1)(\beta+2)}-\frac{\alpha t_1t_2^{\beta+1}}{(\beta+1)}-$$
$$\frac{\alpha\beta t_1^{\beta+2}}{(\beta+1)(\beta+2)}-\frac{\alpha^2t_2^{2\beta+1}t_1}{2(2\beta+1)}+\frac{\alpha^2t_2^{2\beta+2}(\beta+3)}{2(2\beta+2)(\beta+1)}+\frac{\alpha^2t_1^{2\beta+2}(5\beta+3)}{2(2\beta+1)(2\beta+2)(\beta+1)}+\frac{\alpha t_2t_1^{\beta+1}}{(\beta+1)}-\frac{\alpha^2t_2^{\beta+1}t_1^{\beta+1}}{(\beta+1)^2}-\frac{rt_2^2}{2}+rt_2t_1-\frac{rt_1^2}{2}-\frac{r\alpha t_2^{\beta+2}}{(\beta+1)}+\frac{r\alpha t_2^{\beta+1}t_1}{(\beta+1)}$$
$$+\frac{r\alpha t_1^{\beta+3}\beta}{(\beta+3)(\beta+1)}-\frac{\alpha^2t_2^{2\beta+2}r}{2(2\beta+1)}+\frac{r\alpha t_2^{\beta+3}(2\beta+3)}{(\beta+1)(\beta+2)(\beta+3)}-\frac{3\alpha^2\beta t_2^{2\beta+3}r}{2(2\beta+1)\,(\beta+2)(2\beta+3)}-\frac{\alpha^2(5\beta+3)\,t_1^{2\beta+3}r}{2(2\beta+1)\,(\beta+1)(2\beta+3)}-\frac{\alpha t_2t_1^{\beta+2}}{(\beta+2)}+\frac{\alpha^2t_2^{\beta+1}t_1^{\beta+2}}{(\beta+1)(\beta+2)}] \quad (19)$$

Total number of deteriorated items over the period $[0,t_2]$ is given by

Production in $[0,\mu]$+ Production in $[\mu,t_1]$ – Demand in $[0,\mu]$ –Demand in $[\mu, t_2]$

$$=\gamma\int_o^{\mu}Dote^{-rt}dt+\gamma\int_o^{t_1}D_0\,\mu e^{-rt}dt-D_0\int_o^{\mu}te^{-rt}dt-\int_{\mu}^{t_2}Do\mu e^{-rt}dt$$
$$=\gamma D_0\,[\frac{\mu^2}{2}-\frac{r\mu^3}{3}]+\gamma D_0\mu\left(t_1-\mu-r\frac{t_1^2}{2}+r\frac{\mu^2}{2}\right)-D_0\left[\frac{\mu^2}{2}-\frac{r\mu^3}{3}\right]-D_0\mu[(t_2-\mu)-r\frac{t_2^2}{2}+r\frac{\mu^2}{2}]$$
$$=\frac{1}{2}\gamma D_0\mu\left[2t_1-\mu-rt_1^2+\frac{r\mu^2}{3}\right]-\frac{1}{2}D_0\,\mu\left[2t_2-\mu-rt_2^2+\frac{r\mu^2}{3}\right] \quad (20)$$

The cost of production in $[u,u+du]$ *is*

$$\text{Kv d}u=\frac{\alpha_1\gamma}{R^{s-1}} \quad (21)$$

Hence the production cost over the period $[o, t_1]$ is given by

$$\int_o^{t_1}\text{Kv}e^{-ru}du=\int_0^{\mu}\text{Kv}e^{-ru}du+\int_{\mu}^{t_1}\text{Kv}e^{-ru}du$$
$$=\int_o^{\mu}\frac{\alpha_1\gamma}{R^{s-1}}e^{-ru}du+\int_{\mu}^{t_1}\frac{\alpha_1\gamma}{R^{s-1}}e^{-ru}du$$
$$=\alpha_1\gamma{D_0}^{1-s}[\int_o^{\mu}[u^{1-s}(1-ru)+\int_{\mu}^{t_1}u^{1-s}(1-ru)du]$$
$$=\alpha_1\gamma{D_0}^{1-s}[\frac{\mu^{2-s}}{(2-s)}-\frac{\mu^{1-s}}{(1-s)}]+\mu^{1-s}[(t_1-\mu)-\left(\frac{t_1^2}{2}-\frac{\mu^2}{2}\right)]$$
$$=\alpha_1\gamma{D_0}^{1-s}\left[\frac{(s-1)\mu^{2-s}+(2-s)\mu^{1-s}t_1}{(2-s)}\right]+\alpha_1\gamma rD_0{}^{1-s}[\frac{\mu^{3-s}}{2}-\frac{{t_1}^2}{2}\mu^{1-s}-\frac{\mu^{1-s}}{(1-s)}] \quad (22)$$

The total average inventory cost X is given by

X= Inventory Cost + Deterioration Cost+ Production Cost+ Shortage Cost + Lost Sale Cost

$$\text{X}=\frac{1}{t_2}[\,c_1\{(\gamma\text{-}1)\,D_0\,(\frac{\mu^3}{6}-\frac{\alpha\beta\mu^{\beta+3}}{2(\beta+2)(\beta+3)}-\frac{\alpha^2(3\beta+2)\mu^{2\beta+3}}{2(\beta+2)(2\beta+2)(2\beta+3)}-\frac{r\mu^2}{8}+\frac{r\alpha\beta\mu^{\beta+4}}{2(\beta+2)(\beta+4)}-\frac{r\alpha^2(3\beta+2)\mu^{2\beta+4}}{2(\beta+2)(2\beta+2)(2\beta+4)})+D_0\,\mu(\gamma\text{-}1)\,(\frac{{t_1}^2}{2}-$$
$$\frac{\mu t_1}{2}-\frac{\alpha\beta{t_1}^{\beta+2}}{(\beta+1)(\beta+2)}+\frac{\alpha\beta\mu^{\beta+2}}{2(\beta+1)(\beta+2)}-\frac{\alpha^2(3\beta+1){t_1}^{2\beta+2}}{2(2\beta+1)(2\beta+2)(\beta+1)}+\frac{\alpha^2\beta\mu^{2\beta+2}}{2(2\beta+1)(2\beta+2)(\beta+2)}-\frac{\alpha\mu^{\beta+1}t_1}{(\beta+1)(\beta+2)}-\frac{\alpha^2\mu^{2\beta+1}t_1}{2(2\beta+1)(2\beta+2)}+\frac{\alpha^2{t_1}^{\beta+1}\mu^{\beta+1}}{(\beta+1)^2(\beta+2)}-\frac{r{t_1}^3}{3}+$$
$$\frac{r\mu^3}{12}+\frac{r{t_1}^2\mu}{4}+\frac{\alpha r{t_1}^{\beta+3}\beta}{(\beta+1)(\beta+3)}+\frac{r\alpha\beta\mu^{\beta+3}}{2(\beta+1)(\beta+3)(\beta+2)}-\frac{r\alpha^2{t_1}^{2\beta+3}}{2(2\beta+1)(2\beta+3)}+\frac{r\alpha\mu^{\beta+1}{t_1}^2}{2(\beta+1)(\beta+2)}+\frac{r\alpha^2\mu^{2\beta+1}{t_1}^2}{4(2\beta+1)(2\beta+2)}+\frac{r\alpha{t_1}^{(\beta+3)}}{(\beta+3)}-\frac{\alpha\mu r{t_1}^{\beta+2}}{2(\beta+2)}+$$
$$\frac{\alpha^2r{t_1}^{2\beta+3}}{(\beta+1)(2\beta+3)}-\frac{r\alpha^2{t_1}^{\beta+2}\mu^{\beta+1}}{(\beta+1)(\beta+2)^2}+\frac{r\alpha^2\mu^{2\beta+3}(-7\beta^2-12\beta-4)}{8(\beta+1)\,(\beta+2)^2(2\beta+3)}+\frac{\alpha\mu t_1^{\beta+1}}{2(\beta+1)})+D_0\,\mu(\frac{t_2^2}{2}-t_2t_1+\frac{t_1^2}{2}+\frac{(\alpha)t_2^{\beta+2}\beta}{(\beta+1)(\beta+2)}-\frac{\alpha t_1t_2^{\beta+1}}{(\beta+1)}-\frac{\alpha\beta t_1^{\beta+2}}{(\beta+1)(\beta+2)}-$$
$$\frac{\alpha^2t_2^{2\beta+1}t_1}{2(2\beta+1)}+\frac{\alpha^2t_2^{2\beta+2}(\beta+3)}{2(2\beta+2)(\beta+1)}+\frac{\alpha^2t_1^{2\beta+2}(5\beta+3)}{2(2\beta+1)(2\beta+2)(\beta+1)}+\frac{\alpha t_2t_1^{\beta+1}}{(\beta+1)}-\frac{\alpha^2t_2^{\beta+1}t_1^{\beta+1}}{(\beta+1)^2}-\frac{rt_2^2}{2}+rt_2t_1-\frac{rt_1^2}{2}-\frac{r\alpha t_2^{\beta+2}}{(\beta+1)}+\frac{r\alpha t_2^{\beta+1}t_1}{(\beta+1)}+\frac{r\alpha t_1^{\beta+3}\beta}{(\beta+3)(\beta+1)}-$$
$$\frac{\alpha^2t_2^{2\beta+2}r}{2(2\beta+1)}-\frac{r\alpha t_2^{\beta+3}(2\beta+3)}{(\beta+1)(\beta+2)(\beta+3)}-\frac{3\alpha^2\beta t_2^{2\beta+3}r}{2(2\beta+1)\,(\beta+2)(2\beta+3)}-\frac{\alpha^2(5\beta+3)\,t_1^{2\beta+3}r}{2(2\beta+1)\,(\beta+1)(2\beta+3)}-\frac{\alpha t_2t_1^{\beta+2}}{(\beta+2)}+\frac{\alpha^2t_2^{\beta+1}t_1^{\beta+2}}{(\beta+1)(\beta+2)})\}+c_3\{\frac{1}{2}\gamma D_0\mu\left(2t_1-\mu-rt_1^2+\right.$$
$$\left.\frac{r\mu^2}{3}\right)-\frac{1}{2}D_0\,\mu(2t_2-\mu-rt_2^2+\frac{r\mu^2}{3})\}-c_4D_0\,\mu\left\{Tt_2-\frac{t_2^2}{2}-\frac{T^2}{2}-\delta\left(\frac{3t_2T^2}{2}-\frac{3t_2^2T}{2}-\frac{T^3}{2}+\frac{t_2^3}{2}\right)-r\frac{T^2t_2}{2}+r\frac{t_2^3}{2}+r\frac{T^3}{3}+\right.$$
$$\left.\delta r\,(T-t_2)\left(\frac{t_2T^2}{2}-\frac{t_2^3}{6}-\frac{T^3}{3}\right)\right\}+c_5D_0\,\mu\delta\left(\frac{3T^2t_2}{2}-\frac{3t_2^2T}{2}-\frac{T^3}{2}+\frac{t_2^3}{2}-\frac{5rt_2T^3}{6}-\frac{rt_2^4}{6}+\frac{rTt_2^3}{6}+\frac{rT^4}{3}+\frac{rt_2^2T^2}{2}\right)+$$
$$\alpha_1\gamma{D_0}^{1-s}\{\frac{(s-1)\mu^{2-s}+(2-s)\mu^{1-s}t_1}{(2-s)}\}+\alpha_1\gamma rD_0{}^{1-s}\{\frac{\mu^{3-s}}{2}-\frac{{t_1}^2}{2}\mu^{1-s}-\frac{\mu^{1-s}}{(1-s)}\}] \quad (23)$$

Optimum values of t_1 and t_2 for minimum average cost x are the solutions of the equations

$$\frac{\partial X}{\partial t_1}=0\ and\ \frac{\partial X}{\partial t_2}=0$$

Provided they satisfy the sufficient conditions

$$\frac{\partial^2X}{\partial t_1^2}>0,\frac{\partial^2X}{\partial t_2^2}>0\text{ and }\frac{\partial^2X}{\partial t_1^2}\frac{\partial^2X}{\partial t_2^2}-(\frac{\partial^2X}{\partial t_1\partial t_2})^2>0$$

$\frac{\partial x}{\partial t_1} = 0$ *and* $\frac{\partial x}{\partial t_2}$=0 gives

$$C_1\left\{D_0\mu(\gamma-1)\left(t_1-\frac{\alpha\beta t_1^{\beta+1}}{(\beta+1)}-\frac{\alpha^2(3\beta+1)t_1^{2\beta+1}}{2(2\beta+1)(\beta+1)}-\frac{\alpha\mu^{\beta+1}}{(\beta+1)(\beta+2)}-\frac{\alpha^2\mu^{2\beta+1}}{2(2\beta+1)(2\beta+2)}-\frac{\mu}{2}+\frac{\alpha^2t_1^{\beta}\mu^{\beta+1}}{(\beta+1)(\beta+2)}-rt_1^2+\frac{rt_1\mu}{2}+\frac{\alpha rt_1^{\beta+2}\beta}{(\beta+1)}-\frac{r\alpha^2t_1^{2\beta+2}}{2(2\beta+1)}+\frac{r\alpha\mu^{\beta+1}t_1}{(\beta+1)(\beta+2)}+\frac{r\alpha^2\mu^{2\beta+1}t_1}{2(2\beta+1)(2\beta+2)}+r\alpha t_1^{\beta+2}-\frac{\alpha\mu rt_1^{\beta+1}}{2}+\frac{\alpha^2rt_1^{2\beta+2}}{(\beta+1)}-\frac{r\alpha^2t_1^{\beta+1}\mu^{\beta+1}}{(\beta+1)(\beta+2)}+\frac{\alpha\mu t_1^{\beta}}{2}\right)+D_0\mu\left(-t_2+t_1-\frac{\alpha t_2^{\beta+1}}{(\beta+1)}-\frac{\alpha\beta t_1^{\beta+1}}{(\beta+1)}-\frac{\alpha^2t_2^{2\beta+1}}{2(2\beta+1)}+\frac{\alpha^2t_1^{2\beta+1}(5\beta+3)}{2(2\beta+1)(\beta+1)}+\alpha t_2t_1^{\beta}-\frac{\alpha^2t_2^{\beta+1}t_1^{\beta}}{(\beta+1)}+rt_2-rt_1+\frac{r\alpha t_2^{\beta+1}}{(\beta+1)}+\frac{r\alpha t_1^{\beta+2}\beta}{(\beta+1)}-\frac{\alpha^2(5\beta+3)t_1^{2\beta+2}r}{2(2\beta+1)(\beta+1)}-\alpha t_2t_1^{\beta+1}+\frac{\alpha^2t_2^{\beta+1}t_1^{\beta+1}}{(\beta+1)}\right)\right\}+c_3\gamma\,D_0\mu(1-rt_1)+\alpha_1\gamma D_0^{1-s}\mu^{1-s}-\alpha_1\gamma rD_0^{1-s}t_1\mu^{1-s}=0 \quad (24)$$

and

$$C_1D_o\mu\left(t_2-t_1+\frac{\alpha\beta t_2^{\beta+1}}{(\beta+1)}-\alpha t_1\,t_2^{\beta}-\frac{\alpha^2t_1t_2^{2\beta}}{2}+\frac{\alpha^2(\beta+3)t_2^{2\beta+1}}{2(\beta+1)}+\frac{\alpha t_1^{\beta+1}}{(\beta+1)}-\frac{\alpha^2t_1^{\beta+1}t_2^{\beta}}{(\beta+1)}-rt_2+rt_1-\frac{r\alpha(\beta+2)t_2^{\beta+1}}{(\beta+1)}+r\alpha t_1t_2^{\beta}-\frac{r\alpha^2(2\beta+2)t_2^{2\beta+1}}{2(2\beta+1)}+\frac{r\alpha(2\beta+3)t_2^{\beta+2}}{(\beta+1)(\beta+2)}-\frac{3\alpha^2\beta rt_2^{2\beta+2}}{2(2\beta+1)(\beta+2)}-\frac{\alpha t_1^{\beta+2}}{(\beta+2)}+\frac{\alpha^2t_2^{\beta}t_1^{\beta+2}}{(\beta+2)}\right)-C_3D_o\mu(1-rt_2)-C_4D_o\mu\{(T-t_2)-\delta(\tfrac{3}{2}T^2-3\,t_2T+\frac{3t_2^2}{2})-\frac{rT^2}{2}+\frac{3rt_2^2}{2}-\delta rt_2T^2+\frac{2}{3}\delta rt_2^3+\frac{5}{6}\delta rT^3-\frac{\delta rTt_2^2}{2}\}+C_5D_o\mu\delta\left(\frac{3}{2}T^2-3t_2T+\frac{3}{2}t_2^2-\frac{5rT^3}{6}-\frac{2}{3}rt_2^3+\frac{1}{2}rTt_2^2+rt_2T^2\right)-X=0 \quad (25)$$

Case–II ($t_1 \le \mu \le t_2$)

The production starts with zero stock level at t=0. Production begins at t=0 and continues up to t=$t_1$and stops as soon as the stock level becomes L at t= t_2. Because of reasons of market demand and deterioration of items, the inventory level decreases till it becomes again zero at t= t_2 .After time t= t_2, another important factor occurs which is shortages. After that period, the cycle repeats itself.

Let Q(t) be the inventory level of the system at any time t ($0 \le t \le t_2$). The differential equations governing the system in the interval [0, t_2] are given by

$$\frac{dQ(t)}{dt}+\alpha\beta\,t^{\beta-1}\,Q(t)=K-F(t) \qquad 0\le t\le t_1 \qquad (26)$$

with the condition Q (0) = 0, Q (t_1) = L

$$\frac{dQ(t)}{dt}+\alpha\beta\,t^{\beta-1}\,Q(t)=-F(t) \qquad t_1\le t\le \mu \qquad (27)$$

with the condition Q (t_1) = L

$$\frac{dQ(t)}{dt}+\alpha\beta\,t^{\beta-1}\,Q(t)=-F(t) \qquad \mu \le t \le t_2 \qquad (28)$$

with the condition Q (t_2) = 0

$$\frac{dQ}{dt}=-e^{-\delta(T-t)}\,\mathrm{F(t)} \qquad t_2 \le t \le T \qquad (29)$$

with the condition Q (t_2) = 0

using ramp type function F (t) equations (26),(27)(28),(29) become respectively

$$\frac{dQ(t)}{dt}+\alpha\beta\,t^{\beta-1}\,Q(t)=(\gamma-1)D_0t \qquad 0\le t\le t_1 \qquad (30)$$

with the condition Q (0) = 0, Q (t_1) = L

$$\frac{dQ(t)}{dt}+\alpha\beta\,t^{\beta-1}\,Q(t)=-D_0t \qquad t_1\le t\le \mu \qquad (31)$$

with the condition Q (t_1) = L

$$\frac{\mathrm{d}Q(t)}{\mathrm{d}t}+\alpha\beta\,t^{\beta-1}\,Q(t)=-D_0\,\mu \qquad \mu \le t \le t_2 \qquad (32)$$

with the condition Q (t_2) = 0

$$\frac{dQ}{dt}=-e^{-\delta(T-t)}\;\;D_0\mu \qquad t_2\le t\le T \qquad (33)$$

with the condition Q (t_2) = 0

The solution of equation (30) is given by the expression (11) and we have

$$e^{\alpha t^{\beta}}\ Q(t) = (\gamma-1)\ D_0\left(\frac{t^2}{2} + \frac{\alpha t^{\beta+2}}{\beta+2} + \alpha^2\ \frac{t^{2\beta+2}}{2(2\beta+2)} + \right) + C$$

With the condition Q(0) = 0, we get

$$Q(t) = (\gamma-1)\ D_0\, e^{-\alpha t^{\beta}}\left(\frac{t^2}{2} + \frac{\alpha t^{\beta+2}}{\beta+2} + \frac{\alpha^2\, t^{2\beta+2}}{2(2\beta+2)} + \right) \qquad 0 \le t \le t_1 \qquad (34)$$

Using boundary condition $Q(t_1) = L$ in (34) we get

$$L = (\gamma-1)\ D_0\, e^{-\alpha t_1^{\beta}}\left(\frac{t_1^2}{2} + \frac{\alpha t_1^{\beta+2}}{\beta+2} + \frac{\alpha^2\, t_1^{2\beta+2}}{2(2\beta+2)} + \right) \qquad (35)$$

Therefore the solution of equation (31) is given by

$$\frac{dQ(t)}{dt} + \alpha\beta\ t^{\beta-1}\ Q(t) = -D_0 t$$

$$e^{\alpha t^{\beta}}\ Q(t) = -\mathrm{D}_0 \int \mathrm{te}^{\alpha+\beta}\ \mathrm{dt} + \mathrm{C}$$

$$= -D_0 \int t\left(1+\alpha t^{\beta} + \frac{\alpha^2 t^{2\beta}}{2} + \right)dt + C$$

$$= -D_0\left(\frac{t^2}{2} + \frac{\alpha\ t^{\beta+2}}{\beta+2} + \frac{\alpha^2 t^{2\beta+2}}{2(2\beta+2)} + \right) + C$$

Using condition $Q(t_1) = L$

$$Le^{\alpha t_1 \beta} = -D_0\left(\frac{t_1^2}{2} + \frac{\alpha t_1^{\beta+2}}{\beta+2} + \frac{\alpha^2\, t_1^{2\beta+2}}{2(2\beta+2)} + \right) + C$$

$$C = Le^{\alpha t_1^{\beta}} + D_0\left(\frac{t_1^2}{2} + \frac{\alpha\, t_1^{\beta+2}}{\beta+2} + \frac{\alpha^2\, t_1^{2\beta+2}}{2(2\beta+2)} + \right)$$

$$e^{\alpha t^{\beta}}\ Q(t) = -D_0\left(\frac{t^2}{2} + \frac{\alpha t^{\beta+2}}{\beta+2} + \frac{\alpha^2\, t^{2\beta+2}}{2(2\beta+2)} + \right) + (\gamma-1)\ D_0 e\left(\frac{t_1^2}{2} + \frac{\alpha t_1^{\beta+2}}{(\beta+2)} + \frac{\alpha^2\, t_1^{2\beta+2}}{2(2\beta+2)} + \right)$$

$$e^{\alpha t^{\beta}}\ Q(t) = -D_0\left(\frac{t^2}{2} + \frac{\alpha t^{\beta+2}}{\beta+2} + \frac{\alpha^2\, t^{2\beta+2}}{2(2\beta+2)} + \right) + Le^{\alpha t_1^{\beta}} + D_0\left(\frac{t_1^2}{2} + \frac{\alpha t_1^{\beta+2}}{(\beta+2)} + \frac{\alpha^2\, t_1^{2\beta+2}}{2(2\beta+2t)}\right)$$

$$+ D_0\left(\frac{t_1^2}{2} + \frac{\alpha t_1^{\beta+2}}{(\beta+2)} + \frac{\alpha^2\, t_1^{2\beta+2}}{2(2\beta+2)} + \right)$$

$$Q(t) = -D_0 e^{-\alpha t^{\beta}}\left(\frac{t^2}{2} + \frac{2t^{\beta+2}}{(\beta+2)} + \frac{\alpha^2\, t^{2\beta+2}}{2(2\beta+2)} + \right) + \gamma D_0 e^{-\alpha t^{\beta}}\left(\frac{t_1^2}{2} + \frac{\alpha t_1^{\beta+2}}{(\beta+2)} + \frac{\alpha^2\, t_1^{2\beta+2}}{2(2\beta+2)} + \right),\ t_1 \le t \le \mu \qquad (36)$$

Using boundary condition $Q(t_2) = 0$, the solution of equation (32) is given by

$$\frac{\mathrm{d}Q(t)}{\mathrm{dt}} + \alpha\beta\, t^{\beta-1}\ Q(t) = -D_0\mu \qquad \mu \le t \le t_2$$

$$Q(t)\, e^{\alpha t^{\beta}} = -D_0\mu \int e^{\alpha+\beta}\ dt + C$$

$$= -D_0\mu\left(t + \frac{\alpha\, t^{\beta+1}}{(\beta+1)} + \frac{\alpha^2\, t^{2\beta+1}}{2(2\beta+1)} + \right) + C$$

$$\mathrm{C} = D_0\mu\left(t_2 + \frac{\alpha t_2^{\beta+1}}{(\beta+1)} + \frac{\alpha^2\, t_2^{2\beta+1}}{2(2\beta+1)} + \right)$$

$$Q(t)\, e^{\alpha t^{\beta}} = -D_0\mu\left(t + \frac{\alpha t^{\beta+1}}{(\beta+1)} + \frac{\alpha^2\, t^{2\beta+1}}{2(2\beta+1)} + \right) + D_0\mu\left(t_2 + \frac{\alpha t_2^{\beta+1}}{(\beta+1)} + \frac{\alpha^2\, t_2^{2\beta+1}}{2(2\beta+1)} + \right)$$

$$Q(t) = D_0 \mu e^{-\alpha t^{\beta}} \left((t_2 - t) + \frac{\alpha}{(\beta+1)} (t_2^{\beta+1} - t^{\beta+1}) + \frac{\alpha^2}{2(2\beta+1)} (t_2^{2\beta+1} - t^{2\beta+2}) + \right), \quad \mu \le t \le t_2 \tag{37}$$

The solution of equation (33) is given by

$\frac{dQ(t)}{dt} = -D_0 \mu e^{-\delta(T-t_2)}$ $\qquad t_2 \le t \le T$

with boundary condition $Q(t_2) = 0$

$Q(t) = -D_0 \mu \int [1 - \delta(T - t_2)] dt + c$

$Q(t) = -D_0 \mu [t - \delta(T - t_2)t] + c$

By using $Q(t_2)$ =0, we get

$$Q(t) = D_0 \mu[(t_2 - t) - \delta(T - t_2)(t_2 - t)] \tag{38}$$

$$\frac{dQ(t)}{dt} = -e^{-\delta} (T - t_2) D_0 \mu$$

Total inventory over the period $[0, t_2]$ is

$$\int_0^{t_2} Q(t) e^{-rt} dt = \int_0^{t_1} Q(t) e^{-rt} dt + \int_{t_1}^{\mu} Q(t) e^{-rt} dt + \int_{\mu}^{t_2} Q(t) e^{-rt} dt$$

$$\int_0^{t_1} Q(t) e^{-rt} dt = (\gamma - 1) D_0 \int_0^{t_1} e^{-\alpha t^{\beta}} \left(\frac{t^2}{2} + \frac{\alpha t^{\beta+2}}{\beta+2} + \frac{\alpha^2 t^{2\beta+2}}{2(2\beta+2)} + \right) e^{-rt} dt$$

$$\int_0^{t_1} Q(t) e^{-rt} dt = (\gamma - 1) D_0 \int_0^{t_1} (1 - \alpha t^{\beta}) \left(\frac{t^2}{2} + \frac{\alpha t^{\beta+2}}{\beta+2} + \frac{\alpha^2 t^{2\beta+2}}{2(2\beta+2)} + \right) e^{-rt} dt$$

$$= (\gamma - 1) D_0 \int_0^{t_1} \left(\frac{t^2}{2} - \frac{\alpha\beta t^{\beta+2}}{2(\beta+2)} - \frac{\alpha^2 (3\beta+2) t^{2\beta+2}}{2(2\beta+2)(\beta+2)} - \frac{rt^3}{2} + \frac{\alpha\beta r t^{\beta+3}}{2(\beta+2)} + \frac{r\alpha^2 (3\beta+2) t^{2\beta+3}}{2(2\beta+2)(\beta+2)} \right)$$

$$= (\gamma - 1) D_0 \left[\frac{t_1^3}{6} - \frac{\alpha\beta t_1^{\beta+3}}{2(\beta+2)(\beta+3)} - \frac{\alpha^2 (3\beta+2) t_1^{2\beta+3}}{2(2\beta+2)(\beta+2)(2\beta+3)} - \frac{rt_1^4}{8} + \frac{\alpha\beta r t_1^{\beta+4}}{2(\beta+2)(\beta+4)} + \frac{r\alpha^2 (3\beta+2) t_1^{2\beta+4}}{2(2\beta+2)(\beta+2)(2\beta+4)} \right]$$

$$\int_{t_1}^{\mu} Q(t) e^{-rt} dt = D_0 \int_{t_1}^{\mu} e^{-\alpha t^{\beta}} \left(\frac{\gamma t_1^2}{2} + \frac{\alpha\gamma t_1^{\beta+2}}{(\beta+2)} + \frac{\gamma\alpha^2 t_1^{2\beta+2}}{2(2\beta+2)} - \frac{t^2}{2} - \frac{\alpha t^{\beta+2}}{(\beta+2)} - \frac{\alpha^2 t^{2\beta+2}}{2(2\beta+2)} \right) (1 - rt)$$

$$= D_0 \int_{t_1}^{\mu} \left[\frac{\gamma t_1^2}{2} + \frac{\alpha\gamma t_1^{\beta+2}}{(\beta+2)} + \frac{\gamma\alpha^2 t_1^{2\beta+2}}{2(2\beta+2)} - \frac{t^2}{2} - \frac{\alpha t^{\beta+2}}{(\beta+2)} - \frac{\alpha^2 t^{2\beta+2}}{2(2\beta+2)} - \frac{\alpha\gamma t^{\beta} t_1^2}{2} - \frac{\alpha^2\gamma t^{\beta} t_1^{\beta+1}}{(\beta+2)} + \right.$$

$$\frac{\alpha t^{\beta+2}}{2} + \frac{\alpha^2 t^{2\beta+2}}{(\beta+2)} - \frac{r\gamma t_1^2 t}{2} - \frac{r\alpha\gamma t_1^{\beta+2}}{(\beta+2)} - \frac{r\gamma\alpha^2 t_1^{2\beta+2} t}{2(2\beta+2)} + \frac{rt^3}{2} + \frac{r\alpha t^{\beta+3}}{(\beta+2)} + \frac{r\alpha^2 t^{2\beta+3}}{2(2\beta+2)} +$$

$$\left. \frac{r\alpha\gamma t^{\beta+1} t_1^2}{2} + \frac{r\alpha^2\gamma t^{\beta+1} t_1^{\beta+2}}{(\beta+2)} - \frac{r\alpha t^{\beta+3}}{2} - \frac{r\alpha^2 t^{2\beta+3}}{(\beta+2)} \right]$$

$$= D_0 \left[\frac{\gamma\mu t_1^2}{2} - \frac{\gamma t_1^3}{2} + \frac{\alpha\gamma t_1^{\beta+2}\mu}{(\beta+2)} + \frac{\alpha\beta\gamma t_1^{\beta+3}}{2(\beta+2)(\beta+1)} + \frac{\alpha^2\gamma t_1^{2\beta+2}\mu}{2(2\beta+2)} - \frac{\alpha\beta t_1^{\beta+3}}{2(\beta+2)(\beta+3)} - \frac{\mu^3}{6} + \frac{t_1^3}{6} + \right.$$

$$\frac{\alpha\beta\mu^{\beta+3}}{2(\beta+2)(\beta+3)} + \frac{\alpha^2(3\beta+2)\mu^{2\beta+3}}{2(2\beta+2)(2\beta+3)(\beta+2)} - \frac{\alpha\gamma\mu^{\beta+1} t_1^2}{2(\beta+1)} - \frac{\alpha^2\gamma\mu^{\beta+1} t_1^{\beta+2}}{(\beta+1)(\beta+2)}$$

$$- \frac{\alpha^2(3\beta+2) t_1^{2\beta+3}}{2(2\beta+2)(2\beta+3)(\beta+2)} - \frac{\alpha^2\gamma(\beta-2) t_1^{2\beta+3}}{4(\beta+1)(\beta+2)} - \frac{r\gamma t_1^2 \mu^2}{4} + \frac{r\gamma t_1^4}{4} - \frac{r\alpha\gamma t_1^{\beta+2}\mu^2}{2(\beta+2)} + \frac{r\alpha\gamma t_1^{\beta+4}}{2(\beta+2)}$$

$$- \frac{r\gamma\alpha^2 t_1^{2\beta+2}\mu^2}{4(2\beta+2)} + \frac{r\gamma\alpha^2 t_1^{2\beta+4}}{4(2\beta+2)} + \frac{r\mu^4}{8} - \frac{rt_1^4}{8} + \frac{r\alpha\mu^{\beta+4}}{(\beta+2)(\beta+4)} - \frac{r\alpha t_1^{\beta+4}}{(\beta+2)(\beta+4)} + \frac{r\alpha^2\mu^{2\beta+4}}{8(\beta+1)(\beta+2)}$$

$$- \frac{r\alpha^2 t_1^{2\beta+4}}{8(\beta+1)(\beta+2)} + \frac{r\alpha\gamma\mu^{\beta+2} t_1^2}{2(\beta+2)} - \frac{r\alpha\gamma t_1^{\beta+4}}{2(\beta+2)} - \frac{r\gamma\alpha^2 t_1^{2\beta+4}}{(\beta+2)^2} + \frac{r\gamma\alpha^2 t_1^{\beta+2}\mu^{\beta+2}}{(\beta+2)^2} - \frac{r\alpha\mu^{\beta+4}}{2(\beta+4)} + \frac{r\alpha t_1^{\beta+4}}{2(\beta+4)} - \frac{r\alpha^2\mu^{2\beta+4}}{(2\beta+4)(\beta+2)} +$$

$$\left. \frac{r\alpha^2 t_1^{2\beta+4}}{(2\beta+4)(\beta+2)} \right]$$

$$=D_0\ [\frac{\gamma\mu t_1^2}{2}-\frac{\gamma t_1^3}{2}+\frac{\alpha\mu\gamma t_1^{\beta+2}}{(\beta+2)}-\frac{\alpha\gamma\beta t_1^{\beta+3}}{2(\beta+2)(\beta+1)}+\frac{\alpha^2\gamma t_1^{2\beta+2}\mu}{2(2\beta+2)}-\frac{\alpha\beta t_1^{\beta+3}}{2(\beta+2)(\beta+3)}-\frac{\mu^3}{6}+\frac{t_1^3}{6}+\frac{\alpha\beta\mu^{\beta+3}}{2(\beta+2)(\beta+3)}+\frac{\alpha^2(3\beta+2)\mu^{2\beta+3}}{2(2\beta+2)(2\beta+3)(\beta+2)}-\frac{\alpha\gamma\mu^{\beta+1}t_1^2}{2(\beta+1)}-$$
$$\frac{\alpha^2\gamma\mu^{\beta+1}t_1^{\beta+2}}{(\beta+2)(\beta+2)}-\frac{\alpha^2(3\beta+2)t_1^{2\beta+3}}{2(2\beta+2)(2\beta+3)(\beta+2)}-\frac{\alpha^2\gamma(\beta-2)t_1^{2\beta+3}}{2(\beta+1)(\beta+2)}-\frac{r\gamma\mu^2t_1^2}{4}+\frac{r\gamma t_1^4}{4}-\frac{r\alpha\gamma t_1^{\beta+2}\mu^2}{2(\beta+2)}-\frac{r\gamma\alpha^2t_1^{2\beta+2}\mu^2}{8(\beta+1)}+\frac{r\gamma\alpha^2(\beta^2-4\beta-4)t_1^{2\beta+4}}{8(\beta+1)\,(\beta+2)^2}+\frac{r\mu^4}{8}-$$
$$\frac{rt_1^4}{8}-\frac{r\alpha\beta\mu^{\beta+4}}{2(\beta+2)(\beta+4)}+\frac{r\alpha\beta t_1^{\beta+4}}{2(\beta+2)(\beta+4)}-\frac{r\alpha^2(3\beta+2)\mu^{2\beta+4}}{8(\beta+1)(\beta+2)^2}+\frac{r\alpha^2(3\beta+2)t_1^{2\beta+4}}{8(\beta+2)^2(\beta+1)}+\frac{r\alpha\gamma\mu^{\beta+2}t_1^2}{2(\beta+2)}+\frac{r\alpha^2\gamma t_1^{\beta+2}\ \mu^{\beta+2}}{(\beta+2)^2}]$$

$$\int_{\mu}^{t_2} Q(t)\,e^{-rt}\,dt\ = D_0\mu\int_{\mu}^{t_2} e^{-\alpha t^\beta}\,[(t_2-t_1)\frac{\alpha}{(\beta+1)}\left(t_2^{\beta+1}-t^{\beta+1}\right)+\frac{\alpha^2}{2(2\beta+1)}\left(t_2^{2\beta+1}-t^{2\beta+1}\right)]\,e^{-rt}\,dt$$

$$=D_0\mu\int_{\mu}^{t_2}\left(1-\alpha t^\beta\right)[(t_2-t_1)\frac{\alpha}{(\beta+1)}\left(t_2^{\beta+1}-t^{\beta+1}\right)+\frac{\alpha^2}{2(2\beta+1)}\left(t_2^{2\beta+1}-t^{2\beta+1}\right)]\ e^{-rt}\,dt$$

$$=D_0\mu\int_{\mu}^{t_2}[(t_2-t)+\frac{\alpha}{(\beta+1)}\left(t_2^{\beta+1}-t^{\beta+1}\right)+\frac{\alpha^2}{2(2\beta+1)}\left(t_2^{2\beta+1}-t^{2\beta+1}\right)-\alpha t^\beta t_2+\alpha t^{\beta+1}-\frac{\alpha^2}{(\beta+1)}t^\beta t_2^{\beta+1}+\frac{\alpha^2}{(\beta+1)}t^{2\beta+1}-$$
$$rt(t_2-t)-\frac{r\alpha}{(\beta+1)}\ (t_2^{\beta+1}-t^{\beta+2})\frac{r\alpha^2}{2(2\beta+1)}\left(t_2^{2\beta+1}t-t^{2\beta+2}\right)+\alpha\gamma t^{\beta+1}t_2-r\alpha t^{\beta+2}+\frac{\gamma\alpha^2}{(\beta+1)}t^{\beta+1}t_2^{\beta+1}-\frac{\gamma\alpha^2}{(\beta+1)}t^{2\beta+2}]$$

$$=D_0\mu[t_2^2-t_2\,\mu-\frac{t_2^2}{2}+\frac{\mu^2}{2}+\frac{\alpha}{(\beta+1)}\left(t_2^{\beta+2}-t_2^{\beta+1}\mu-\frac{t_2^{\beta+2}}{\beta+2}+\frac{\mu^{\beta+2}}{(\beta+2)}\right)+\frac{\alpha^2}{2(2\beta+1)}\left(t_2^{2\beta+2}-t_2^{2\beta+1}\mu-\frac{t_2^{2\beta+2}}{(2\beta+2)}+\frac{\mu^{2\beta+2}}{(2\beta+2)}\right)-\frac{\alpha t_2^{\beta+2}}{(\beta+1)}+$$
$$\frac{\alpha\mu^{\beta+1}t_2}{(\beta+1)}+\frac{\alpha t_2^{\beta+2}}{(\beta+2)}-\frac{\alpha\mu^{\beta+2}}{(\beta+2)}-\frac{\alpha^2t_2^{2\beta+2}}{(\beta+1)^2}+\frac{\alpha^2\mu^{\beta+1}t_2^{\beta+1}}{(\beta+1)^2}+\frac{\alpha^2t_2^{2\beta+2}}{(\beta+1)(2\beta+2)}-\frac{\alpha^2\mu^{2\beta+1}}{(\beta+1)(2\beta+2)}-\frac{rt_2^3}{2}+\frac{r\mu^2t_2}{2}+\frac{rt_2^3}{3}-\frac{r\mu^3}{3}-\frac{r\alpha}{(\beta+1)}\left(\frac{t_2^{\beta+3}}{2}-\right.$$
$$\left.t_2^{\beta+1}\frac{\mu^2}{2}-\frac{t_2^{\beta+3}}{\beta+3}+\frac{\mu^{\beta+3}}{(\beta+3)}\right)-\frac{r\alpha^2}{2(2\beta+1)}\left(\frac{t_2^{2\beta+3}}{2}-t_2^{2\beta+1}\frac{\mu^2}{2}-\frac{t_2^{2\beta+3}}{2\beta+3}+\frac{\mu^{2\beta+3}}{2\beta+2}\right)+\frac{\alpha rt_2^{\beta+3}}{\beta+2}-\frac{\alpha r\mu^{\beta+2}t_2}{(\beta+2)}-\frac{r\alpha^2(\beta-2)t_2^{2\beta+3}}{4(\beta+2)(2\beta+3)}+\frac{r\alpha^2\mu^2t_2^{2\beta+1}}{4(2\beta+1)}+\frac{r\alpha^2(3\beta+1)\mu^{2\beta+3}}{2(\beta+1)(2\beta+3)(2\beta+1)}]$$

$$=D_0\mu[(\frac{t_2^2}{2}-t_2\,\mu+\frac{\mu^2}{2})-\frac{\alpha\mu t_2^{\beta+1}}{(\beta+1)}+\frac{\alpha\beta t_2^{\beta+2}}{(\beta+1)(\beta+2)}-\frac{\alpha\beta\mu^{\beta+2}}{(\beta+1)(\beta+2)}+\frac{\alpha^2(\beta-1)t_2^{2\beta+2}}{2(2\beta+2)(\beta+1)}-\frac{\alpha^2\mu t_2^{2\beta+1}}{2(2\beta+1)}-\frac{\alpha^2(3\beta+1)\mu^{2\beta+2}}{2(2\beta+1)(2\beta+2)(\beta+1)}+\frac{\alpha^2\mu^{\beta+1}t_2^{\beta+1}}{(\beta+1)^2}+$$
$$\frac{\alpha\mu^{\beta+1}t_2}{(\beta+1)}-\frac{rt_2^3}{6}+\frac{r\mu^2t_2}{2}-\frac{r\mu^3}{3}-\frac{\alpha\beta rt_2^{\beta+3}}{2(\beta+2)(\beta+3)}+\frac{r\alpha\mu^2t_2^{\beta+1}}{2(\beta+1)}+\frac{\alpha r\beta\mu^{\beta+3}}{(\beta+1)(\beta+3)}-\frac{r\alpha\mu^{\beta+2}t_2}{(\beta+2)}-\frac{r\alpha^2(\beta-2)t_2^{2\beta+3}}{4(\beta+2)(2\beta+3)}+\frac{r\alpha^2\mu^2t_2^{2\beta+1}}{4(2\beta+1)}+\frac{r\alpha^2(3\beta+1)\mu^{2\beta+3}}{2(\beta+1)(2\beta+3)(2\beta+1)}]$$

Total inventory over the period $[0, t_2]$ *is* given by

$$\int_{\circ}^{t_2} Q(t)e^{-rt}\,dt=(\gamma-1)\,D_0\left[\frac{t_1^3}{6}-\frac{\alpha\beta t_1^{\beta+3}}{2(\beta+2)(\beta+3)}-\frac{\alpha^2(3\beta+2)t_1^{2\beta+3}}{2(2\beta+2)(\beta+2)(2\beta+3)}-\frac{rt_1^4}{8}+\frac{\alpha\beta rt_1^{\beta+4}}{2(\beta+2)(\beta+4)}+\frac{r\alpha^2(3\beta+2)t_1^{2\beta+4}}{2(2\beta+2)(\beta+2)(2\beta+4)}\right]+D_0\left[\frac{\gamma\mu t_1^2}{2}-\right.$$
$$\frac{\gamma t_1^3}{2}+\frac{\alpha\gamma t_1^{\beta+2}\mu^2}{\beta+2}-\frac{\alpha\gamma\beta t_1^{\beta+3}}{2(\beta+2)(\beta+1)}+\frac{\alpha^2\gamma t_1^{2\beta+2}\mu}{2(2\beta+2)}-\frac{\alpha\beta t_1^{\beta+3}}{2(\beta+2)(\beta+3)}-\frac{\mu^3}{6}+\frac{t_1^3}{6}+\frac{\alpha\beta\mu^{\beta+3}}{2(\beta+2)(\beta+3)}+\frac{\alpha^2(3\beta+2)\mu^{2\beta+3}}{2(2\beta+2)(2\beta+3)(\beta+2)}-\frac{\alpha\gamma\mu^{\beta+1}t_1^2}{2(\beta+1)}-$$
$$\frac{\alpha^2\gamma\mu^{\beta+1}t_1^{\beta+2}}{(\beta+1)(\beta+2)}-\frac{\alpha^2(3\beta+2)t_1^{2\beta+3}}{2(2\beta+2)(2\beta+3)(\beta+2)}-\frac{\alpha^2\gamma(\beta-2)t_1^{2\beta+3}}{4(\beta+1)(\beta+2)}-\frac{r\gamma t_1^2\mu^2}{4}+\frac{r\gamma t_1^4}{4}-\frac{r\alpha\gamma t_1^{\beta+2}\mu^2}{2(\beta+2)}-\frac{r\gamma\alpha^2t_1^{2\beta+2}\mu^2}{8(\beta+1)}+\frac{r\gamma\alpha^2(\beta^2-4\beta-4)t_1^{2\beta+4}}{8(\beta+1)(\beta+2)^2}+\frac{r\mu^4}{8}-\frac{rt_1^4}{8}-$$
$$\frac{r\alpha\beta\mu^{\beta+4}}{2(\beta+2)(\beta+4)}+\frac{r\alpha\beta t_1^{\beta+4}}{2(\beta+2)(\beta+4)}-\frac{r\alpha^2(3\beta+2)\mu^{2\beta+4}}{8(\beta+1)(\beta+2)^2}+\frac{r\alpha^2(3\beta+2)\,t_1{}^{2\beta+4}}{8(\beta+2)^2(\beta+1)}+\frac{r\alpha\gamma\mu^{\beta+2}t_1^2}{2(\beta+2)}+\frac{r\alpha^2\gamma t_1^{\beta+2}\mu^{\beta+2}}{(\beta+2)^2}\Big]+D_0\mu\left[\frac{t_2^2}{2}-t_2\,\mu+\frac{\mu^2}{2}\right]-\frac{\alpha\mu t_2^{\beta+1}}{(\beta+1)}+$$
$$\frac{\alpha\beta t_2^{\beta+2}}{(\beta+1)(\beta+2)}-\frac{\alpha\beta\mu^{\beta+2}}{(\beta+1)(\beta+2)}+\frac{\alpha\mu^{\beta+1}t_2}{(\beta+1)}+\frac{\alpha^2(\beta-1)t_2^{2\beta+2}}{2(2\beta+2)(\beta+1)}-\frac{\alpha^2\mu t_2{}^{2\beta+1}}{2(2\beta+1)}-\frac{\alpha^2(3\beta+1)\mu^{2\beta+2}}{2(2\beta+1)(2\beta+2)(\beta+1)}+\frac{\alpha^2\mu^{\beta+1}t_2^{\beta+1}}{(\beta+1)^2}-\frac{rt_2^3}{6}+\frac{r\mu^2t_2}{2}-\frac{r\mu^3}{3}-$$
$$\frac{\alpha\beta rt_2^{\beta+3}}{2(\beta+2)(\beta+3)}+\frac{r\alpha\mu^2t_2^{\beta+1}}{2(\beta+1)}+\frac{\alpha r\beta\mu^{\beta+3}}{(\beta+1)(\beta+3)}-\frac{r\alpha\mu^{\beta+2}t_2}{(\beta+2)}-\frac{r\alpha^2(\beta-2)t_2^{2\beta+3}}{4(\beta+2)(2\beta+3)}+\frac{r\alpha^2\mu^2t_2^{2\beta+1}}{4(2\beta+1)}+\frac{r\alpha^2(3\beta+1)\mu^{2\beta+3}}{2(\beta+1)(2\beta+3)(2\beta+1)}]\quad(39)$$

The number of deteriorated items over the period $[0,\ t_2]$ is given by

Production in $[0, t_1]$ - Demand in $[0, \mu]$ – Demand in $[0, t_2]$

$$=\gamma\ D_0\int_0^{t_1} te^{-rt}\,dt-\ D_0\int_0^{\mu} te^{-rt}\,dt-\quad D_0\mu\int_{\mu}^{t_2} e^{-rt}\,dt$$

$$=\gamma\ D_0\ (\frac{t_1^2}{2}-\frac{rt_1^3}{3})-\ D_0\ (\frac{\mu^2}{2}-\frac{r\mu^3}{3})-D_0\mu\,[\ (t_2-\mu)-r\ (\frac{t_2^2}{2}-\frac{\mu^2}{2})]\qquad(40)$$

Hence the production cost over the period $[0, t_1)$]is given by

$$\int_0^{t_1} Kve^{-ru}\,du\ =\int_0^{t_1}\frac{\alpha_1\gamma}{R^{s-1}}\,e^{-ru}\,\mathrm{du}$$

$$=\ \alpha_1\gamma\ \ D_0^{1-s}\ \int_0^{t_1} u^{1-s}\ (1\text{-ru})\ \mathrm{du}$$

$$=\ \alpha_1\gamma\ \ D_0^{1-s}\ [\frac{t_1^{2-s}}{(2-s)}-\frac{rt_1^{1-s}}{(1-s)}]\qquad(41)$$

\Shortage cost over the period [0, T] is given by

$$\int_{t_2}^{t}\theta(t)dt = -\int_{t_2}^{T} D_0\,\mu[\{t_2-t)-\delta[T-t_2](t_2-t)]dt\,e^{-rt}$$
$$= -D_0\,\mu\left[\left\{Tt_2-\frac{t_2^2}{2}-\frac{T^2}{2}-\delta\left(\frac{3t_2T^2}{2}-\frac{3t_2^2T}{2}-\frac{T^2}{2}+\frac{t_2^3}{2}\right)-r\frac{T^2t_2}{2}+r\frac{t_2^3}{2}+r\frac{T^3}{3}+\delta r\,(T-t_2)\left(\frac{t_2T^2}{2}-\frac{t_2^3}{6}-\frac{T^3}{3}\right)\right\}\right] \quad (42)$$

Lost sale cost per cycle is

LS= $D_0\,\mu\int_{t_2}^{T}(1-e^{-\delta(T-t_2)})\,dt$

LS= $D_0\,\mu\int_{t_2}^{T}\delta(T-t_2)(t_2-t)\,dT$

sale cost over the period [0,T] is

$$= D_0\,\mu\delta\left[\frac{3T^2t_2}{2}-\frac{3t_2^2T}{2}-\frac{T^3}{2}+\frac{t_2^3}{2}-\frac{5rt_2T^3}{6}-\frac{rt_2^4}{6}+\frac{rTt_2^3}{6}+\frac{rT^4}{3}+\frac{rt_2^2T^2}{2}\right] \quad (43)$$

From (39)(40),(41),(42),(43), the total average inventory cost X of the system is Lost

$$X=\frac{1}{t_2}[c_1\{D_0(\gamma-1)[\frac{t_1^3}{6}-\frac{\alpha\beta t_1^{\beta+3}}{2(\beta+2)(\beta+3)}-\frac{\alpha^2(3\beta+2)t_1^{2\beta+3}}{2(2\beta+2)(\beta+2)(2\beta+3)}-\frac{rt_1^4}{8}+\frac{\alpha\beta rt_1^{\beta+4}}{2(\beta+2)(\beta+4)}+\frac{r\alpha^2(3\beta+2)t_1^{2\beta+4}}{2(2\beta+2)(\beta+2)(2\beta+4)}]+D_0[\frac{\gamma\mu t_1^2}{2}-\frac{\gamma t_1^3}{2}+$$
$$\frac{\alpha\gamma t_1^{\beta+2}\mu^2}{\beta+2}-\frac{\alpha\gamma\beta t_1^{\beta+3}}{2(\beta+2)(\beta+1)}+\frac{\alpha^2\gamma t_1^{2\beta+2}\mu}{2(2\beta+2)}-\frac{\alpha\beta t_1^{\beta+3}}{2(\beta+2)(\beta+3)}-\frac{\mu^3}{6}+\frac{t_1^3}{6}+\frac{\alpha\beta\mu^{\beta+3}}{2(\beta+2)(\beta+3)}+\frac{\alpha^2(3\beta+2)\mu^{2\beta+3}}{2(2\beta+2)(2\beta+3)(\beta+2)}-\frac{\alpha\gamma\mu^{\beta+1}t_1^2}{2(\beta+1)}-\frac{\alpha^2\gamma\mu^{\beta+1}t_1^{\beta+2}}{(\beta+1)(\beta+2)}-$$
$$\frac{\alpha^2(3\beta+2)t_1^{2\beta+3}}{2(2\beta+2)(2\beta+3)(\beta+2)}-\frac{\alpha^2\gamma(\beta-2)t_1^{2\beta+3}}{4(\beta+1)(\beta+2)}-\frac{r\gamma t_1^2\mu^2}{4}+\frac{r\gamma t_1^4}{4}-\frac{r\alpha\gamma t_1^{\beta+2}\mu^2}{2(\beta+2)}-\frac{r\gamma\alpha^2t_1^{2\beta+2}\mu^2}{8(\beta+1)}+\frac{r\gamma\alpha^2(\beta^2-4\beta-4)t_1^{2\beta+4}}{8(\beta+1)(\beta+2)^2}+\frac{r\mu^4}{8}-\frac{r\alpha\beta\mu^{\beta+4}}{2(\beta+2)(\beta+4)}+$$
$$\frac{r\alpha\beta t_1^{\beta+4}}{2(\beta+2)(\beta+4)}-\frac{r\alpha^2(3\beta+2)\mu^{2\beta+4}}{8(\beta+1)(\beta+2)^2}+\frac{r\alpha^2(3\beta+2)\,t_1{}^{2\beta+4}}{8(\beta+2)^2(\beta+1)}+\frac{r\alpha\gamma\mu^{\beta+2}t_1^2}{2(\beta+2)}+\frac{r\alpha^2\gamma t_1^{\beta+2}\mu^{\beta+2}}{(\beta+2)^2}]+D_0\mu[\frac{t_2^2}{2}-t_2\,\mu+\frac{\mu^2}{2}-\frac{\alpha\mu t_2^{\beta+1}}{(\beta+1)}+\frac{\alpha\beta t_2^{\beta+2}}{(\beta+1)(\beta+2)}-$$
$$\frac{\alpha\beta\mu^{\beta+2}}{(\beta+1)(\beta+2)}+\frac{\alpha\mu^{\beta+1}t_2}{(\beta+1)}+\frac{\alpha^2(\beta-1)t_2^{2\beta+2}}{2(2\beta+2)(\beta+1)}-\frac{\alpha^2\mu t_2{}^{2\beta+1}}{2(2\beta+1)}-\frac{\alpha^2(3\beta+1)\mu^{2\beta+2}}{2(2\beta+1)(2\beta+2)(\beta+1)}+\frac{\alpha^2\mu^{\beta+1}t_2{}^{\beta+1}}{(\beta+1)^2}-\frac{rt_2^3}{6}+\frac{r\mu^2t_2}{2}-\frac{r\mu^3}{3}-\frac{\alpha\beta rt_2^{\beta+3}}{2(\beta+2)(\beta+3)}+$$
$$\frac{r\alpha\mu^2t_2^{\beta+1}}{2(\beta+1)}+\frac{\alpha r\beta\mu^{\beta+3}}{(\beta+1)(\beta+3)}-\frac{r\alpha\mu^{\beta+2}t_2}{(\beta+2)}-\frac{r\alpha^2(\beta-2)t_2^{2\beta+3}}{4(\beta+2)(2\beta+3)}+\frac{r\alpha^2\mu^2t_2^{2\beta+1}}{4(2\beta+1)}+\frac{r\alpha^2(3\beta+1)\mu^{2\beta+3}}{2(\beta+1)(2\beta+3)(2\beta+1)}\}+c_3\{\gamma D_0(\frac{t_1^2}{2}-\frac{rt_1^3}{3})-D_0(\frac{\mu^2}{2}-\frac{r\mu^3}{3})-$$
$$D_0\mu((t_2-\mu)-r\left(\frac{t_2^2}{2}-\frac{\mu^2}{2}\right))\}-c_4D_0\,\mu[\left\{Tt_2-\frac{t_2^2}{2}-\frac{T^2}{2}-\delta\left(\frac{3t_2T^2}{2}-\frac{3t_2^2T}{2}-\frac{T^3}{2}+\frac{t_2^3}{2}\right)-r\frac{T^2t_2}{2}+r\frac{t_2^3}{2}+r\frac{T^3}{3}+\delta r\,(T-$$
$$t_2)\left(\frac{t_2T^2}{2}-\frac{t_2^3}{6}-\frac{T^3}{3}\right)\right\}+c_5\,D_0\,\mu\delta\{\frac{3T^2t_2}{2}-\frac{3t_2^2T}{2}-\frac{T^3}{2}+\frac{t_2^3}{2}-\frac{5rt_2T^3}{6}-\frac{rt_2^4}{6}+\frac{rTt_2^3}{6}+\frac{rT^4}{3}+\frac{rt_2^2T^2}{2}\}+\alpha_1\gamma\;D_0^{1-s}\{\frac{t_1^{2-s}}{(2-s)}-\frac{rt_1^{1-s}}{(1-s)}\}] \quad (44)$$

Optimum values of t_1 *and* t_2 for minimum average cost are obtained as in Case 1 which gives

$$C_1\left\{(\gamma-1)D_o\left(\frac{t_1^2}{2}-\frac{\alpha\beta t_1^{\beta+2}}{2(\beta+2)}\right)-\frac{\alpha^2(3\beta+2)t_1^{2\beta+2}}{2(2\beta+2)(\beta+2)}-\frac{rt_1^3}{2}+\frac{\alpha\beta r(3\beta+2)t_1^{\beta+3}}{2(\beta+2)}+\frac{r\alpha^2(3\beta+2)t_1^{2\beta+3}}{2(2\beta+2)(\beta+2)}\right)+D_o\left(\gamma\mu t_1-\frac{3}{2}\gamma t_1^2+\alpha\gamma\mu^2t_1^{\beta+1}-\right.$$
$$\frac{\alpha\gamma\beta(\beta+3)t_1^{\beta+2}}{2(\beta+2)(\beta+1)}+\frac{\alpha^2\gamma t_1^{2\beta+1}\mu}{2}-\frac{\alpha\beta t_1^{\beta+2}}{2(\beta+2)}+\frac{t_1^2}{2}-\frac{\alpha\gamma\mu^{\beta+1}t_1}{(\beta+1)}-\frac{\alpha^2\gamma\mu^{\beta+1}t_1^{\beta+1}}{(\beta+1)}-\frac{\alpha^2(3\beta+2)t_1^{2\beta+2}}{2(2\beta+2)(\beta+2)}-\frac{\alpha^2\gamma(\beta-2)t_1^{2\beta+2}(2\beta+3)}{4(\beta+1)(\beta+2)}-\frac{r\gamma t_1{}_1\mu^2}{2}+r\gamma t_1^3-$$
$$\left.\frac{r\alpha\gamma\mu^2t_1^{\beta+1}}{2}-\frac{r\gamma\alpha^2\mu^2(2\beta+2)t_1^{2\beta+1}}{8(\beta+1)}+\frac{r\gamma\alpha^2(\beta^2-4\beta-4)t_1^{2\beta+3}}{4(\beta+1)(\beta+2)}-\frac{rt_1^2}{2}+\frac{r\alpha\beta t_1^{\beta+3}}{2(\beta+2)}+\frac{r\alpha^2(3\beta+2)t_1^{2\beta+3}}{4(\beta+2)(\beta+1)}+\frac{r\alpha\gamma\mu^{\beta+2}t_1}{(\beta+2)}+\frac{r\alpha^2\gamma t_1^{\beta+1}\mu^{\beta+2}}{(\beta+2)}\right\}+$$
$$C_3\gamma D_o(t_1-rt_1^2)+\alpha_1\gamma D_o^{1-s}(t_1^{1-s}-rt_1-s)=0 \quad (45)$$

and

$$C_1D_0\mu\left(t_2-\mu-\alpha\mu t_2^{\beta}+\frac{\alpha\beta t_2^{\beta+1}}{(\beta+1)}+\frac{\alpha\mu^{\beta+1}}{(\beta+1)}+\frac{\alpha^2(\beta-1)t_2^{2\beta+1}}{2(\beta+1)}-\frac{\alpha^2\mu t_2^{2\beta}}{2}+\frac{\alpha^2\mu^{\beta+1}t_2^{\beta}}{(\beta+1)}-\frac{rt_2^2}{2}+\frac{r\mu^2}{2}-\frac{\alpha\beta rt_2^{\beta+2}}{2(\beta+2)}+\frac{r\alpha\mu^2t_2^{\beta}}{2}-\frac{r\alpha\mu^{\beta+2}}{(\beta+2)}-\right.$$
$$\left.\frac{r\alpha^2(\beta-2)t_2^{2\beta+2}}{4(\beta+2)}+\frac{r\alpha^2\mu^2t_2^{2\beta}}{4}\right)-C_3D_0\mu(1-rt_2)-C_4\;D_0\mu\left\{(T-t_2)-\delta\left(\frac{3T^2}{2}-3t_2T+\frac{3t_2^2}{2}\right)-\frac{rT^2}{2}+\frac{3}{2}rt_2^2+\frac{2}{3}\delta rt_2^3-\right.$$
$$\left.\delta rt_2T^2+\frac{5}{6}\delta rT^3-\frac{1}{2}\delta rTt_2^2\right\}+C_5D_0\mu\delta\left(\frac{3}{2}T^2-3t_2T+\frac{3}{2}t_2^2-\frac{5rT^3}{6}-\frac{2}{3}rt_2^3+\frac{rTt_2^2}{2}+rt_2T^2\right)-X=0 \quad (46)$$

Numerical Examples

Lets us consider the inventory system with following data:

For case I ($\boldsymbol{\mu \le t_1 \le t_2}$)

$D_0 = 15,\ s = 1.6,\ \mu = 2,\ \alpha_1 = 2,\ \beta = 0.06,\ \alpha = 0.08,\ \gamma = 2,\ r = 0.03,\ c_1 = 4,\ C_4 = 18,\ C_5 = 1.2,\ \ C_3 = 2.5,\ \delta = 0.5\ and\ T = 5$

output results are

$$t_1 = 2.20206, \qquad t_2 = 3.88174,\ \ T.C = 133.177$$

Graphical representation of the converities of $\boldsymbol{t_1}$ ***and*** $\boldsymbol{t_2}$ ***w.r.t. T.C for case*** $\boldsymbol{1.(\mu \le t_1 \le t_2)}$

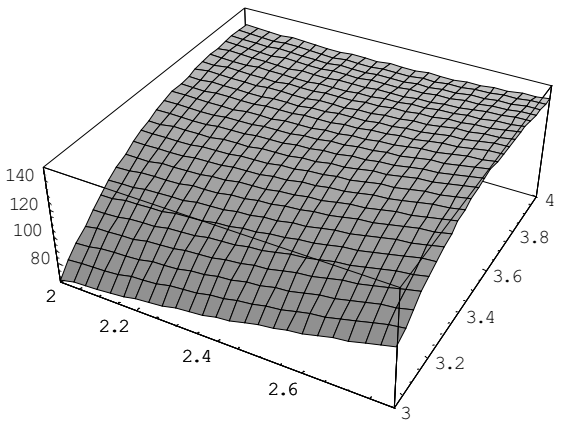


Figure 1. Convexity of t_1 and t_2 $w.r.t$ $T.C$

Table 1. Sensitivity Analysis: The Sensitivity analysis of the key parameter s, r, α, D_0 are given in the below Table for case I.

Parameters		t_1	t_2	**T.C**
r	0.031	2.21034	3.87322	138.003
	0.032	2.21401	3.87539	139.89
	0.033	2.21773	3.87758	141.777
	0.034	2.22149	3.87979	143.665
α	0.09	2.24434	3.8.4784	133.39
	0.10	2.35915	3.79136	125.091
	0.11	2.28212	3.82882	130.644
	0.12	2.32042	3.81021	127.878
D_0	16	2.20702	3.86844	144.888
	17	2.2069	3.86943	153.663
	18	2.20679	3.8703	162.442
	19	2.2067	3.87107	171.223
S	1.3	2.20762	3.87302	134.531
	1.4	2.20716	3.87604	133.797
	1.5	2.20658	3.87786	133.4
	1.6	2.20605	3.879	133.176

The following points are observed

1. t_1 & t_2 increase and T.C also increase with the increase in value of the parameter r
2. t_1 increases while t_2 & T.C decrease with the increase in value of the parameter α.
3. t_1 decreases while t_2 & T.C increase with the increase in value of the parameter Do.
4. t_1 & T.C decrease while t_2 increases with the increase in value of the parameter s.

For case 2 ($t_1 \leq \mu \leq t_2$)
$T = 3,\ s = 1.3,\ \mu = 0.9,\ \alpha_1 = 2.2,\ \beta = 0.04,\ \alpha = 0.08,\ \gamma = 2.2,\ r = 0.02,\ D_0 = 14,\ C_1 = 2,\ C_4 = 4,\ C_5 = 0.4,\ C_3 = 2.4,\ \delta = 0.5,$

Output results are
t_1= 0.0158244, t_2 =2.666, T.C=94.5816

Conclusion

In this study, an EOQ model with ramp type demand rate and unit production cost under inflationary condition has been developed. The quality and quantity of goods decrease in course of time due to deterioration it is a natural phenomena .Hence consideration of Weibull distribution time varying deterioration function defines a significant meaning of perishable, volatile and failure of any kind of item. Shortages are allowed and partially backlogged. The another considered phenomenon viz

inflation plays an important role in realistic scenario. A mathematical model has been developed to determine the optimal ordering policy cost which minimizes the present worth of total optimal cost. Thus the model concludes with numerical examples.

Equation (24) and (25) are non- linear equation in t_1 and t_2. These simultaneous non-linear equations can be solved for suitable choice of the parameters c_1, c_3, c_4, c_5, $\alpha, \beta, r, \gamma, \mu, \delta$, D_0, α_1 and s ($\neq 2$). If t_1^* and t_2^* are the solution of (24) and (25) for Case I, the corresponding minimum cost $c^*(t_1, t_2)$ can be obtained from (23). It is very difficult to show analytically whether the cost function C (t_1, t_2) is convex. That is why, C (t_1, t_2) may not be global minimum. If C (t_1, t_2) is not convex, then C (t_1, t_2) will be local minimum.

Similarly, solution of equations (45) and (46) for Case II can be obtained corresponding minimum cost C (t_1, t_2) can be obtained from (44).

References

[1] Berrotoni, J.N. (1962), Practical application of Weibull distribution. In ASQC Technical Conference Transactions, PP.303-323.

[2] Covert, R.P., Philip, G.C. (1973), An EOQ model for items with Weibull distribution deterioration. AIIE Transactions, 5(4), 323-326.

[3] Deng, P.S., Lin, R.H.J., Chu, P.(2007), A note on the inventory models for deteriorating items with ramp type demand rate, European Journal of Operational Research ,178(1),112-120.

[4] Misra, R.B.(1975),Optimum production lot-size model for a system with deteriorating inventory, International Journal of Production Research, 13(5), 495-505.

[5] Hill, R.M. (1995), Inventory model for increasing demand followed by level demand, The Journal of the operational Research Society, 46,1250-1259.

[6] Wagner, H.M., Whitin, J.M. (1958), Dynamic version of economic lot size model, Management Science, 5(1), 89-96.

[7] Manna, S.K and Chaudhuri, K.S.(2006), An model with ramp type demand rate, time dependent deterioration rate, unit production cost and shortages, European Journal of Operational Research, 171 (2), 557-566.

[8] Giri, B.C., Jalan, A.K. and Chaudhuri, K.S.(2003), Economic order quantity model with Weibull deterioration distribution shortage and ramp type demand, International Journal of Systems Science, 34 (4), 237-243.

[9] Panda, S., Senapati, S., and Basu, M.(2008), Optimal replenishment policy for perishable seasonal products in a reason with ramp type time dependent demand, Computers and Industrial Engineering,54(2),301-314.

[10] Buzzacott, J.A. (1975), Economic order quantities with inflation, Operation Research Quarterly, 26(3), 553-558.

[11] Yang, H.L., Teng, J.T. and Chern, M.S. (2001), Deterministic inventory lot size model under inflation with shortages and deterioration for fluctuating demand, Naval Research Logistics, 48 (3), 144-158.

[12] Liao, H.C., Tsai, C.H, SU, C.T. (2000), An inventory model with deteriorating items under inflation when delay in Payment is permissible. International Journal of Production Economics , 63 (2),207-214.

[13] Roy, T., Chaudhuri, K.S. (2011), A finite time horizon EOQ model with ramp-type demand under inflation and time discounting. International Journal of Operational Research , 11(1) 100-118.

[14] Sarkar, B., Sana, S.S., Chaudhuri, K. (2011), An imperfect production process for time varying demand with inflation and time value of money an EMQ model, Expert Systems with Application,38 (11), 13543-13548.

[15] Abad, P.L. (1996), Optimal pricing and lot sizing under conditions of perishability and partial backlogging, Management Science, 42 (8) , 1093-1104.

[16] Chang, H.J. and Dye, C.Y.(1999), An EOQ model for deteriorating items with the varying demand and partial backlogging, Journal of the Operational Research Society, 50 (11), 1176-1182.

[17] Wang, S.P. (2002). , An inventory replenishment policy for deteriorating items with shortages and partial backlogging, Computers and Operational Research, 29, 2043-2051.

[18] Wu, K.S., Ouyang, L.Y. and Yang, C.T. (2006), An optimal replenishment policy for non-instantaneous deteriorating items with stock dependent demand and partial backlogging, International Journal of Production Economics, 101, 369-384.

[19] Singh, S.R. and Singh, T.J. (2007), An EOQ inventory model with Weibull distribution deterioration, ramp type demand and partial backlogging rate,Indian Journal Of Mathematics and Mathematical Science, 3(2),127-137.

[20] Dye, C.Y., Ouyang, L.Y. and Hsieh, T.P. (2007), Deterministic inventory model for deteriorating items with capacity constraint and time proportional backlogging rate, European Journal of Operational Research, 178(3), 789-807.

[21] Singh, T.J., Singh, S.R. and Singh, C. (2008), Perishable inventory model with quadratic demand, partial backlogging and permissible delay in payments, International Review of Pure and Applied Mathematics, 53-66.

[22] Hariga, M.(1995),An EOQ model for deteriorating items with shortages and time varying demand, The Journal of the Operational Research Society, 46, 398-404.

[23] Ruxian, L., Hongjie, L. and Mawhinney, J.R.(2010),A review on deteriorating inventory study, Journal of Service Science and Management, 3, 117-129.

[24] Sicilia, J., San-Jos, L.A. and Garca- Laguna, J.(2009), An optimal replenishment policy for an EOQ model with partial backlogging, Annals of Operations Research, 169,93-115

[25] Yang, H.L., Teng, J.T. and Chern, M.S.(2001),Deterministic inventory lot –size models under inflation with shortages and deterioration for fluctuating damand, Naval Research Logistics, 48, 144-158.

[26] Chen, H.L., Ouyang, L.Y., and Teng, J.T. (2006), On an EOQ model with ramp type demand rate and time dependent deterioration rate, International Journal of Information and Management Sciences, 17(4), 51-66.

[27] Goyal, S.K. and Giri, B.C. (2001), Recent trends in modeling of deterioration inventory, European Journal of Operational Research, 134,1-16.

[28] Raafat, F.(1991),Survey of literature on continuously deterioration inventory model, The Journal of the Operational Research Society, 42, 27-37.
[29] Goyal, S.K., Singh, S.R. and Dem, H. (2013), Production policy for ameliorating / deteriorating items with ramp type demand, International Journal of procurement Management, 6(4),444-465.
[30] Manna, P., Manna, S.K., Giri, B.C. (in press), An economic order quantity model with ramp type demand rate, constant deterioration rate and unit production cost, Yugoslav Journal of Operations Research.

Parameter Centric XML Parsing Techniques for Embedded Systems

Rashmi Sonar*, Sadique Ali** and Amol Bhagat***
*Department of Computer Science and Engineering,
Prof Ram Meghe College of Engineering and Management, Badnera, Amravati, India
rashmi.sonar@gmail.com
**Department of Computer Science and Engineering,
Prof Ram Meghe College of Engineering and Management, Badnera, Amravati, India
softalis@gmail.com
***Innovation and Entrepreneurship Development Center,
Prof Ram Meghe College of Engineering and Management, Badnera, Amravati, India
amol.bhagat84@gmail.com

Abstract: Embedded systems are dedicated to special purpose so they are different than conventional system. XML is known standard for communication because of the features like scalability, openness, flexibility. XML must be processed efficiently when it is used as a communication language. Parsing is one of the major tasks in processing XML. XML parsing itself is the challenge to achieve especially when the recourses are limited and tasks are with deadline. XML parsing in embedded systems has dedicated purpose which is vital in accomplishment of required tasks. In this paper different parameters for XML parsing are analyzed. The parameters for enhancing the efficiency of real time embedded systems are presented in this paper. This paper discusses the various XML parsing techniques available for embedded systems. The presented three basic approaches of parsing can be used for devising new approaches for XML parsing in embedded systems. As evaluated using experiments, VTD can handle large size XML documents. Therefore VTD can be considered for development of XML parsers in real time embedded systems.

Keywords: Document object model, embedded systems, parameter based parsing, real time systems, simple API for XML, virtual token descriptor, XML parsing.

Introduction

XML processing consists of four stages namely parsing, access, modification, and serialization. The parsing consists of three-steps. First step is character conversion where xml document character bit sequence is converted into character set of host programming language. Second step is lexical analysis in which character stream is converted into tokens like start element, attribute, end element. Final step is syntactic analysis where actual well formedness of document is verified. The character conversion and lexical analysis are invariant through any parsing model. The syntactic analysis represents data for access and modification with application of different parsing models. Some basic models are document object model (DOM), simple API for XML (SAX), and virtual token descriptor (VTD). DOM creates tree based API to represent XML document. SAX creates event based API with push and pull based parsing respectively, and VTD parsing method parses the XML document and creates 64-bit binary format VTD record (token) for each event [1].

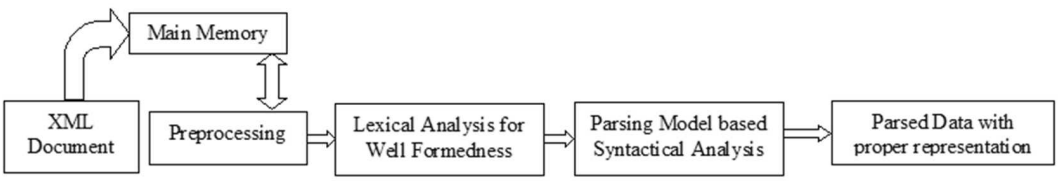


Figure 1. XML parsing process in embedded system

XML is de-facto standard for communication in message passing from one desktop to other. Many researches are there for desktop performance. XML web services is the most used technology for realizing service oriented architecture (SOA) because of the features like easy use, compose-ability, modularity, support, low cost, and commonality. For optimum utilization of SOA, XML behavior in web services also plays important role [2]. The XML parser implementation in

embedded systems differs from conventional computing systems primarily for the limited resources on embedded systems. Embedded systems are dedicated for particular purpose so XML parser should also be designed for accomplishing that purpose. In embedded systems primary requirement is to transfer XML syntax to internal format such as c [3, 4]. The overall architecture of XML parser in embedded system is shown in figure 1. Here pre-parsing is an important stage. As the embedded system having the limited memory, there is necessity to manage the data such a way that the better results must be produced efficiently. The processing time of XML comprised of parsing time and preprocessing time. In this paper section 2 compare different approaches of XML parsing for real time embedded systems, section 3 presents identified parameters for enhancing the performance of XML parsing in embedded system, section 4 evaluates three basic parsing techniques which can be extended in real time embedded system environment.

Related Work

In performance evaluation of XML web services for real-time applications, [2] defines use of SOA and XML web services in real time business application and system with the example of two banking scenario. For the integration of different types of software application, service oriented architecture (SOA) is one of the methods. The XML web services evaluation is performed so that each adopter can get SOA for large and complex system [2]. For high performance web services when typical schema is required, an adaptive XML parser, based on table driven XML (TDX) technology, plays an important role [5]. In this parser scanning, parsing and validation are performed in one pass with tabular representation of schemas and push down automaton at runtime. The parser achieves high performance without reconstruction and redeployment of web services which is five times greater than traditional parser.

A schema specific parser [6], combines parsing and validation into a single pass and achieves better response time. The technique used is table driven streaming XML parser in tabular form. For TDX, a toolkit is constructed for developing parser in C which comparable with DFA based parser. When asynchronous message passing required, a non-blocking schema specific parsing method for XML with two stack push down automaton [7], without backtracking gives the better performance. Parallel processing is also the performance enhancement approach. In [8] a DFA based approach is identified which recognizes XML specification and converts DFA to NFA for any of the input. The parser is based on parallel processing of the large scale xml documents with performance scale as 6-7 cores. Identifying the XML limitation, a cache parser is presented in [9], with the goal as to cache XML document information. Like DOM, cache parser exploits the syntactical tree of an XML document. It is strictly based on sender receiver and takes the advantage of xml document syntactic tree stored in cache.

XML DOM parsers memory characteristics are identified in [10] such as heap composition, object size, and type distribution, object life with JikesRVM2.4.1 and Merlin algorithm. With the proper garbage collector method, the parsing performance can be improved. BNF tree based BNFParser is proposed in [11] in which XML formal grammar is represented by extended Backus Naur form. BNFParser parser carries the idea with run time adaptive and code size efficient parsing which can be suitable for embedded system where storage restriction and memory limitations is the main issue. The only drawback is that, it is applicable to XML data less than 100 KB. Many times the structure of XML document is similar and only the values of attribute changes. The concept of structural encoding is introduced in [12]. Exploring identical XML structure, processing time can be reduced. Coupling is loose when algorithm is less aware of interfaces of parser and generally java is less efficient than C and C++. How to loosely couple the algorithm is analyzed in [13]. In [14], processing rules generator for Java phone is described with use of techniques like the DTD parser, XSLT, Java. This type of processing rule if applied to mobile database [24], mobile data can be read as manipulated.

XML Parsing Performance Enhancement in Embedded Systems

As the embedded systems are dedicated to particular task, there is some typical type of document processing which is having same structure, here values of attribute changes with same skeleton of document. The key point for parsing is no need to parse the whole document but storing the skeleton of XML, only identify values of attribute. In [15] XML structural data is reduced to format strings and arguments are sent as they are generated using modifications of real-time compression techniques specific to each data type. The portions of the XML structure which are common to many packets are generated on the fly or a priori and the values which vary from packet to packet are compressed using techniques specific to the type of data being sent. In this approach correlated and uncorrelated numeric data and short and long text strings are used. Format strings generation for each type of packet is carried out. The format string expresses the structure of the XML data in the packet and the portions which differ from packet to packet (arguments) become all that must be transmitted for subsequent packets. For example, assume a target tracking application generated the following two data packets for a target's location at separate times:

```
<target><lat>45</lat><lon>50</lon></target>
<target><lat>43</lat><lon>55</lon></target>
```

The format string could be expressed as <target><lat>[arg1]</lat><lon>[arg2]</lon></target> and the wireless device could just send the arguments after the format string was established. New entries were compressed in the same manner as long strings and the index positions were sent with the minimum number of bits required. Format strings are simply the element structure of the XML packet with the escape characters. If written well, this will be optimal and allow for the highest compressibility; however this would require more training than many users may want to do. While insert messages are simply replace messages with a zero length and delete messages are replace messages with an empty fragment. In [15] comparison of the TinyPack XML against Deflation, XMill, XMLPPM, and PAQ over the four datasets in both delay tolerant and real time experiments measuring compression, latency, processor usage, RAM requirements, and energy consumption is presented. All the data was collected prior to compression and compression was done on the entire dataset at once. TinyPack XML slightly outperforming XMill and XMLPPM and slightly underperforming the expensive "ideal" PAQ algorithm. The SNAResult and track data contained more static structure than the other datasets and required less RAM for TinyPack since the static portions of the structure are only stored in one place and are only compressed once.

The SCBXP technique architecture [16] makes use of a content-addressable memory that must be configured with a skeleton of the XML document being parsed, a finite state machine that controls FIFOs; in order to align XML data properly, multiple state machines acting on the multilevel nature of XML, and dual-port memory modules. In SCBXP technique the production of well-aligned data, at the end of the matching stage, is naturally accompanied with well-formedness checking and validation, without the need to rescan and reprocess the same XML document in terms of well formedness and validation. Prior to processing a new XML document, the SCBXP must perform the task of configuring the CAM with a skeleton derived from the XML document to be parsed. Due to the structure of the skeleton, a successful match of a tagged XML string against any of the CAM contents implies that this string is well formed and validated.

A new method for designing and implementing a manual XML parser named BNFParser is presented in [17], which is based on the mechanism of XML document matching against a BNF tree that built on XML formal grammar represented by extended Backus-Naur form (EBNF) notation. BNFParser is parsing an XML document with size less than 100 KB BNF tree, which embodies all of BNF generations, is composed of three kinds of BNF nodes: non-terminal symbol node, terminal symbol node and symbol group node, and one BNF expression is represented by one or more BNF nodes. The parsing process of BNFParser is matching XML document input against BNF Nodes in BNF tree. The main steps are: during the process, BNF stack is used to keep the latest status of traveling XML BNF tree to assist matching and possible backtracking actions. Compared with YACC/LEX-assisted parsers, BNFParser can work immediately without recompiling when changing XML syntax sets.

Most structure-related processing is identical for data items with identical structure; it is thus evident that the overall performance of XML processing will improve if redundancy in structure related processing can be reduced [18]. Structure encoding and the approaches to quickly identifying recurring structures, including one relying on collision-resistant hash function is introduced in [18]. The techniques to improve the performance of XML transmission, tokenization, parsing, and transformation by using structure encoding is also described in [18]. The implementation of prototypes of structure encoding based XML parser and transformer is carried out by extending the kXML parser and the XT XSLT processor. In structure encoding, the structure of an XML document is derived from the serialized text of the document, after removing non-whitespace text nodes and after "canonicalizing" element tags. "Canonicalizing" element tags includes "whitespace canonicalization" and "attribute canonicalization". "Whitespace canonicalization" removes any optional whitespaces in element start tags and end tags, and replaces any required whitespace with a single space character. "Attribute canonicalization" removes '=' character, the attribute value, and any delimiters surrounding attribute value. Efficient transmission of XML documents is particularly important in mobile environments. Structure recurrence can be exploited for compression and efficient transmission, by avoiding the redundant transmission of structural in formation and by compressing recurring test nodes and attribute values. Implementation does not support documents with variable-length arrays such as lists of identically structured elements with non-fixed lengths.

A conventional serial XML parsing architecture is described in [19] along with an improvement of architecture, using speculative pipeline structure to parsing XML, and speedup the parsing rate. The validator will examine: the validity of particular element or attribute in the document; the element under consideration is a child element; the order and number of child elements at any particular level of hierarchy in the document; the contents of the elements and attributes conform to the specified data-type; the contents of the elements and attributes conform to the specified valid range of values; a particular element or attribute is an empty element or can include text, and the value is default/fixed value. The data flow of XML parsing contains: first, data are loaded from either network or local hard disk. Then, data flow into the memory subsystem: main memory, L2 and L1 caches [20]. At the end, the processor fetches data from cache and performs the actual computation.

The performance bottleneck of XML parsing is in the memory data loading stage, rather than the disk data loading stage or the network exchange stage. In other words, the overhead introduced by the memory subsystem really plays on the weakness of the XML data parsing. Therefore, to speed up the XML parsing execution, it is imperative to focus less on acceleration and instead reduce the overhead incurred by the memory subsystem loading an XML document into memory and reading it prior

to parsing may take even longer than the actual parsing time. Consequently, instead of optimizing the specific computation of parsing, acceleration from the memory side; that is to say, accelerate the XML data loading stage memory-side accelerators deliver considerable effectiveness across existing parsing models is explored. They are able to reduce cache misses by up to 80 percent, which translates into up to 20 percent of performance improvement [20]. For code generation and parsing of XML schema, the open source software EXIficient is used in [21]. EXIficient is the Java implementation of W3C EXI specification. It is able to convert XML files using XML Schema to EXI streams and vice versa. EXIficient XML Schema parser and grammar builder is used to leverage VHDL code generation. EXIficient was extended with additional classes that use the internal representation of EXI structure to generate VHDL code. Then, the VHDL is used for hardware synthesis.

Table 1. Various parameters analyzed for performance improvement of XML parsing in embedded systems

Data Size	Token Descriptor	Memory Size	Parsing Time	Throughput
480 Byte [14]	Binary representation of the information and the structure of the document	-	0.066sec	
Tinypack compression [15]	structure of the XML data in one packet		1.4% of time for processing packet	
CAM is Configured with a Single Skeleton- 15 character or 15228bytes CAM is Configured with a New Skeleton for Each Loaded XML Document 15 character or 15228 bytes [16]	Token Id			25MHz-449Mbps 25MHz-400Mbps
<100 Kb on embedded system [17]	BNF Node	Heap memory <200 KB		
Large Documents, 30.40MB [18]	VTD array and MED Structure	2 times the size of XML document		35 MB/sec
With increment of EXI stream length [19]	EXI (Embedded XML Encoding) Stream		Hardware EXI get 0-10 μs	
Blooming filter length1024, [22]	ordered labeled trees in which the nodes represent elements or values, and the edges represent a relationship between two nodes.	false positive rate estimation $p.$=0.25 then 150 Kbytes. i.e $5 \times 10{-}5$ times the amount of space	Matching Time 32μs	
< 666KB [23]		PSDXP with 2 threads uses 5% Slice Register, 9% Slice LUT and 8% Block RAM. PSDXP with 4 threads uses 11% Slice Register, 19% Slice LUT and 17% Block RAM.		Percentage of post well-formed checking under overall CPB is 0.0868% and 0.1949% in PSDXPx2 and PSDXPx4 individually. PSDXP can achieve 0.5004 CPB and 0.2505 CPB with two threads and four threads individually.

The models do not explicitly deal with namespace processing, but instead use attribute processing to mimic namespace value processing in [22]. The models are based on the parameters: M- the number of blocks. A block is a structure in the XML document that consists of a parent tag delimited by a start and end tags. Several child elements also delimited by start and end tags, are contained within the parent tag. n- the number of child elements in a block. w- the number of attribute in the parent element of each block. z- the number characters in each text element in a block. x- the number of attributes in each child element in a block. Length (child_name)- the length of a child element name(<address>, length is 7 characters). It can be considered as a function f (child_name) = Length (child). Following the preceded convention f (text)- the length of a text node. f (att_name)- the length of attribute name. f (att_value)- the length of the attribute value. f (block_name)- the length of the block element. Pd- the number of characters corresponding to decorating tags which are the XML processing instructions and the root element.

Table 2. Comparison of available XML parsers

Parser	Implementation Language	Usage	Remarks
LibroXml	C	The library **libroxml** for XML parsing inside applications. The binary **roxml**, an xpath resolver that can be used from shell. The module **fuse.XML** that can be used to mount an XML file as a file system	library is minimum, easy-to-use
PugXML	C	The PugXML parser performs string scanning, tokenization, parsing, and construction of the document tree structure in a single pass.	Presented is a small, fast, non-validating DOM XML parser, contained in a single header.
RomXML	RomXML AE Parsing and Framing toolkit with C-language structures	Translate XML syntax to and from embedded internal C-language structures	More resources available for application features
ElectricXML	Java-based XML parser	Designed to have a small memory footprint and an intuitive operation as part of the GLUE distributed computing platform.	Electric XML parses DTDs but does not perform validation or implement the DOM. The ability to parse SOAP messages significantly faster than popular DOM-based parsers was a significant design goal
MinML	Java	MinML is an XML parser written in Java which implements nearly all of the XML language (it ignores DTDs).	It was developed for use in small embedded systems and has a code footprint of less than 10Kb.
MinML-RPC	Java	MinML-RPC 0.1 is a minimal eXtensible Markup Language Remote Procedure Call (XML-RPC) implementation that will run on small embedded systems (about 512Kb of RAM).	
NanoXML	Java	NanoXML is a very small (5KB) XML parser for Java.	SAX is a SAX adapter for NanoXML
Expat XML parse	C, Expat is an XML parser library written in C.		It is a stream oriented parser that requires setting handlers to deal with the structure that the parser discovers in the document.
TinyXML	C++	TinyXML is a simple, small, minimal, C++ XML parser that can be easily integrating into other programs.	It reads XML and creates C++ objects representing the XML document. The objects can be manipulated, changed, and saved again as XML.
SAX	Java	You can process the XML document in a linear fashion from the top down	
RapidXML	C++	RapidXml is an attempt to create the fastest XML parser possible, while retaining usability, portability and reasonable W3C compatibility	RapidXml achieves its speed through use of several techniques.

Compact structure representation is proposed in [23] using bloom filter, which also provides an easy solution for separation of the parsing process from the matching process so as to relief the burden of parsing from the matching time is significantly reduced due to the separation of parsing and matching, and the space for indexing structure is tremendously reduced due to the compactness of bloom filter. In [23] two issues are addressed. First, the parsing and matching are tightly bound together and cannot be easily separated. As the parsing takes a tremendous amount of time, the time efficiency is severely compromised. Second, the complex indexing and matching algorithms adopted by these works impose the risk of memory (or

storage) space overuse. The preprocessing module is comprised of a parser and a bloom-filter creator. It will be used to process the raw XML documents and XPath filters before the filtering process. They can be embedded at the client side. In this way the parsing processes are distributed over the individual clients themselves, providing a naturally balanced distribution of the computation throughout the system. When a subscription/notification has been generated, it will be first processed by a preprocessing module, which is comprised of an XPath/XML parser and the bloom-filter generator. A bloom filter is created after the processing for each subscription/XML document. The bloom filter combined with a subscriber identifier will be sent to the filtering engine, then indexed and stored for the filtering process. As to the notification, the XML documents will be sent together with the corresponding bloom filters. Once a document has arrived, its bloom-filter will be used for the evaluation against the subscription bloom filters stored previously. The un-subscription process is a simple instruction to remove the record in the filter index, identical to that in a conventional pub-sub system. When a subscription/notification has been generated, it will be first processed by a preprocessing module, which is comprised of an XPath/XML parser and the bloom-filter generator.

Generic XML parser [25] can be used to parse and reconfigure any valid XML file. The application of this parser is particularly useful in Software Communication Architecture where XML files represent the properties and parameters of hardware components and devices. A change in the parameters of the XML files changes the behavior of the hardware components. By using this Generic Parser, the XML files can be reconfigured dynamically at the middleware level and hence control the behavior of the hardware through software. This parser, when used at the middleware level to parse the XML files allows the user application to focus on the application logic itself, without dwelling on the tedious details of parsing the XML. The generic parser creates an info object for each node in the XML file and the structure of the info object is then referred by the DTD or schema. If the input XML file does not have schema, is not correct at the semantics level. The parsing technique used is the recursive descent parsing. A recursive descent parsing is a top down parsing technique where parsing starts from the parent node and proceeds down till it reaches the innermost child node.

After getting the root element of the XML document instance (topmost parent), a hash table is created. A hash table or hash map is a data structure that uses a hash function to map identifying values, known as keys, to their associated values. Thus, a hash table implements an associative array. The hashmap is used to store the key-value pair for all the elements of the instance of the XML document. Once the hashmap is created, the element is traversed till the end to find the attribute list. If an attribute is found, it is added to the attribute list hashmap in the form of (name, value) pair. All the available attributes for an element is added to the hashmap for each element.

All the child nodes are obtained along with their properties starting from the parent node. For this, the child is first checked if it is an instance of the parent node. If it is an instance of the parent element, then the name of the child is obtained. The attributes of the child node are added to a newly created hashmap. This child is treated as a new parent and it is parsed similar to the root node. If it does not have any instance, the name and the value associated with the child node is parsed and printed. A counter is used and hence it is possible to jump to the next child node from the root element directly without having to traverse through the visited nodes all over again.

The design is based on Document Object Model) to alter or change the parameters of the hardware devices through the parser. DOM is a tree-based interface that models an XML document as a tree of various nodes such as elements, attributes, texts, comments, entities, and so on. A DOM parser maps an XML document into such a tree rooted at a Document node, upon which the application can search for nodes, read their information, and update the contents of the nodes. Once it is found, the new value may be entered and it is updated in the document instance. Suitable error handling has been implemented to ensure that the algorithm does not deviate from its original flow. After the updation is completed, the entire document is written back to the original XML file and the updations are reflected in the XML file.

Evaluation of Extensive Parsing Techniques for Real Time Embedded Systems

XML parsers can be classified as heavy and light processors. Processors used for high computation called as heavy parsers which are not suitable for limited memory resulting in inconvenient for embedded systems where limited memory is available. Processors used for limited memory space and limited processing power are called as light processor. Some heavy parses are JDOM, JAXP, Xerces, or Xalan and most of them offer support for both DOM and SAX. Some light parsers are NanoXML, kXML, Xparse-J, ASXMLP, WoodStox, and TinyXML. The DOM and SAX parsers are generally used in conventional platforms. Embedded implementations of DOM and SAX are also available in C and Java. DOM's requirement for storing tree structure is high as compare to SAX's requirement with simple operation and little buffer space is more suitable for embedded system. MinML and NanoXML are examples of embedded parser. VTD-XML is also one of the basic XML parser which is not the object oriented and enables random access of document with the feature to run the application. VTD- XML can be implemented in software or hardware implementation with low resource usage. Though the light weight processor is available, problem of memory fragmentation is there which creates garbage collection in the environment like Java and C# with runtime overhead. Table 2 shows some of the available XML parsers with their usage, advantages and languages in which they are implemented. Any XML Parsing approach is the extension of basic approach like SAX, DOM, VTD. Therefore these approaches are evaluated in this paper for giving the directions for designing and developing novel

XML parsers and extending them for real time embedded system environment. Table 3 shows the experimental result for medium size XML document parsing with these basic approaches.

Table 3. Parsing results of three basic parsing approaches for small XML datasets

File Size (Bytes)	Parsing Time Required (milliseconds)		
	SAX	DOM	VTD
1102	63	57	31
5079	93	63	36
8167	94	63	47
120283	219	78	62
135386	250	93	62
156293	359	97	47
1294444	2324	124	70

The graph shown in figure 2 indicates that when the size of XML document is minimum all the approaches requires near about same time. When the size increases, at some points DOM and VTD require the same time. But for the big size xml document, VTD performs better than DOM and SAX. When SAX parser gets XML input, it generates the events. SAX is event-based and stores nothing in memory. SAX access the small part of document at a time. DOM parser converts the XML input to in memory objects. DOM converts the whole XML document in memory as object tree. IN VTD XML input is converted to tokens based on binary encoding specification. Each Virtual Token Descriptor record is a 64-bit integer that encodes the token length, starting offset, type, and nesting depth of a token in XML.

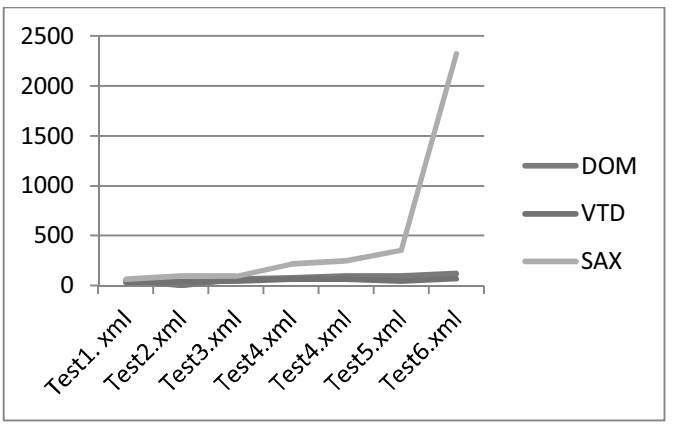


Figure 2. Comparison of SAX, DOM, and VTD on the basis of parsing time

Conclusion

XML parsing is always performance bottleneck. For parsing in embedded system, focus should be on preparsing methods such as reusable objects, garbage collection method, XML document serialization, structure encoding in future. The various performance measures can be concentrated for improving the performance of XML parsing in embedded systems. The various parsers available for minimum memory requirement as essential in real time embedded systems are discussed and compared in this paper. The experimental results show that VTD is the better solution as basic parsing approach. This approach can be extended in future for paring in real time embedded environment by combining it with existing strategies such as view management for approximated data, pattern tokenization, XML filtering schemes, etc. [22-24].

References

[1] Tak Cheung Lam, Jianxun Jaso Ding and Jyh - Charn Liu, "XML Document Parsing: Operational and Performance Characteristics", in IEEE Computer Society, 2008.

[2] Hazem M., El-Bakry and Nikos Mastorakis, "Performance Evaluation of XML Web Services for Real-Time Applications", in International Journal of Communications, Volume 3, Number 2, 2009.

[3] Esther Mınguez Collado, M. Angeles Cavia Soto, Jose A. Perez Garca, Ivan M.Delamer, and Jose L.Martınez Lastra, "Embedded XML DOM Parser: An Approach for XML Data Processing on Networked Embedded Systems with Real-Time," EURASIP Journal on Embedded Systems, Volume 2008.

[4] M. F. L Hufkens, "XML for embedded systems: the role of XML in distributed embedded networks", Master's thesis, Technische Universiteit Eindhoven, Department of Mathematics and Computing Science, August 2003.

[5] Wei Zhang, Robert A. Van Engelen, "An Adaptive XML Parser for Developing High-Performance Web Services" in Fourth IEEE International Conference on eScience , 2008.

[6] Zhang W. and R. van Engelen, "A Table-Driven Streaming XML Parsing Methodology for High-Performance Web Services", in IEEE International Conference on Web Services (ICWS'06), Sep 2006, pp. 197–204.

[7] Zhang W. and R. van Engelen, "High-Performance XML Parsing and Validation with Permutation Phrase Grammer Parsers", in IEEE International Conference on Web Services (ICWS'08), Sep 2008, pp. 286-294.

[8] Michael R Head and Madhusudhan Govindaraju, "Parallel Processing of Large-Scale XML–Based Application Documents on Multi-core Architectures with PiXiMaL", in IEEE Fourth International Conference on eScience, (eScience'08), 2008, pp. 261-268.

[9] Lelli F., Maron G. and Orlando S. , "Improving the performance of XML based technologies by caching and reusing information" in International Conference on Web Services, (ICWS '06), Chicago, Sep 2006, pp. 689-700.

[10] Gang WANG, Cheng XU, Ying LI, Ying CHEN, "Analyzing XML Parser Memory Characteristics: Experiments towards Improving Web Services Performance" in IEEE International Conference on Web Services (ICWS'06), 2006.

[11] Zhou Yanming and Qu Mingbin, "A Run-time Adaptive and Code-size Efficient XML Parser," in Proceedings of the 30th Annual International Computer Software and Applications Conference (COMPSAC'06), IEEE Computer Society, 2006.

[12] Zhou Dong, "Exploiting Structure Recurrence in XML Processing", in Proceedings of the Eighth IEEE International Conference on Web Engineering (ICWE'08), July 2008, pp. 311–324.

[13] Eric Jui-Lin Lu and Ying-Sheng Lee, "A Processing Rules Generator for an XML-based Mobile Database" in Proceedings of the IEEE International Conference on Networking, Sensing & Control Taipei, Taiwan, March 2004, pp. 106- 111.

[14] Giuseppe Psaila "Loosely Coupling Java Algorithms and XML Parsers: a Performance-Oriented Study" in Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW'06) IEEE Computer Society , 2006.

[15] T. Szalapski, S. Madria, and M. Linderman, "TinyPack XML: Real Time XML Compression for Wireless Sensor Networks", in Wireless Communications and Networking Conference (WCNC), IEEE, 2012, pp. 3165-3170.

[16] Fadi El-Hassan, and Dan Ionescu, "SCBXP: An Efficient CAM-Based XML Parsing Technique in Hardware Environments," IEEE Transaction on Parallel and Distributed Systems, Vol. 22, Nov 2011, pp.1879- 1887.

[17] Zhou Yanming and Qu Mingbin, "A Run-time Adaptive and Code-size Efficient XML Parser," in Proceedings of the 30th Annual International Computer Software and Applications Conference (COMPSAC'06), IEEE Computer Society, 2006.

[18] Zhou Dong, "Exploiting Structure Recurrence in XML Processing", in Proceedings of the Eighth IEEE International Conference on Web Engineering (ICWE'08), July 2008, pp. 311-324.

[19] Ma Jianliang, Shaobin Zhang, Tongsen Hu, Minghui Wu and Tianzhou Chen, "Parallel Speculative Dom-based XML Parser", 14th IEEE International Conference on High Performance Computing and Communications, 2012.

[20] Jie tang, Shaoshan Liu, Chen Liu, Zhimin Gu, and Jean-Luc Gaudiot, "Acceleration of XML Parsing through Prefetching", in IEEE Transactions on Computers, Volume 62, Number 8, Aug 2013.

[21] Vlado Altmann, Jan Skodzik, Peter Danielis, Nam Pham Van, Frank Golatowski, and Dirk Timmermann, "Real-Time Capable Hardware-based Parser for Efficient XML Interchange" in 9th International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP), 2014, pp. 395-400.

[22] G. Suddul, N. Nissanke, and N. Mohamudally, "A pattern based tokenization model for XML parsing on mobile devices", in IEEE transactions, Sep 2013, pp. 1-5.

[23] Yu Xiaochuan, C. S. Alvin, "A time/space efficient XML filtering system for mobile environment", 12th IEEE International Conference on Mobile Data Management (MDM), Volume 1, 2011, pp. 184-193.

[24] Amol Bhagat and Bhaskar Harle, "Materialized view management in peer to peer environment," Proceedings of the ACM International Conference & Workshop on Emerging Trends in Technology, 2011, pp. 480-484.

[25] Akhil Rangan C K, Jayanthi J, "A Generic Parser to Parse and Reconfigure XML files", in IEEE conference on Recent Advances in Intelligent Computational Systems (RAICS), pp. 823-827, 2011.

Emotion Extraction using Rule-based and SVM-KNN Algorithm

Mohini Chaudhari * and Sharvari Govilkar **
* Department of Computer Engg. PIIT, New Panvel, India
3mohinichaudhari89@gmail.com
** Department of Computer Engg. PIIT, New Panvel, India
sgovilkar@mes.ac.in

Abstract: Language and emotions generally goes hand in hand. Language forms a firm base for communication and expression of emotion. Computers are not capable of understanding and taking decisions on its own. In order to make it analyze the language and detect emotions, Natural Language Processing (NLP) techniques have been applied to automatically identify the information content in text. As emotions influences human thinking, perception and behavior, they play an important role in decision making, and learning, and can even overcome reason under stress conditions. Extraction of emotions from a given input text is one of the most popular research areas today and involves lots of challenges. We need computational approaches that would successfully analyze the online emotion rich content, recognize and aggregate relevant information, and draw useful conclusions. Recognizing emotions conveyed by a text can provide an insight into the author's intent and sentiment, and can lead to better understanding of the text's content. Thus, the idea is to propose a system that will accept single document as input in English NLP and processes the input and extract emotions by using a hybrid system.

Keywords: List five or more keywords related to your work.Emotion Detection, Natural Language Processing,Machine Learning, Support Vector Machine (SVM), Rule-based, K-Nearest Neighbor (KNN).

Introduction

Emotions are our subjective feelings and thoughts. People express emotions as part of everyday communication. Emotions can be judged by a combination of clues viz. facial expressions, prosodies, gestures, and actions. Emotions are also articulated by written texts. Considering the inspiration by work done in sentiment analysis, the thesis explores hybrid approach for detection of emotions in text. Emotion recognition in text is just one of the several dimensions of the task of making the computers make sense of emotions. It is essentially a content-based
Classification problem involving concepts from the domains of Natural Language Processing as well as Machine Learning.
Understanding and expressing emotions between people is a complex interactional phenomenon that forms an intricate web. People are able to perfectly distinguish the expressed emotions because they are capable of understanding the meaning of the words and phrases. They also are able to generate expressions and sentences for various emotions. But, developing a computer system that would analyze and interpret different emotions in a given text is a different task.
The paper presents, a combination of three machine learning techniques to detect the emotion labels present in the given text in pure English language. In section 2, related work has been discussed. Section 3 deals with the detailed proposed system architecture for the emotion detection system. Working of system is mentioned in detail in section 4. Section 5 explores accuracy obtained by system. Finally, paper is concluded in section 6.

Related work

In this section we cite the relevant past literature that use the various emotion extraction techniques. Most of the system extracts the emotions from sentences, blogs, and text document.
Ruchika Sharma and Amit Arora [1] have extended and compared the work proposed by Mullen and Collier (2013). The system comprises of two phases viz., a feature Extraction phase and a learning phase, which forms the basis of the analyzing the overall sentiment of the document. Machine learning approaches such as Naives Bayes, Maximum Entropy, SVM and Kernels were explored and out of all of these approaches Multiple Kernel outperforms. It produces an accuracy of 90% and for cross validation in 5 fold and 10 fold respectively is 92%.
Arti Buche, Dr. M. B. Chandak and Akshay Zadgaonkar [2] proposed the technique to detect and extract subjective information in text document that is opinion mining. Sentiment classification or Polarity classification is the binary classification process that labels an opinionated document as expressing either an overall positive or an overall negative opinion.

V.S. Jagtap and Karishma Pawar [3] focus on different approaches used in sentiment classification for sentence level sentiment classification. It aims at analyzing a solution for sentiment classification at a fine-grained level in which the polarity of the sentence can be given as positive, negative or neutral. According to them, Sentiment Analysis is the process of extracting knowledge from the peoples' opinions, appraisals and emotions towards the entities, events and their attributes. With the evolution of web technology, there is large amount of data present in web for the internet users. These users use the available resources in the web, and also give their feedback, thus generating additional useful information.

Pravesh Kumar Singh and Mohd. Shahid Husain [4] concludes that though opinion mining is in an inchoate stage of development, the ground is still set for dense growth for researchers. They made an attempt to appraise different techniques involved in feature extraction. According to their convergent point of view, textual classification can be easily done by Naïve Bayes, clustering is suitable for consumer services and for biological reading and interpretation SVM gives better results.

Azadeh Nikfarjam, Ehsan Emadzadeh, [5] has proposed a system which is a combination of machine learning and rule based techniques. They had described this approach as a hybrid approach for sentiment analysis. The system consists of a rule-based engine and a trained Support Vector Machine (SVM) classifier for each possible emotion. A set of syntactic and semantic features are extracted from sentences to build the rules and train the classifier. In order to generate the sentence features we propose a new approach to identify a sentence‘s clauses and its constitutive grammatical elements and to use them to measure the polarity (a quantitative measure of the positive or negative feelings reflected in it) of a given sentence. In order to handle the problem of multiclass classification, they implemented a component emotion detector for all of the 15 emotion categories.

G. Vinodhini [6] has proposed the techniques used for sentiment classification which includes Naïve Bayes (the basic idea is to estimate the probability of categories given a test document by using the joint probability of words and categories), SVM (Method of statistical classification based on the structural risk minimization principle), Centriod Classification, K-nearest neighbor Method, Winnow (Well-known as online mistaken-driven method), and Ensemble technique (Combines several base classification output to generate an integrated output).

Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka [7] describe a lexical rule-based approach to recognition of emotions from text and an application of the developed Affect Analysis Model in Second Life. Based on the result of the Affect Analysis Model, the developed EmoHeart (—object‖ in Second Life) triggers animations of avatar facial expressions and visualizes emotion by heart-shaped textures. They propose a twofold focus in their research: (1) recognition of affective content conveyed through text, and (2) automatic visualization of emotional expression of avatars, which allows avoiding manual control by the user and enriching remote communications effortlessly.

Tony Mullen and Nigel Collier [8] introduces an approach to sentiment analysis which uses support vector machines (SVMs) to bring together diverse sources of potentially pertinent information, including several favorability measures for phrases and adjectives and, where available, knowledge of the topic of the text.

Joy, sadness, anger, surprise, hate, fear, so on are some of the emotion expressions. W. Gerrod Parrot [9] wrote a book in the year 2001, named — Emotions in Social Psychology, in which he explained the emotion system and formally classified the human emotions through an emotion hierarchy in six classes at primary level i.e., Love, Sadness, Joy, Anger, Fear and surprise as shown in following table. There are many other words that also fall in secondary and tertiary levels.

Proposed system

We propose a system to extract the emotions using a Grimm's dataset which combines different fairy tales to identify nine artistic emotions as described as: pleasure, mirth, anger, energy, fear, disgust, astonish, serenity, sorrow.

The input to the system is a single text document in English Natural Language and the output is the emotion label for each sentence. The proposed approach undergoes two phases viz.,

1 Pre-processing Phase
2 Emotion Extraction Phase

Figure 1 shows the proposed architecture of the system. The system consists of two phases, i.e., Pre-processing phase and Emotion Extraction Phase. The input given to the system is in the form of text document in pure English Language. This text document in pure English Language when given to the system first enters the pre-processing phase and once the preprocessing is done, it is further sent to the emotion extraction phase.

Pre-processing Phase

The pre-processing phase consists of following modules.

Sentence Segmentation: The given input text is chunked into sentences and these sentences are stored in different text file for further use.

Tokenization: It is the process of splitting the sentences into words by identifying the spaces, comma and special symbols between the words. So list of sentences and words are maintained for further processing.

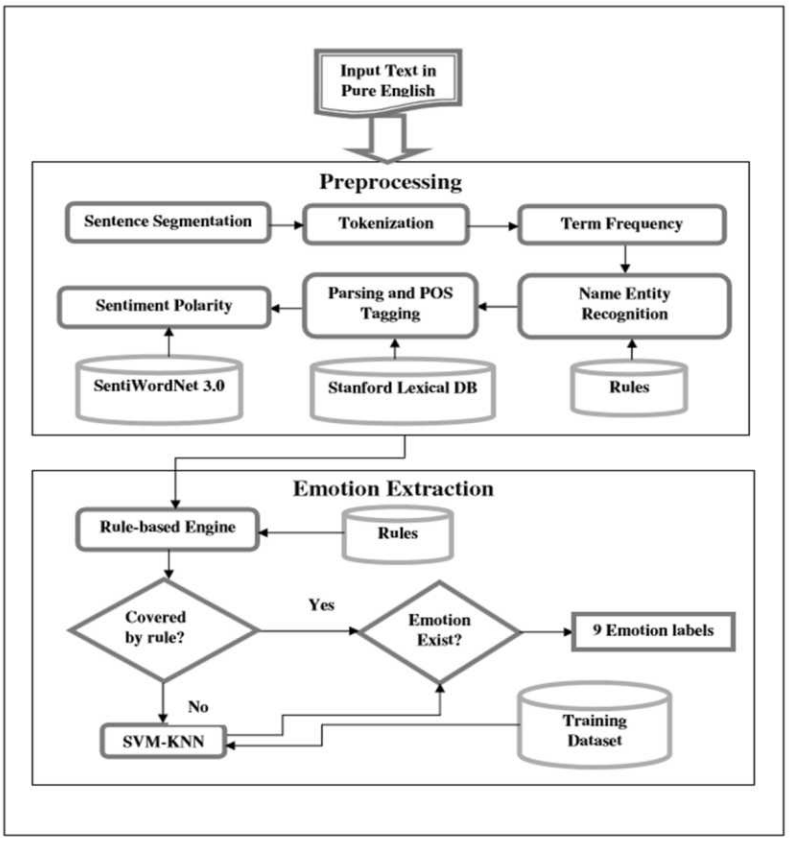


Figure1. Proposed System

Term Frequency: Term frequency is defined as the number of times a word occurs in the document. Term frequency generally depends on the length of the document, i.e., a term may occur more frequently in a large document as compared to a small document.

Name-Entity Recognition: Named Entity Recognition (NER) labels sequences of words in a text which are the names of things, such as person and company names, or gene. It comes with well-engineered feature extractors for Named Entity Recognition, and many options for defining feature extractors.

POS Tagging: Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word, such as noun, adjective, verb, etc.

Sentiment Polarity Calculation: Sentiments are generally classified into two groups such as positive or negative. But determining a sentiment as positive or negative is not just enough. It is also necessary to analyze the intensity of that sentiment i.e. how positive or how negative the sentiment is. This is done with the help of SentiWord Net 3.0 [15]. As per SentiWordNet, Sentiment Polarity is given as follows.

$$\text{Sentiment Polarity} = 1 - (\text{Pos Score} + \text{Neg Score})$$

Emotion Extraction Phase

In this phase, the preprocessed document is given as a input. This phase works in three modules, viz., Rule-based engine, SVM classifier and KNN. This is the main phase of the system wherein the emotion labels are extracted using rule-based engine, SVM and KNN. In this phase set of rules are applied to the input text document. If the rule based engine does not extract the emotional label then such sentences are forwarded to the SVM. SVM build a hyper-plane based on the polarity calculated. Here the positive values are at one side of the hyper-plane while the negative values are at the other side. While implementing this system on the Grimm's dataset it was observed that there are some polarity values that lie on the hyper-plane and the emotion labels for such valued sentences are not extracted. In order to overcome this problem, KNN algorithm is used. Those sentences whose polarity value belong to the hyper-plane where thereafter given as an input to the KNN algorithm which considers an optimal value of k and finds all the nearest neighbors and finally provides the emotion label.

Working of System

This is the main phase of the system wherein the emotion labels are extracted using rule-based engine, SVM and KNN. In this phase set of rules are applied to the input text document. If the rule based engine does not extract the emotional label then such sentences are forwarded to the SVM. SVM build a hyper-plane based on the polarity calculated. Here the positive values are at one side of the hyper-plane while the negative values are at the other side. While implementing this system on the Grimm's dataset it was observed that there are some polarity values that lie on the hyper-plane and the emotion labels for such valued sentences are not extracted. In order to overcome this problem, KNN algorithm is used. Those sentences whose polarity value belong to the hyper-plane where thereafter given as an input to the KNN algorithm which considers an optimal value of k and finds all the nearest neighbors and finally provides the emotion label.

Any text document in pure English language can be given as an input to the system. At the initial stage, when the system starts, we need to open the text document for which we wish to extract emotions. Once the text document is open, the system first checks for script validation and opens the document by eliminating those script which are not in English language. The following figure shows how the system accepts the input document.

Figure 2. GUI of the system

Considering the following sample text document, let us understand the working of the system.

Sample text document

Cinderella thanked him, went to her mother's grave, and planted the branch on it, and she wept so much that her tears fell upon it and watered it. It grew and became a beautiful tree.

Step1: In order to accept the input text document, the system provides an open option in the File menu from where the user can open any text document for which he requires the emotion detection. The figure below shows the GUI for accepting a text document in pure English language as input.

Figure 3. GUI for opening the input text document

Step2: Once the text document is given as input, the input needs to be preprocessed. The preprocessing step consists of Sentence Segmentation, Tokenization, Parsing and POS Tagging. The process of breaking a string of text in sentences is known as sentence segmentation. In English language this can be done particularly by analyzing the full stop. The figure below shows the GUI for Sentence Segmentation.

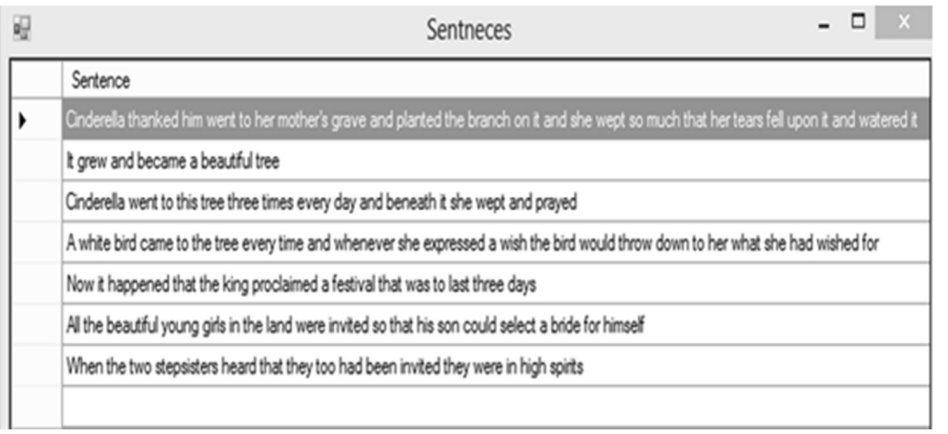


Figure 4. GUI for Sentence Segmentation

Step 3: After segmenting the complete text into sentences, tokenization is performed. It is the process of breaking the string of text into words (Tokens). This is done by analyzing the space between to the two tokens. These tokens are then saved to a different file for further use. The figure below shows the GUI for tokenization.

Figure 5. GUI for Tokenization

Step 4: After performing tokenization, Parsing is done. Parsing is the process of recognizing the sentence and how it is constructed. Constructing was a bit difficult for the system as the sentences provided in the input text document are too long, so the system has assigned different colors to each token is the sentence. The figure below shows the GUI for Parsing.
Step 5: In this step, the system assigns part-of-speech tag to each token in the sentence. This process of assigning Pos to each token is known as POS Tagging. The figure below shows the GUI for POS Tagging.
Step 6: After completing POS Tagging, the Feature Extraction phase begins. This phase consists of Term frequency calculation, Name-Entity Recognition and Sentiment Polarity Calculation. This step determines how frequently a term occurs in the document. The figure below shows the GUI for Term Frequency Calculation.
Step 7: In this step the system assigns the atomic text elements to the predefined categories such as name of persons, organization, location, etc. This process is known as Name-Entity Recognition. The figure below shows the GUI for NER.
Step 8: After recognizing the atomic elements of the text, the system calculates the sentiment polarity of each sentence in the document with the help of SentiWordNet. The figure below shows the GUI for Sentiment Polarity Calculation.

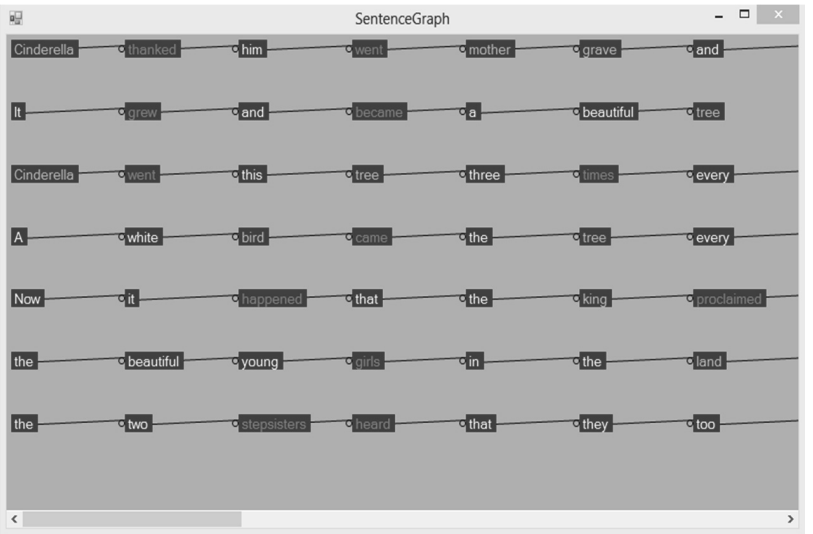


Figure 6. GUI for Parsing

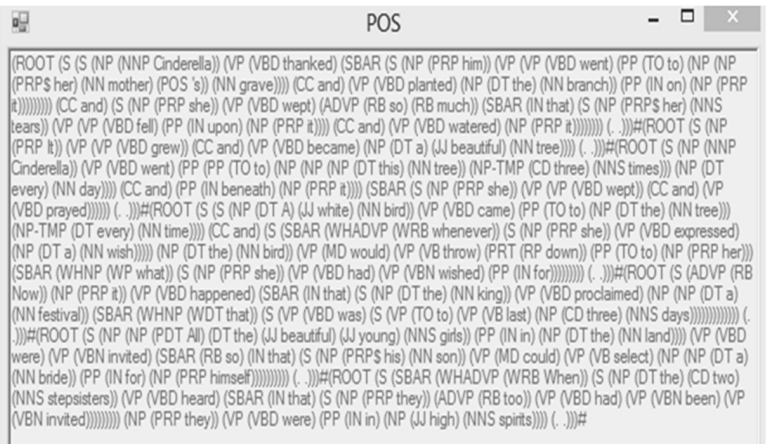


Figure 7. GUI for POS Tagging

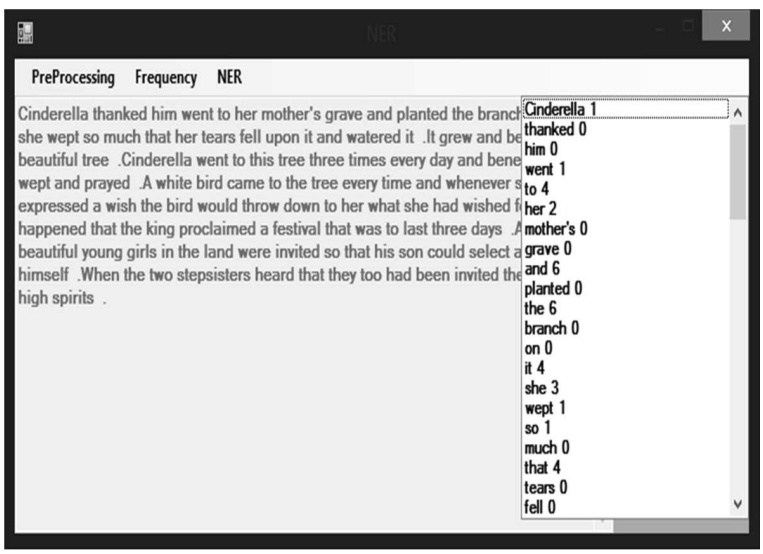


Figure 8. GUI for Term Frequency Calculation

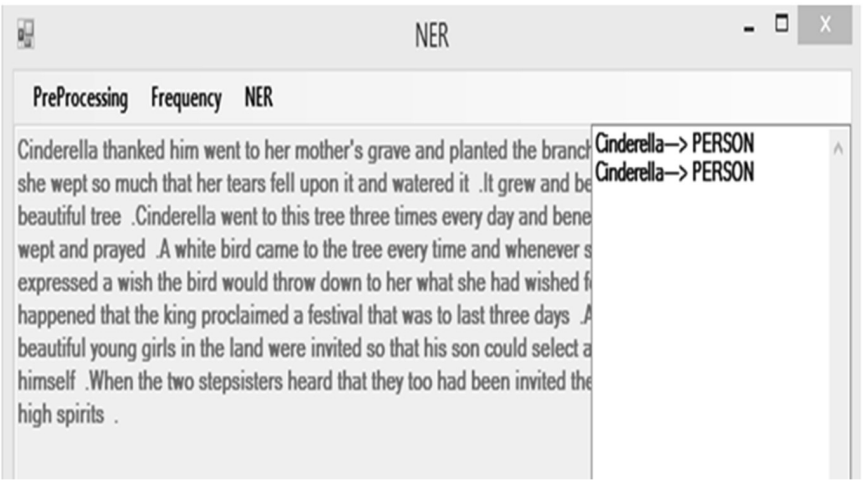
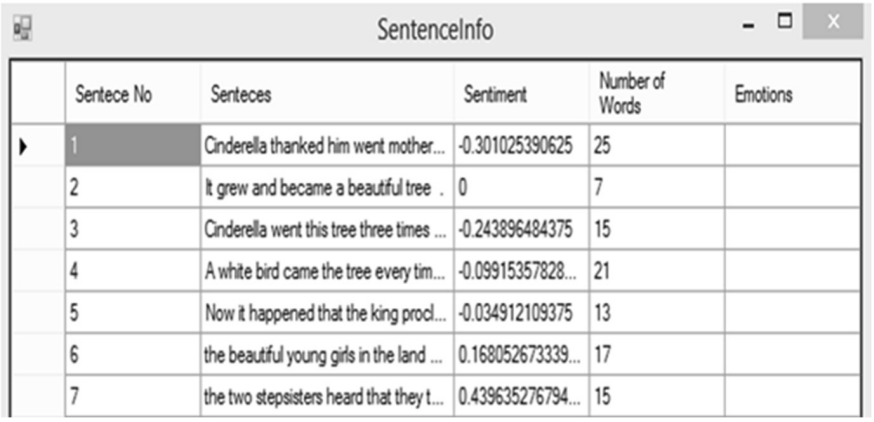


Figure 9. GUI for NER



	Sentece No	Senteces	Sentiment	Number of Words	Emotions
▶	1	Cinderella thanked him went mother...	-0.301025390625	25	
	2	It grew and became a beautiful tree .	0	7	
	3	Cinderella went this tree three times ...	-0.243896484375	15	
	4	A white bird came the tree every tim...	-0.09915357828...	21	
	5	Now it happened that the king procl...	-0.034912109375	13	
	6	the beautiful young girls in the land ...	0.168052673339...	17	
	7	the two stepsisters heard that they t...	0.439635276794...	15	

Figure 10. GUI for Sentiment Polarity Calculation

Step 9: In order to train the classifier, the system need to load the training set before analyzing the emotion labels which is done in this step. The figure below shows the GUI for loading the training set.

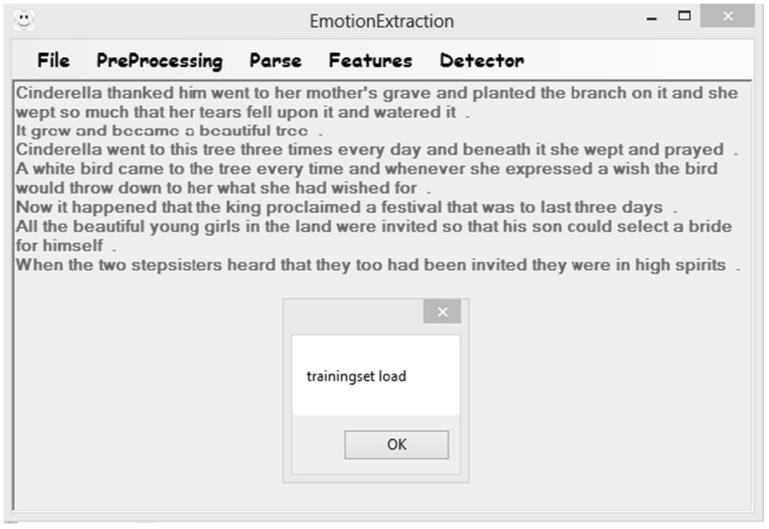


Figure 11. GUI for Loading the Training set

Step 10: This is the last step where the system analyzes each and every sentence of the text document in order to assign them with the correct emotion label. The figure below shows the GUI for Emotion Detection.

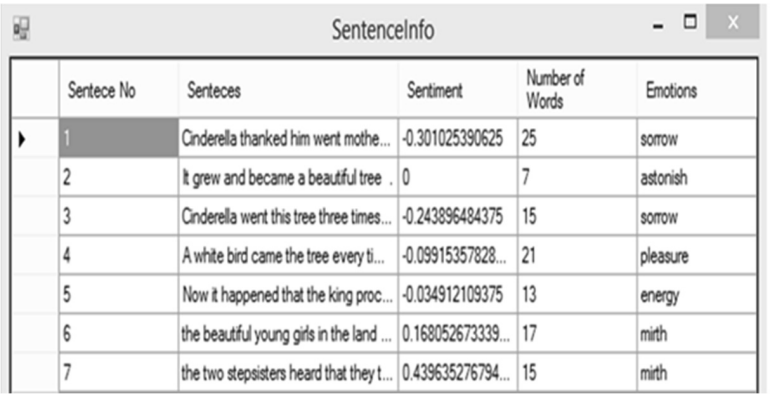


Figure 12. GUI for Emotion Detection

Performance Analysis

The system is evaluated to check whether the output generated by the current system is efficient than that of the existing system or not. Various performance measures such as accuracy, precision, recall and f-measure score has been used to check the same. To check the accuracy and efficiency of the system number of sentence in the document are used as unit for evaluation. In order to perform these evaluations, some notations are used. They are as follows: D1, D2, D3.. Represents input documents from Grimm's test dataset used for testing. DG1, DG2,.. Represents input documents based on general topics such as facebook posts, news, movie reviews, comments, etc. C1, C2,C9 - Represents the class of emotions such as Anger, Astonish, Disgust, Energy, Fear, Mirth, Pleasure, Serenity and Sorrow respectively, in which the sentences are classified. PC1, PC2, …….., PC9 Represents the predicted class of emotions such as Anger, Astonish, Disgust, Energy, Fear, Mirth, Pleasure, Serenity and Sorrow respectively, in which the sentences are classified.

1) Accuracy:

Accuracy is defined as the ratio of number of sentences for emotion labels are extracted correctly to the total number of sentences in the document.

$$\text{Accuracy} = \frac{number\ of\ sentences\ for\ which\ emotion\ labels\ are\ extracted\ correctly}{total\ number\ of\ sentences\ for\ which\ emotion\ labels\ are\ extracted\ by\ the\ algorithm} \times 100$$

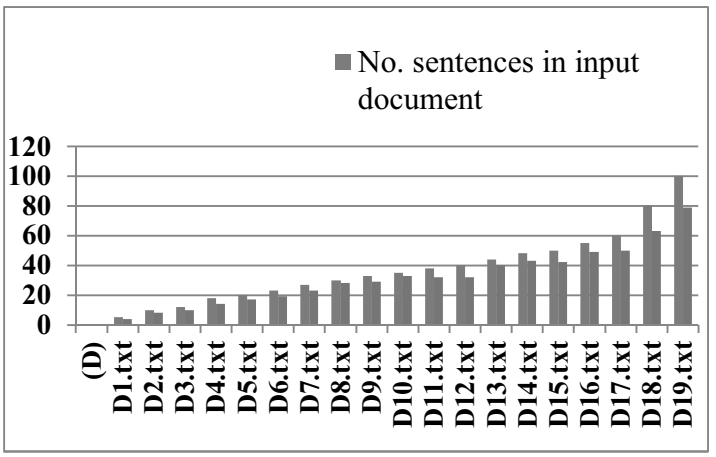


Figure 14. Evaluation Graph of the system for documents based on General Topic test dataset

2) Efficiency:

To evaluate the efficiency of the system, the accuracy of the existing system is compared with the accuracy of the current system. Table 5.2 describes the efficiency of the current system i.e. Rule-based + SVM-KNN.

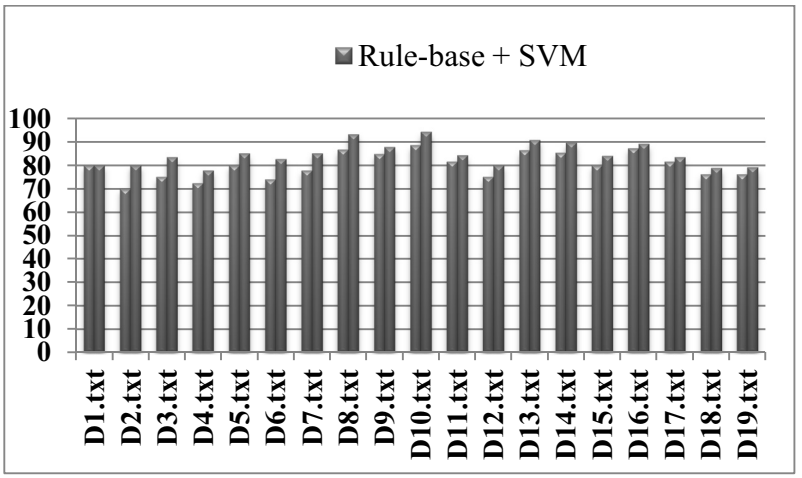


Figure 15. Efficiency comparison graph

3) Precision and Recall:
As the system deals with multi-class classification problem, the only way to evaluate the performance of the system is for each class label compute the precision and recall and analyze the individual performance on those class labels or average on class labels or average the values to get the overall precision and recall. For evaluating precision and recall for the system, 10 sample sentences are considered for each class of emotion.

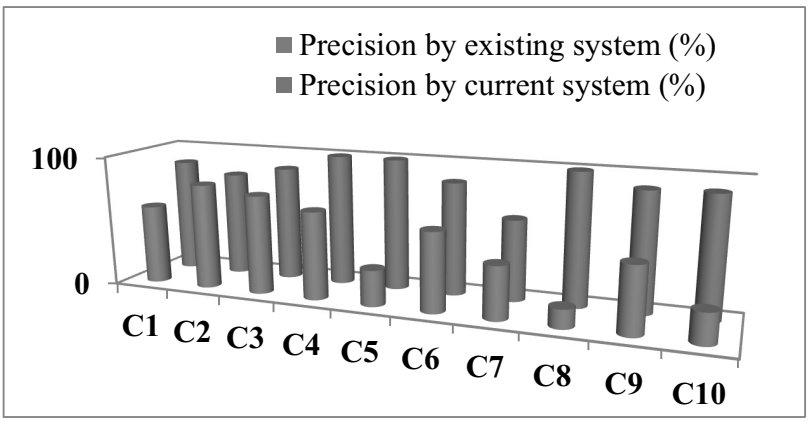


Figure 16. Precision comparison graph

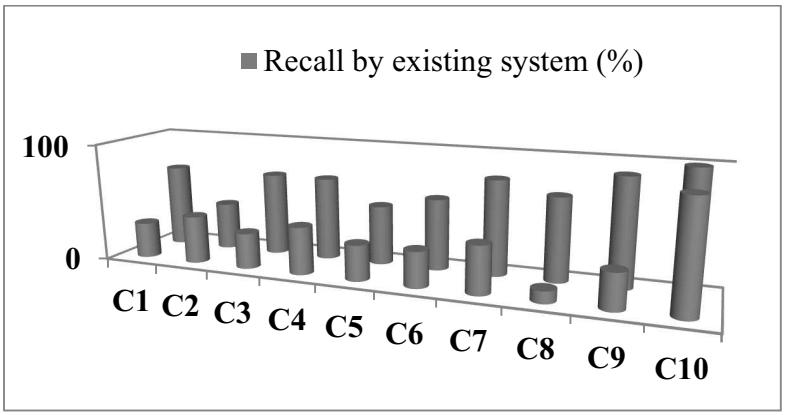


Figure 17. Recall comparison graph

Conclusion

The system has performed extremely well in terms of accuracy while showing the signs that it can be further scaled to much bigger and different dataset with better performance. In this research, the main focus was on analyzing emotions at sentence level which can further explored to paragraph level and topic level. The work can also be expanded in future by introducing methods that increase the accuracy by handling problems like change of emotions when the personal pronoun changes which still needs to be evolved properly.

Acknowledgment

I am using this opportunity to express my gratitude to thank all the people who contributed in some way to the work described in this paper. My deepest thanks to my project guide for giving timely inputs and giving me intellectual freedom of work. I express my thanks to the head of computer department and to the principal of Pillai Institute of Information Technology, New Panvel for extending his support.

References

[1] Ruchika Sharma and Amit Arora, ―Improve Sentiment Analysis Accuracy using Multiple Kernel Approach‖ International Journal of Computer Applications (0975 – 8887) Volume 71– No.20, June 2013
[2] Arti Buche, Dr. M. B. Chandak, Akshay Zadgaonkar, "Opinion Mining and Analysis: A Survey" – International Journal of Natural Language Computing (IJNLC), Vol 2, No.3, June 2013
[3] Jayashri Khairnar, Mayura Kinikar, "Machine Learning Algorithms for Opinion Mining and Sentiment Classification" – International Journal of Scientific and Research Publications, Volume 2, Issue 6, June 2013.
[4] V.S.Jagtap, Karishma Pawar, Analysis of different approaches to Sentence-level Sentiment Classification – International Journal of Scientific Engineering and Technology, Volume 2, Issue 3, April 2013.
[5] Parvesh Kumar Singh, Mohd Shahid Husain, "Analytical Study of Feature Extraction Techniques in Opinion Mining" – CS & IT – CSCP 2013.
[6] Azadeh Nikfarjam, Ehsan Emadzadeh and Graciela Gonzalez, "A Hybrid System for Emotion Extraction from Suicide Notes", Biomedical Informatics Insights 2012:5 (Suppl. 1) 165–174, doi: 10.4137/BII.S8981
[7] G. Vinodhini and RM. Chandrashekharan, "Sentiment Analysis and Opinion Mining: A Survey" – International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6, June 2012.
[8] Neviarouskaya A, Prendinger H, Ishizuka M. EmoHeart: Conveying Emotions in Second Life Based on Affect Sensing from Text. Advances in Human-Computer Interaction. 2010;(2):1–13.
[9] T. Mullen and N. Collier, ―Sentiment analysis using support vector machines with diverse information sources,‖ in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '04), pp. 412–418, Barcelona, Spain, July 2004.
[10] W.G. Parrot, ―Emotion in Social Psychology,‖ in Psychology Press, Philadelphia 2001, p.p. 102-105.
[11] P. Ekman, (1992), ―An Argument for Basic Emotions‖ International Journal of Cognition and Emotion, Vol. 6(3), published by Lawrence Associates Ltd, US, Jan 1992, p.p. 169- 200.
[12] Daniel Dichiu, Ana Lucia Pai, Sunita Andreea Moga, Catalin Buiu, ―A Cognitive System for Detecting Emotions in Literary Texts and Transposing them into Drawings‖, IEEE, 2010.
[13] Teng, F. Ren, and S. Kuroiwa, ―Recognition of Emotion with SVMs‖, in Lecture Notes of Artifical Intelligence 4114, D-S. Hung, K. Li, and G. W. Irwin, EDS. Springer, Berlin Heidelberg, 2006, p.p. 701-710. 33
[14] N. Anitha, B. Anitha, S. Pradeepa, "Sentiment Classification Approaches – A Review" – International Journal of Innovations in Engineering and Technology, Volume 3, Issue 1, October 2013.
[15] Walaa Medhat, Ahmed Hassan, Hoda Korashy, "Sentiment Analysis Algorithms and Application: A Survey", 2090-4479, 2014 Production and hosting by Elsevier B.V. on behalf of Ain Shams University.
[16] SentiWordNet 3.0, http://sentiwordnet.isti.cnr.it/ (Last accessed on Spetember 13, 2015 at 3.05 P.M).
[17] Stanford Parser, http://nlp.stanford.edu:8080/parser/index.jsp, June 15, 2012. (Last accessed on September 15, 2015 at 4.00 P.M).
[18] http://www.gutenberg.org/files/2591/old/grimm10.pdf (last accessed on September 15, 2015 at 4.10 P.M).
[19] http://nlp.stanford.edu/downloads/ (last accessed on October 10, 2015 at 8.10 P.M).
[20] http://www.redbooks.ibm.com/redpapers/pdfs/redp3937.pdf (last accessed on October 10, 2015 at 10.0 P.M)

Sentiment Classification in Hindi by using a HindiSentiWordNet (HSWN)

Pooja Pandey*, Sharvari Govilkar** and Sagar Kulkarni***
*Department of Computer Engg. PIIT, New Panvel, India
pooja.clx@gmail.com
**Department of Computer Engg. PIIT, New Panvel, India
sgovilkar@mes.ac.in
***Department of Computer Engg. PIIT, New Panvel, India
skulkarni@mes.ac.in

Abstract: Sentiment Analysis is a natural language processing task that deals with finding orientation of opinion with respect to a given topic. It deals with analyzing emotions, feelings, and the attitude of a speaker or a writer from a given piece of text or any other form of media. The target of sentiment classification system is to find opinions, identify the sentiments they express, and then classify them according to their polarity. The proposed system for sentiment classification of Hindi documents uses Hindi SentiWordNet (HSWN) to find the overall polarity of the Hindi text document where the final aggregated polarity calculated by system can be positive, negative or neutral. Existing HSWN is enhanced by adding more number of sentiment bearing words. The proposed System also handles negation and discourse relations which influence sentiment associated with a given input.

Keywords: Sentiment Analysis (SA), SentiWordNet, HindiSentiWordNet (HSWN), Polarity, Synset Replacement., Natural language processing (NLP).

Introduction

Sentiment Analysis is a task under natural language processing which finds orientation of a person opinion or feelings over an entity [1]. It deals with analyzing personal emotions, feelings, attitude and opinion of a speaker or a writer over an object. The primary target of SA is to find the sentiments expressed by person over an information or entity [2].

Sentiment analysis helps to find sentiment associated with the given input which can be in the form of single line or paragraph or a full document about a given subject. SWN consists of words present in specific language with its associated polarity. For the given input overall polarity or sentiment can be calculated by extracting and aggregating polarity of each sentiment word in the input.

There are different classification levels in SA: document-level, sentence-level and aspect-level. Document-level SA aims to classify an opinion of the whole document as expressing a positive or negative sentiment. Sentence-level SA aims to classify sentiment expressed in each sentence which involves identifying whether sentence is subjective or objective. Aspect-level SA aims to classify the sentiment with respect to the specific aspects of entities which is done by identifying the entities and their aspects.

The paper presents, sentiment analysis system in Hindi language where overall sentiment is classified as positive or negative. In section 2, proposed system is discussed in detail. Working of system is mentioned in detail in section 3. Section 4 explores accuracy obtained by system. Finally, paper is concluded in section 5.

Related Work

In this section we cite the relevant past literature of research work done in the field of sentiment analysis for Hindi language.

Namita mittal et al [1] developed an efficient approach based on negation and discourse relation to identifying the sentiments from Hindi content . They developed an annotated corpus for Hindi language and improve the existing Hindi SentiWordNet (HSWN) by incorporating more opinion words into it.

Aditya Joshi and Pushpak Bhattacharyya [2] proposed a fallback strategy for Hindi language. Authors proposes use of, In-language Sentiment Analysis, Machine Translation and Resource Based Sentiment Analysis to find sentiment in Hindi text. Hindi SentiWordNet (HSWN) was developed using two lexical resources (English SentiWordNet and English-Hindi WordNet Linking .78.14% accuracy was obtained using SVM classifier for in-language sentiment analysis.

Akshat Bakliwal and Piyush [6] present a method of building a subjective lexicon for Hindi. Authors discussed a method of building a subjective lexicon for Hindi. Using WordNet and Breadth First Graph traversal method, they construct the subjectivity lexicon. Main contribution of their work is developing a lexicon of adjectives and adverbs with polarity scores

using Hindi WordNet and developing an annotated corpora of Hindi Product Re-views. The limitation of this system is that algorithm does not perform Word Sense Disambiguation. The proposed algorithm achieved ~79% accuracy on classification of reviews and 70.4% agreement with human annotators.
Rekha Jain [5] proposed a Hindi language opinion mining system. In this paper a Hindi language based Opinion Mining System is proposed named as "Hindi Sentiment Orientation System" based on an unsupervised dictionary approach that determine the polarity of user reviews in Hindi language. Negation is also handled in the proposed system. The experiments have been performed by using 50 sentences of movie reviews and achieved the accuracy of 65%.

Proposed System

To extract sentiment associated with Hindi documents, HindiSentiWordNet (HSWN) will be used which consists of Hindi sentiment words and their associated positive and negative polarity. Here existing HSWN is improved by adding missing sentimental words related to Hindi. For the input overall polarity is calculated; which can be positive, negative or neutral.
The proposed system consists of two stages:

1. Improving HindiSentiWordNet (HSWN) .
2. Sentiment extraction.

Our proposed approach performs Sentiment Analysis of Hindi documents using HindiSentiWordNet (HSWN). During the first stage we are improving the existing HSWN with the help of English SentiWordNet, where sentimental words which are not present in the HSWN are translated to English and then searched in English SentiWordNet to retrieve their polarity. In the second stage, sentiment is extracted by finding the overall polarity of the document; which can be positive, negative or neutral. Here during pre-processing tokens are extracted from sentence and stop words are removed. Rules are devised for handling negation and discourse relation which highly influence the sentiments expressed in the document. Finally, overall sentiment orientation of the document is determined by aggregating the polarity values of all the sentimental words in the document.

Improving HindiSentiWordNet

In this phase existing version of HindiSentiWordNet is improved, as it consists of limited numbers of words. HSWN is created using the Hindi WordNet and English SentiWordNet (SWN). During the creation of this resource for Hindi language, it is assumed that all synonyms have the same polarity while all antonyms have the reverse polarity of a word. HSWN is improved in the same way as it was developed initially.
HSWN is improved in two phases by mixing the process of automatic and manual updating of the existing HSWN.
In the first phase all the words which are tagged as adjectives, adverbs and verbs are extracted from English SentiWordNet and then converted to Hindi using BING translator [4]. Now these converted words with attached polarity are added into existing HSWN if they do not already exist in HSWN, to create new updated HSWN named as Improved Hindi SentiWordNet "Improved-HSWN".
In the second phase a dataset is created which consists of Hindi reviews extracted from web. From this dataset sentiment words are extracted which are not present in Improved-HSWN. Around 700 such words where extracted and polarity was manually assigned and then added to Improved-HSWN.
Finally an improved HSWN is available which consists of 28703 words as compared to 11941 words present in HSWN provided by IIT Bombay [3]. Improving existing SentiWordNet is important part as more the count of sentiment words with polarity then there is better chance for system to find accurate sentiment associated with the input text.

Sentiment Extraction

In this stage overall sentiment of the input document will be extracted by using Improved-HSWN. By using Improved-HSWN polarity of words present in the document are extracted one by one and aggregated together to calculate the overall polarity associated with the document. The overall polarity is then classified as positive, negative or neutral to specify the sentiment associated with the document.
This stage consists of three sub modules:

1. Pre-processing
2. Apply Negation and Discourse rules
3. Extracting Polarity

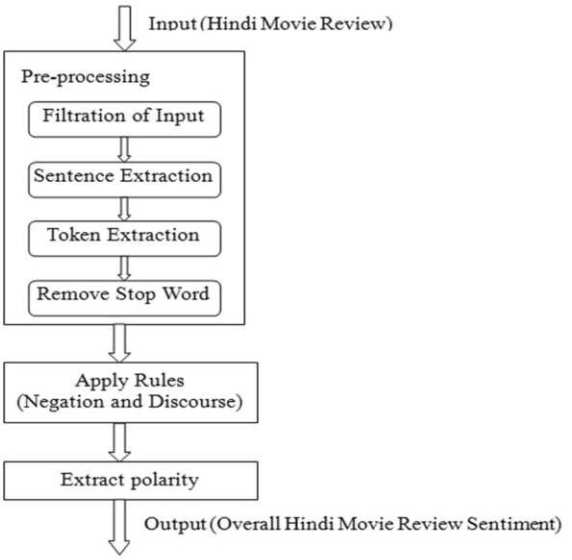


Figure 1. Proposed System

Algorithm for extracting polarity from the text

Step 1: For each token in the document.
Step 2: Check If (word is present in Improved-HSWN)
Then Retrieve polarity (POL) from Improved-HSWN and go to Step 5
Else using synset replacement algorithm replaces the sentiment words with closest meaning word present in the Improved-HSWN.
Step 4: If (no polarity assigned to word)
Then fetch next token and go to Step 2
Step 5: If (word is negated)
Then word POL= - (POL);
Else word POL remains the same
Step 5: If (word is marked for discourse)
Then word POL= weight*(POL);
Else word POL= (POL);
End For Loop and go to Step 6 when all tokens are processed.
Step 6: Compute the aggregate polarity of the document (doc POL) by adding the polarities values of the entire tokens.
Step 7: If (doc POL > zero)
Then label the document as positive
Else If (doc POL<zero)
Then label the document as negative
Else classify the document as neutral

Working of System

The input to the system is a single text document in Hindi. The system accepts a .txt file as input or user can write there reviews in the provided text area using any offline or online Hindi input tool. Working model of system is shown in Figure 2. The input to the system is a single text document in Hindi. The system accepts a .txt file as input or user can write there reviews in the provided text area using any offline or online Hindi input tool.

Improving HSWN

To improve the existing HSWN provided by IIT Bombay first complete existing English SentiWordNet is translated into Hindi using Bing translator and if translated word doesn't exist in existing HSWN then it's added to the existing HSWN.
As our target domain to extract sentiment is of Hindi Movie reviews so to improve existing HSWN by adding more movie domain related words into it we have collected overall 235 Hindi movie reviews from different online movie reviewing sites. Batch processing is performed where all unwanted characters and symbols and stop word are removed from it initially. Then words which are not found in Improved- HSWN are assigned polarity manually and added in the Improved-HSWN.

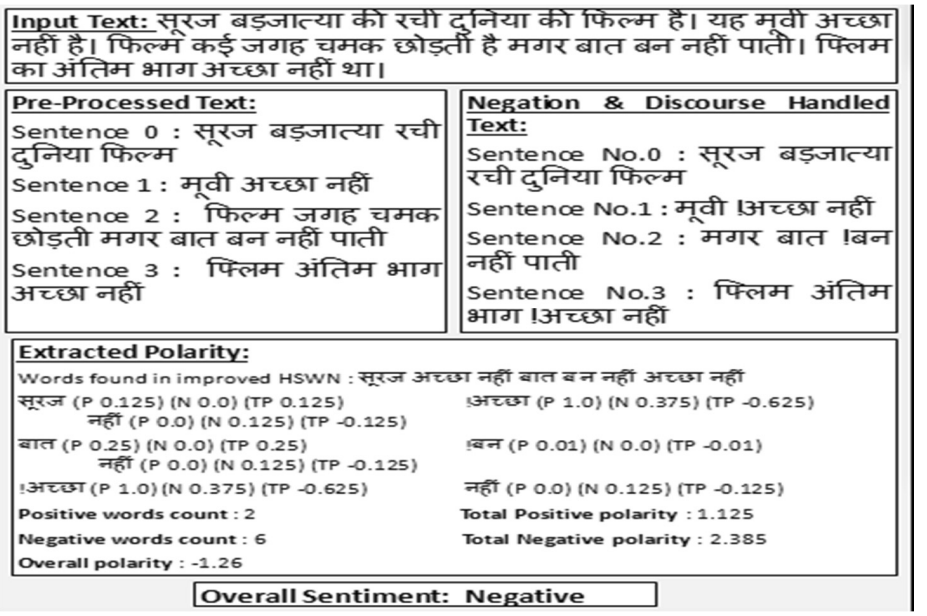


Figure 2. Working Model

Sentiment Extraction

Here in the first step input text is preprocessed to remove unwanted characters symbols and stop words from the input text. Here in filtration of input text non Devanagari Unicode characters are removed from the input text. Also symbols and punctuation and numbers expect "," and "|" are removed from the input text. Here individuals sentence are separated and stored in a list from the filtered input text. Each sentence is given a Sentence ID to it as shown in Figure 3.

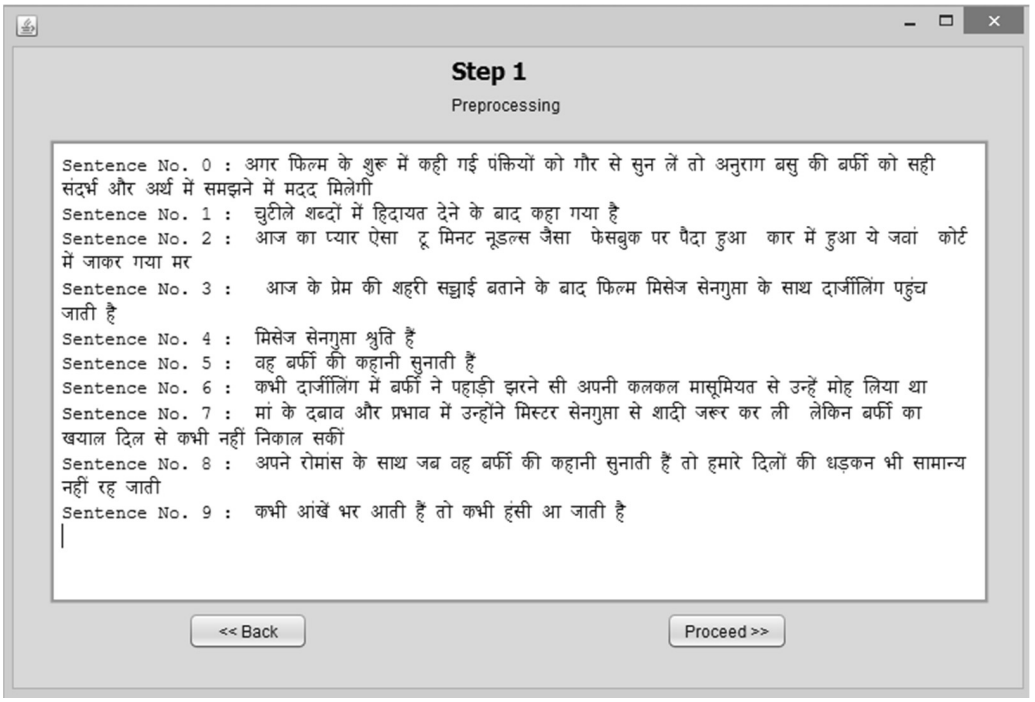


Figure 3. Filtration and extraction of sentence from input text

Stop words which don't provide any relevance to extract overall sentiment are removed from the tokenized input text.

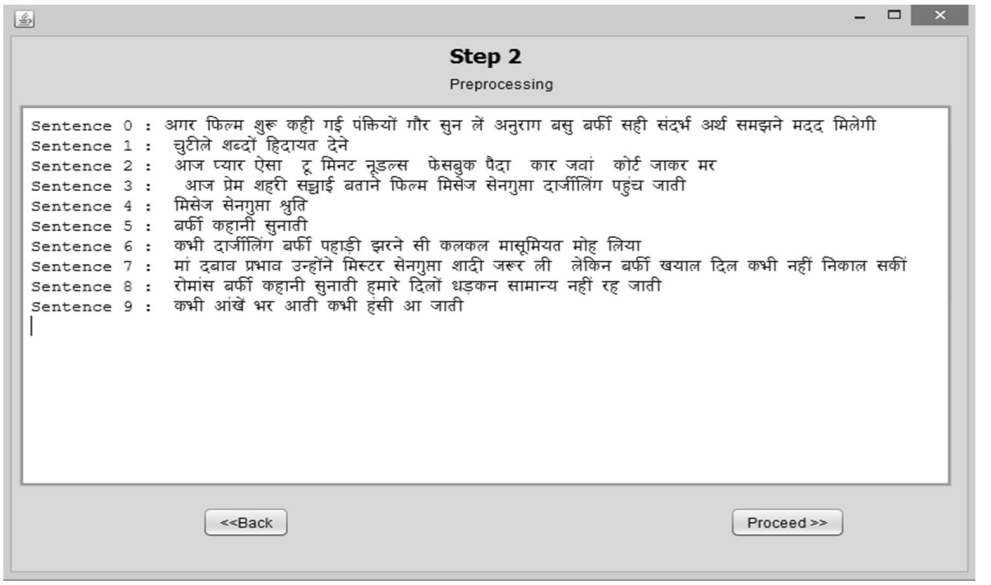


Figure 4. Stop words removed

Words which are candidate to be negated are assigned "!" mark next to it. Such negated words polarity will be reversed in further stages.

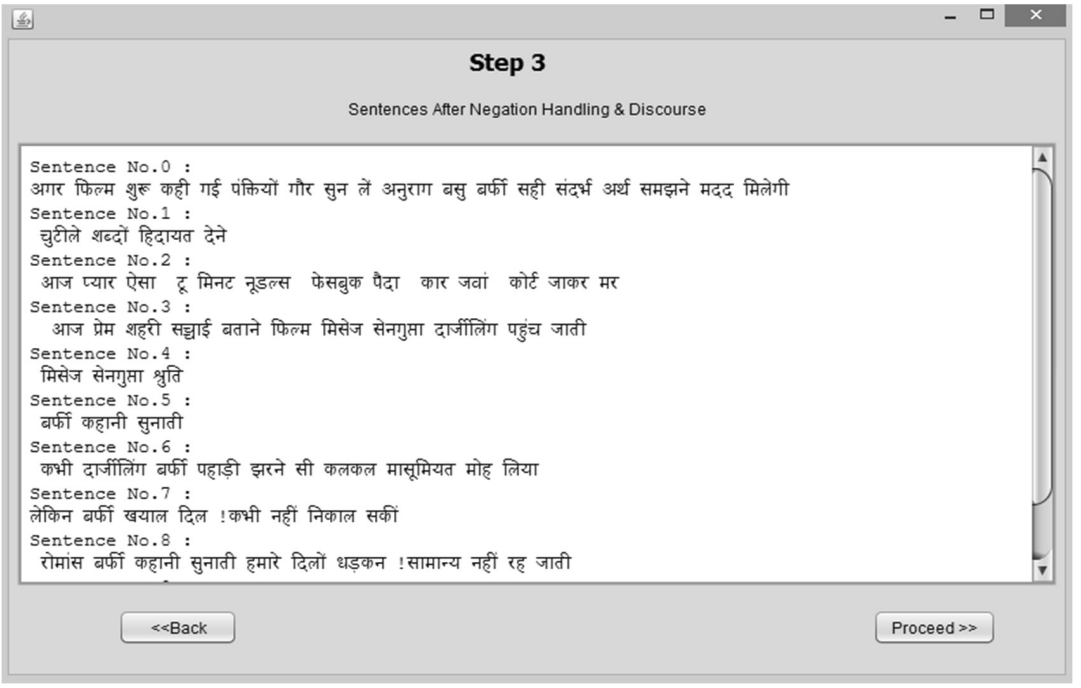


Figure 5. Negation and discourse handling

Words of which polarity is not found in Improved-HWSN are then processed using sense based synset replacement algorithm to find polarity of word by finding the word having polarity with the same synset ID as the input word.

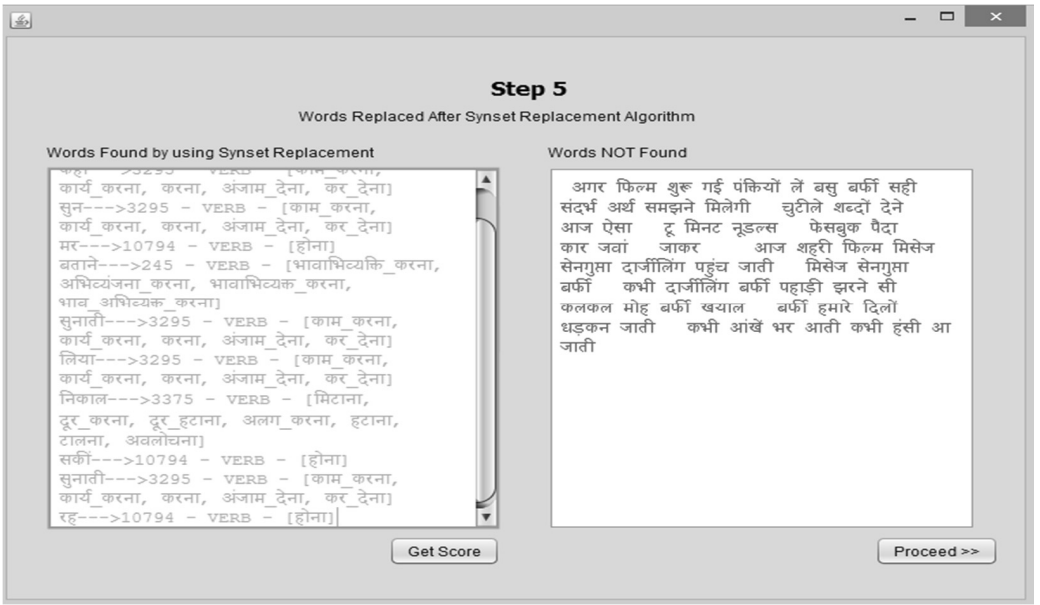


Figure 6. Synset replacement

Words are assigned polarity to it by fetching of polarity from the Improved-HSWN. Word polarity value can be positive and negative like a word “X” is assigned polarity value as positive “0.0” and negative “0.1”.

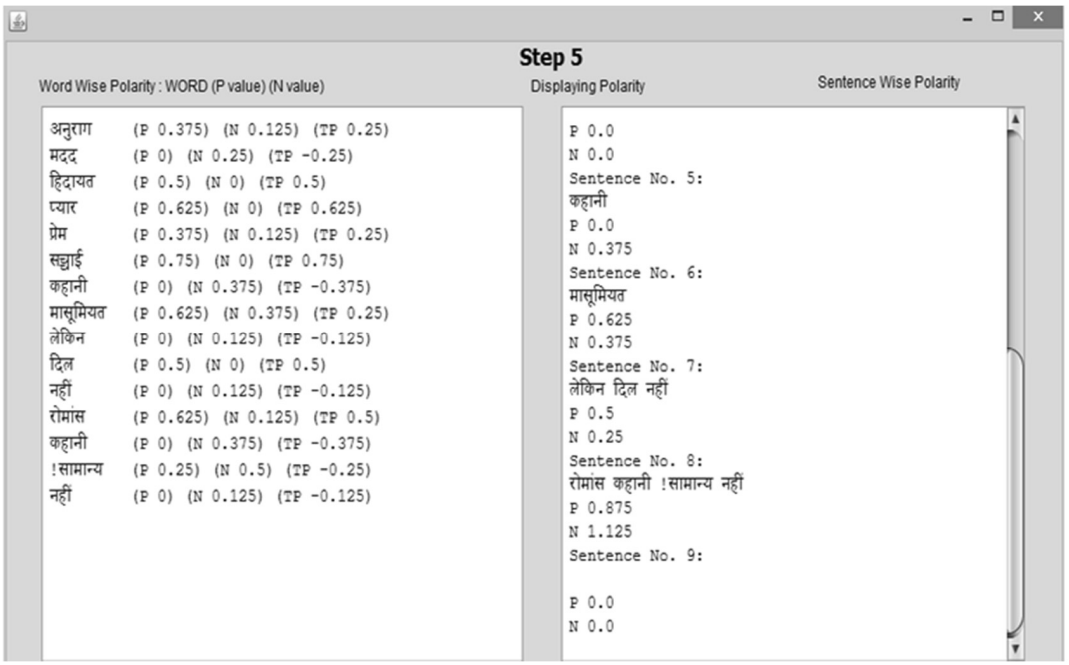


Figure 7. Polarity of each token in the input text

Finally polarities of all words are aggregated to find total overall polarity of the input text and sentiment associated with the input text is presented by classifying the polarity value as positive negative or neutral.

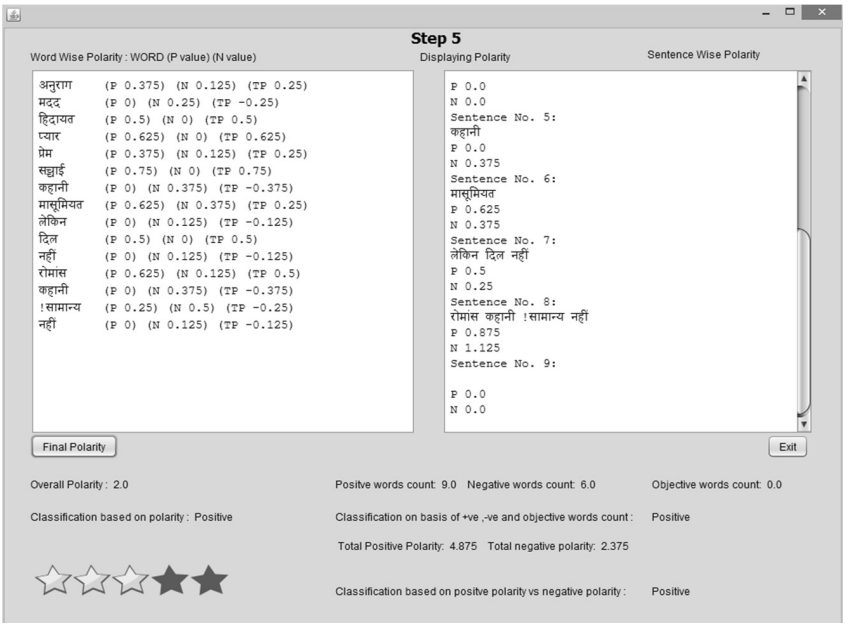


Figure 8. Sentiment classification of input text

Performance Analysis

The system is evaluated to check whether the output generated is correct or not. The calculations are performed to check whether the given input is classified correctly or not. A given input can be classified as positive, negative or neutral by the system. Performance of system is evaluated by analyzing how much accurately system extract sentiment associated with the given input.

Accuracy is the measure used here is to evaluate the system. Accuracy of the system is calculated by using following formula: Accuracy = ((Total number of documents which are classified correctly) / (Total number of document present))*100.

To calculate accuracy of system overall 90 Hindi documents having opinions or sentiments are considered of varying domain like news, people, place and technology. Also Hindi tweets are processed to extract sentiment. Out of this 90 documents around 73 documents where correctly classified.

Comparison of accuracy for 90 Hindi documents between original HSWN and Improved HSWN is show in below figure.

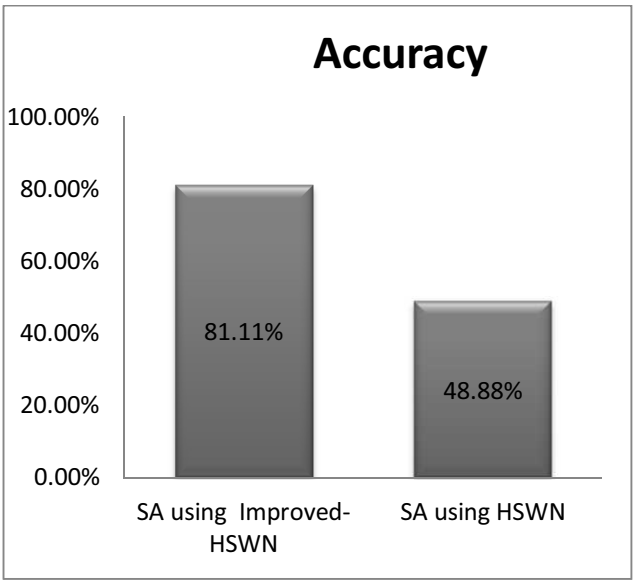


Figure 9. Comparison of system accuracy using HSWN vs. Improved-HSWN for movie reviews

Conclusion

Sentiment analysis has lead to determine the attitude or inclination of a communicator through the contextual polarity of their speaking or writing. Sentiments can be mined from texts, tweets, blogs, social media, news articles, comments or from any source of information.

Sentiment Analysis has been quite popular and has lead to building of better products, understanding user's opinion, executing and managing of business decisions. People rely and make decisions based on reviews and opinions. This research area has provided more importance to the mass opinion instead of word-of-mouth, with the system in their daily spoken natural language.

In future system can be updated to add more sentimental words to improve accuracy; also other sentiment analysis challenges like sarcasms and implicit opinion can be handled. Also tools like word sense disambiguation can be used which can help in correct extraction of word polarity based on word actual sense. The work can be extended to support other regional languages like Marathi, Kannada, Guajarati, Manipur etc.

Acknowledgment

I am using this opportunity to express my gratitude to thank all the people who contributed in some way to the work described in this paper. My deepest thanks to my project guide for giving timely inputs and giving me intellectual freedom of work. I express my thanks to the head of computer department and to the principal of Pillai Institute of Information Technology, New Panvel for extending his support.

References

[1] Namita Mittal, Basant Agarwal, Garvit Chouhan, Nitin Bania, Prateek Pareek"Sentiment Analysis of Hindi Review based on Negation and Discourse relation",International Joint Conference on Natural Language Processing,Nagoya, Japan, October 2013.

[2] Joshi, Aditya, A. R. Balamurali, and Pushpak Bhattacharyya, "A fall-back strategy for sentiment analysis in Hindi: a case study," Proceedings of the 8th ICON (2010).

[3] http://www.cfilt.iitb.ac.in/Sentiment_Analysis_Resources.html last accessed on September 13, 2015 at 5.35 p.m.

[4] http://www.microsoft.com/en-us/translator/translatorapi.aspx last accessed on October 6, 2015 at 4.02 p.m .

[5] Richa Sharma, Shweta Nigam and Rekha Jain, "Polarity Detection Of Movie Reviews In Hindi Language" International Journal on Computational Sciences & Applications (IJCSA), August 2014.

[6] Akshat Bakliwal, Piyush Arora, Vasudeva Varma, " Hindi Subjective Lexicon: A Lexical Resource for Hindi Polarity Classification", Search and Information Extraction Lab, LTRC,2012.

Regression Analysis for Stock Market Prediction using Weka Tool Without Sentiment Analysis

Sudip Padhye* and Karuna Gull**
*K.L.E. Institute of Technology, Hubli, Karnataka
sudip.padhye@gmail.com
**K.L.E. Institute of Technology, Hubli, Karnataka
karunagull74@gmail.com

Abstract: The paper deals with prediction of stock value for upcoming week, for various companies by knowing the previous weeks stock data. Stock data for different companies are collected from NSE India, which are given as input to the various regression methods using weka tool, the output of which is tabulated. By applying the numerical methods on this tabulated data we get aggregation. Thus aggregated data which is the predicted value for the successive week, is then compared with the actual data picked from NSE after completion of that successive week. Finally statistical analysis is depicted for comparison with the help of graphs.

Keywords: Regression Analysis, Weka Tool, Stock Prediction, Numerical Formula, Technical Analysis, Sentiment Analysis, Pre-processing, Weka Forecasting.

Introduction

Introduction to stock market

The Indian stock market is controlled by two regulatory bodies – NSE (National Stock Exchange) and BSE (Bombay Stock Exchange). This paper retrieves stock data only from NSE. The National Stock Exchange (NSE) is India's leading stock exchange covering various cities and towns across the country [1]. NSE was set up by leading institutions to provide a modern, fully automated screen-based trading system with national reach.NSE offers trading in Equities, Derivatives and Debt (Corporate bonds). This paper deals with only NSE Equities. The Equity stocks can be classified on the basis of stock price (Small-cap, Mid-cap & Large-cap), based on Sector of their operation (Banking, IT, Pharmaceutical, Automobile etc.) or based on stock's fundamentals (past record, turnovers etc.). Irrespective of classification, this paper put forth an approach to predict the stock prices.

Introduction to machine learning

The Machine Learning paradigms are of three types – Supervised Learning, Unsupervised Learning and Reinforcement Learning. The paper uses Supervised Learning approach as the basis for prediction. In supervised learning, initially training is given for later testing. Supervised are classified as Classification and Regression problems. The classification problem determines discrete set of values and thus is used to classify the given data. On the other hand, regression problem are used for prediction. They use continuous set of values for the same. Hence, several regression methods are used in this paper, for predicting the stock values.

Regression analysis [2] is used when anyone wants to predict a continuous dependent variable from a number of independent variables. Regression models are often constructed based on certain conditions that must be verified for the model to fit the data well, and to be able to predict accurately.

Application of machine learning to stock prediction

Fundamental and Technical methods can be used to do the prediction in stock market for different companies.

Fundamental method [3] depends on the past records and news updates of a company which is under prediction. Fall and rise of the stock of a particular company mainly depends on the good and bad news. So consideration of these strong and weak records spread by the news bulletins drastically affect the company's stock market. So more precisely we can say sentiment analysis on data sources collected from business news channels and newspaper headlines etc. with past record analysis of the company under prediction yields more accurate results.

Technical method depends on the weekly data [7]. Thus the stock value can be predicted using Weka forecast package. More specifically doing comparison of different algorithms over considerable amount of data (33 weeks data in our case) to predict future position of stock market.

Even though inclusion of fundamental with technical methods are giving the most accurate results, but here we are applying only the Technical method to predict the stock value for a set of various companies irrespective of any type (like large cap, small cap, medium cap, sectors etc.) to start with.

Literature review

As per the related work done by us, it tells why we have chosen the regression methods for analyzing and predicting the stock data collected. The summary of the literature is given below.

"Generally, higher highs and higher lows indicate an uptrend, whereas lower highs and lower lows mean a downtrend. So, look at the trend. Look at news related to the stock", says Shrikant Chouhan, senior vice president, technical research, Kotak Securities [13]. Thus, he tries to convince that Stock Predictions can be done using both Technical and Sentiment Analysis for better accuracy. The stock not only depends upon its movement trends but also on the news affecting that particular stock or the news regarding a sector(s) related to that stock.

Analysts and market experts take the help of various parameters to confirm if a stock is a trade pick. These include 200-day Moving Average, Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Fibonacci Retracement and Candle-stick price chart (Support and Resistance). Thus, onlyfew of the existing methods use Machine Learning – Regression Analysis as a technique for prediction. Moreover, unlike Regression Analysis, the existing techniques don't provide 90%+ accuracy.

According to Rupinder Kaur and Ms. Vidhu Kiran [4], the prediction was done using ANN (Artificial Neural Networks) for forecasting features like current stock price, opening price, closing price, volume, change and change percent. The System is trained using NARX Neural Networks where the errors are back propagated to the system from output layer. At the end, the predicted values are compared with the actual values using Mean Squared Error (MSE) technique by plotting the graph.

Farhad Soleimanian Gharehchopogh, Tahmineh Haddadi Bonab and Seyyed Reza Khaze [5], carried out the work by undertaking a Case Study that uses Linear Regression to predict stock market. They have discussed how the trading volume is one of the important parameter affecting the dynamics of the stock market. Using this approach, they observed 61.35% accuracy of results. Hence, linear regression approach for prediction does not always give the correct results i.e. with greater accuracy of greater than 90%.

K.K. Sureshkumar and Dr.N.M. Elango [6], forecasted Indian Stock Market Price and obtained their performance analysis by describing the usage of different neural classifiers of Weka tool. The actual stock price of Infosys technology over 1000 days had been compared with the predicted value of the Gaussian, Isotonic regression, Least mean square, Linear regression, Multilayer Perceptron functions, Pace regression, Simple linear regression and SMO regression values. Later, they applied the error evaluation rate indicators such as mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE) and root relative squared error (RRSE). They concluded that isotonic regression function offers the ability to predict stock price more accurately.

Vishal Parikh and Parth Shah [8] described Stock Prediction using Automated Trading System that shows how Stock value prediction can be done using Naïve-Bayes classifier and Random Forest classifier and then the robustness is evaluated using precision, recall and accuracy.

Raj Kumar and Anil Balara [10], made Time series forecasting of Nifty Stock Market using Weka in which SMO regression offered ability to predict more accurately than other functions such as Gaussian processes, linear regression and multilayer perception.

Methodology

Concept of Regression

In Regression problems, we take input variables and try to map the output onto a continuous expected result function [11]. Regression with one variable is known as "Uni-variate Regression" whereas regression with multiple variables is called as "Multi-variate Regression". "Uni-variate regression is used when you want to predict a single output value from a single input value. On the other hand, Multi-variate regression is used to predict from multiple input values [12].

Regression models involve the following variables [15]:

- The unknown parameters, denoted as β, which may represent a scalar or a vector.
- The independent variables, X.
- The dependent variable, Y.

In various fields of application, different terminologies are used in place of dependent and independent variables.

A regression model relates Y to a function of X and β.

$$Y \approx f(X, \beta) \quad \text{--------- (1)}$$

The approximation is usually formalized as

$$E(Y \mid X) = f(X, \beta) \quad \text{--------- (2)}$$

Various regression techniques used in this paper are,

Regression By Discretization
Regression By Discretization[16] is a regression technique that requires the class attribute of the data to be discretized or with equal width. The predicted value is the expected value for each discretized interval based on the predicted probabilities for each interval.

RBF Regressor
RBF (Radial Basis Function) Regressor[17] implements RBF Networks by training in a supervised manner by minimizing squared error with the BFGS (Broyden–Fletcher–Goldfarb–Shanno algorithm) method. All attributes are normalized into the [0,1] scale, including the class. The ridge parameter is used to penalize the size of the weights in the output layer, which implements a simple linear combination. It is also possible to use conjugate gradient descent rather than BFGS updates, which is faster for cases with many parameters, and to use normalized basis functions instead of un-normalized ones. Nominal attributes are processed using the unsupervised Nominal to Binary filter and missing values are replaced globally using Replace Missing Values.

Linear Regression
The Linear Regression[11] is a popular technique, wherein the hypothesis (decision-boundary) obtained is in the form of line.The hypothesis function has the general form:

$$h_\theta(x)=\theta_0x_0 + \theta_1x_1 + \theta_2x_2 + \ldots + \theta_nx_n= \theta^Tx \qquad \text{------ (3)}$$

where, Θ is parameter vector x is the feature vector.
The cost function, which is the difference between the predicted value and the actual value, is given by,

$$J(\theta_0, \theta_1 \ldots \theta_n) = \left(\frac{1}{2m}\right) \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 \qquad \text{------- (4)}$$

where, n is no. of features & m is no. of training sets.
Finally, the cost function can be minimized by finding minimum value of θ, which can be determined either using Gradient Descent or Normal Equation. If we compare Gradient Descent and Normal Equation then Gradient Descent requires correct choice of α and it also requires many iterations. Conversely, Normal equation doesn't contain α and also does not iterate. On the other hand, Gradient Descent can be used on any number of features (input variables) in training data, unlike the normal equation, which can be used to calculate Θ values for limited number of features(upto 10,000 features) due to complexity in computation of $(X^TX)^{-1}$. Also, the Gradient Descent finds the local minima whereas Normal Equation helps in finding Global minima.

Mutilayer Perceptron
Multilayer perceptron [18] is a type of Neural Network. It is a Classifier that uses back-propagation to classify instances. This network can be built by hand, created by an algorithm or both. The network can also be monitored and modified during training time. The nodes in this network are all sigmoid/logistic.

MLP Regressor
MLP Regressor[19] trains a multilayer perceptron with one hidden layer. All attributes, including the target attribute, are standardized. There are several parameters - the ridge parameter is used to determine the penalty on the size of the weights. The number of hidden units can also be specified using ridge parameter. Finally, it is possible to use conjugate gradient descent rather than BFGS updates, which may be faster for cases with many parameters. To improve speed, an approximate version of the logistic function is used as the default activation function for the hidden layer, but other activation functions can also be specified. In the output layer, the sigmoid function is used for classification. If the approximate sigmoid is specified for the hidden layers, it is also used for the output layer. For regression, the identity function is used activation function in the output layer. Also, if delta values in the back-propagation step are within the user-specified tolerance, the gradient is not updated for that particular instance, which saves some additional time. For large datasets, parallel calculation of loss function and gradient is possible when multiple CPU cores are present. Thus in this case, the data is split into batches and processed in separate thread.

SMOReg
The SVM (Support Vector Machine) is used for implementing SMOReg[20]. The SMOReg implements sequential minimal optimization algorithm for training a support vector regression using polynomial or RBF kernels. This implementation globally replaces all missing values and transforms nominal attributes into binary ones. It also normalizes all attributes by default. Thus, the coefficients in the output are based on the normalized/standardized data, not the original data.

Gaussian Processes
Gaussian processes[21]implements regression without hyper-parameter-tuning. This method applies normalization or standardization to the attributes. Any missing values are replaced by the global mean/mode. Also, Nominal attributes are converted to binary ones.

M5P
M5P[22]is a reconstruction of Quinlan's M5 algorithm for inducing trees of regression models.M5P combines a conventional decision tree with the possibility of linear regression functions at the nodes.M5P can deal effectively with enumerated attributes and missing values. First, a decision-tree induction algorithm is used to build a tree, for which a splitting criterion is used that minimizes the intra-subset variation in the class values down each branch, instead of maximizing the information gain at each inner node. The splitting procedure in M5P stops if the class values of all instances that reach anode vary very slightly, or only a few instances remain. Second, the tree is pruned back from each leaf. When pruning, an inner node is turned into a leaf with a regression plane. Third, to avoid sharp discontinuities between the sub-trees, a smoothing procedure is applied that combines the leaf model prediction with each node along the path back to the root, smoothing it at each of these nodes by combining it with the value predicted by the linear model for that node. Smoothing substantially increases prediction accuracy.

Packages required
Weka is a data mining tool. The Weka workbench is a collection of state-of-the-art machine learning algorithms and data pre-processing tools. It is designed so that users can quickly try out existing machine learning methods on new datasets in very flexible ways. The required packages can be installed using “Package Manager” in “Tools” menu of “Weka GUI Chooser”. The various packages required on Weka 3.7.2 or above versions, to make regression on data are:

1. Forecast package (for finding results for building and evaluating Time series forecasting models)
2. Arff converter (to convert given .xlsx file to .arff file format)
3. Linear regression
4. Gaussian regression
5. RegressionByDiscretization
6. RBFRegressor
7. MultilayerPerceptron
8. MLPRegressor
9. SMOReg
10. M5P

Note: If any of the above mentioned package is not installed, then they can be installed using Package Manager available from Weka version 3.7.2 onwards.

Steps for weka prediction

1. Open Weka in “Explorer” Mode and choose the pre-processed file.
2. Later, click on “Forecast” menu and, in “Basic configuration”, choose the outputs to predict (Avg. High and Avg. Low) in “Target Selection”. Then, select “Number of time units to forecast” as 1 (in order to make prediction of upcoming 1 week), “Time stamp” as Week and “Periodicity” as Weekly.
3. In “Advanced configuration”, choose different regression methods and click on “Start” to start the prediction by keeping the rest of the settings as default.
4. On predicting, the values for the next week will be presented, along with the Time stamp marked with *.
5. Weka also provides Visualization of the data along with the predicted values in the form of Line graph.

Flow Diagram
Collect the required data from the NSE for different companies like KPIT, Bajaj Auto, L&T etc. Pre-process the data as per our requirement like converting the daily stock data to weekly stock data. Then create the files containing pre-processed data in the form needed by regression methods. The output of different regression methods yield the predicted stock values for upcoming week. Now use the numerical method to normalize the data using the formula defined by the author in the equation (5).

Estimated value =

$$\frac{7(\text{RegressionByDiscretization}) + 6(RBFRegressor) + 5(Linear\ Regression + MultiLayerPerceptron + MLPRegressor + SMOReg) + 4(GaussianProcesses + M5P)}{41} + \delta \quad \text{----(5)}$$

Where, δ is the stochastic error factor

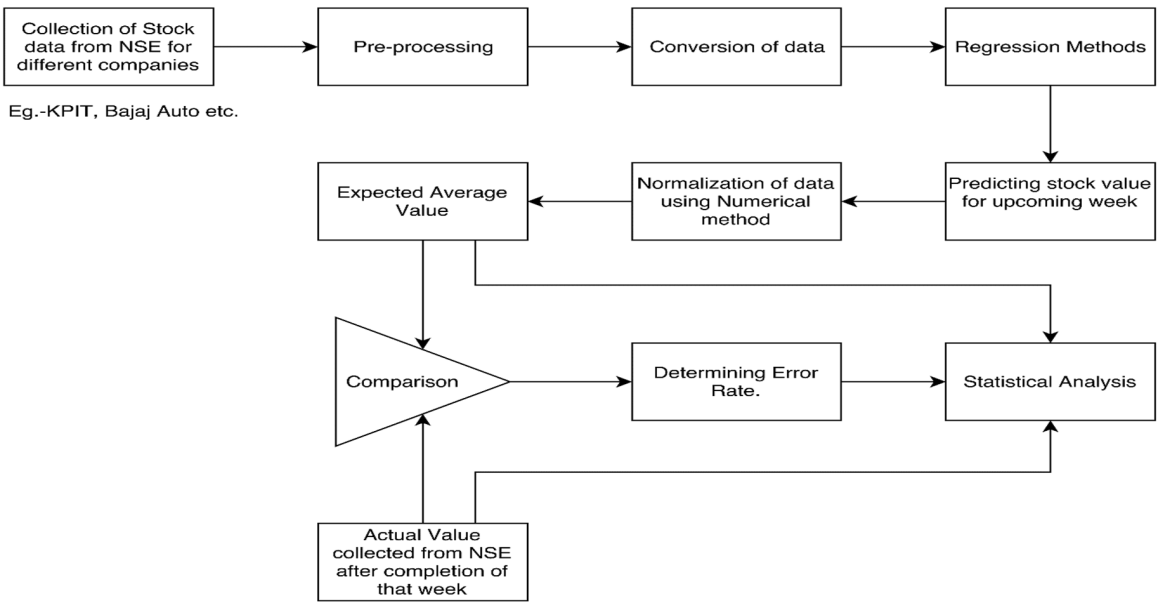


Figure 1. Technical Analysis Flow-Chart

The result of which is an expected average values. These average values are now compared with actual values picked from the NSE after completion of that week. Comparison results lead to a deviation in percentage as an error rate. Finally the graph is drawn for the average values vs. actual values and vs. the error rates.

Analysis of data

Analysis is done in following steps,

Collection of data

Raw data collected from NSE India for different companies is depicted in Table 1.

Table 1. Raw data collected from NSE for different companies

Company		KPIT		L&T		Bajaj Auto		Ashok Leyland		Axis Bank		LIC Hsg. Fin.	
Week	**Date**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**
Week 1	29/06/15	97.25	89.50	1792.7	1720	2574.05	2480.45	71.25	68.5	567.3	546	454	431.2
	30/06/15	95.80	92.95	1807.9	1768.15	2584	2531.1	72.7	71.05	566	553	456	448
	01/07/15	96.50	93.10	1818.15	1783	2555	2516	74.9	72.9	580.5	556.85	459.8	450.8
	02/07/15	95.70	93.10	1817.6	1792.15	2586.2	2528	76.4	73.65	586.65	576.1	461	450.05
	03/07/15	94.40	93.05	1823.8	1792.3	2609	2575	73.95	72.4	588.5	581	454.75	443.4
:	:	:	:	:	:	:	:	:	:	:	:	:	:
Week 33	15/02/16	123	115.5	1154.7	1067.95	2372.3	2305	90	84.7	420	395.65	432.8	414.2
	16/02/16	124	119.1	1165	1101.1	2368.7	2328.55	91	86.9	422.5	405	431.1	408.85
	17/02/16	121.3	115.2	1133.1	1082.65	2369	2263.8	90.5	85.1	406.9	394.75	414	402.05
	18/02/16	123.85	118.25	1174.65	1133.05	2417.65	2364.1	91.1	86.75	412.1	391.6	417	405.1
	19/02/16	126	117.95	1168.9	1131.4	2467	2406.15	91.5	88.85	396.9	390	421.2	406.75

Pre-processing
Raw data is pre-processed and shown in table 2.

Table 2. Pre-processing of the Raw data collected from NSE for different companies

Company	KPIT		L&T		Bajaj Auto		Ashok Leyland		Axis Bank		LIC Hsg. Fin.	
Week	Avg-High	Avg-Low	Avg-High	Avg-Low	Avg-High	Avg-Low	Avg-High	Avg-Low	Avg-High	Avg-Low	Avg-High	Avg-Low
1	95.67	93.13	1836.9	1800.68	2581.05	2533.66	73.84	71.7	577.79	562.59	457.11	444.69
2	105.68	99.78	1879.25	1855.75	2539.51	2487.59	73.15	71.18	584.39	573.39	451.39	439.65
3	114.79	108.32	1866.02	1828.04	2585.19	2505.16	74.86	73.11	596.78	583.21	468.3	456.61
4	111.44	105.93	1798.79	1755.25	2531.65	2483.64	81.15	78.76	598.04	581.71	491.31	473.06
5	130.42	122.98	1808.82	1765.94	2573.27	2511.98	84.63	82.18	573.33	559.26	497.2	483.05
	:	:	:	:	:	:	:	:	:	:	:	:
29	159.85	149.47	1175.4	1131.87	2376.54	2327.07	92.96	91.05	410.67	391.72	497.47	480.66
30	146.26	136.92	1128.4	1085.22	2308.04	2246.28	92.09	87.91	402.57	385.94	476.02	460.21
31	146.5375	141.4125	1121.85	1092.9	2337.875	2279.825	85.99	82.67	420.76	408.66	483.26	475.57
32	142.1	135.4	1144	1112.05	2382.03	2308.65	88.44	86.82	399.28	383.97	460.49	443.76
33	131.07	123.91	1128.48	1084.82	2376.49	2313.59	88.45	83.77	402.91	386.95	433.98	415.49

Output of Weka tool for various regression methods
The output of the various regression methods for highs and lows of various companies are shown in table 3 and table 4.

Table 3. Output (**High values**) ofWeka tool for various regression methods

Stock	Regression ByDiscretization	RBF Regressor	Linear Regression	MultiLayer Perceptron	MLP Regressor	SMOReg	Gaussian Processes	M5P
KPIT	133.49	130.9485	133.5284	123.0326	131.6938	127.4935	118.1059	131.0725
LT	1158.63	1115.452	1493.4073	1118.9145	1123.462	1083.2275	1078.6941	1108.3348
BAJAJ AUTO	2369.9853	2319.142	2478.3914	2406.3816	2397.206	2375.7498	2378.303	2412.5801
ASHOK LEYLAND	88.7886	92.8258	88.6225	87.8416	87.8343	89.1433	88.418	91.3648
AXIS BANK	403.8575	407.5133	402.1328	413.0255	406.7175	398.3615	387.7654	395.3627
LIC HSG. FIN.	454.5125	420.1207	359.9999	433.3918	446.707	456.1832	453.7118	432.2268

Apply the numerical method on final data collected from Weka tool
No single method gave correct predictions always, thus there was a necessity to take average of all methods. The averages allocates equal priorities to all methods, thus we decided to go for weighted averages. We carried out41 experiments, out of which results were provided 7 times by Regression By Discretization, 6 times by RBF Regressor, 5 times by Linear Regression, 5 times by Multilayer Perceptron, 5 times by MLP Regressor, 5 times by SMOReg, 4 times by Gaussian Processes and 4 times by M5P. Hence, we assigned weights to each method using their success frequency and derived the Numerical Method formula (Equation 6) shown below.

Table 4. Output (**Low values**) of Weka tool for various regression methods

Stock	RegressionBy Discretization	RBF Regressor	Linear Regression	MultiLayer Perceptron	MLP Regressor	SMO Reg	Gaussian Processes	M5P
KPIT	135.88	124.473	127.9339	115.8912	119.0986	115.1329	110.9317	129.8308
LT	1121.7567	1073.2648	1460.0694	1064.9707	1077.3879	1027.3228	1044.4092	1062.6313
BAJAJ AUTO	2267.1926	2254.574	2434.0913	2309.141	2308.8375	2313.1226	2312.4154	2332.3748
ASHOK LEYLAND	86.7543	77.4038	87.5913	78.27	82.4829	87.7574	84.9615	87.4764
AXIS BANK	387.145	390.3497	376.9289	386.3938	390.5927	380.2483	371.8556	387.9576
LIC HSG. FIN.	451.391	401.9892	333.3012	410.1261	432.5221	436.4709	438.6235	421.5869

Estimated value =

$$\frac{7(\text{RegressionByDiscretization}) + 6(RBFRegressor) + 5(LinearRegression+MultiLayerPerceptron+MLPRegressor+SMOReg) + 4(GaussianProcesses + M5P)}{41} + \delta \quad \text{----(6)}$$

Here, δ represents small variation in data which can be minimized by considering large dataset and by using Sentiment Analysis.

Comparison of aggregate value with actual NSE value in the table form

Find the aggregation of the output for all regression methods for lows and highs and also find the deviation of result from their actual values by error as shown in table 5 and 6.

Table 5. Average high and actual values and error rate for various companies

Stock	Estimated Avg. High	Actual Avg. High	% Error
KPIT	129.16	123.63	4.47
LT	1162.10	1159.27	0.24
BAJAJ AUTO	2389.19	2398.93	-0.41
ASHOK LEYLAND	89.39	90.82	-1.58
AXIS BANK	402.58	411.68	-2.21
LIC HSG. FIN.	432.38	423.22	2.16

Table 6. Average low and actual values and error rate for various companies

Stock	Estimated Avg. Low	Actual Avg. Low	% Error
KPIT	123.20	117.2	5.12
LT	1118.75	1103.23	1.41
BAJAJ AUTO	2312.27	2333.52	-0.91
ASHOK LEYLAND	83.95	86.46	-2.90
AXIS BANK	384.44	395.4	-2.77
LIC HSG. FIN.	416.45	407.39	2.22

Experimental Results

The statistical analysis of the experimental results Vs Actual values and Error rates were depicted using various figures Fig 2 to Fig 9.

Graphs for various regression method

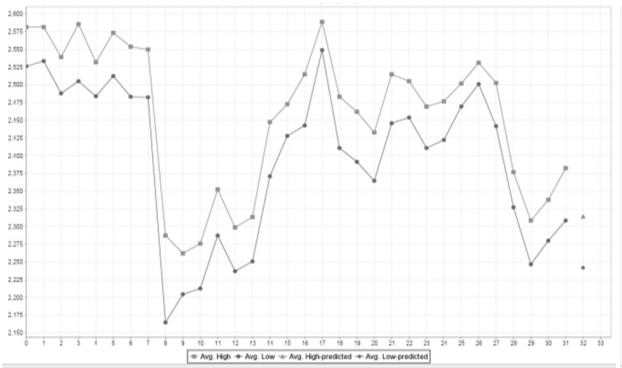


Figure 2. Bajaj auto predicted data – RBFRegressor

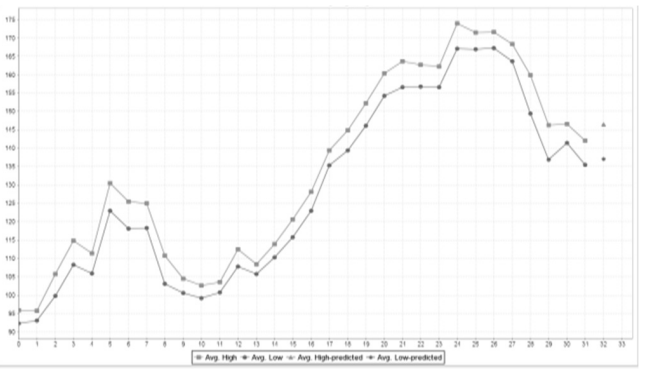


Figure 3. KPIT predicted data – RegressionByDiscretization

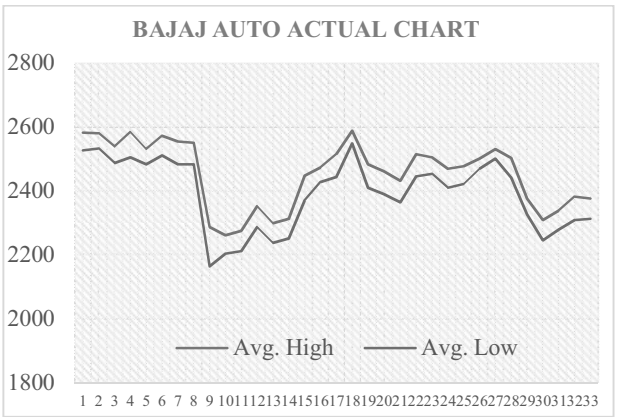


Figure 4. Bajaj Auto Actual Data

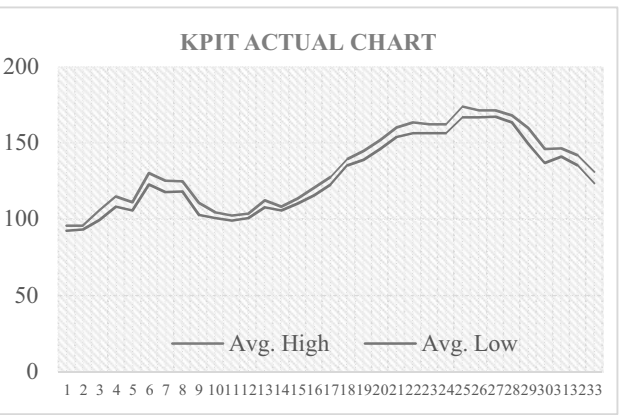


Figure 5. KPIT Actual Data

Bar chart for comparison of estimated value and actual value

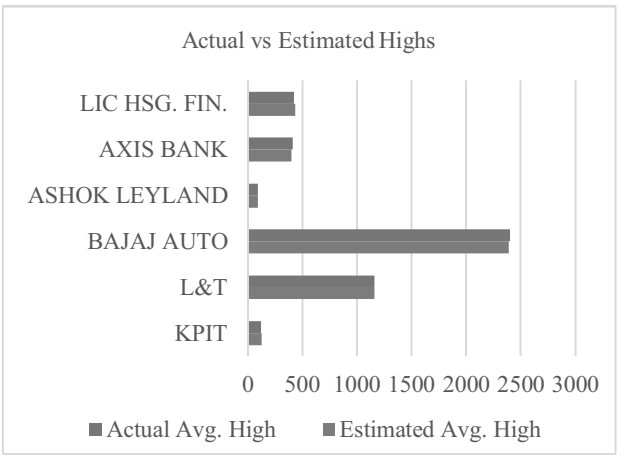


Figure 6. Actual vs. Estimated Highs

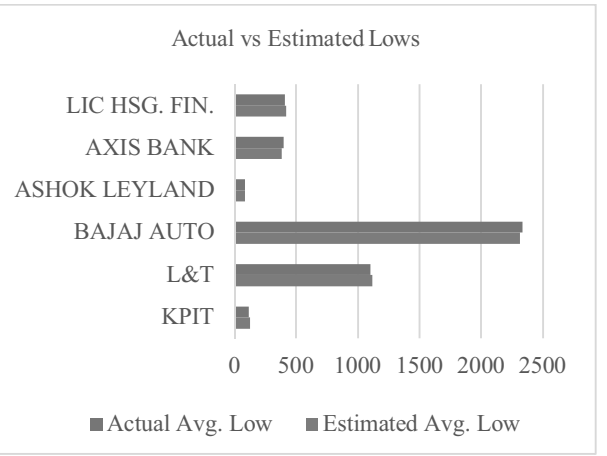


Figure 7. Actual vs. Estimated Lows

Column chart for %Error

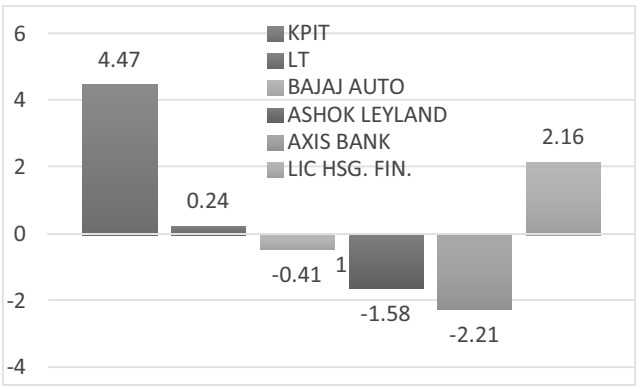


Figure 8. Error% for Highs

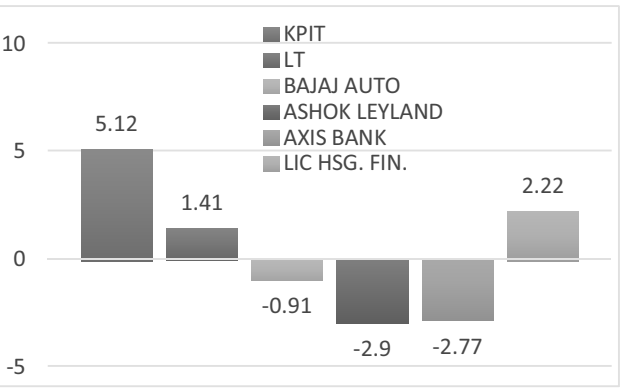


Figure 9. Error% for Lows

Performance Evaluation

Table 7. Performance Evaluation by computing error

Functions	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)	Relative Absolute Error (RAE)	Root Relative Squared Error (RRSE)
RegressionByDiscretization	20.1616	28.9586	2.68%	2.64%
RBF Regressor	21.8251	34.9426	2.89%	3.27%
Linear Regression	52.2119	106.5071	6.93%	9.69%
MultiLayer Perceptron	16.3219	22.9255	2.17%	2.09%
MLP Regressor	14.3147	19.4126	1.90%	1.78%
SMOReg	21.2613	31.9352	2.82%	2.91%
Gaussian Processes	23.5843	21.7965	3.13%	3.01%
M5P	15.1037	32.557	2.01%	1.98%
Numerical estimation	**10.1396**	**13.1145**	**1.35%**	**1.19%**

At the end, the performance of the predictor is calculated by comparing the estimated value with the actual value. There are various methods for evaluation such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE) and Root Relative Squared Error (RRSE). Formulas for calculating these errors are as given below

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - x_i'| \quad \text{..... (7)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x_i')^2}{n}} \quad \text{..... (8)}$$

$$RAE = \frac{\sum_{i=1}^{n}|x_i - x_i'|}{\sum_{i=1}^{n}x_i} \quad \text{..... (9)}$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x_i')^2}{\sum_{i=1}^{n}x_i^2}} \quad \text{..... (10)}$$

where, x_i is actual stock value and x_i' is the predicted stock value.
The Relative errors are represented w.r.t. the standard value (actual stock value) whereas the rest are computed w.r.t. total no. of values (n). Out of these four methods, RMSE and RAE are very common among several literatures. The MAE and RMSE are represented in decimal format while RAE and RRSE are represented in percentage. Lesser the error, more is the accuracy of it. Thus, our Numerical Estimation Formula shows least error for all four methods as compared to other individual regression functions.

Conclusion and future work

To conclude we say that only few of the existing methods use Machine Learning – Regression Analysis as a technique for prediction. But however the expected results from them is not as accurate as we got i.e. those techniques don't provide 90%+ accuracy. So taking these as main drawbacks of existing ones, we started by considering two methods fundamental and

technical to predict the stock values for upcoming week. But we end up with a completion ofonly the Technical method for prediction in depth.

Technical method depends on the weekly data and the Weka forecast package. More specifically the comparison is done on the data collected for different companies using various regression methods to predict the future position of the stock market. This paper began by stating that irrespective of the type or the sector of the stock, predictions with less errors can be made. In our consideration of 6 stocks, KPIT belongs to IT Sector, Ashok Leyland and Bajaj Auto belongs to automotive sector, Axis Bank is a banking stock whereas LIC Housing Finance is a Finance Stock & L&T is a diversified stock company. Thus, in spite of this variation in nature of stocks, our Numerical method formula proved successful with minimal error. The comparison result of the algorithms had given more accurate results i.e. by showing less deviation from actual values collected after completion of the week under prediction. Statistical analysis drawn show the predicted results and accuracy of the methods defined.

As a future work if Fundamental and Technical methods are combined and used for prediction of stock market position of any company they give the most accurate results. Because Fundamental method depends on the past records and news updates of a company which is under prediction. Fall and rise of the stock of a particular company mainly depends on the good and bad news. So consideration of news data really affect the company's stock market. Thus addition of sentiment analysis [9] on data collected from different news media along with past history analysis of the company under prediction will give way to most accurate results.

References

[1] Oscar Alsing, OktayBahceci, "Stock Market Prediction Using Social Media Analysis", Thesis,Computer Science, stockholm, sweden 2015.

[2] Kim-Georg Aase, "Text Mining of News Articles for Stock Price Predictions", Thesis, Master of Science in Computer Science, Norwegian University of Science and Technology, 2011.

[3] Ruchi Desai, Prof.Snehal Gandhi, "Stock Market Prediction Using Data Mining", International Journal of Engineering Development and Research (IJEDR) - 2014 , Volume 2, Issue 2, ISSN: 2321-9939, pp 2780-2784.

[4] Rupinder kaur1, Ms.VidhuKiran, "Efficient Artificial Neural Network based Practical Approach of Stock Market Forecasting", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 8, August 2015, DOI 10.17148/IJARCCE.2015.4855,ISSN (Online) 2278-1021, ISSN (Print) 2319 5940. Pp-263-267.

[5] FarhadSoleimanianGharehchopogh, TahminehHaddadiBonab and Seyyed Reza Khaze, "A Linear Regression Approach To Prediction Of Stock Market Trading Volume: A Case Study", International Journal Of Managing Value And Supply Chains (IJMVSC) Vol.4, No. 3, September 2013, DOI: 10.5121/ijmvsc.2013.4303, pp-25-31.

[6] K.K.Sureshkumar ,Dr.N.M.Elango , "An Efficient Approach to Forecast Indian Stock Market Price and their Performance Analysis", International Journal of Computer Applications (0975 – 8887) Volume 34– No.5, November 2011, pp-44-49.

[7] PierpaoloDondio "Stock Market Prediction Without sentiment Analysis: Using a Web-Traffic Based Classifier and User-Level Analysis", 46th Hawaii International Conference on System Sciences, 2013, DOI 10.1109/HICSS.2013.498 ,1530-1605/12 $26.00 © 2012 IEEE. pp- 3135- 3144.

[8] Vishal Parikh, Parth Shah, "Stock Prediction and Automated Trading System", IJCSC, Vol- 6 • Issue-1 Sep - Mar 2015 pp.104-111 available at www.csjournalss.com.

[9] Yoosin Kim, SeungRyulJeong, Imran Ghani, "Text Opinion Mining to Analyze News for Stock Market Prediction", Int. J. Advance. Soft Comput. Appl., Vol. 6, No. 1, March 2014, ISSN 2074-8523; Copyright © SCRG Publication, 2014

[10] Raj Kumar, Anil Balara, "Time Series Forecasting Of Nifty Stock Market Using Weka", JRPS International Journal for Research Publication & Seminar Vol 05 Issue 02 March -July 2014, pp-1-6.

[11] https://www.coursera.org/learn/machine-learning/home/welcome

[12] https://onlinecourses.nptel.ac.in/noc16_cs03/preview

[13] http://www.businesstoday.in/moneytoday/stocks/technical-analysis-tools/story/21155.html

[14] http://www.nseindia.com

[15] https://en.wikipedia.org/wiki/Regression_analysis

[16] http://weka.sourceforge.net/packageMetaData/regressionByDiscretization/index.html

[17] http://wiki.pentaho.com/display/DATAMINING/RBFRegressor

[18] http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/MultilayerPerceptron.html

[19] http://weka.sourceforge.net/doc.packages/multiLayerPerceptrons/weka/classifiers/functions/MLPRegressor.html

[20] http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/SMOreg.html

[21] http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/GaussianProcesses.html

[22] http://www.opentox.org/dev/documentation/components/m5p

Component based Face Recognition using Hybrid Approach

R Srinivasa Perumal* and Chandra Mouli P.V.S.S.R**
*School of Information Technology and Engineering
r.srinivasaperumal@vit.ac.in
**School of computing Science and Engineering
chandramouli@vit.ac.in

Abstract: This paper presents a hybrid framework for face recognition based on the component features that are extracted by the salient points from the facial components of the face. Face recognition is still challenging under different pose, illumination conditions etc. The Proposed method detects the face from the original image. The components (eyes, nose and mouth) are extracted from the detected face through Viola-Jones algorithm. Extract the component features using Speeded up Robust Features (SURF). Extracted features are concatenated into a single feature vector to represent the face and compute the similarity between the query face and feature vectors stored in the database. The proposed hybrid method is implemented on FERET and ORL datasets and it improves the recognition rate and decreases the system running time.

Keywords: Viola Jones Algorithm, SURF, Face Detection, Hybrid face detection method, component based face detection.

Introduction

Face detection plays a vital role in automated human recognition. The methods are broadly categorized into two categories namely holistic approaches and Component based approaches. The Holistic face approaches utilize global information from face which is represented by features. Local Binary Patterns [7], biologically inspired features [16], Scale Invariant Feature Transform [21] and appearance-based representations [2] are some of the holistic approaches. In component based representations, features are extracted from specific facial components. Using the features from the facial components, the similarity measure is calculated and the decision is chosen based on the similarity score. Component based face recognition is the order of the day.

Schwaninger et al. [14] proved that a human has the ability to identify the faces even with different facial expressions, pose invariant and occluded image through natural intelligence. The current evidence [5] shows that idea supports the face processing to integrate each individual component of the person. The facial components like eyes, nose and mouth are integrated by human and compared with optimal Bayesian integrator based on component Processing. The comparison results show that the Bayesian framework performs slightly better than human performance. In addition, the behavioral characteristics of human components will be taken into account, such as outline of the face, eyes, nose and mouth. The reason behind the selection of these characteristics is that the upper part of the face contains more useful information than the lower part. The existing approaches that fall under this category are Template Matching [1], Elastic Bunch Graph Matching (EBGM) [3], Linear Discriminant Analysis (LDA) [8, 11], Cascaded LDA [10], Local Binary Pattern (LBP) [13], and Appearance based Method [23].

Template matching techniques fall under the component based approaches. R. Brunelli and T. Poggio [1] proposed a template matching based face recognition. In this method, facial components (eyes, eyebrows, nose and mouth) are extracted using geometrical features. From each component, features are extracted based on the intensity value lying in the template and generate set of templates for each face. A set of 35 features are extracted for each facial component and finally face is recognized using Bayes classifier. EBGM [3] computes the fiducial points for each face that is characterized by Gabor wavelet coefficients. The fiducial points and their relative positions are extracted and formed into an image graph. This method recognizes the person from known class and the performance will be decreased if huge variation occurs between the images.

Local Discriminant Analysis (LDA) [8, 11] explores the feature space which maximizes between-class scatter and minimizes within-class scatter. The face is divided into five regions such as lower part and upper part of mouth, nose and left and right eye region. Each region describes the discriminant information. Then it constructs the component bunches in order to discriminate information. It provides higher recognition rate under pose change and lighting conditions. Cascaded LDA [10] method divides the face into four overlapped regions. The features are extracted from overlapped regions of face through Eigen vector and determine the discriminating information of overlapped regions.

Lin and Tang [12] introduced a framework for high resolution face recognition manipulating features in multiple scales. Here the face image is factorized into four layers: global appearance, facial organs, skins, and irregular details. This system applied Multilevel PCA followed by Regularized LDA to model global appearance and facial organs. However, the system is not suitable for representing of irregular details and texture of skin. Lin and Tang address the issue by using SIFT and Multiscale texton features are used to describe the texture of skin and irregular details. To increase performance the systems combine the information conveyed by all layers. K. Pan et al. [13] proposed Local Binary Patterns for extraction of local features. The purpose of LBP is used to extract the texture information and to enhance the feature vectors. LBP feature vectors are adjusted with monotonic transform for higher recognition rate at different illumination conditions. In case of varying pose and facial expressions, the accuracy of LBP fails.

Heisele et al. [15] proposed a component based face detection using 3D models, it determine a set of templates which is specific to faces consisting of the face components and their relations. Component based face detection and verification using online learning based on unsupervised clustering. It provides more accurate results and reduces the execution time. This system is more flexible to use to describe the features in appearance based method as well as geometry based method. Appearance-based [23] used anthropometric metric to extract the components. Active shape model identify the location of facial components such as eyes, eyebrows, mouth and nose. The features are extracted by multi scale local binary pattern. This method improves the recognition rate in pose variation as well as for occluded images.

The above existing methods extract the facial features approximately by predefined position of the face. Instead of using predefined position of the face the hybrid approach identifies the most appropriate method for determining facial component. In component based representation, face images are divided into different facial components and then compute the individual feature vectors for each facial component. The extracted features are finally concatenated. The component based framework demonstrates the robustness in orientation, occlusions and recognition accuracies by learning algorithms. Table.1 summarizes some studies done in component based representation approach for face recognition. The rest of the paper is organized as follows. In Section 2, the proposed method is explained. Experimental results and performance analysis are presented in Section 3 and finally Section 4 concludes the work.

Table 1. Summarization of component based face recognition methods

Description	Component Extraction Method	Component(s) used	Component representation	Database
Template Matching [1]	Rigid	Eyes with Eyebrows, nose and mouth	Pixel representation and Laplacian	Private [6]
EBGM [3]	EBGM	N/A	Gabor wavelet Coefficients	FERET, Bochum
Component based LDA method with component Bunches [8]	Rigid	Left eye, Right eye, nose, left mouth and right mouth	Pixel representation	FERET
Component based Cascade LDA [10]	Rigid	Four overlapped regions from whole face	Pixel representation	FERET
Component based LDA Face Description [11]	Rigid	14 components across whole face	Pixel representation	MPEG-7, XM2VTS, FERET
Framework for High resolution face recognition [12]	Rigid	Left eye, Right eye, nose, mouth, forehead, skin, irregularities	Pixel representation	Private
Part based Face Recognition [13]	Rigid	10 components	LBP	Near Infrared
Component based face identification by 3DModels [15]	Reference points on 3D head models	14 learned components	Histogram equalized gray values	3D face database
Component based representation in automated face recognition [23]	Rigid	Eyebrows, Eyes, nose and mouth	MLBP	LFW, FERET

Proposed System

The proposed system is considered as a hybrid approach as it combines the Viola Jones algorithm and SURF method. In the first step, the face and the facial components are detected by Viola Jones algorithm. In the second step, features are extracted from each component through SURF descriptor. The Key points are described for facial components. The feature vectors of facial components are concatenated into one feature vector. The concatenated feature vector is matched with everyone feature vector that was stored in the database. Finally the decision is taken based on the similarity measure. The architecture of the proposed system is shown in Fig 1. The process in detail is discussed in the subsequent subsections.

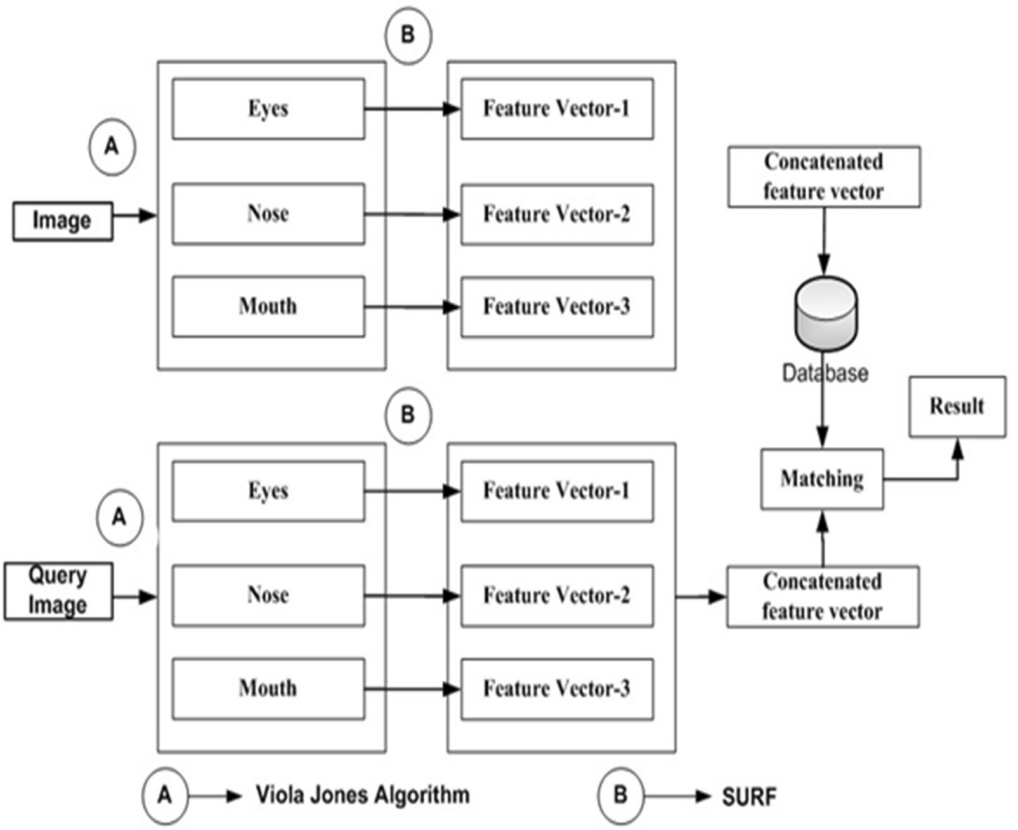


Figure 1. Architecture of the proposed method

The architecture of the proposed system is shown in Fig 1. In Fig 1.A shows the three components that are detected by Viola jones algorithm. In Fig 1.B, SURF descriptor is applied in each component to extract the feature vectors from image. Interest points are extracted from face image such as histogram equalization and normalization. SURF extracts 15 -25 fiducial points per component. Similarly for each component 15-25 fiducial points are extracted. Finally the feature vectors of every component are fused to form a single feature vector for the face image and it is matched with the feature vectors stored in the database to identify the person is recognized person or unknown person.

Component classification by Viola Jones Algorithm

The basic idea of Viola Jones algorithm is to scan a sub-window capable of detecting faces. The traditional image processing could be rescale into different sizes and then run the constant size detector through these images. It is time consuming approach. In Viola Jones algorithm rescale the detector instead of the input image and execute the detector 'n' number of times through the image, each time with a different size. Viola Jones algorithm is scale invariant and time consuming approach. Viola Jones algorithm is a real time facial component detection method based on simple features. They used rectangle features for computing the sum and difference of the pairs in the rectangular sub-image. The windows are scaled to detect the components from face [9]. This method is invariant to geometrical operations like rotation, scaling as well as invariant to illumination conditions. The algorithm has three major steps. They are

(i) Scale invariant detection
(ii) Feature selection using Learning Classification Function
(iii) Cascading classifiers to detect the facial component

Scale invariant detector
The first step of the Viola-Jones face detection algorithm is to convert the input image into an integral image. The conversion of input image into integral image is by making each pixel equal to the entire sum of all pixels above and the left of the concerned pixel. It computes the sum of all pixels inside any given rectangle using only four values. These values are the pixels in the integral image that coincide with the corners of the rectangle in the input image. To achieve scale invariant detection, rectangular features are used, with a new image representation defined as integral image. The integral image is defined in equation (1).

$$ii(x,y) = \sum_{x' \le x, y' \le y} i(x,y) \tag{1}$$

where ii(x,y) is the integral image and i(x,y) is the original image. For example

1	1	1
1	1	1
1	1	1

Input Image

1	2	3
2	4	6
3	6	9

Integral Image

Selecting the features with Learning Classification function
The feature set can be classified using Adaboost algorithm. AdaBoost is a machine learning boosting algorithm capable of constructing a strong classifier through a weighted combination of weak classifiers. It is strong and guaranteed classified algorithm. The feature set is classified based on the weighted majority vote. A weak classifier (h(x, f, p, θ)) is defined in equation (2).

$$h(x,f,p,\theta) = \begin{cases} 1 & if\ pf(x) < p\theta \\ 0 & Otherwise \end{cases} \tag{2}$$

where x is a sub window, f is the feature, θ is threshold and p is polarity indicating the direction of inequality.
Select the best weak classifier with respect to their weight in order to detect the strong classifier because it is time consuming process rather than selecting a strong classifier.

Cascade classifiers to detect the face and facial component
The detector scans the image several times with different size. The cascaded classifier discards the non-face region instead of determining the face. The cascaded classifier is composed of stages each containing a strong classifier. Each stage is to determine whether a given sub-window is definitely not a face or non- face region. When a sub-window is classified to be a non-face by a given stage it is immediately discarded. Conversely a sub-window classified as a non-face is passed on to the next stage in the cascade. It follows that the more stages a given sub-window passes, the higher the chance the sub-window actually contains a face. Similarly the components are also determined. This algorithm will help to reconstruct the classifier which achieves good performance and reduces the time.

Feature Extraction for component by SURF
SURF [18] is a method to extract features. The feature has 64 dimensions and it is invariant to rotation, contrast, brightness and scale. It is based on Haar wavelet and use integral images. It utilizes an integer approximation to the determinant of Hessian matrix, which can be computed extremely fast with the use of integral image. For features, it uses the sum of Haar wavelet responses around the point of interest. Feature vectors can be detected through salient points from the image. The Algorithm contains following steps.

1. Finding interest point from the image.
2. Find most important interest points in scale space by approximate Laplacian of Gaussian.
3. Find Feature description in horizontal and vertical direction
4. Generate feature vector

Interest point from the image

The interest points can be extracted through the SURF for facial component. The original image is convolved with Gaussian kernel to determine the determinant of Hessian Matrix is defined in equation (3).

$$H(x,y,\sigma) = \frac{\mathrm{Gxx(x,y,\sigma).Gyy(x,y,\sigma)} - Gxy(x,y,\sigma)^2}{\sigma^2} \quad (3)$$

where Gxx(x,y,σ) is defined in equation (4).

$$Gxx(x,y,\sigma) = \frac{\partial N(0,\sigma)^2}{\partial i.\partial j} * image(x,y) \quad (4)$$

The local maxima filter occurs in facial component with gradient in multiple directions. The integral image location x = (x, y)^T contains the aggregate sum of pixels within the bounded (rectangular) region and opposite corners from the current position of the pixels. The integral image is defined in equation (5).

$$I(x) = \sum_{i=0}^{i\leq x}\sum_{j=0}^{j\leq y} I(x,y) \quad (5)$$

The interest points are localized in scale space. Once the integral image is computed, it takes three additions to calculate the sum of intensities over rectangular area because the calculation time is independent of its size.

Scale Space Representation

Using scale space value, the feature vectors are generated from the most significant point from the image. These points can be calculated rapidly through the integral image. Due to use of box filter and integral images, the filter can be any size at exactly the speed on the image. Hence the scale space is analyzed by up-scaling the filter size rather than iteratively reducing the image size. The resultant of 9 x 9 filter, is considered as the initial scene layer, to which we will refer as scale s = 1.2 (corresponding to Gaussian derivatives with σ = 1.2). The layers are attained by filtering the image with steadily bigger mask. At larger scales, the step between consecutive filter sizes should also scale accordingly. The motivation for this sampling is computational efficiency.

The scale space is divided into octaves. An octave characterises of filter response maps achieved by convolving the original input image with a filter of increasing size. Each octave is subdivided into a constant number of scale levels because of the discrete nature of integral images; the minimum scale difference between 2 subsequent scales depends on the length l0 of the positive or negative lobes of the partial second order derivative in the direction of derivation (x or y). To localize the most interest point in the original image over scales, non-maximum suppression in a 3 X 3 X 3 neighbourhood is applied. The maxima of the determinant of the Hessian matrix are then interpolated in scale and image space [25].

Feature Description

SURF describes the features by sum up the pixel information of neighbors. First, it determines the orientation of each feature by convolving in its neighborhoods with horizontal and vertical Haar wavelet filter. The filter is shown in Fig 2. These filters help to determine the directional derivatives of the image's intensity. From the edge responses, the descriptor describes the specific orientation of images. The rotation invariant property allows identifying the descriptor accurately than existing methods.

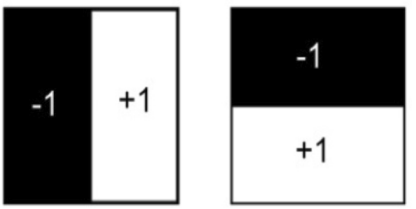


Figure 2. Haar Wavelet Filters

The edge gradient responses are taken from dark versus light, therefore the responses to generate a feature vector representing each component feature and its neighborhood. This property allows improving the recognition rate in illumination conditions. So the neighborhood is divided into smaller regions randomly. From each region, the features are extracted and it is summed up to form single feature vector.

Feature Vector

SURF describes the feature vector of image based on sum of Harr coefficients. Construct a 4 x 4 rectangular region that preserves the information. Apply Haar wavelet transform in scale space to extract the facial component descriptor in both directions (i.e. horizontal and vertical direction). The sum of Haar responses dx and dy for each region are generated and considered as a feature vector. The Sum of pixels for each region feature vector is denoted as |dx| and |dy|. Concatenate the Haar response vector and sum of pixel vector into a single feature vector is defined in equation (6).

$$V = \left(\sum d_x \sum d_y \sum |d_x| \sum |d_y|\right) \tag{6}$$

Feature Matching

Face Recognition determine a database features for each image in the standard database. When given a query image at runtime, the system generates the features of query image and its compare the features within the database using distance metrics. If the features are similar then the system concludes that the query image is present in the database. In component based representation the features are extracted from each component. The extracted features of component are concatenated into a feature vector for face. The features are stored in the database. Given a query image, the features are generated based on the process of extraction. Euclidean distance measure is used for determining the distance between features of query image and features of database images. If the distance between the query image features and the database image features is very close then the person is considered as matched. If there are "n" persons are identified as closer, then it is difficult to identify the exact person. In this case Euclidean distance gives ambiguous results. To overcome such ambiguities, the ratio of the number of features pairs found to the sum of the squared distance is employed for accurate detection. This ratio is defined in equation (7).

$$S_R = \frac{N_i}{\sum_{j=1}^{j=N_i} d_{ij}^2} \tag{7}$$

where Ni is no of feature pairs identified and d^2_{ij} is sum of squared distance.

Experimental Results and Performance Analysis

The experimental results show the effectiveness of component processing in automated face recognition. In this work the face images are detected in two ways. First, the global approach which represents the feature vector from whole face and second, the component approach which represents the separate feature vector for each facial component. The current research is mostly based on the global approach in face recognition. If the face image size increased in global approach then the accuracy is low comparatively with component approach. The proposed system used component based representation for detecting the components. The facial component is considered as a block and the features are extracted from each block. In Fig 3, the first row represents some sample images taken from FERET [4] database. The second row represents the face detected using Viola Jones algorithm. The third, fourth and fifth rows represent the components extracted from each face. The proposed method is implemented on benchmark datasets like FERET and ORL database.

Table 2. Recognition rate on FERET Database

Method	FERET 1a	FERET 1b
PCA	0.83	0.85
LDA	0.89	0.96
SIFT	0.93	0.94
PFD-SIFT	0.97	0.98
RS-LDA	0.99	0.99
Proposed Method	0.99	0.99

Analysis on FERET database

The FERET database [4] consists of 2388 images and the dataset comprised into 1194 persons. Select 2 images from Fa and Fb per person. Faces are cropped by viola jones algorithm and reshape the image size into 100 X100. From the dataset 250 people images are selected randomly for training and rest of the images are used for testing. (FERET 1a) [24] In FERET 1b, 497 images are randomly selected and it is used for training and the rest of the images are used for training. The proposed method applied to the database to estimate the performance. Table 2 and Fig. 4 show the recognition rate of holistic and component based approaches compared with the proposed system. From Table 3, it is observed that RS-LDA and the proposed method have similar recognition rates. However, the proposed method has lower running time than RS-LDA. Fig.5

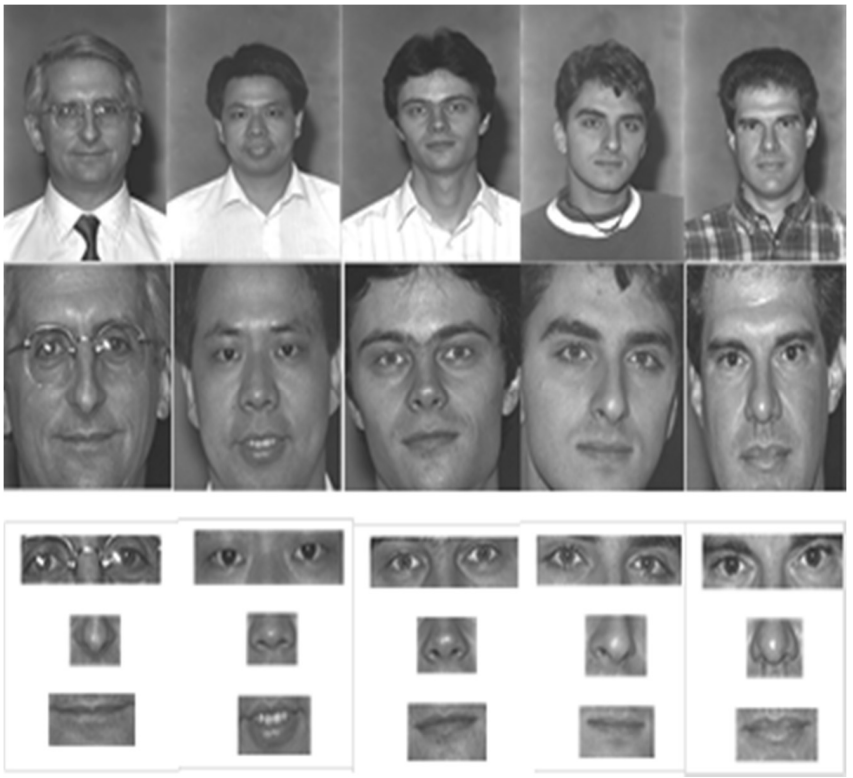


Figure 3. Sample images of FERET database

and Table 3 shows the comparative analysis of the component based method (RS-LDA). In the case of individual components, both the methods have similar recognition rate for nose component and slightly better in rest of the components.

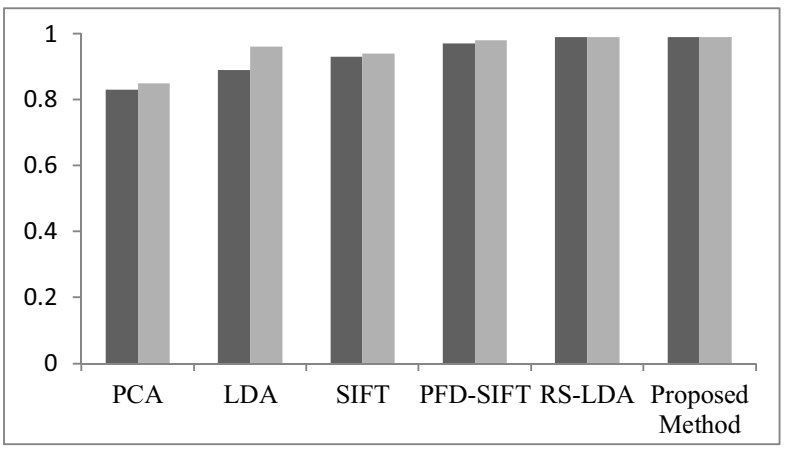


Figure 4. Rank 1 Recognition rate on FERET Database

Table 3. Performance of Each Component on FERET database

Component	RS-LDA	Proposed Method
Eyes	0.86	0.87
Nose	0.99	0.99
Mouth	0.79	0.80

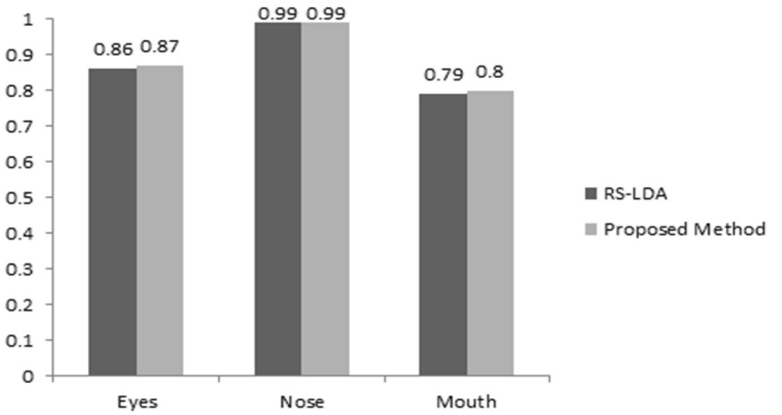


Figure 5. Performance of Each Component on FERET database

Analysis on ORL database

The ORL database contains 40 different persons image. Each subject 10 images with different variation (varying the lighting, facial expressions and facial details). Each subject five images are randomly selected for training and rest 5 images for testing. In ORL database 200 images are used as gallery images and remaining 200 images are used as probe set. The recognition rate of the proposed system is compared with existing holistic approaches and is shown in Fig.6 and Table 4.

Table 4. Percentage of Recognition rate on ORL database

Method	Recognition Rate
PCA	0.85
LDA	0.92
SIFT	0.90
PFDM-SIFT	0.99
Proposed Method	0.99

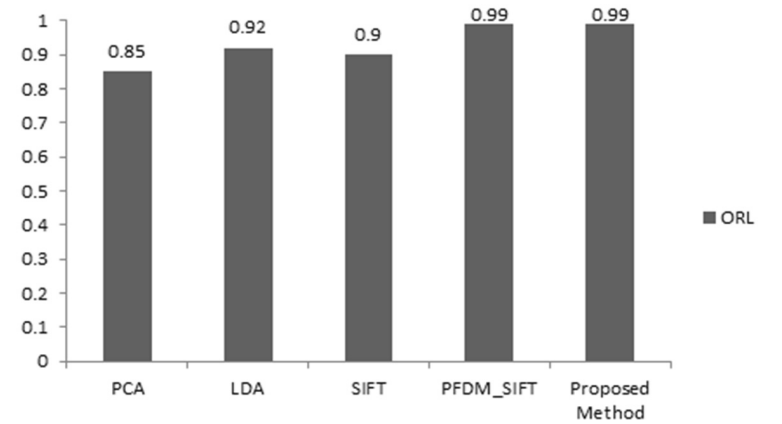


Figure 6. Percentage of Recognition rate on ORL database

Conclusion

This work presents a hybrid approach for automated face recognition based on the facial components. The hybrid approach demonstrates the efficiency in pose variation and illumination condition. The experimental results show that the proposed method provides good results comparatively with holistic approaches. The system slightly improves the recognition rate in component. The proposed method reduces the running time of system.

References

[1] Brunelli and T. Poggio, "Face recognition: Features versus templates," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 10, pp. 1042–1053, 1993.
[2] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, pp. 711–720, 1997.
[3] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," In Proc. IEEE Int. Conference on Image Processing, vol. 1, 1997.
[4] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The feret evaluation methodology for face recognition algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, 2000.
[5] N. Sagiv and S. Bentin, "Structural encoding of human and schematic faces: Holistic and part-based processes," Journal of Cognitive Neuroscience, vol. 13, no. 7, pp. 937–951, 2001.
[6] H. Jie, H. Yu, and J. Yang, "A direct lda algorithm for high-dimensional data with application to face recognition," Pattern Recognition, vol. 34, pp. 2067–2070, 2001.
[7] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multi resolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Analysis & Machine Intelligence, vol. 24, no. 7, pp. 971–987, 2002.
[8] J. Huang, P. Yuen, W.-S. Chen, and J. Lai, "Component-based lda method for face recognition with one training sample," IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp.120–126, 2003.
[9] Paul Viola and Michael J.Jones ," Robust Real Time Face Detection", International journal of Computer vision.57(2), pp. 137-154,2004.
[10] W. Zhang, S. Shan, W. Gao, Y. Chang, and B. Cao, "Component-based cascade linear discriminant analysis for face recognition," Advances in Biometric Person Authentication Lecture Notes in Computer Science, vol. 3338, pp. 19–79, 2005.
[11] T.-K. Kim, H. Kim, W. Hwang, and J. Kittler, "Component-based lda face description for image retrieval and mpeg-7 standardisation," Image and Vision Computing, vol. 23, pp. 631–642, 2005.
[12] D. Lin and X. Tang, "Recognize high resolution faces: From macrocosm to microcosm," Computer Vision and Pattern Recognition, pp. 1355– 1362, 2006.
[13] K. Pan, S. Liao, S. Zhang, S. Z. Li, and P. Zhang, "Part-based face recognition using near infrared images," IEEE conference on Computer Vision and Pattern Recognition, 2007.
[14] Schwaninger, S. Schumacher, C. Wallraven, and H. Buelthoff, "Using 3d computer graphics for perception: The role of local and global information in face processing," Applied Perception in Graphics and Visualization, 2007.
[15] Heisele, T. Serre, and T. Poggio, "A component-based framework for face detection and identification," International Journal of Computer Vision, vol. 74, no. 2, 2007.
[16] E. Meyers and L. Wolf, "Using biologically inspired features for face processing," Int. Journal of Computer Vision, vol. 76, no. 1, pp. 93–104, 2008.
[17] X. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," Pattern Analysis and Machine Intelligence, IEE Transactions on, vol. 30, no. 3, pp. 383–394, 2008.
[18] C. Evans, "Notes on the opensurf library," University of Bristol, Tech. Rep. CSTR-09-001, January 2009. [Online]. Available: http://www.chrisevansdev.com
[19] S. Xie, S. Shan, X. Chen, and J. Chen, "Fusing local patterns of gabor magnitude and phase for face recognition," IEEE Transactions on Image Processing, vol. 19, 2010.
[20] B. Klare and A. Jain, "On taxonomy of facial features," in Proc. of IEEE Conference on Biometrics: Theory, Applications and Systems, 2010.
[21] B. Klare, Z. Li, and A. Jain, "Matching forensic sketches to mugshot photos," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 33, no. 3, pp. 639–646, 2011.
[22] J. Richler, O. Cheung, and I. Gauthier, "Holistic processing predicts face recognition," Psychological Science, vol. 22, no. 4, pp. 464–471, 2011.
[23] Kathryn Bonnen, Brendan Klare, and Anil Jain," Component-Based Representation in Automated Face Recognition," IEEE Transactions on Information Forensics and Security, 2013.
[24] C. Geng and X. Jiang, "Face recognition based on the multi-scale local image structures", Pattern Recognition, vol. 44, pp. 2565–2575, 2011.
[25] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, L. Van Gool, H. Baya, A. Essa, T. Tuytelaarsb, and L. Van Goola, "Speeded-up robust features (SURF)," Comput. Vis. image Underst., vol. 110, no. 3, pp. 346–359, 2008.

Framework for Surplus Food Management using Data Analytics

M. Sridevi* and B.R. Arunkumar**
*Assistant Professor, Dept. of MCA, BMSIT&M, Bengaluru, India
muthyala.sridevi1@gmail.com
**Professor & Head, Dept. of MCA, BMSIT&M, Bengaluru, India
arunkumar.mcahod@gmail.com

Abstract: Food wastage is a major problem faced by the entire world in various facets. The food waste not only impacts a country's economy financially, but also environmentally. The food waste leads to wasteful use of fertilizers and pesticides. More fuel is used for transporting the food to markets and surplus of it to garbage dumps. The wedding parties and functions in India stand as a prestige icon spending money lavishly on large number of variety dishes leading to food wastage. The rotten food creates methane – one of the most harmful greenhouse gases that contribute to climate change and global warming. Methane is twenty-three times more effective than CO_2 as a green house gas. A problem properly addressed is a problem half solved. Moreover, a solution to one problem can indirectly solve many other potential problems. Research is being done on how to minimize the food loss and wastage at the supply chain level but no models are developed to channelize the surplus food generated for its prompt utilization. There is an immense need of models that would facilitate the surplus food management to minimize the food loss and wastage with the community contribution and collaboration.

Keywords: Surplus Food, Supply Chain, Food Wastage, Data Analytics, Optimum Utilization, Cluster Pivot, Clustering, Decentralization.

Introduction

Food – the basic necessity of every living being has become a barely available luxury for many of the people across the world. According to the World Food Program's (WFO) hunger statistics, approximately 795 million people in the world do not have enough food to lead a healthy and active life. It is literally one in nine people across the world suffer from hunger [1]. The large volume of the world's hungry population lives in developing countries. Two thirds of the total hungry people live in Asia. 66 million primary school age children attend their classes hungry across the developing world, with 23 million in Africa alone [2]. [3] stated that the global food security crisis of 2008-09 led to deep and growing concerns about how to sustainably increase food productivity in the present scenario of continuous rapid population growth and worsening climate change. The most landmark study on international food waste, the UN Food and Agriculture Organization claims that "roughly one-third of the edible parts of food produced for human consumption gets lost or wasted globally, which is about 1.3 billion ton per year [4]." All the world's nearly one billion hungry people could be lifted out of malnourishment on less than a quarter of the food that is wasted in the Unites States [5].

Indians by nature are known to spend lavishly on weddings and parties on a large scale. In some cases, the waste is to the extent of 20-25 percent when the number of dishes exceeds the number of guests. A survey shows that annually, Bengaluru alone wastes 943 tons of quality food during weddings which is enough to feed 2.6 crore people a normal Indian meal.

The surplus food with function halls, restaurants, college hostels etc. if properly channelized would satiate the hunger of many people those who doesn't have access to proper food. This paper proposes a model for decentralized surplus food distribution which makes the entire process optimal, quick, and enables maximum utilization. Next sections cover about the literature review, proposed framework, the process overview, the role of data analytics in the optimization of the application, limitations of the model, future enhancements and thus bringing us to the conclusion.

Literature Review

According to the CSR Journal, Indians waste as much food as the whole of United Kingdom consumes. Food wastage is an alarming issue in India. According to the UNDP (United Nations Development Programme), up to 40% of the food produced in India is wasted. In fact, according to the estimates of the agriculture ministry, Rs. 50,000 crore worth of food produced is wasted every year in the country. India ranks 63 among 88 countries in GHI (Global Hunger Index) published by International Food Policy Research Institute (IFPRI). 300 million barrels of oil are used to produce food that is ultimately wasted.

[6] The way of functioning of the food supply chain results in the food loss and waste, culturally, economically, and technically. Loss occurs in storage, transport, and processing in the low income countries and investments are required to develop the value chain. Loss is at the retailer and consumer levels in high-income countries and requires distinct strategies to mitigate waste. There are a number of ongoing private-public, and public partnerships tackling wastage of food at the country level demonstrating the concern towards addressing the issue.

In [7], the authors emphasized the Micro economic conditions like the production theory and consumer theory, and Macro economic conditions like infrastructure, urbanization and socio-economic growth, trade and globalization, and unemployment that would culminate to food losses and wastage. They also stressed on the non economic conditions like culture and societal norms, social aspects, environment and climate, policy, legislation, and private standards that would result in food losses and wastage.

Besides the micro and macro economic conditions that lead to food loss and waste, its impact on economy and environment is negative and threatening. With respect to economy, they represent a wasted investment that can reduce farmers' incomes and also increase consumers' expenses [8]. With respect to environment, food loss and waste pose a host of impacts, including unwanted greenhouse gas emissions and inefficiency in usage of water and land which in turn can lead to degraded natural eco systems and the quality of services they provide.

Research and subsequent actions are in progress to address the food loss and wastage at the food supply chain level, and to mitigate its affect on the environment and economy. In [9], the author analyses the role of food banks in Canada in terms of advancing the human right to food, their effectiveness in achieving food security and the extent to which they contribute to the increasing emphasis by governments on welfare reform policies informed by neo-conservative ideology.

The India Food Banking Network (IFBN) is evolving an ecosystem for food security to support number of feeding programmes in India by bringing the private sector, government sector, and NGOs together under one roof to fight hunger and malnutrition in India. IFBN aims to achieve this by establishing an efficient and strong network of food banks throughout the country, so that every district has access to minimum one food bank by 2020 [10]. IFBN's stakeholders constitute global, domestic, and local community partners who voluntarily contribute to support its development and humanitarian projects. IFBN invites organizations and individuals to collaboratively work with them to eliminate hunger in India. The IFBN runs on donations from companies involved in large scale processed food production. They support IFBN by giving food which is stocked as unsaleable and can't be put into supply chain. This stock is perfectly good for consumption and put across all food quality checks. Sam Pitroda, Chairman, Food Security Foundation India, India FoodBanking Network, urges the need to focus on hunger with a community driven responsibility.

The above studies discuss about the food loss and wastage at different levels of food supply chain but the problem of dealing with the surplus food that is being thrown into the garbage dumps which would otherwise feed lots of needy people is not being addressed adequately. The food wastage in mostly not because it is stale but it is made in excess sometimes because of wrong predictions about the intake quantity and sometimes as a result of status icon that increases the number of dishes than the number of guests. There is a clear research gap for the optimization of this surplus food distribution process in India. There are some organizations that collect surplus food and distribute it to the needy but they are dedicated to specific localities and most of the work is being done manually. This makes the distribution process slow, transportation costs higher and the food may be spoiled before it reaches the targeted people. Moreover, the amount of surplus food generated can't be predicted before hand to make necessary preparations for its distribution.

This paper proposes a framework for surplus food management as an easy to use mobile application that decentralizes the entire process with minimal administration but maximum outcome.

The Proposed Framework

In this paper, we refer to the excess food as not waste food but surplus food as it deserves a respectable mention because it can satiate the hunger of people who are in real need of it. The model proposed by this system makes the distribution process easier and faster before the food gets spoilt and enables proper utilization of the same. It involves the citizen volunteers as part of the process who can understand the value of food.

The model constitutes four modules – the registration module, the communication module, the logistics module and the data analytics module.

The Registration Module

The module allows three types of users to register namely, the Donor, the Volunteer and the Receiver. Figure 1 depicts the users of this application.

The ***Donor*** can register with the details like name of the organization or function hall or restaurant or hostel, address with PIN code, contact number, volume of surplus food that would be available and the pickup time at which it can be collected. The advantage of this application is that the donor need not register long before to utilize the service of the volunteers. They can check the approximate amount of food that will be left out in next one or two hours and then register for the services. Already registered users can utilize services by submitting context specific details.

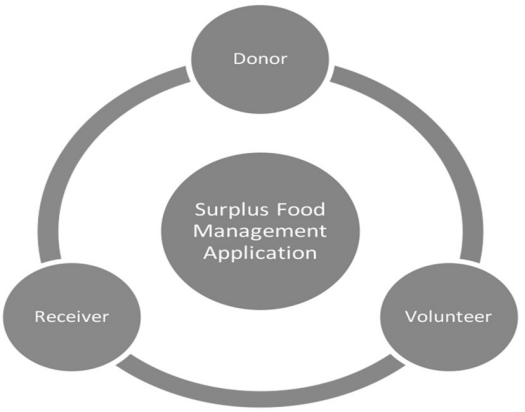


Figure 1. The users of the application

The ***Volunteer*** is any common man who is interested to serve voluntarily with his time as a valuable investment. The volunteer registers with his/her name, address with PIN code, contact number, contact time (during which he/she would be able to serve a delivery), and vehicle type (two-wheeler or a four-wheeler) which they possess to estimate the capacity to deliver.
The ***Receiver*** which may be Non Government Organizations (NGOs), Non Profit Organizations (NPOs), orphanages, old age homes, food banks etc. can register with their organization name, address with PIN code, contact number, and number of benefiters in their organization or through their organization.

The Communication Module
The module communicates the availability of surplus food with any organization, to the volunteers after some data analysis (discussed in the next sections) so that all the nearest volunteers get the "Accept Delivery" message with the source and delivery addresses. Depending on the number of volunteers required, when the module receives the number of required "Request Accepted" notifications from the volunteers, it will immediately communicate to the rest of the volunteers in the cluster that the delivery is accepted and their services will be utilized next time. The communication module also notifies the receivers through a message about the food to be supplied so that they can make necessary arrangements to utilize it within the organization or dispatch it externally to the targeted people after performing quality checks for the food received.

The Logistics Module
The logistics module deals with minimizing the transportation cost for the volunteers by selecting the nearest volunteer and connecting him to the nearest receiver, and ensures maximum comfort while delivering by assigning the delivery quantity to the volunteer based on the vehicle type or capacity the volunteer posses and the address to which it has to be delivered.

The Data Analytics Module
The crucial part of the entire system is the data analytics module which needs special emphasis in the next sections.

The Process Overview

The role of Data Analytics in decentralization and optimization process of Surplus Food Management is of ample importance as the delegation and distribution process makes the model more effective with good impact on the users and benefiters of the application.
Let us consider pin code value around which the cluster is built as the **cluster pivot.** After each registration by the user is completed, the algorithm for the clustering process will run creating clusters as follows:

Algorithm: Cluster_Formation
Input: Record of the registration details of the donor or volunteer or receiver.
Output: New cluster is formed or existing cluster is updated.

1: The PIN code attribute is compared with existing cluster pivots.
2: If the PIN code = = cluster pivot then go to step 3 else go to step 4.
3: Assign cluster ← input record.
4: Create a new cluster with PIN code value as the cluster pivot.

The clusters are activated when minimum one donor, one receiver and multiple volunteers are added to it. Otherwise they remain passive till they reach the minimum requirement. The outcome of the algorithm can be depicted as shown in Figure 2a.

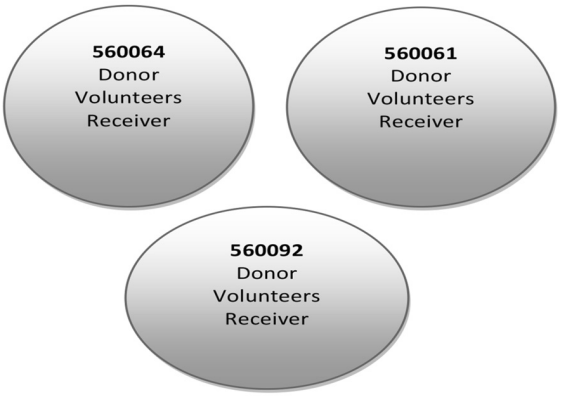


Figure 2a. The Clustering Process output based on PIN code as the cluster pivot

It is ensured that the members of a cluster belong to the same PIN code. This helps in minimizing the pickup and delivery time and also to minimize the cost of transportation for the volunteer. Further the Volunteers are formed into sub clusters as shown in Figure 2b, based on their contact time so that they will not be disturbed with any SMS in other timings. The same cluster formation algorithm discussed above applies to the sub-clustering process also, this time the cluster pivot being the contact time of the volunteers.

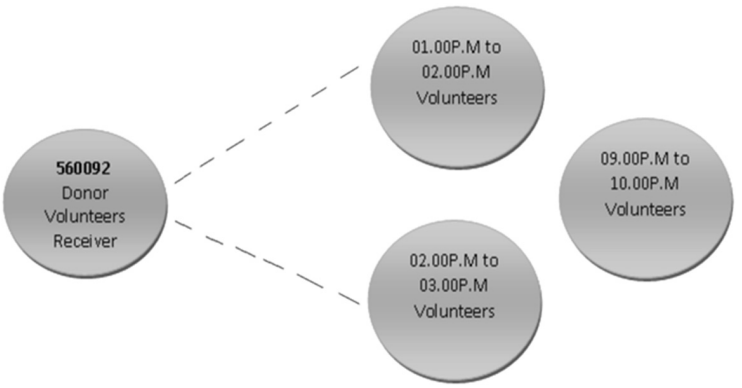


Figure 2b. The Sub Clustering Process based on contact times of the Volunteers

The volunteer sub-clusters will have a separate mechanism deployed for receiving the messages. When the message is received from the donor about the surplus food availability, the application links it to the appropriate cluster. The cluster generates a call to the Message_Volunteer procedure of the sub-cluster. It sends a notification to all the volunteers matching the pickup time with the donor to the contact time of the volunteers with a request to accept the delivery. The number of volunteers required per request is decided by the amount of surplus food available with the donor and the vehicle type available with the volunteer.

Procedure: Message_Volunteer
Input parameters: Pick-up time, Volume of surplus food

```
1: [Initialize] i=1;
2: Repeat
        If ((pick_up = = contact_time) && (volume = = delivery_capacity)) then
                Notify the volunteer;
        Else
```

Continue with the next iteration;
Until the total number of volunteers are notified;

Once the required number of acceptances from the volunteers is received, rest of the volunteers will receive another notification that the request is responded and their services will be utilized for the next time. This enables prompt usage of the volunteer services avoiding multiple responses. The application simultaneously messages the receiver about the surplus food availability so that they make necessary preparations for using it internally or distributing it externally. The receivers make the food received pass through quality checks before distributing the same.

The process flow is depicted by Figure 3.

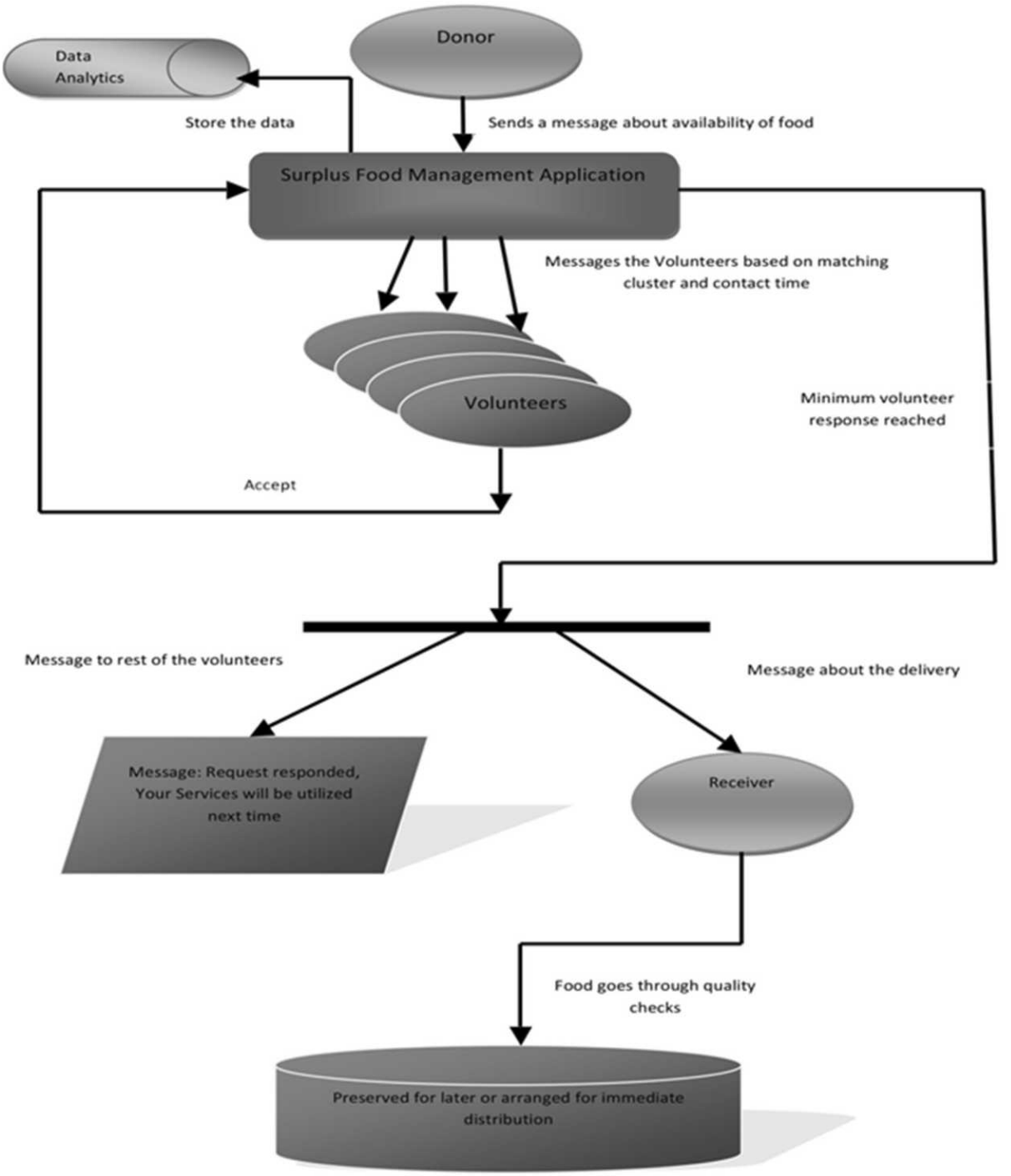


Figure 3. The Process Flow

Role of Data Analytics in Optimization of the Application

The data about the donors, volunteers, receivers, and each of the transactions is recorded for further analysis. The framework uses Multidimensional Online Analytical Processing (MOLAP) server which uses multidimensional storage engines that are array-based for multi-dimensional views of the data. The storage utilization will be low with multidimensional data stores if the data set is sparse. This works efficiently with passive clusters and saves the storage till they become active. The data analytics part uses various OLAP operations such as Roll-up, Drill-down, Slice and Dice, and Pivot to unveil and answer interesting questions like–

- Which season of the year is maximum surplus food generated?
- Who are the volunteers that contribute most of the time and what is their profile?
- Which type of donor generates most of the surplus food?
- How much food is being saved from being dumped into garbage?
- How well the process can be optimized to allow inter-cluster contribution if excess surplus food is being generated within one cluster?

These questions can help the application to reorganize the clusters for improving the efficiency of the system. For inter-cluster contribution, the data points can be temporarily reorganized to share the surplus food with other clusters if more surplus food is available within one cluster. The model thus refines itself periodically to give optimal performance and optimum utilization of surplus food.

Limitations of the Model

- Distributed architecture works well only if there are sufficient number of data points in each cluster.
- Application success depends on the committed participation of the citizen volunteers.
- The clusters remain passive till the minimum requirement for the cluster formation is met.

Future Enhancements

More work has to be done on the cluster reorganization and inter-cluster sharing of the surplus food. Inter-cluster communication and sharing becomes a complex objective to be achieved as the pin code values which serve as cluster pivots do not exactly represent adjacent or nearby areas. This makes balancing surplus food among the clusters a challenging task. In this scenario, an alternative for cluster pivots have to be selected that enable inter-cluster sharing easier and cluster reorganization smoother.

Conclusion

The model proposed in this paper tries to address the problem of surplus food management in a decentralized and efficient way with the community collaboration. The entire process is decentralized to ensure quick and optimum distribution and utilization of the surplus food. The clustering concept used for grouping the donors, volunteers, and receivers will help the application to reduce the transportation cost, and delivery time. Moreover it requires minimal administration as the process is decentralized. The data analytics help the application to iteratively refine itself for reorganizing the clusters to enable inter-cluster collaborations in instances when some of the clusters generate more of the surplus food than what is utilized within the cluster. The application has to be intensively publicized before it is released so that the community collaboration is maximized for the expected outcome.

References

[1] State of Food Insecurity in the World, FAO, 2015.
[2] Two Minutes to Learn About School Meals, WFP, 2012.
[3] United Nations. A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development. United Nations Publications, New York, 2013.
[4] Global Food Losses and Food Waste, FAO 2011.
[5] Tristram Stuart, “Waste: Uncovering the Global Food Scandal,” 2009.
[6] A brochure on “No more food to waste – Global Action to stop Food Losses and Food Waste” in partnership with CGIAR Research Program on Climate Change, Agriculture and Food Security and Others.
[7] Segre A., Falasconi L., Politano A., Vittuari M., Background paper on the Economics of Food Loss and Waste, Rome, FAO, 2014.
[8] Brian Lipinski, Craig Hanson, James Lomax, Lisa Kitinoja, Richard Waite, Tim Searchinger, “Reducing Food Loss and Waste”- A working paper, World Resources Institute, May, 2013.
[9] Graham Riches, “Food Banks and Food Security: Welfare Reform, Human Rights and Social Policy. Lessons from Canada?” International Journal of Policy and Research, Vol 36, Issue 6, Pages 648-663, December 2002.
[10] CSR Mandate – Enabling Sustainability, “Alleviating Poverty and Malnutrition: Towards a Hunger-Free India,” Vol 2, Issue 1, 2014.

Group Key Management Protocol: Secured Transmission in Compliant Groups

Amol Bhagat* and Lovely Mutneja**
* Innovation and Entrepreneurship Development Center,
Prof Ram Meghe College of Engineering and Management, Badnera, Amravati, India
amol.bhagat84@gmail.com
** Department of Computer Science and Engineering,
Prof Ram Meghe College of Engineering and Management, Badnera, Amravati, India
lmutneja@gmail.com

Abstract: Popularity of group-based application and related protocols leads to exponential growth of group communication. One form of group communication is Cooperative group. Cooperative group consist of multiple nodes/hosts connected together in resource sharing environment. Types of Cooperative groups are Wireless Mesh Network and Mobile ad hoc Network. The specified communication is totally unsecure. Secured transmission is one of the complex issues when dealing with the groups. A proper management of keys is needed in the form of generation, exchange, storage and replacement of keys i.e. a key management is needed for the management of cryptographic keys in a cryptosystem. Group Key Management Protocol (GKMP) provides an effective key management solution which deals with generation of keys and distributes them among different communication peers. This paper addresses the use of GKMP protocol for providing security of data transmission in cooperative group. The process of creating group, distributing keys to each member, controlling transmission, implementing security, and deleting group members is described in this paper. The comparison recent approaches are presented in this paper. This paper also describes the proposed methodology for group management. The performance evaluation of the proposed methodology is presented on the basis throughput, time required for key generation and time required for key distribution.

Keywords: Cooperative Groups, Group Key Management Protocol, Key Distribution, Key Generation, Secured Transmission.

Introduction

Group based applications and protocols have gained their popularity as they are used to provide efficient packet delivery from source to destination/receivers. Security is one of the main concern here as these involves communication over open networks, therefore proper key management is needed for providing security for a communication in groups. The main building block for achieving security in group communication scenarios is management of the secret information, this information is known only to the participants and therefore it is known as secret group key. In secure communication two entities should communicate without the intervention of a third party to listen in. Proper care must be taken so that eavesdropping or interception should not occur in such communication. In secure communication people can share information by certain means with varying degrees of certainty that third parties cannot intercept the data exchanged during communication.

The basic working of GKMP [1] is to provide group wise key in groups, support to dynamic joining and leaving in group, support to scalability, centralized operation. With the help of these features of GKMP protocol a foolproof solution for secure transmission can be designed. The basic purpose of GKMP is to deal with security related issues. It first creates keys for cryptographic groups for security. It also provides capability for distribution of keys, provides keys access control, known compromised host's access denial, and also provide controlling of group actions. The key generation methodology is cooperative generation which is implemented between two protocol entities. As the keys are generated then GKMP distributes the keys to verified GKMP entities. Verification means are provided for the means of distribution. GKMP implements review concept in which protocol entities passes permission certificates (PC), this process is perform during key distribution and is considered as a part of key distribution. The PC supported provides means of access control information, which leads to checking of permissions and comparing the level of service requested by GKMP entity. The service is denied if the permissions is less than or equal the request. These features of creating, distributing and providing means of access control information are the key concept for the proposed algorithm.

The proposed work with help of GKMP consist of compromise recovery feature in which GKMP distributes a list of compromised entities to group members during key management actions. The use of Compromise Recovery List (CRL)

allows members of groups to deal with compromised entities by implementing dropping of the connection action. GKMP consist of inbuilt feature to control group actions. In several situations while dealing with certain networks it is mandatory for a higher authority to control the generation of groups. Identification of group key controller, group key creation, group key distribution and group rekey [1] are the main objectives of the proposed approach in this paper.

Related Work

There are various approaches for providing security to data being transfer and for effective key management. Cryptosystem helps in sharing secret keys among users in an insecure channel e.g. satellite broadcasts, IP multicasting, TV subscription etc. Behzad Malek and Ali Miri [2] proposed secure broadcast encryption scheme with the use of short cipher texts. The algorithm generates fixed size cipher text. Leaving and joining of members in the group is quite easy, it does not affect any public parameters or private keys of existing members. This algorithm specifically works with small cipher text and provides greater security as compared to other models.

B Rong [3] proposes a pyramidal security model to deal with all these issue in which multi security level are maintained. This model consists of hierarchical security groups and multicast groups. For efficient key management solution in multicast groups, three schemes has been proposed Star key. Graph, separated tree key graph and integrated star key graph for key management in multicast group schemes used are: Diffie Hellman algorithm extended contributory key management, computational number theoretic approach, and logical key hierarchy (LKH).

Now a day's network applications are based upon group communication, therefore providing security and other necessary service is main concern. For addressing this, a novel solution has been proposed which addresses scalability problem in group/multicast key management [4]. Recommendation is to treat group as a triple (U, K, R) where U is a set of users, K represent set of keys, and R is a user-key relation. Representation of secure group is given by means of key graphs. Three strategies have been used for securely distributing rekey message. Rekeying strategies and join/leave protocols are implemented. These strategies have the advantage that they are scalable to large groups and provide frequent joins and leaves.

Broadcast channel deals with transmission of data from a source station to every node within the network. Examples are LAN, Packet radio network, etc. Need arises to send secret message to various users at a same time. Application provides such service is called secure broadcasting applications. This kind of applications is useful when dealing with document distribution, teleconferencing, etc. Guang-Huei Chiou and Wen-Tsuen Chen [5] proposed concept of locking using Chinese Remainder Theorem (CRT). Secure lock is used to lock the session key. This lock system has various advantages: single cipher text copy is sent, efficient deciphering and secret keys held by each user are minimized. This secure lock helps in solving the secure broadcasting problem.

Broadcast encryption [6] deals with secure data transmission over insecure channel with changing set of user in the group. The public key broadcast encryption scheme is designed for to provide encryption in way that the public key is stored in user's device, or it is provided to the receiver along with cipher text. Here two fully collusion-resistant broadcast encryption schemes has been proposed (PKBE public key broadcast encryption) which are used by the stateless receivers. This scheme assures that it is secure against numbers of colluders. Again this scheme is also scalable as transmission cost not depends on the number of users. The proposed schemes are referred as Decision modified Bilinear Diffie-Hellman. There is controlled number of users available beyond the specific set who sometime receives the multicast. This kind of relaxation leads to the development of f-redundant establishment key allocations [7]. The scheme provides guarantee that the receivers are not more than f times to intended recipients.

Increase in group-orientation applications and protocols leads to increase in group communication e.g. Multicasting, videoconferencing application etc. Security services are necessary to all this aspects for maintaining communication privacy and integrity. For dealing with this key agreement in dynamic peer groups has been designed. Dynamic Peer groups require both initial key agreement (IKA) and auxiliary key agreement (AKA) operations, such as member addition, deletion, group fusion etc. For this the concept of CLIQUES protocol [8] was introduced which offers complete key agreement services. A CLIQUE is based on Diffie-Hellman key exchange method. Versa Key architecture [9] provides group wide keys and scalability. This architecture is suited where there is dynamicity in group for joining and leaving. Key distribution time is very less in Versa Key architecture. These are the various features of Versa Key Architecture.

Attack-Resilient Security Architecture [10], called ARSA. The need for bilateral roaming agreement establishment is eliminated by the use of ARSA and it has real-time interactions between numerous WMN operators. By using ARSA, each end user is not bounded to any specific network operator. Key agreement between a user and a serving WMN domain and efficient mutual authentication is supported by ARSA. It is also resilient to a wide range of attacks. ARSA is a homeless solution i.e. not bound to specific WMN operator. It is designed to be resistant to various attacks against WMN access.

In [11] an algorithm called OFT (One-way Function trees) is presented, its main purpose is to established cryptographic keys in large groups having dynamicity. It is a centralized algorithm based on one-way function trees. It is a bottom-up algorithm, it approximately halves the number of bits needed to broadcast to group members for the purpose of rekeying when a member is added or evicted. This algorithm provides complete forward and backward security, which means newly joined

node cannot read previous message, and deleted members cannot read future message. In OFT members are allowed to contribute entropy to the group key. It has the capability that even working on Pentium II processor, it can handle groups up to 10 million members.

As technology get updated, rapid increase in multimedia application and data also get increased, internet also allows for wide distribution of digital media data. Duplicity also got increased on digital data. It leads to many threats as digital documents are now easy to copy and distribute. [12, 13] focuses on three algorithms to provide security to data based on nine factors for achieving efficiency, flexibility, and security. Author claims AES [22] is better than DES and 3DES, by comparing on factors such as key length, cipher type, block size, developed, cryptanalysis resistance, security, possibility key, possible ASCII printable character keys, time required to check all possible key. Session keys and transport session keys are the two types of keys chosen by Group Key Agreement (GKA) [14], Key transfer protocol rely on GKA for choosing those keys. Each entity shares secret key during registration use for the purpose of encrypting session keys. Author has proposed an authenticated key transfer protocol which is based on secret sharing scheme; GKA [14, 15] broadcasts group key information to all authorized members, unauthorized users cannot recover the group key.

Group key agreement [16, 17] protocol is used for the establishment of cryptographic keys dedicated to group participants; key is based on each member contribution, over a public network. This key works as a secure bridge between participants. The proposed protocol is unauthenticated and secure against passive adversaries only. Various methods are available for reducing transmission overheads in broadcast encryption, which are generally based on assigning one key per partition by implementing partitioning the users one way key-chains. One of the method adapts 'skipping' chains on partition which consist of 'p' revoked users and other methods works with cascade chains on partition with layer structure, here 'r' represents number of revoked users. Scheme one has some transmission overhead and the same transmission overhead with the Subset Difference (SD) is inherent to scheme two. Author has proposed a new broadcast encryption scheme [18] with same transmission overhead is same as that of SD scheme for small r and will become smaller than the SD with growing r. Scheme has small Transmission overhead (TO) if 'r' is not very small, scheme has same TO with SD when 'r' is very small, Smallest TO for all 'r' when both schemed are combined.

Important Quality of service aspect when dealing with large groups when concentrated on dynamic membership is the time cost performing key updates in events of member joins and leaves. Join-exit-tree (JET) [19] key management framework has been proposed by the author. Initially, special key tree topology for key updates with join and exit subtrees is introduced, then for determining the capacity of join and exit subtree optimization techniques are employed. Confidential and authenticated data transmission over an insecure network is one of an important goal of cryptographic research. An algorithm is considered to be secure if it cannot be broken with standard resources. For an efficient cryptosystem key distribution is also very important. This work focuses on comparison of various algorithms with TACIT Encryption Algorithm [20] by comparing parameters like key length, block size, type and features and also this work investigates HDL implementation of TACIT Encryption Algorithm. By highlighting capabilities of various algorithm author claims RSA is the most secure algorithm whereas DES is the least secure algorithm. Considering the speed factor TACIT Encryption Algorithm with Dual Port RAM is the fastest whereas RSA is slowest. Considering area TACIT Encryption Algorithm consumes more area.

Confidentiality is one of the important aspects of data which is to be consider, Data needs protection. Security is also needed for avoiding unauthorized access or modification of data. Fast changes in technologies leads to more and more multimedia data generation and transmission with possibility of data being edited, modified, deleted and duplicated other threats are also been applied. Therefore accuracy and sensitivity of information is a big security and privacy issue, for this a firm and robust solution is needed. [21] Cryptography is a technique useful to provide protection to user information. It can be achieve by means of implementing encryption and decryption.

Proposed Approach for Group Key Management in Compliant Groups

Group communication becomes one of the important technologies. As far as security is concern it is to be considered for securely transmission of important and private data in groups. Various techniques are available for secure transmissions in group all are having their own pros and cons. This motivated to design a proper, effective, cost efficient, easy to implement design for securely transmission of data. The proposed approaches for group key creation, group key distribution, group rekeying, member initiated joining and hostile member deletion are described in the following sections.

Group Keys Creation

In the proposed model first the command is transferred to group. The group controller (GC) listen this command. The future group keys are created with the coordination of the first member and the GC. The appropriate cryptographic properties must be possessed by the resulting key this is achieved creating group keys on two separate hosts. The information is exchanged between the controller and the group members to simplify a well ordered group creation, management. This information is utilized for distinctively identifying the GC identity, its permissions, authorization to create keys, the future group's permissions, the compromise list current state, and information management relating to the keys being created. Asymmetric signature methodology is utilized for protecting this information from counterfeit. To verify net wide parameters such as host

permissions, public keys are widely maintained. To verify locally generated information (e.g. peer identity) the public key is sent with the messages. Due to this there is no need to store the host's public key. The key creation process has following goals:

1. cooperatively generate a GTEK and GKEK,
2. to verify the identity of the key creation companion by verifying the messages signatures by the key creators,
3. share public keys,
4. by signing the GC identification, group identification, and group permissions, validate the GC
5. allow the first member, signed by the group commander to receive the GC identity, group member identities, group identity, group permissions, and group rekey interval for remotely managed grouped.

Distributing Group Keys to Members

The other group members must get the group keys before the group is fully operational. The purpose of other group member initialization is as follows:

1. cooperatively generate a session key encryption key (SKEK) for the transmission of the GTEK and GKEK from the GC,
2. allow each member to verify the identity of the controller and vice versa,
3. to create the group allow each member to verify the authorization of controllers,
4. send the key packet (KP) (consisting of the GTEK, GKEK), GC identity, group member identities, group identity, group permissions, and group rekey interval to the other members,

Group Rekey

Rekey is a two-step function that involves message exchange between the GC and a first member and other members. For group creation messages between the first member and GC must be exactly described. For the purpose of distributing the new GTEK and the new GKEK messages between the other members and the GC are utilized. These functions are

1. send the new GTEK and new GKEK to the other members,
2. allow each member to verify the identity of the controller,
3. to rekey the group, group identification, and GC identification allow each member to verify the authorization of controllers,
4. the other members should receive the GC identity, group member identities, group identity, group permissions, and group rekey interval.

Join Initiated By Member

Joins initiated by member to the group is supported by the GKMP. When the group initiator does not need to control group membership except to verify that all members of the group follow to some previously agreed protocols such type of service is most striking. A potential group member must request the key from the GC, unambiguously identify themselves, pass their permissions, and receive the keys before joining group operations. Several messages are pass between the joining member and the GC during this process. The purposes of these messages are as follows.

1. Request group join from controller
2. for the transmission of the group traffic encryption cooperatively generate a SKEK and GKEK from the GC,
3. allow each member to verify the identity of the controller and vice versa,
4. to create the group allow each member to verify the authorization of controllers,
5. the other members must receive the KP, GC identity, group member identities, group identity, group permissions, and group rekey interval to,

Cooperative and Hostile Member Deletion

There are two types of member deletion scenarios - cooperative and hostile. The cooperative deletion scenarios is the removal of a trusted group member for some management reason (i.e., reduce group size; prepare the member for a move). The hostile deletion usually results in a loss of secure state at the member's site (i.e., compromise, equipment breakage). The two scenarios present different challenges to the network. Minimization of network impact is paramount in the cooperative scenario.

In the case of a hostile deletion, the goal is to return to a secure operating state as fast as possible. In fact there is a trade-off. The compromised group can be eliminated as soon as the compromise is discovered, but this may cripple an important asset. So security concerns need to be balanced with operational concerns. The cooperative deletion function occurs between a trusted member and the GC. It results in a reliable deletion of the group key encryption and GTEKs at the deleted member. This deletion is intended to be an administrative function.

The essence of the issues involves a tradeoff between security susceptibility and operational stability. All traffic on the network is stopped if a member is found to be vulnerable, from a security point of view. The group may prefer to live with the security leak if group traffic is supporting a critical operation, instead of interrupting the group communication. To

restrict access of compromised members the proposed approach provides two mechanisms. First, a Certificate Revocation List (CRL) is created for utilizing it during the group creation process. This list will not allow a vulnerable member to be encompassed in a new group. Second, the proposed approach facilitates creation of another group without inclusion of the vulnerable member(s). The proposed approach does not dictate whether or not the group may continue to operate with a vulnerable member. The proposed approach uses a mechanism to remove a vulnerable member by key that member out. This involves switching group operations in newly created group, without the vulnerable member. A group delete message is multicasts to remove old group.

Proposed System Working

Group Head and Group Members are the two building blocks. The main responsibility of the group head is to form key and distribute this secret key to members for joining the group. The Group head can create any number of groups and can send a broadcast signal to all members of the particular group. After the secret key distribution is over the members authenticate themselves with the group header using the secret key. Once the authentication is done, the group is being form. Now the group head can perform data distribution to all members in the group. New member can join the group for this purpose the group head generates keys and perform re-keying. Group head can remove any of the members from group. Once the group gets form, server can send any kind of data to nodes. The complete working of the group head is as shown. The process gets start by forming a network, once the network gets form, group head get decided which will generate the secret key. Once the secret key get distributed, the members who acquire the key by any means can participate in group forming, for that they have to provide this secret key to the group head as an authentication means.

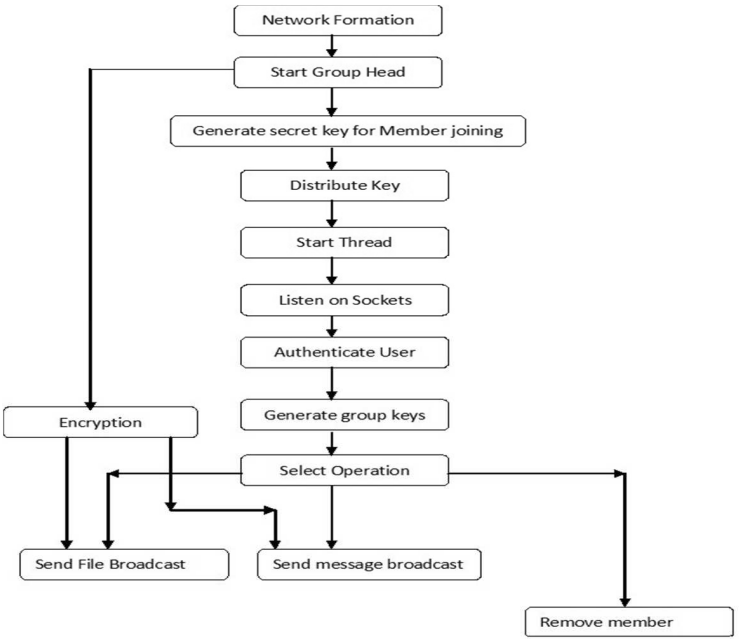


Fig. 1. Flow graph of proposed group head design

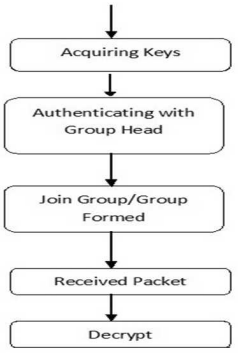


Fig. 2. Flow graph of proposed group member design

Group head observe the key and compares with its generated key if it matches it authenticate the member and joins in a particular group. Members in the network first acquire key from the head and then authenticate themselves to the group head to form group once the group gets formed they are able to receive the message in that particular group. The secret number from 1 to 99 as secret key is generated for joining of members to form a group. This method is linked with a Group Head class where group Head is used to generate numbers randomly as specified in code. Once the key been generated it is distributed to the other nodes for the purpose of group formation. The Group head authenticate with this particular key and listen on port number 19999 for all communication. After authenticating nodes a group is being formed with these authenticated nodes. Now the group head can perform any of the task i.e. can send a text message, can perform file transmission in group, or can remove the members from the group. All the data transmission either text or file transmission get performed with AES encryption and Decryption. Group Members acquires key for the purpose of group formation. This is the secret number generated by the group head and used for the purpose of authenticating the group members for group formation. Once the key being submitted the group head forms group. Group Head now can perform its task. The overall process is shown in figure 1 and figure 2.

Proposed System Implementation and Experimental Evaluation

The overall implementation of the proposed system is carried out in Java, on 2.8 GHz i3 processor, 4GB RAM and 3Mbps transmission rate. The Eclipse IDE is used for carrying out the implementation. The complete class hierarchy implemented in java is shown in figure 3. The main purpose of the cryptoClass is to provide the encryption and decryption mechanism. For providing this capability it consist of two methods encrypt() and decrypt(). It also consists of variable Plaintext and encryptionKey. The encrypt() is used in class GroupHead for the purpose of sending the message to the connected node. The decrypt() method is use for the retrieval of original message from the encrypted message. This decrypt() method is used in the GroupMember class for the decrypting purpose. The GKP class is used for the purpose of generating keys. These keys are distributed among different nodes. This class consist of various methods such as getCurrentKey(), setCurrrentkey(), getNextKey(), setNextKey(), setNextKey() and other methods. It consists of two main variables CurrentKey and Nextkey. The CurrentKey is a string variable which is used for the purpose for providing the Current key used. Next key is the generated key after the current key. First step is to identify the variable currentkey if it is less than 16 bit then padding is performed and if it is more than 16 bit then it is made 16 bit for operation.

The KeyGenerator class is a secret key generator used for the purpose of generating secret key used at the time of joining the group. It consist of generateKey() method which takes an integer argument to decide the length of the key. For generating the random key it used an object of a Random class. GroupHead class is a basic building block of this particular work. It consists of various data members as well as function. The first function is to start itself i.e. GroupHead(), it causes the Grouphead() to generate the secret keys for the purpose of group joining. Port 19999 is chosen to be a default port number for proposed work. It gets started on port number 19999 and is waiting to listen on this particular port number. Once it gets started it listens on port 19999, and will authenticate members according to secret key generated.

The GroupMember class first task is to provide the secret key to the group head for the purpose of forming groups. It first acquire secret key which is any secret number generated from the group head for the purpose to form group. This secret number is entered by the group member is used as an authentication by a group head. All broadcast are encrypted at group head and are decrypted at the group Members.

Proposed system works with the key length of 128 bit. Various performance metrics has been calculated and it has been identified 128 bit key is efficient from performance and security point of view. Key length has been decided by performing simulation in NS3 on different encryption algorithm and their key generation and distribution time has been calculated, throughput has been calculated by considering the total bandwidth of 3 Mbps as shown in table 1.

Table 1. Comparative analysis of key management and key distribution strategies

Key Size (bits)	Throughput (Mbps)	Time required for group key creation (seconds)	Time required for group key distribution (seconds)
64	2.916	0.523	1.352
128	2.892	0.793	1.788
256	2.809	1.493	2.480
512	2.769	2.456	2.982

Experimental result for Encryption algorithm AES, DES and RSA are shown in table 2, which shows the comparison of three algorithm AES, DES and RSA using same text file for five experiments i.e. 32 KB, 64 KB, 128 KB, 256 KB and 512 KB. The computational time is considered the time that an encryption algorithm takes to produces a cipher text from a plain text. It is used to calculate the throughput of an encryption scheme, and is calculated as the total plaintext in bytes encrypted

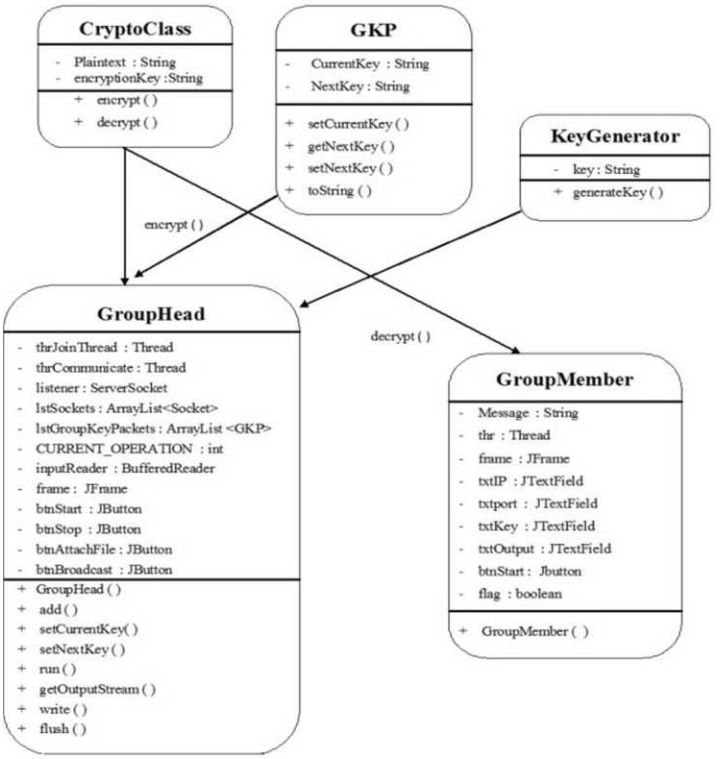


Fig. 3. Classes developed for carrying out experimentations using proposed approach

divided by the encryption or decryption time. It can be observed that time taken by RSA is very much higher as compare to DES and AES. DES is not as secure as compare to AES. Therefore AES is selected for carrying out the proposed approach.

Table 2. Comparative analysis of AES, DES and RSA using time required for encryption

Data in KB	Time Required for encryption (Seconds)		
	DES	AES	RSA
32	1.81	2.02	9.45
64	1.83	2.13	10.53
128	2.03	2.29	11.41
256	2.14	2.47	16.27
512	2.43	2.63	24.44

Conclusion and Future Scope

In this paper the existing literature on group communication and various security mechanism and key management concepts from previous literature is presented for formulating the proposed approach. Proposed work is based on GKMP features for group security. It provides Group management; i.e. creation of group, adding members, deletion of members. Proposed work addresses Key management concept. It provides security to data being transferred by means of implementing encryption and decryption techniques. Encryption and decryption in this work is done with the help of AES implementation which is consider to be a highly secured, and fast in operation.

Group communication is considered to be an important aspect in today's networking environment. Proposed work looks after security by means of implementing GKMP concepts involving group management, and key management, but still there are various aspects which need to be address in future. Proposed work addresses only few group communication features, complete full duplex mode is needed to be address in future. In future multiple group communication needed to address. AES is highly secured and effective, more effective and more reliable technology can be implemented. Key Distribution can be more enhanced. In Future work can be extended to form a Group, initiated by Group member.

References

[1] Mutneja, L. S. and Bhagat, A. P.: Secured Transmission in Cooperative Groups Using Group Key Management Protocol. In: IEEE International Conference on Communication and Network Technologies, pp. 139-144. IEEE Sivakasi, India (2014).

[2] Malek, B. and Miri A.: Adaptively Secure Broadcast Encryption with Short Ciphertexts: IJ Network Security, vol. 14, no. 8, pp. 71-79, (2012).
[3] Rong B., Chen H., Qian Y., Lu K., Hu R., and Guizani S.: A Pyramid Security Model for Large-scale Group-oriented Computing in Mobile ad-hoc: IEEE Transactions on Vehicular Technology, vol. 58, no. 1, pp. 36-47, (2009).
[4] Kei C., Gouda M., and Lam S.: Secure Group Communications using Key Graphs. IEEE/ACM Transactions on Networking, vol. 8, no. 1, pp. 16-30 (2000).
[5] Chiou G., and Chen W.: Secure Broadcasting using the Secure Lock. IEEE Transactions on Cyber Security, vol. 15, no. 8, (1989).
[6] Park J. H., Kim H. J., Sung M. H. and Lee D. H.: Public Key Broadcast Encryption Schemes with Shorter Transmissions. IEEE Transactions on Security, vol. 54, no. 3, (2008).
[7] Abdalla M., Shavitt Y., and Wool A.: Key Management for Restricted Multicast using Broadcast Encryption. IEEE/ACM Transactions on Networks, vol. 8, no. 4, (2000).
[8] Steiner M., Tsudik G., Waidnere M.: Key Agreement in Dynamic Peer Groups. IEEE Transactions on Networks, vol. 11, no. 8, (2000).
[9] Waldvogel M., Caronni G., Sun D., Weiler N., Plattner B.: The Versa-key Framework: Versatile Group Key Management. IEEE Journal Networks, vol. 17, no. 9, (1999).
[10] Zhang Y., and Fang Y.: ARSA: An Attack-resilient Security Architecture for Multi-hop Wireless Mesh Networks. IEEE Journal on Networks, vol. 24, no. 10, (2006).
[11] Sherman A. T., and McGrew D. A.: Key Establishment in Large Dynamic Groups using One-way Function Trees. IEEE Transactions Software Engineering, vol. 29, no. 5, pp.444-458, (2003).
[12] Hamdan A.: New Comparative Study between DES, 3DES and AES within Nine Factors. (2010).
[13] Harney H., and Muckenhirn C.: RFC 2093: Group key management protocol (GKMP) specification: Network Working Group, Website http://www.ietf.org/rfc/rfc 2093.txt (1997).
[14] Masood W., and Rasheed T. A Flexible Communication of Group Key Agreement.
[15] Lin Y.: A Group Key Management Protocol Based on Weight-Balanced 2-3 Tree for Wireless Sensor Networks. (2011).
[16] Luca V.: A novel batch-based group key management protocol applied to the Internet of Things. In: Ad Hoc Networks vol. 11, no. 8, pp. 2724-2737 (2013).
[17] Augot D.: An Efficient Group Key Agreement Protocol for Ad-hoc Networks. Sixth IEEE International Symposium on World of Wireless Mobile and Multimedia Networks, 2005.
[18] Cheon J. H.: Skipping, Cascade, and Combined Chain Schemes for Broadcast Encryption. IEEE Transactions on Information Theory, vol. 54, no. 11, pp. 5155-5171 (2008).
[19] Mao Y.: JET: Dynamic Join-Exit-Tree Amortization and Scheduling for Contributory Key Management. IEEE/ACM Transactions on Networking, vol. 14, no. 5, pp. 1128-1140, (2006).
[20] Kaur M.: Comparison of TACIT Encryption Algorithm with Various Encryption Algorithms. International Journal of Electronics and Computer Science Engineering, ISSN-2277-1956.
[21] Mathur M., and Kesarwani A.: Comparison between DES, 3DES, RC2, RC6, Blowfish And AES. Proceedings of National Conference on New Horizons, (2013).
[22] Juremi J.: Enhancing Advanced Encryption Standard S-Box Generation Based on Round Key. International Journal of Cyber-Security and Digital Forensics vol. 1, no. 3, pp. 183-188 (2012).

A Scheduling Approach in Cloud Environment using Deadline and VMs MIPS and Bandwidth

Anamika* and Pawan Luthra**
*Research Student
anamika3534@gmail.com
**Assistant Professor
pawanluthra81@gmail.com

Abstract: Cloud computing is the new innovation which is presently being utilized by PC users. Cloud registering gives more solace in the advanced time as there is no compelling reason to convey substantial gadgets with you since everything is accessible over the Internet. A cloud is gathering of number of virtual machines which can deal with different procedures to execute parallel. As cloud is growing there is a huge increase in number of cloud users. When there are various undertaking demands over the cloud then there is a need to choose which assignment will be executed first so that QOS parameters give better results. Scheduling of the cloudlets implies discovering the suitable and the required resources for that assignment and after that schedules the cloudlet onto that resource to achieve the objective for that cloudlet. Credit based scheduling algorithm considers only cloudlet length and cloudlet priority to schedule the cloudlet but in this approach a calculation has been characterized expecting cloudlet length, cloudlet priority, cloudlet dynamic submission time and cloudlet deadline, VM MIPS and VM bandwidth to decrease response time and waiting time using CloudSim 3.0.3 simulator.

Keywords: Cloud computing, scheduling, deadline, cloudlet, bandwidth.

Introduction

Cloud computing implies controlling, getting to and arranging the applications online and benefiting applications, infrastructure and data storage over the Internet. The clients are not required to introduce a product on their PCs, they can have the administrations of programming over the system through cloud. It is less costly in light of the fact that clients don't have to buy the entire framework, they will pay according to their utilization. Virtualization is a method which is utilized behind the cloud computing. It permits to share single physical object of a resource or application among numerous number of clients. Cloud design comprises of two parts. Front end is the customer piece of cloud computing framework which is made out of uses and framework that are fundamental to get to the cloud computing stage for instance web browser. Back end is cloud comprising of the considerable number of cloud resources crucial to give cloud computing benefits that are data storage, virtual machines, security systems, servers, deployment models. In public cloud model the administrations are effortlessly open to overall population for example Google. In private cloud model the clients connected with the specific association can get to the administrations of private cloud. The hybrid cloud model is combination of public and private cloud. Decisive cloudlets are taken care of in private cloud and indecisive are taken care of in public cloud. In community cloud gathering of associations is permitted to get to its administrations. Cloudlet scheduling is the procedure of dispensing cloudlets onto accessible resources in time taking after the limitations portrayed by the client and the cloud supplier. Research on cloudlet scheduling in cloud computing has turned into an interesting issue. The main motive of cloudlet scheduling algorithm is to allocate resources to cloudlet and minimize completion time, maximize resource utilization [1]. Load balancing is one of the fundamental challenge in cloud computing. It is a method which is required to circulate the dynamic workload over different virtual machines to guarantee that no single virtual machine is overloaded. Load balancing procedures help in ideal usage of resources and thus in upgrading the performance of the framework. The objective of load balancing is to minimize the resource consumption which will facilitate decrease power usage and carbon emanation rate that is the critical need of cloud computing [2]. Despite the fact that cloud computing has an extensive advantages over customary registering models, yet it raises serious security worries that point of confinement its across the board selection loss of administration, information security and unreliable information deletion are a few examples [3]. Datacenters associated with a cloud system are typically geographically distributed, yet connected together with dedicated high-bandwidth communication links. Latest trends demonstrate that cloud computing is developing to traverse more globally distributed datacenters. For geo-distributed datacenters, there is an expanding requirement for calculations to place assignments crosswise over datacenters, by together considering information and computation [4]. Power management approaches which mean to diminish total energy consumption in datacenters challenges in both hardware and resource administration policies. The ideal usage level of a host to execute a specific number of cloudlet to minimize energy consumption of the host is

required [5]. Cloud computing offers diverse resources over network appears an option answer for software testing. In any case, how to assess and allot these effective assets to particular testing projects are still required further examination [6]. It is critical to precisely appraise the reliability of a cloud computing framework keeping in mind the end goal to better reduce faults and accordingly successfully use its execution to accomplish the SLA of cloud users [7]. Hence an optimal cloudlet scheduling algorithm is needed to decrease the completion time and increase resource utilization. In this paper an algorithm has been designed in which cloudlets get scheduled on the cloud resources according to cloudlet priority, cloudlet length, cloudlet dynamic submission time and cloudlet deadline, VM (Virtual Machine) bandwidth and VM MIPS. The rest of the paper is sorted out as follows. The segment 2 specify some related work on cloudlet scheduling and part 3 portrays the proposed work, part 4 describes experimental setup and results and part 5 concludes the whole process.

Related Work

The main motive of cloudlet scheduling in cloud computing is to schedule cloudlets onto the available resources so that the resources are utilized and user requests are satisfied. The datacenter broker plays the main role for scheduling cloudlets onto virtual machines as it knows in advance both the cloudlets and available datacenters. The performance of cloudlet scheduling is evaluated by calculating execution time, response time, waiting time, turnaround time and cost etc by assuming QOS parameters like cloudlet length, cloudlet priority, cloudlet deadline, VM MIPS and VM bandwidth.

The paper [8] defines scheduling algorithm according to cloudlet priority and completion time. The priority of the cloudlet is determined by indicated properties like client benefits, need expected by the client, length of the assignment and its workload. At that point cloudlet with least completion time is assigned first so that every one of the cloudlet could be finished in less time and load on the framework is balanced. In the paper [9] novel static planning calculation has been proposed taking into account time-sharing virtual machines. It can advance the makespan of cloudlets inside of a datacenter. In [10] two level scheduling algorithm is used taking into account load balancing. The methodology considers both the dynamic requirements of client and load over the cloud resources, thus it meets both the dynamic client requirements and expansion usage of resources. The paper [11] highlights another scheduling strategy with suitable load adjusting system that aides in circulating the cloudlets to the VMs just as prone to their capacity which makes the framework more dynamic, alive, and adjusted. This reduces the completion time of the cloudlets and in addition decreases the makespan of the VMs and the hosts of a datacenter. The algorithm presented in [12] ensures the resource utilization and the completion of cloudlets before its deadline by utilizing two methodologies EDF (Earliest Deadline First) and LWF (Largest Weight First).If the cloud has enough resources to execute then the VM will be apportioned to the occupation and occupation will be expelled from the holding up line and machine status will be changed to busy from idle. But if the required resources are not accessible then this calculation will utilize the LWF as backfilling algorithm, to choose the cloudlet with most extreme waiting time. In [13] authors presented a heuristic scheduling strategy which schedules the cloudlets on resources and aims at improving overall performance of the system. Paper [14] presented a heuristic which aims at reducing the cost of the system by addressing all the servers. The authors of [15] also presented a monitoring scheme of resources to check the usage of resources and migrate them dynamically to improve the overall throughput of the system. Similar algorithm is proposed in [16] which too aim at utilizing the resources and minimize the processing time of the customer requests. Beloglazov and R. Buyya [17] proposed a resource management policy which monitors the resources of the system, migrates them from one cloudlet to another dynamically and switches the idle resources off. This policy saved the energy of the system to a significant extent. Authors of [18] also proposed an approach for resource management which saves the wastage of resources in cloud. D. Warneke and O. Kao [19] presented a framework Nephele which schedules the cloudlets on the different types of VMs by exploiting the resources of the cloud dynamically. In [20] authors have proposed a scheduling algorithm which reduces the turnaround time of the cloudlets in heterogeneous resource environment. The algorithm utilizes the greedy approach for allocation of cloudlets on the resources of cloud. Jun Wu [21] proposed an algorithm BATS (Blocking-aware two-speed) which synchronize the cloudlets on the shared resources dynamically to saves the energy of the resources and aids the cloudlets to meet their deadlines. Another energy efficient scheduling algorithm VILCF(Voltage Island Largest Capacity First) [22] schedules the periodic real-time cloudlets on the multi-core processors and utilizes the resource energy efficiently. A similar work is done in [23] to improve the energy efficiency of system. The authors in [24] proposed a cloudlet scheduling model that minimizes the energy consumption by reducing the number of VMs. In [25] authors proposed an Adaptive and hierarchical cloudlet scheduling scheme (AHS) for dynamically scheduling cloudlets in cloud. The authors in [26] proposed a multi-objective scheduling (MOS) scheme which aims at minimizing makespan and at the same time reducing the resource cost and preserving fault tolerance. In [27] the authors state that deadline of a task is an important factor and highest priority is assigned to the task having earliest deadline. In [28] the authors describe the scheduling algorithm in which tasks are kept in queue and task with least completion time gets executed first and cost matrix is being generated.

Proposed Work

Introduction

Cloudlet scheduling in cloud environment means deciding which resource will be available to which cloudlet at a particular time. There are various QOS parameters which are considered while scheduling cloudlets over datacenters e.g. cloudlet priority, cloudlet length, cloudlet deadline, VM bandwidth, VM MIPS. While scheduling cloudlets the main motive is to reduce response time, cost, execution time, waiting time and turnaround time. Deadline is the final time by which the cloudlet must complete its execution, waiting time is the time for which the cloudlet waits in the ready queue, response time is the time by which the cloudlet gives its first response in case of time sharing system and dynamic submission time is the time at which the cloudlet gets submitted to the broker. The existing algorithm only calculates the makespan while considering two QOS parameters cloudlet length and cloudlet priority. But the proposed algorithm calculates and reduces waiting time and response time while considering cloudlet length, cloudlet priority, cloudlet dynamic submission time, cloudlet deadline, VMs bandwidth and VMs MIPS.

Problem Statement

The proposed algorithm considers QOS Parameters like cloudlet length, cloudlet priority, cloudlet deadline, cloudlet dynamic submission time, VMs bandwidth and MIPS. The cloudlet having lower length, higher priority and earlier deadline gets scheduled first onto the VM that is having highest value of bandwidth and MIPS. The main motive to use dynamic submission time is to reduce requests on a single resource by various cloudlets. Hence there are less number of cloudlet requests for resources and the scheduler can take quick decisions and waiting time is reduced.

Assigning Priority and Deadline to Cloudlets

The proposed algorithm assigns priority and deadline to the cloudlet as defined by the user. In the simulation process unique priority and deadline are randomly generated by the system.

Sorting Cloudlets according to QOS parameters

Credit is assigned to each task according to cloudlet length, cloudlet priority and cloudlet deadline; to assign credit to cloudlets according to deadline Algorithm 1 is followed. To assign credit according to cloudlet length and cloudlet priority CBSA [29] is followed; firstly average length is calculated and compared with individual cloudlet length and then credit is assigned to each cloudlet in the range of 1 to 5 following CBSA [29] and the total credit is calculated by multiplying all three credits and cloudlets are sorted according to descending order of total credit.

Algorithm 1: To sort cloudlets according to credit Assigned

```
Create tasks form i to n with length Li.
Assign priority Pi.
Assign deadline Di.
Assign credit to task based on length and priority according to CBSA
To Assign credit according to deadline
For(i to n)
{
   If(deadline<=9)
      {
          deadline factor=10
      }
   else
      {
         deadline factor=100
       }
Deadline[i]=deadline/deadline factor
}
Total credit=length credit*priority credit*deadline credit;
Sort tasks in descending order according to total credit
```

Creation of VMs
MIPS and bandwidth are defined for VMs and they are sorted in descending order according to combined value of VMs bandwidth and MIPS.

Binding of cloudlets to VMs
Cloudlet with the lower value of credit gets bind to VM with highest MIPS and bandwidth, cloudlets are submitted to the broker with the difference in submission time and then they have to wait lesser time in the queue as a result of which waiting time and response time has been reduced.

Algorithm 2: To Sort VMs according to bandwidth and MIPS and bind cloudlet to VM

```
Generate VMS for j to n

Assign bandwidth and MIPS

  For(j to n)

   {

     Sort VMs according to bandwidth and MIPS in descending order

   }

Dynamic submission time difference = 4ms +5( i) ;

bindcloudletToVm (cloudlet_i, VM_j)

When VM in VM list are finished then next cloudlet is assigned to first VM in VM list.
Repeat until all the Cloudlets in cloudlet list get executed.
```

Experimental Setup and Results

The simulation was done on CloudSim 3.0.3 simulator on eclipse platform. The operating system used for this simulation was Window 7 with core 2 duo processor with 3GB ram and 320 GB hard disk. In this simulation 50 cloudlets having length 1000-50000 are processed over 10 VMs starting from one VM and then increasing the number of VMs one by one with MIPS ranges from 300 to 1000. Waiting time and response time has been calculated for varying number of VMs.

Table 1. Comparison Table

No. of VMs	Waiting Time(in ms)		Response Time(in ms)	
	CBSA	Proposed Algorithm	CBSA	Proposed Algorithm
1	416	324	441	341
2	202	127	227	155
3	130	63	156	91
4	94	40	115	59
5	73	29	99	56
6	59	20	84	52
7	49	11	74	44
8	41	8	67	40
9	35	6	61	37
10	31	4	54	35

As form the comparison table it is clear that waiting time and response time is improved as compared to CBSA [29].

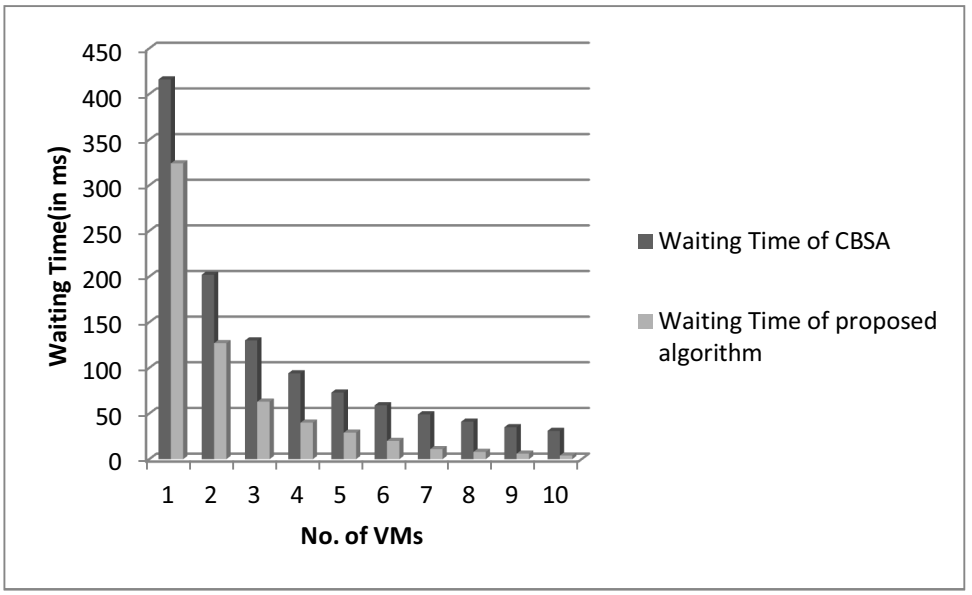


Fig 1.Waiting Time Comparison

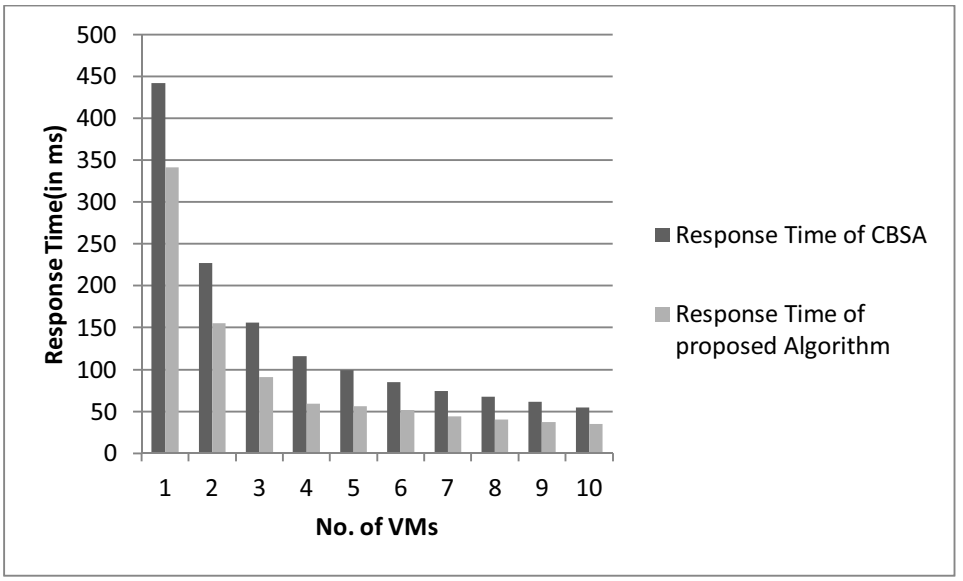


Fig 2. Response Time Comparison

Conclusion

In this paper, a scheduling algorithm is proposed so as to reduce the waiting time and response time of the cloudlets and the experiments conclude that the algorithm outperforms CBSA[29] and the waiting time and response time are reduced significantly through this algorithm. The outputs of the algorithm covey that cloudlet deadline, cloudlet dynamic submission time, VMs MIPS and bandwidth are also important to be considered while working with cloud. Future work would involve improving these factors with the help of more QoS parameters.

References

[1] Liping Zhang, Weiqin Tong, Shengpeng Lu, "Task scheduling of cloud computing based on Improved CHC algorithm," IEEE, july 2014, pp.574-577.

[2] Velagapudi Sreenivas, Prathap.M, Mohammed Kemal, "Load balancing techniques: Major challenge in Cloud Computing - a systematic review,"IEEE, Feb 2014, pp.1-6.
[3] Fatimah M. Alturkistani, Ahmed Z. Emam, "A Review of Security Risk Assessment Methods in Cloud Computing," Springer, vol.275, 2014, pp 443-453.
[4] Lingyan Yin, Jizhou Sun, Laiping Zhao, Chenzhou Cui, Jian Xiao, Ce Yu, "Joint scheduling of data and computation in Geo-distributed cloud systems,"15th International Symposium on Cluster, Cloud and Grid Computing IEEE, may 2015 ,pp.657-666.
[5] Seyedmehdi Hosseinimotlagh, Farshad Khunjush, Rashidaldin Samadzadeh, "SEATS: smart energy-aware task scheduling in real-time cloud computing, "Springer, 2015, vol. 71, pp.45-66.
[6] Chorng-Shiuh Koong, Tzu-I Yang, Chang-Chung Wu, "An Implementation of Automatic Software Testing on Cloud Computing Environment," Springer, 2015,vol.330, pp.1253-1259.
[7] Thanadech Thanakornworakij, Raja F. Nassar, Chokchai Leangsuksun, Mihaela Paun, "A reliability model for cloud computing for high performance computing applications," Springer, 2013, vol.7640, pp.474-483.
[8] Xiaonian Wu, Mengqing Deng, Runlian Zhang, Bing Zeng and Shengyuan Zhou, "A task scheduling algorithm based on QOS driven in cloud computing," Science direct, Procedia computer science, 2013, pp.1162-1169.
[9] Xuejun Li,Siyu Zhou,Jian Wang,Xiangjun Liu,Yiwen Zhang,Cheng Zhang,Yun Yang, "Time sharing virtual machine based efficient task level scheduling in scientific cloud workflows," Springer, 2014, vol.309, pp.121-126.
[10] Yiqiu Fang Fei Wang, Junwei Ge, "A task scheduling algorithm based on load balancing in cloud in cloud computing," Springer, 2010, vol.6318, pp.271-277.
[11] Soured Banerjee, Mainak Adhikari,Sukhendu Kar, Utpal Biswas, "Development and Analysis of a New Allocation Strategy for QoS Improvement in Cloud," Springer, May 2015, pp.1409-1425.
[12] SaeMi Shin, , Yena Kim and SuKyoung Lee, "Deadline guaranteed scheduling algorithm with improved resource utilization for cloud computing," IEEE, 2015, pp.814-819.
[13] Sameer Singh Chauhan, R. C. Joshi, "A waited mean time min-min max-min selective scheduling strategy for independent cloudlets on grid," IEEE, Feb 2010, pp.4-9.
[14] Benjamin Speitkamp, Martin Bichler, "A mathematical programming approach for server consolidation problems in virtualized datacenters," IEEE, 2010, vol.3, pp.266-278.
[15] Jaideep Moses, Ravi Iyer, Ramesh Illikkal, Sadagopan Srinivasan, Konstantinos Aisopos, "Shared resource monitoring and throughput optimization in cloud computing datacenters," IEEE, July 2011, pp.192-196.
[16] S. Gomathi, D.Manimegalai, "An adaptive grouping based job scheduling in grid computing," IEEE, May 2011, pp.1024-1033.
[17] Anton Beloglazov, Rajkumar Buyya, "Energy efficient allocation of virtual machines in cloud datacenters," IEEE, May 2010, pp.577-578.
[18] Yagız Onat Yazır et al, "Dynamic Resource Allocation in computing clouds using distributed multiple criteria decision analysis," IEEE, 2010, pp.91-98.
[19] Daniel Warneke, Odej Kao, "Exploiting dynamic resource allocation for efficient parallel data processing in the cloud," IEEE, June 2011, vol.22, pp.985-997.
[20] Saumitra Singh, Krishna Kant, "Greedy grid scheduling algorithm in dynamic job submission environment," IEEE, 2011, pp.933-936.
[21] Jun Wu, "Energy-efficient scheduling of real-time tasks with shared resources," Elsevier, 2016, pp.179-191.
[22] Jun Liu, Jinhua Guo, "Energy efficient scheduling of real time tasks on multi core processors with voltage island," Elsevier, 2016, pp.202-210.
[23] QingZhao, Congcong Xiongn, CeYub,ChuanleiZhan,XiZhao, "A new energy aware task scheduling method for data-intensive applications in cloud," Elsevier, 2016, pp.14-27.
[24] YonghuaXiong, ShaoyunWan, JinhuaShe, MinWu, YongHe, KeyuanJiang, "An energy-optimization-based method of task scheduling for a cloud video surveillance center," Elsevier, 2016, pp.63-73.
[25] Yizhuo Wanga, Yang Zhang,Yan Su,Xiaojun Wanga,Xu Chen,Weixing Ji, Feng Shi, "An adaptive and hierarchical task scheduling scheme for multi core clusters," Elsevier, 2014, pp.611-627.
[26] Fan Zhang, ,Junwei Cao,Keqin Li,Samee U. Khand, Kai Hwange, "Multiobjective scheduling of many tasks in cloud platform," Elsevier, 2014, pp.309-320.
[27] E. Iniya Nehru, Saswati Mukherjee and Abhishek Kumar, "Deadline based priority management in cloud," Springer, 2015, pp. 745-751.
[28] Mousumi Paul, Goutam Sanyal, "Task scheduling in cloud computing using credit based assignment problem," IJCSE,2011,pp. 3426-3430.
[29] Antony Thomasa, Krishnalal Ga, agathy Raj V P, "Credit based scheduling algorithm in cloud computing environment," Elsevier, 2015, vol.46, pp.913-920.

Prioritized Schedule Algorithm : Time Table Generation

Rutuja Popale*, Veeresh Hiremath**, Nityanand B Nishanimath***, Anil Rathod****, Vandana Kale*****, G N Jayalaxmi****** and C Sujata*******
* ******* Department of Computer Science Engineering, BVB college of Engineering and Technology,Hubli Karnataka, India
{rutujapopale, muttu1789, nbnishanimath, rathoda125, kalevandana002}@gmail.com
jaya_gn@bvb.edu, sujata_c@bvb.edu

Abstract: Author's have taken into consideration issues of task allocation and resource selection problem with respect to the minimum processing cost to be an NP-hard problem. The problem of timetable scheduling is described as a highly constrained NP-hard problem. A lot of complex constraints need to be addressed for development of an efficient algorithm to solve this problem. The hand operated system of Time table preparation in college is very monotonous and time consuming which results in either the same teacher ending up with more than one class at a time or a number of classes conflicting at the same classroom. Due to non-automatic perspective absolute utilization of resource has proven ineffective , In order to deal with such problem , mechanized system can be designed with computer aided Time table generator which will serve in many applications .The proposed design of algorithm has the ability to manage Students ,Lecturers, venues, slot and courses of the department, the data given as input is saved in a MYSQL database using a configurable database module and report generation module to generate the best combination of courses and slot for creating a feasible timetable for working days of week, making excellent application of all resource in a way which will be best suited for the constraints.

Keywords: Genetic algorithm, Heuristic algorithm, Prioritized schedule algorithm (PSA).

Introduction

Resource Scheduling and Managing is part of every system and it governs the system. Resource scheduling and management is a NP- hard problem, such problems do not have hundred percent solution. Gaming project assignment, decision subset sum problem, time table formation are examples for NP hard problem. In this article time table formation is taken as the case study for Resource Scheduling problem with the proposed novel algorithm. The main purpose of this research work is to generate a time table with available resources, and has to manage the resources effectively in allocating the time slots.

Even though most college administrative work has been computerized, the lecture timetable scheduling is still mostly done manually due to its inherent difficulties. Most of the institutions are spending a lot of time in preparing the time table for their academic activities. Generating a time table manually is a tedious task and is a time consuming process, automating it will reduce the human efforts involved and will save enormous amount of time. In order to overcome this problem the author has designed an algorithm named PSM Algorithm to automate the process of time table generation.

The automated time table generator can be used in various fields may it be a school, college, IT Company or may be for a personal purpose. As a case study, the algorithm has been tested using the database of Computer Science and Engineering Department of author's institute. Provision is provided to modify the application based on the requirements of the user.

The application handles issues such as collision of resources, fixed slot allocation, allocation of labs and optimal utilization of the resources is done while generating the time table. At first the fixed slots in the time table are allocated manually, the provision of doing this is given only to the administrator. After the completion of this assignment the theory classes are allocated and then the labs are assigned. In the next section of the article the literature survey made during the research is discussed.

Organization of article is as follows: the ordering of article is as follows, section 2 describes survey on time table generation algorithms. The proposed novel algorithm is narrated in section 3 comparative study of different Algorithm in section 4 and article is concluded in section 5.

Literature Survey

Literature survey for this research paper is made on NP-hard problem, Resource management and Scheduling problem.

Anisha Jain, Ganapathy S C Aiyer et el. [1] Methodologies of genetic algorithm is used in combination for the implementation in this article. Genetic algorithms (GAs) are search algorithms that begin with a set of potential solutions. This set then evolves towards a set of more optimal solutions. The proposed paper takes four parameters as input: Person – name of lecturers; Subject – name of courses in the class, Room – name of classes and capacity of each and Time interval – starting time and end time. The system selects one rule with the highest priority to fire, or arbitrarily selects one rule to fire if

there is more than one with the same priority. Intermediate and final reports are generated, the workload is equally distributed among the lecturers, prioritizes time slots according to customized priority are few of the issues handled.
Dipti Shrinivasan [2] This article presents an Evolutionary Algorithm (EA) based approach in solving a heavily constrained university timetabling problem. The approach uses a problem-specific chromosome representation. Heuristics and context-based reasoning have been used for obtaining feasible timetables in a reasonable computing time. An intelligent adaptive mutation scheme has been employed for speeding up the convergence. The comprehensive course timetabling system presented in this paper has been validated, tested and discussed using real world data from a large university.
Mei Rui[3] mathematical model for the course timetable system is proposed. At the same time, through the use of the pattern recognition technology in artificial intelligence, aiming at this mathematical model a new university course timetable system design program is proposed and realized. This program not only can well solve the shortages of the existing course timetable system, but also is simple and easy to operate, has strong versatility.
Bhaduri A [4] evolutionary techniques have been used to solve the time table scheduling problem. Methodologies like Genetic Algorithms (GAs), Evolutionary Algorithms (EAs) etc have been used with mixed success. In this paper, we have reviewed the problem of educational time table scheduling and solving it with genetic algorithm. We have further solved the problem with a mimetic hybrid algorithm, genetic artificial immune network (GAIN) and compare the result with that obtained from GA. Results show that GAIN is able to reach the optimal feasible solution faster than that of GA
The survey motivates the author to recognize the issues and challenges to solve time table generation on NP hard problem and it is presented in next section.

Proposed System

In order to deal with timetable generation issues the Author's are proposing a system which would mechanically generate timetable for institute. After doing detail study about the literature survey Author's have designed an algorithm which is effective as compared to the one that already exists in [1]. PSM Algorithm is component of research work which produces he HTML based timetable even / odd semester sheet as the output. The project takes various inputs from the user such as Teacher List, Course List, Semester List, Room List, Day List and Timeslot as well as various rules, facts and constraints using web based forms. During design of algorithm the challenges faced by author in designing the algorithm are as follows'

- First problem was, from where to start?
- Second problem was, does it really going to work?

PS Algorithm is generally designed to automate the generation of time table preparation in order to reduce the manual effort and time required in manual preparation. The proposed system is web based to automate the time table preparation. The main modules of the proposed system are Administrator, Data base management & Report generation which are explained in further subsections 3.1, 3.2 and 3.3 respectively

Administrator module
The security feature is very strong, therefore record can be only inserted, updated by Admin & assigning related faculty to that subject. Admin has to login with Password screen includes the username and password. On being entered, the password is evaluated and the entry is given only to the correct password entry and can perform operations and also view Time table. Whereas Staff and Student are not provided with the login and can directly view the master time table & staff time table.
3.1.1 PS Algorithm.
Explanation: In Line 1 Semester and Division is loaded for which the time table has to be generated, line 2 takes all the fixed slots ad calculates the contact hours ,Line 3 takes initial weekday as Monday and In line 4 while loop is taken that should execute until weekday=sat ,In line 7 initialization of daycount variable takes place, In line 8 it is checking whether daycount is lesser than 4,which means maximum 4 subjects can be allocated on a day. From line 9 to 20 it checks whether slot is free, faculty is free and Resource is free, if all of them are free than allocate it to time table, If either of them is not free than swap it with some other slot or faculty or resource, In line 21 if daycount>4 than increment the weekday to Tuesday and continue from line 4.

Data base management module
The database module[7] handles all Input and output operations to MYSQL database backend such as read, write and delete. It contains all the Tables used by PSM for the storage and retrieval of courses, lectures, venues and other necessary data.
ER Diagram for PSM.

```
Step 1:   Load Semester and Division.
Step 2:   Assign all fixed slots an calculate the contact hours
   3     Weekday=Mon
   4     While (Weekday=SAT)
   5     Load Courses
   6     Load Resources
   7     Initialize daycount   variable
   8     If (daycout<4)
   9             If (slot[i] ==free)
   10                 If (isFaculty Free (subject[i]))
   11                     If((isResourceFree(res[i]))
   12                                {
   13                                  Allocate  subject and resource to time table
   14                                }
   15                      Else
   16                           Resource=resource+1
   17                  Else
   18                    Swap subject[i] with subject [i+1]
   19              Else
   20                Slot=slot+1
   21     Else
   22       Weekday++
   23    End while
```

Figure 3.2 PS Algorithm

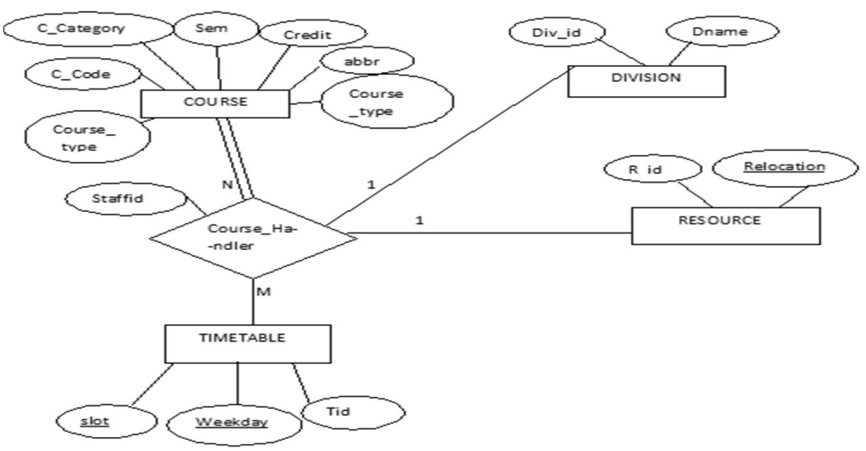


Figure 3.3 ER Diagram for PSA

TABLES USED FOR PSA

COURSE

Ccode	Sem	Credit	Cname	category	ABBR	Contact_ hours	Course_type

COURSE_HANDLE

Course_code	Staff_id	Sem	D_Divname

RESOURCE

Rid	Rlocation

DIVISION

Div_id	Dname

TIME_TABL

Tid	Weekday	Sem	Div_name	Slotc_code	F_sid	Res_id

Figure 3.4 Tables used for storing data in PS Algorithm

Author has designed tables (Course, Course_handler, Division, Resourse and timetable)as per the requirement for storing the data before and after procesing the Algorithm.In course table Primary key considered is Ccode since the code for particular subject is unique ,In couse_handler table ccode,sem and div are taken primary key in order to uniquely identify particular course taken by staff for instance same staff may engage same course for both the division so here staffid will not act as primary key,In timetable table all the fields are taken as Primary key except tid in order to uniquely identify each record which represents record after the Algorithm processing which performs the assingment ,here each record shoud be unique to take care of clashes and tid is used in case of deleting any record.

Report generation module

Report generation module takes the inputs and processes it in order to consider the following features and constrainst

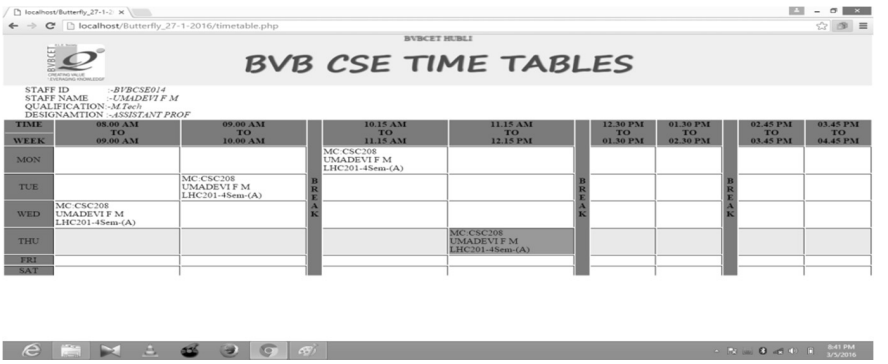
- Hard constraints-
  - A Student should have only one class at a Time
  - A Teacher should have only one class at a time
  - A classroom should be booked only for one class at a time
- Soft Constraints-
  - Courses must be eventually distributed
  - Scheduling of teachers should be well spread in the week.
- Fair distribution of classes to staff
- NO slot clashes
- Always considers the other department slots first
- Considers all the constraints such as resource available, staff available and contact hours.

Snapshots of generated timetable

TIME / WEEK	08.00 AM TO 09.00 AM	09.00 AM TO 10.00 AM		10.15 AM TO 11.15 AM	11.15 AM TO 12.15 PM		12.30 PM TO 01.30 PM	01.30 PM TO 02.30 PM		02.45 PM TO 03.45 PM	03.45 PM TO 04.45 PM
MON	PCD:CSC222 KARIBASAPPA K G LHC203-4Sem-(B)	JAVA:CSC223 MANJULA PAWAR LHC203-4Sem-(B)		MC:CSC208 A S NAYAK LHC203-4Sem-(B)	ADDS:CSE350 PARIKSHIT P HEGDE LHC203-6Sem-(B)						
TUE	JAVA:CSC223 MANJULA PAWAR LHC203-4Sem-(B)	MC:CSC208 A S NAYAK LHC203-4Sem-(B)		CN:CSC320 VIJAYLAKSHMI M LHC203-6Sem-(A)	ADBMS:CSE308 VIJAY S BIRADAR LHC203-6Sem-(B)						
WED	MC:CSC208 A S NAYAK LHC203-4Sem-(B)	CN:CSC320 VIJAYLAKSHMI M LHC203-6Sem-(A)		OS:CSC321 SHANTALA G LHC203-6Sem-(A)	CN:CSC320 PADMASHREE DESAI LHC203-6Sem-(B)						
THU	CN:CSC320 VIJAYLAKSHMI M LHC203-6Sem-(A)	OS:CSC321 SHANTALA G LHC203-6Sem-(A)		CN:CSC320 PADMASHREE DESAI LHC203-6Sem-(B)	OS:CSC321 NAGARATHNA D KULENAVAR LHC203-6Sem-(B)						
FRI	LI:CSE432 G S HANCHINMANI LHC203-8Sem-(B)	NS:CSE429 S I BATTUR LHC203-8Sem-(B)									
SAT											

Figure 3.5 view of class time table

From the above snapshot it gives a view of number of subjects schedule in the week along with the details of staff who is taking a particular subject,staff who is intended to take,Subject code,duration(slot),semester and division so that one can get an idea of scheduled classes in a particular classroom for the week.



TIME / WEEK	08.00 AM TO 09.00 AM	09.00 AM TO 10.00 AM		10.15 AM TO 11.15 AM	11.15 AM TO 12.15 PM		12.30 PM TO 01.30 PM	01.30 PM TO 02.30 PM		02.45 PM TO 03.45 PM	03.45 PM TO 04.45 PM
MON				MC:CSC208 UMADEVI F M LHC201-4Sem-(A)							
TUE		MC:CSC208 UMADEVI F M LHC201-4Sem-(A)									
WED	MC:CSC208 UMADEVI F M LHC201-4Sem-(A)										
THU					MC:CSC208 UMADEVI F M LHC201-4Sem-(A)						
FRI											
SAT											

Figure 3.6 view of staff time table

From the above snapshot a staff can get a view of the schedule classes which particular staff is intentded to take with information that contain staff id,staff name,staff Qualification,Staff designation, subject name,Subject code,duration(slot),semester ,division and class room number so that one can get an idea of scheduled classes in a particular classroom for the week.

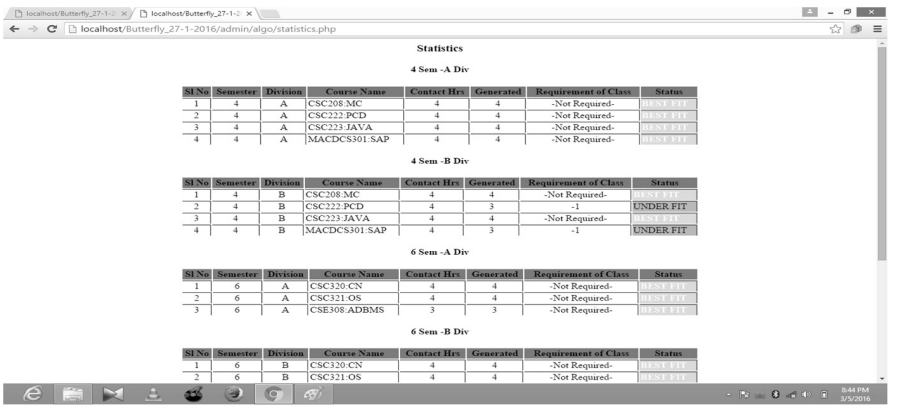


Figure 3.7 Statistics view

Statistics view gives a view of whether number of particular subjects are allocated as per the contact hours specified or its overloaded /under loaded and specifies it as "UNDER FIT" if less classes are allocated than the contact hours, "OVER FIT" if excess classes are allocated than the contact hours and "BEST FIT" allocated classes are equal to contact hours.
Author has tested the Algorithm using data from CSE department in the Authors Institute, Author has solved 80% of the issues which come under automatic time table generator with very least human interaction. The proposed system provides simple data entry, reduces time consumption and the effort in framing the timetable manually. The benefits of this approach are simplified design and reduce development time .After analyzing the result and comparing with existing algorithm ,Algorithm suggested by Author take various advantage over the existing one which can be seen in further page.

Result Analysis

Author's are using various tools[6] during the implementation of the Algorithm which includes the following.

- Data base used is MYSQL, since it is a open source database system, occupies less space on disk and can be installed in all major operating system.
- JavaScript is used for providing alert message and pop up message if any invalid entry.
- PHP is used for server side scripting since one can connect to database easily and it is relatively fast.
- Jquery and CSS are used for creating Animation for WebPages and also border and shadow applied for data submission form.
- Graph for comparison of GA,HA and PSA in terms of performance and working

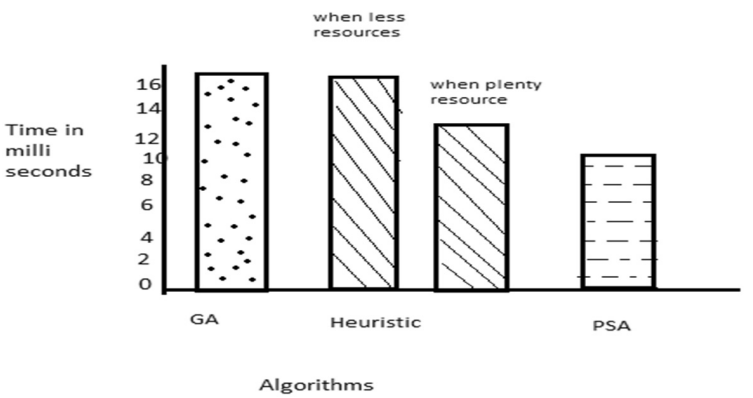


Figure 3.9 Comparison of GA,HA and PSA in terms of time

Table 1. Comparison study of authors algorithm(PSA) with existing ones that is GA and HA

Parameters	Genetic Algorithm	Heuristic Algorithm	PS Algorithm.
Performance	Good	Better	Better
Time taken	High	-Less when plenty resource -high when less resource	Least
Deployment status	NO	Yes	Yes
Availability of number of resource handled	NO	Yes, Two approaches proposed, one to be used when scarce resource and other when plenty of resources available	All resources are consider fixed
Weekday Consider	Generalized	Works for 5 days in a week.	Works for 6 days in a week.
Constraints under consideration	-Most trivial: Teacher must not be present in more than one class during the same time slot - Classroom should not be double booked - Classroom must be large enough to hold the strength for each class -Labs should not be double booked	-Most trivial: Teacher must not be present in more than one class during the same time slot -There is a limit on the number of people that can be accommodated in a room	- The system generates intermediate level as well many final reports including weekly time table, teacher timetable, room wise time table, student time table, department level time table etc. - It distributes workload of lectures equally among all the specified time -avoid the conflict or overlapping class slots. - To take care of assigning classes so that overlapping of class should not happen, since the number of class rooms and labs are fixed -To handle reserved slots (Humanity department, Math's or other departments) which will have reserved time slots so the regular classes should not overlap.

The Genetic algorithm solution is time consuming. Also, in cases where resources are scarce, the time required maybe considerably high. The HA proposes a practically implemented approach that deals with both, the abundance as well as scarcity of resources. The third PSA based algorithm reduces time taken for generation of timetable.

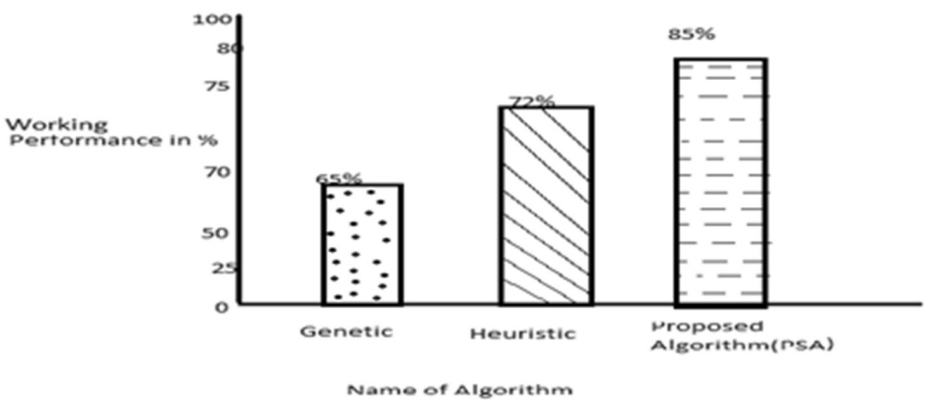


Figure 3.10 Comparison of GA, HA and PSA in terms of performance

This Graph depicts the working efficiency of GA,HA and PS Algorithm, The GA is not deployed and provides only 65% of solution. The HA is providing 72% of working solution and has been deployed. Whereas PSA provides 85% of solution considering all the constraints.

This graph depicts GA does not consider any resource to be handled ,HA provides 2 solution one when scarce resource and another solution when plenty of resources are used, and PSA uses all the available resources.

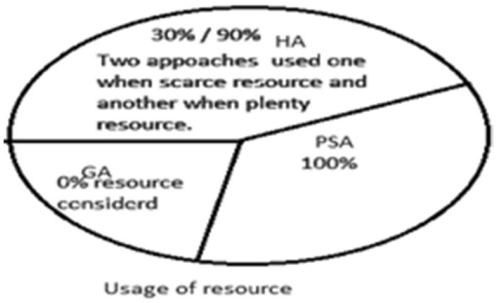


Figure 3.11 comparison study in terms of resource usage

Conclusion

As discussed PS algorithm for automatic time table generation has been proposed. The intention of the algorithm to generate a time- table schedule automatically is satisfied .By automating this process with the help of computer assistance timetable generator can save a lot of precious time of administrator who are involved in creating it and are much more accurate ,precise than the ones created manually. The proposed algorithm considers all the soft and hard constraints and assigns classes so that overlapping of class never happen, since the number of class rooms and labs are fixed and also enables fair distribution of classes to staff.

Author's have used real data of CSE department of Authors Institute to test the method and have derived the conclusion that it is functioning efficiently.

Acknowledgement

We would like to thank Dr.P G Tiwari Principal , Prof. K.P Biradar HOD and Time Table preparation Committee members of Computer Science Engineering ,BVBCET for providing all the data and resource required for carrying out this research work.

References

[1] Anisha Jain, Ganapathy S C Aiyer et el**.** A Literature Review on Timetable generation Algorithms based on Genetic Algorithm and Heuristic approach, Computer Engineering, MPSTME, Mumbai, India

[2] Dipti Shrinivasan "automated time table generation using multiple context reasoning for university modules" Published in: evolutionary computation, 2002. cec '02. proceedings of the 2002 congress on (volume:2)

[3] Mei Rue Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on (Volume.4)

[4] Bhaduri a "university timetable scheduling using genetic algorithm". Advances in Recent Technologies in Communication and Computing, 2009. ARTCom '09. International Conferenc

[5] Simon Robinson, Christian Nagel, Jay Glynn, Morgan Skinner, Karli Watson, Bill Evjen "Professional C#", 3rd Edition, Wiley Publishing,Inc 2004 [7] Silberschatz, H. Korth and S. Sudarshan, "Database System Concepts", 4rth Edition, McGraw-Hill International, 2002.

Handling Sink and Object Mobility in Wireless Sensor Networks

Kulwardhan Singh* and T. P. Sharma**
*National Institute of Technology Hamirpur, India
kulwardhan@yahoo.com
**National Institute of Technology Hamirpur, India
teekparval@gmail.com

Abstract: The rapid progress of wireless communication and embedded micro-sensing MEMS technologies enable the wireless sensor networks (WSN) offer many new possibilities for applications such as remote monitoring objects and environmental surveillance by allowing the sink/user to move around freely. However, monitoring mobile objects and disseminating sensing data to the mobile sink raises significant design challenges for the routing scheme. In addition, limited power supply of sensor nodes and therefore reducing energy consumption in order to prolong the lifetime of the WSN is another challenge. In this paper, in order to monitor the mobile objects and disseminates the data to mobile sinks, we propose a routing scheme for Handling Sink and Object Mobility (HSOM) in Wireless Sensor Networks, which is energy efficient and has minimum rerouting frequency. Analytical and simulation study reveals significant improvement over the existing schemes.

Keywords: Wireless Sensor Network (WSN), Sink Mobility, Object Monitoring, Energy Efficient.

Introduction

The Wireless Sensor Networks (WSNs) are made up of wireless nodes endowed with sensing capabilities that are deployed for implementing a host of different applications. Wireless sensor networks (WSNs) are widely used in many industrial and civilian application areas, including industrial process monitoring and control, machine status monitoring, environment and habitat monitoring, healthcare applications, home automation, and traffic control. Typically, a large number of tiny sensor devices called motes constitute a WSN, where motes are considered as constrained in resources, such as limited on-board memory, short-range radio transceivers and limited battery power. These sensor nodes form a decentralized, multi-hop, self-organized network system. Depending on the application environment, nodes are interfaced with various sensors for monitoring some phenomenon of interest (temperature, humidity, pressure, etc.) and forward the stimulus data to the data centres (Sinks) through multi-hop communication [1]. To prolong network lifetime, minimizing the energy consumption of individual sensor nodes is important, as well as balanced energy consumption among all the sensor nodes is also desired.

Monitoring the mobile objects is one of the most important areas where the advantages of wireless sensor networks can be exploited. These networks may be deployed for the application in military for tracking enemy vehicles, detecting illegal border crossings as well as in civilian for monitoring the movement of wild animals in wildlife preserves [2][3]. Object monitoring consists of detecting and monitoring locations of real-world objects, using several types of sensing such as acoustic, seismic, electromagnetic, etc. WSN can have one or multiple sinks that sends query or control commands to the sensor node and collect the information from the sensor node(s). In traditional WSNs, sensor nodes are distributed in the sensing field whereupon detecting some event of interest, nodes report the sensed event back to some static sink(s) through single-hop or multi-hop communication. One major drawback of such communication infrastructures is occurrence of hot-spot or sink-hole problem in the neighborhood of the sink(s). This is because sensor nodes close to the static sink will consume more energy and thus their energy will deplete quickly. To overcome hot-spot or sink-hole problem, the concept of mobile sink was introduced in WSN [4], that not only results in balanced energy consumption among the nodes but can also be exploited to connect isolated segments of the network. The mobile sink(s) are more energy efficient than the static, but has the additional overhead such as sink's location maintenance, continuous data delivery and dynamic route adjustments with sink mobility [5][6]. There are various routing protocols proposed for WSN in order to deal efficiently with the sink mobility [7,8,9]. The mobile sink has the multifold advantages like hotspot problem removal, energy efficient, longer network lifetime etc., but also include new challenges such as sink location management and dynamic route adjustments.

There are many protocols developed for WSN, which support the mobile sink(s) such as Directed Diffusion [10], GEAR[11], GBR [12]. These protocols maintain the location of the mobile sink by continuously propagating the location of the sink throughout the sensor network, so that all sensor nodes are updated with the recent location of the sink(s). But frequent updating cause traffic increase in WSN, collision in wireless transmission and more power consumption. TTDD[13] provides

two disseminating tiers for large-scale sensor networks with multiple mobile sinks. TTDD architecture exploits the fact that the sensor nodes are stationary and are location aware and queries of multiple mobile sinks are confined within the local only. In TTDD, on event detection each source node proactively constructs a grid throughout the sensor field, so that, only sensors located at the grid points need to acquire the sensing information. However, when multiple events occur sensed by multiple nodes, each node proactively constructs its own grid structure, due to which communication and storage overhead of TTDD architecture increases. Therefore, as number of sources increases, the data dissemination point management overhead increases considerably. Similarly, ALS [14] protocols also uses the virtual grid structure to find the location of mobile sinks. ALS is a grid based protocol that provides sink location information in a scalable and efficient manner. In ALS, the sink agent distribute the location information of the sink using anchor system in four straight orthogonal directions (North, South, East, West) and all the grid nodes that lie along the routing path are recruited as anchor which acts as location server. On occurrence of an event, the source node will register itself with the nearest grid node, which is known as the source agent. The source agent will send four query packets in orthogonal directions to find the location of the sink agent. Once the source agent receives the sink agent's point location, it forwards it to the source, which finally sends the data packets to the sink agent using the GPSR protocol [15]. This protocol has some drawbacks such as source can't identify the location of the sink if it has not constructed the anchor system, detour problem occurs when sink has high mobility, hotspot problem for border nodes because ALS always uses its border nodes when it provides its sink location. Tsai H. et al. proposed Dynamic Object Tracking (DOT) protocol [20] for sensor networks in 2007, which concentrate on the mobile users how to query target tracks and obtain the target position effectively. This protocol guides a mobile user to chase a mobile object and avoid flooding to obtain the present location of the object. In this protocol the mobile user will catch the target along the sequence of beacon sensor nodes. As the target of moving arbitrarily, therefore, track rout never be a straight line. Thus the mobile user never catches the mobile target having velocity more than mobile user. E. Lee et al. proposed IGS [16] in 2011, which exploits a k-level Independent Grid Structure (IGS) for data dissemination from sources detecting the events to sinks. This protocol supports the scalability and mobility against sinks and events in wireless sensor networks. Sources are required to send event reports to grid header in the inner most layer of IGS and sinks can send queries to grid headers in IGS to request event reports. The grid head aggregates report packets of the same event and makes an aggregated report packet. Then, the grid head disseminates the aggregated report packet to its higher level grid head and this process continues till report packets reach at the inner most grid head (i.e. 1-level grid head).

In this paper, we propose a routing scheme for Handling sink and object Mobility (HSOM) in Wireless Sensor Networks, which is a grid-based routing to monitors the mobile objects and disseminates the data to the mobile sinks. In this scheme, the grid is constructed by the sink appearing first in the sensor field or when no valid grid exists. The Grid Nodes (GNs) are used to forward the data/query between source and mobile sink. A node become the source upon detection of an object and report the attributes of the object to the nearest GN called Source Grid Node (SGN). The SGN initiate the path setup message to sink as well as sends the object appearance message to its all neighboring nodes. This will help to continuous monitoring of object while moving across the cells in sensor network. In this scheme, SGN sends the updated location of the object to the sink. Thus sink maintains the complete route map of the object mobility. As object is always in the periphery of SGN, the sink can reach to the target through straight line instead of following the arbitrary path. This scheme handles the mobility of sink efficiently and has the ability to modify partial or full path to avoid detour problem.

Rest of the paper is organized as follows. Section 2 describes the grid construction, object detection, object monitoring and sink mobility. In section 3, performance of the (HSOM) is evaluated. Section 4 concludes the work.

Handling Sink and Object Mobility (HSOM)

The basic assumption considered for HSOM protocol are mentioned below:

- The sensor field is represented as a two-dimensional plane constructed along x-axis and y-axis and divided into equal sized cells.
- The sensor nodes are randomly deployed in two dimensional square field. The Sensor nodes remain stationary and aware of their geographical location using GPS system or localization algorithm [18].
- Data/query is disseminated using single-hop or multi-hop communication.
- HSOM uses the grid that is constructed by the sink appearing first in the sensor field or when no valid grid exists. All other sources and sinks appear thereafter use the same existing grid.
- Each sensor node is aware of its available energy. One or more mobile sinks are deployed in the sensor field to gather data.
- Objects appear field are wondering across the sensor field.

Grid Construction

In HSOM scheme, the grid construction is initiated by the sink appears first in the sensor field or when no valid grid exists. Sink starts grid construction process by keeping itself at one of crossing point (CP) of the grid with coordinates (X_S, Y_S). The

grid is constructed in same way as mentioned in SLDD [17]. The two dimensional geographical coordinates (x, y) of this sink thus become starting point for formation of grid of square sized cells. In this scheme, the node nearest to the CP and within radius r (where r = R/8 and R is the transmission range of a sensor node) from CP is selected as Grid Node (GN). Each GN can communicate with its neighboring GNs in a single hop communication. Thus, in HSOM, the cell size is determined by the radio range of sensor node. As each GN can forward the data to all neighboring GNs in a single hop, therefore, two GNs lying diagonally can't be apart more than their transmission range R. Thus, the cell size α is determined as:

$$\alpha = \frac{3R}{4\sqrt{2}} \tag{1}$$

All other crossing points (CPs) located at $P = (X_P, Y_P)$ are calculated using starting point (X_S, Y_S) and cell size α as:

$$\{X_P = X_S + i * \alpha, Y_P = Y_S + j * \alpha; \} \tag{2}$$

Where i, j= ±0, ±1, ±2, ±3, ……..}

For any GN, all its neighboring GNs are lying within its transmission range. Therefore, it can communicate with its neighboring GNs in a single hop communication.

Object detection process

When an object appears in the sensor field, it is detected by one or more nodes with the cell. This sensor node become the source and sends the report message to the GN that is nearest to the sink called Source Grid Node (SGN). There are chances that the object is detected by one or more sources. All these sources send the stimulated data about the object to the SGN. The SGN aggregates the data if required in order to eliminate the redundant data. The SGN is responsible for path setup and data delivery to sink. The SGN sends the path setup message to upstream GN towards sink, which in turn further forwards the message to its upstream GN towards sink. This process continues till message reaches at the sink as shown in figure 1.

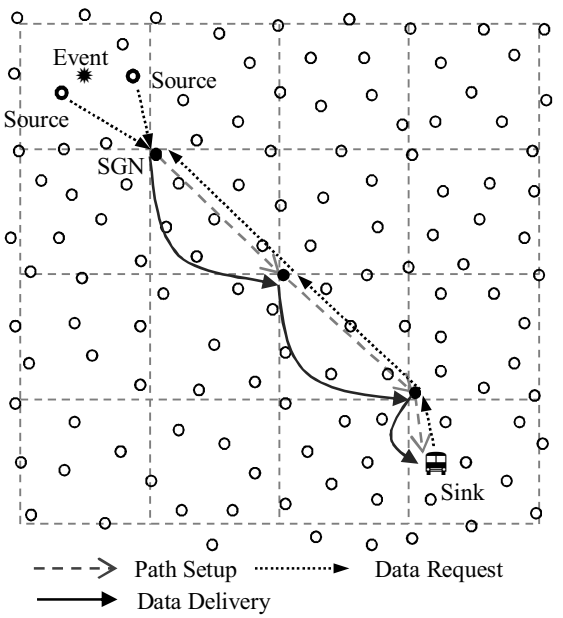


Figure 1. Event Detection and Announcement

Object Monitoring Process

As the object is mobile, there are chances it may go out of site of some source nodes and some new nodes become the sources. These new source nodes send the stimulated data to the SGN. To monitor the moving object continuously, SGN sends the object occurrence message to its all neighboring GNs as shown in figure 2. All these GNs evaluate the data path to the sink in advance to maintain the continuous data delivery. This will help the active SGN to handoff its responsibility to the GN when object enters in its periphery. As the object is moving randomly in the sensor field, therefore, the possibility to selects a GN as new SGN is totally depends upon object mobility. When the object is moving in the cells for which SGN is

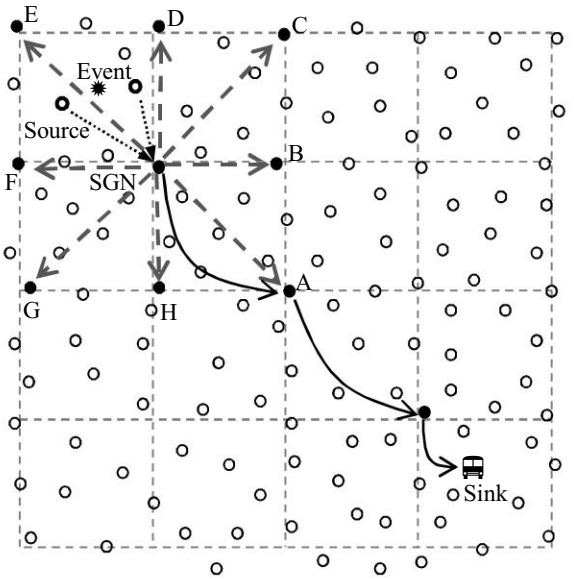


Figure 2. Announcing occurrence of an object to neighbouring GNs

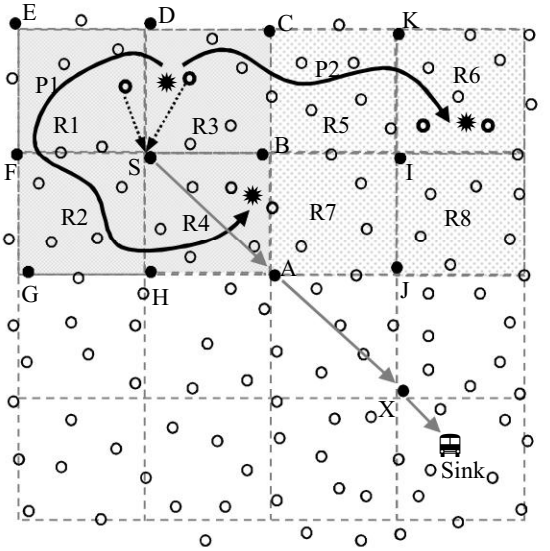


Figure 3. Monitoring object movement and delivering data to sink.

Table 1. Selection of SGN and FN w.r.t. object mobility in different regions

Region in which mobile object enters	GN selected as SGN	Neighbour GN/FN selected for data forwarded
R1	S	A
R2	S	A
R3	S	A
R4	A	X
R5	B	A
R6	I	A
R7	A	X
R8	J	X

common GN, then data is deliver through the same path. As Shown in figure 3, the object is moving along the path P1 over the regions R1, R2, R3 and R4. In this case source grid node 'S' is the common for all these regions. When the object is in the regions R1, R2, R3, the nodes that become the source sends the message to GN 'S'. But when the object is in the region R4, there are two GNs i.e. 'S' and 'A' that are forwarding the data towards the sink. In this case, the node that is nearest to the sink is selected as new SGN. Thus GN 'A' will become the new SGN and all new sources will forwarded the message/data to the newly selected SGN. This SGN then forward the data to sink through same existing path.

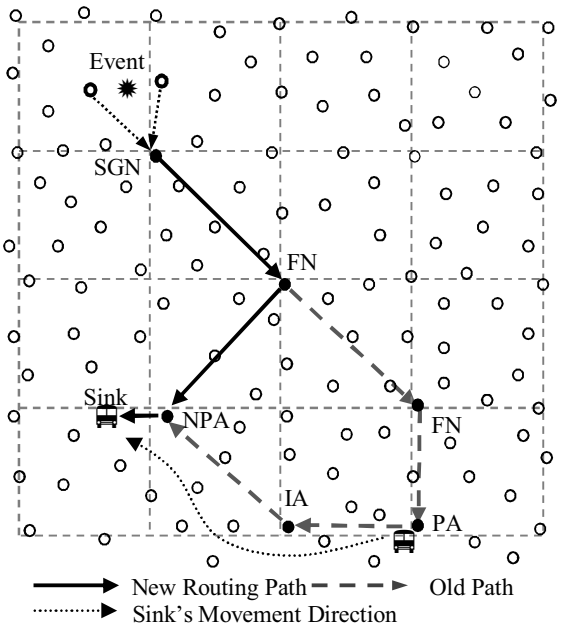


Figure 4. Handling Sink Mobility

When the object moves across the cells where SGN is not common, the alternate GN is selected as new SGN depending upon different scenarios. As Shown in figure 3, the object is moving along the path P2 over the regions R5, R6, R7 and R8. When the object enters in the cell e.g. region R7 where one or more GN is/are acting as SGN or Forwarding Node (FN), then the GN that is nearest to the sink as new SGN (in region R7 the GN 'A' will become the SGN). When object enter in the cell where no GN is acting as SGN/FN, then GN that is nearest to sink and in the neighborhood of previous SGN is selected as new SGN. This new SGN will forward the data to sink through previous SGN or any neighbor FN nearest to sink. Thus depending upon the position of the object as shown in figure 3, the selection of SGN, and the FN through which data is disseminated is summarized in table 1.

Handling Sink Mobility

The proposed HSOM scheme supports the sink and object mobility. Therefore, it is required to monitor the object and maintain the path for continuous data delivery. When sink moves, the node that is nearest initial CP is selected as Primary Agent (PA). PA communicates with the mobile sink while it moves within one hop distance. PA is responsible for receiving the query from the sink and forwards the data to it. When sink moves beyond PA range then it selects the nearest GN as Immediate Agent (IA). If IA is a FN on the existing path, then it become New PA (NPA) and removes the old upstream path from NPA to PA. If IA is not a FN, then it sends a message to its neighbour GNs. If any neighbouring GN is acting as FN on existing path except PA, then IA selects the FN that is nearest to SGN as NPA and removes the upstream path from NPA to PA. Otherwise, IA become NPA and receives the data through old PA. In this situation if NPA also initiates a process to check the detour problem. If NPA discovers the economic path from SGN or intermediate FN, then it initiate new path setup process as shown in figure 4. Once new path is setup, the SGN or FN from where new path is setup, sends path termination message through old path. When NPA receives this message, it stops receiving any more data from PA. This will helps to avoid loss of data flowing through old path.

Performance Analysis

In this section we evaluate the performance of HSOM and compared with some existing schemes such as TTDD and IGS. The performance of HSOM is evaluated by comparing it to TTDD and IGS in terms of in terms of total energy consumption, average delay with varying number of nodes and sinks/event mobility. In this performance evaluation we use the energy model as described by Bhardwaj M. et al. [16]. The key energy parameters are the energy needed to sense a bit (E_{sense}), receive a bit (E_{rx}) and transmit a bit over a distance d (E_{tx}). Assuming path loss in energy model is $1/d\eta$. The default simulation setting has a square sensor field of size 2000 x 2000 m^2 in which 200 sensor nodes are uniformly distributed. Some of these sensor nodes act as sources and generate one data packet per second. Simulation model is run 100 times and the observation is based on the varying numbers of sensor nodes, sink mobility and event mobility. The size of control/query packet is 36 bytes and data packets are 64 bytes. Path loss is set as $\eta = 2$. The transmission range R of each sensor is 100 m and the value of α is evaluated according to equation (1). Table 2 summarizes various simulation parameters.

Effect of node density on total energy consumption and average delay

In this subsection we evaluate the total energy consumption with varying node density. The number of sensor node varies from 200 to 700 and four sinks are moving in the field at a speed of 10 m/s. The total energy consumed by HSOM is less energy as compared to TTDD and IGS as shown in fig 5. It is observed that HSOM consumes 16% and 11% less energy when compared with TTDD and IGS respectively. This is because node density doesn't impact much in case HSOM scheme as the data/query communication through GN only. Also, in HSOM single grid structure is used where as in TTDD one grid per source is constructed and in case of IGS k-level grids are used. The overall delay of HSOM is approximately 18% and 30% less when compared with the TTDD and IGS respectively. This is because HSOM uses the optimal path as it has the ability to modify partial or full path efficiently to avoid any detour problem while sink is mobile.

Effect of sink and event speed on total energy consumption

In this subsection we evaluate the total consumed energy with various sink and event speed. The sink speed is varied from 0, 2, 4, 6, 8 to 10m/s. There are 200 sensor nodes deployed in the field. The total energy consumed by HSOM is considerable less as compared to TTDD and IGS as shown in figure 7. This is because HSOM has the ability to modify partial or full path efficiently to avoid any detour problem. The figure 8 shows the total consumed energy when the speed of object/event is varies from 0, 2, 4, 6, 8 to 10m/s. It is observe that energy consumed by HSOM is approximately 51% and 38% less as compared to TTDD and IGS respectively. This is because HSOM uses single grid structure where as TTDD constructs one grid per source and IGS uses K-level grid structure. Therefore TTDD and IGS consume more energy for constructing and maintaining the grid structure.

Table 2: Simulations parameters

Parameters	Values
Size of Sensor Network	2000 X 2000 m^2
$\propto_1$ ($\propto_1 = \propto_{11} + \propto_{12}$)	180nJ/bit,
$\propto_2$	10pJ/bit/m^2
Data Packet Size	64 Bytes
Query/Control Message Size	36 Bytes
Transmission Range (d)	100 m
Number of Sensor nodes	200
Numbers of Sinks	4
Distribution Type of Sensor Nodes	Uniform

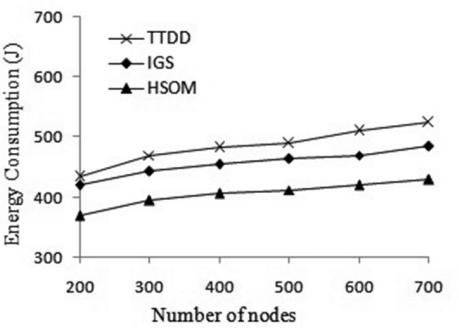


Figure 5. Energy Consumption vs. Node Density

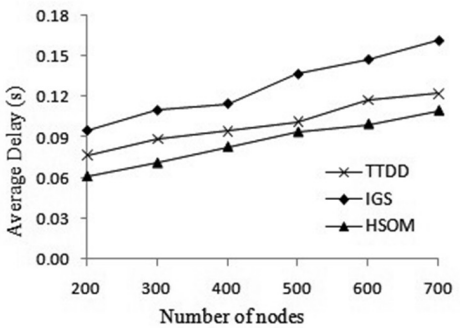


Figure 6. Average Delay vs. Node Density

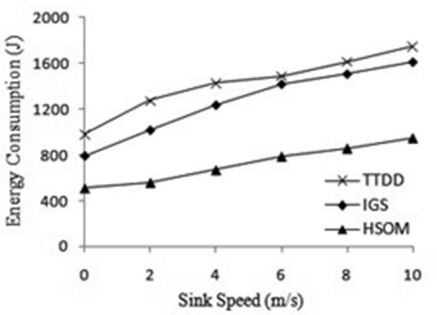


Figure 7. Energy Consumption vs. Sink Speed

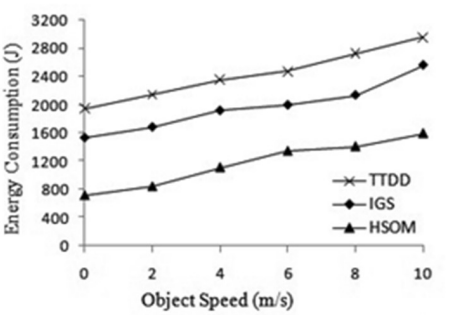


Figure 8. Energy Consumption vs. Object Speed

Conclusion

Proposed routing scheme for Handling Sink and Object Mobility (HSOM) in wireless sensor networks is suitable for scalability and mobility against sinks and events. The HSOM uses the virtual grid structure constructed by the sink appearing first in the sensor field or when there exists no valid grid. The grid nodes are used to disseminate data/query to/from sinks. To monitor the event mobility, Source Grid Node (SGN) sends the announcement message to its neighboring grid nodes (GNs). Each GN evaluates the energy efficient path well in advance. When mobile object enters in the periphery of a GN, it disseminates the data through the path it had evaluated in advance. To handle the sink's mobility, this scheme exploits the location of sink to setup up the shortest path between source and sink. Moreover, HSOM handles mobile sink very efficiently and maintains the path for continuous data delivery. It also construct/update a partial or new path between source and mobile sink if any detour problem occur, thus conserving the sensor nodes energy and increasing the network lifetime. Simulation results also indicate that HSOM consumes less energy and average delay as compared to TTDD and IGS for different numbers of sensor nodes, sinks/event mobility.

References

[1] Al-Karaki J. N. and Kamal A. E. "Routing techniques in wireless sensor networks: a survey", IEEE Wireless Communications, 11(6), 6–28, 2004.

[2] Albert Y. Z., "Handbook of Sensor Networks: Algorithms And Architectures", A John Wiley & Sons, Inc. publication, 2005.

[3] Ramya K., Praveen Kumar K. and Rao V. S., "A Survey on Target Tracking Techniques in Wireless Sensor Networks", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.4, pp 93-108, August 2012.

[4] Chakrabarti A., Sabharwal A. Aazhang B., "Using Predictable Observer Mobility for Power Efficient Design of Sensor Networks", In Proceedings of the Information Processing in Sensor Networks, (IPSN'03), pp. 129–145, CA, USA, 22–23 April 2003;

[5] Ekici E., Yaoyao G., and Bozdag D. , "Mobility-Based Communication in Wireless Sensor Networks", IEEE Communications Magazine, Vol. 44 issue 7, pp. 56 – 62, July 2006.

[6] Gandham S.R., Dawande M, Prakash R., Venkatesan S., "Energy Efficient Schemes for Wireless Sensor Networks with Multiple Mobile Base Stations", In Proceedings of the IEEE Global Telecommunications Conference 2003 (GLOBECOM'03), pp. 377–381, USA, December 2003.

[7] Hamida E. B. and Chelius G. , "Strategies for data dissemination to Mobile Sinks in Wireless Sensor Networks", IEEE Wireless Communication, Vol. 15 No. 6, pp. 31 – 37, December, 2008.

[8] Faheem Y., Boudjit S., Chen K., "Data Dissemination Strategies in Mobile Sink Wireless Sensor Networks". IEEE 2nd Wireless Days Conference, France, pp. 1 – 6, December 2009.

[9] Pantazis N. A., Nikolidakis S. A., and Vergados D. D.; "Energy-efficient routing protocols in wireless sensor networks: A survey"; Communications Surveys & Tutorials, IEEE, Vol.-15, Issue- 2, pp 551-591, May 2013.

[10] Intanagonwiwat C., Govindan R. and Estrin D., "Directed Diffusion: a Scalable and Robust Communication Paradigm for Sensor Networks", ACM International Conference on Mobile Computing & Networking (MobiCom'00), Boston, MA, pp. 56–67, 2000.

[11] Yu Y., Estrin D. and Govindan R., "Geographical and Energy-Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks", UCLA Computer Science Department Technical Report, UCLA-CSD TR-01-0023, May 2001.

[12] Schurgers C. and Srivastava M. B., "Energy Efficient Routing in Wireless Sensor Networks", MILCOM Proceedings Communications for Network-Centric Operations: Creating the Information Force, McLean, VA, 2001.

[13] Luo H., Ye F., Cheng J., Lu S. and Zhang L., "TTDD: Two-Tier Data Dissemination in Large-Scale Wireless Sensor Networks", Springer, Wireless Networks, Vol. 11, pp. 161–175, 2005.

[14] Zang R., Zhao H. and Labrador M., "The Anchor Location Service (ALS) Protocol for Large-scale Wireless Sensor Networks" ACM Proceedings of the First International Conference on Integrated Internet Ad hoc and Sensor Networks, article no. 18, France, May 2006.

[15] Brad Karp and Kung H. T., "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks", In Mobicom proceedings of the 6th annual international conference on Mobile computing and networking, NY, USA, pp. 243 – 254, August 2000.

[16] Lee E. et al. , "Data Dissemination Protocol based on Independent Grid Structure in Wireless Sensor Networks", 22 International symposium on Parallel, indoor and Mobile Radio Communication, Toronto, 11-14 Sept. 2011.

[17] Singh K., Sharma T.P. , "Sink Location-Oriented Data Dissemination using Grid in Wireless Sensor Networks," Elsevier Science and Technology, pp. 346-353, 2013.

[18] Han G., Xu H., Duong T. Q., Jiang J., Hara T., "Localization algorithms of Wireless Sensor Networks: a survey", Springer Telecommunication Systems, Volume 52, Issue 4, pp 2419-2436, April 2013.

[19] Bhardwaj M., Garnett T., Chandrakasan A. P., "Upper bonds on the Lifetime of Sensor Networks", IEEE Conference on Communications (ICC 2001), Finland, , Vol. 3, pp 785 – 890, June 2001.

[20] Tsai H. et al., "Mobile Object Tracking in Wireless Sensor Networks", Science Direct, Computer Communications, volume 30, Issue 8, pp. 1811 – 1825, 2007.

Instance based Multi Criteria Decision Model for Cloud Service Selection using TOPSIS and VIKOR

Deepti Rai * and V Pavan Kumar **
*Sahyadri College of Engineering
dee.rai@gmail.com
**Sahyadri College of Engineering
pavan.cs@sahyadri.edu.in

Abstract: Cloud computing has been incorporated in almost every fold of customer's world today. As more challenging becomes the world of a resource consumer, so does the plethora of numerous cloud vendors. Every vendor tries to sell off their services with different attractions. Some of them are free, some as pay-as-you-go services and some at rented structures. With so many choices available to a random user, making the right choice about the type of vendor is a crucial decision. We present in this paper novel methods for cloud service selection. The methods derives from the traditional selection methods with emphasis on user criteria weights. Inherent comparisons have been conducted amongst the various methods to help in analyzing and finalizing a broker architecture for selection of the best service provider among the contending cloud vendors.

Keywords: cloud computing, cloud service selection, cloud user weights, TOPSIS, VIKOR.

Introduction

Some of the best works done so far cover various Multi Criteria Decision Making methods like TOPSIS[1] , ELECTRE[14], VIKOR[3], PROMETHEE[5], AHP[6] etc. with new age additions to user preferences for fuzzy or trusted requirements[12]. Having come across various work done using these methods, it is highly imperative to understand that all these are time tested and efficient in one scenario or other. However most of these methods are applied to inherent ready set of data on the whole. The data set though complete with past values, seldom explores the individual performances of the cloud providers on an instance basis. In this paper we propose a method where we concentrate on the instance based outcomes of the service provider. To do this, we select the most accepted 2 methods – TOPSIS and VIKOR for our service selection of IaaS Clouds. We shall then categorize the inputs in two directions. First by altering the way the functional requirements are taken and secondly the way how the user weights are assigned. All these shall be realized using a simulation tool – Cloudsim[7] before concluding on the final result.

Existing System

Existing system is a system which has been time tested and proven efficient with any or other major changes. This system may be working or just in process of being setup. In either case, an existing system comes with its own set of shortcomings and problems. Hence there is always scope of improvement to bring about better systems. For the juncture of this project, the existing system is a system where we have basically three parties. As the entire system is based on cloud computing, needless to say we have two major sides - Cloud Providers and Cloud Users. In general terms, Cloud Provider is the party which provides cloud services like hosting, storage, application setup, service setup etc. On the other hand, cloud users is the party which avails these services by either paying or by virtue of free servicing. Now as a facilitator between these 2 parties, we have a broker[18] in the middle. The broker is essentially a third party, which deals with aggregation - bringing together all the required services of the user under one banner, de-duplication - by removal of redundant data available in the cloud data sets and security - by providing a trusted environment for users to upload their applications. The brokers themselves are subjected to trust evaluation as they are third party and have their own gain as primary objective. Some Service Level Agreement negotiations are also done by brokers by drawing up a contract with both the parties. The existing system that uses brokers for all such activities is also extended for the usual service selection decision problem. The cloud service selection problem is a multi criteria problem with many decisions[19]. Hence it is done in multiple layers. While the broker goes about accumulating the metrics from the cloud service providers, the data set becomes huge in nature. Thereafter the broker applies any of the multiple criteria decision modeling methods like TOPSIS, VIKOR, PROMETHEE, AHP etc. to through its software and filters out the best cloud services. This system takes into account data metrics collected over a period of time as whole. Thereafter the decision modeling is applied to the entire data set and result is generated. This system requires maintenance of huge database which holds the criteria metrics. The averaging of the weighted criteria method implemented is basically very general in nature. It has advantages and disadvantages to its credit.

Multi Criteria Decision Making Model

When it comes to the cloud service selection problem, we deal with multiple vendors with multiple services and varied user preferences[2]. Hence they automatically fall into the MCDM category. As with any MCDM problem[16], we can have numerous approaches like MAUT methods, AHP, French Outranking methods and Russian ordinal methods[8]. Every method has its strong and weak areas. While some methods work well with small data set, some are very effective with large data sets like that of a cloud service provider.[13] If we were to take a generic example of cloud services along with the user wanted criteria, then with the inclusion of user preference the problem of cloud service selection becomes very difficult[11]. Also another important factor is the consideration of these aspects both in real time and also past performance[4]. When a comparison is made between clouds, it is highly preferred to have a trust factor built into the decision. This factor can be obtained by observing the performance of the said criteria over a period of time and not in an instant.[10]. For doing this, we have to micro calculate the best cloud services in parts i.e one instance of a time. This method keeps the selection more accurate in comparison to the selection from the average data set values.

To summarize, despite many MCDM methods applied to cloud service selection, the ever changing nature of clouds and their quality of service criteria with time has not been incorporated effectively. Hence the existing approaches are not completely accurate in determining the best service provider. This paper considers the different aspects of time in past and present. While MCDM methods are the most suitable to sort out multi criteria problems[20], they themselves are inefficient in giving real time answers to user requirement. Experiments done to validate our method and outcome are an intelligent, resourceful and practical way of getting the cloud selection from user point of view.

Proposed Model

The proposed method of this selection model is simple in terms of the components involved. To start with we have the computing environment with different cloud Datacenters[9]. These can be considered as either cloud service providers or instances of CSPs. These are simulated as an IaaS with core computing qualities. They are responsible for publishing their services and respective paradigms. This service related information is stored in a database for further use. Next we have the user group who call in the cards by specifying which criteria of the service is of importance to them. This can be realized with either assigning weights with variance method or by fuzzy weights by asking the user his/her level of importance of each criterion in relation to other. In between these 2 categories we have the main decision maker – the broker who is responsible for employing the decision making algorithm to get us the best service as the result. Traditional methods involve simple concept of taking an absolute average of all criteria values and applying MCDM to it, however very accepted, this method is not free of flaws when dealing with a huge data set of criteria values over a considerably long period. Any Cloud provider can vary with its services over a period of time. We may have a trusted cloud server to be performing poorly in the recent past or vice versa. In order to get the most unbiased measure of all the criteria, it is imperative to give more weightage to the recent past. Hence the algorithm is applied on every instance(daily) of the values received and then the overall result is computed as depicted in Fig.2. We can dig further and calculate the best provider on quarter hourly or hourly basis, however with the huge amount of data produced, and citing no major upheavals like the stock market, daily basis is sufficient enough. Once the details of the CSPs are obtained, the users are asked to weigh in the criteria as per their need. Thereafter any of the MCDM algorithms mentioned below are applied to get the best service for daily set of values. For this experiment we have chosen 2 MCDM algorithms - TOPSIS and VIKOR. Both of them have proven their efficiency and worth in the area of multi criteria decision making.

Advantages of Proposed Model

- The disadvantage of normal weighted average method of being sensitive to extreme values, is overcome by calculating instance based averages rather than average of the whole system. Especially when leading to dynamics of cloud, we come across extreme values which when averaged on the whole can give an inconsistent value.
- Unlike normal average method this proposed method is best suited for time series type of data. The concept of time series data comes into picture where there is diversity and volatility is involved. This type of data which is inherent to the cloud platform, is best suited to be calculated instance based rather than as a bundle.
- The proposed method works even when all values are not equally important. Especially when it comes to clouds, we have metrics of criteria ranging over a wide area. Hence Not all are equally important. The criteria have their importance marked by the user given weight. Depending on user preference, the weights are assigned to the metrics. Hence the proposed method is most suited for unequal data sets.
- Proposed method works faster in comparison to the existing method as the average is calculated on a smaller sample of data and aggregated finally in one big step. For huge databases, the minor upheavals in time can result in major performance benefit.

Architecture

The architecture of this selection model is simple in terms of the components involved as in Fig 1. To start with we have the computing environment with different cloud Datacenters[9]. These can be considered as either cloud service providers or instances of CSPs. These are simulated as an IaaS with core computing qualities. They are responsible for publishing their services and respective paradigms. This service related information is stored in a database for further use. Next we have the user group who call in the cards by specifying which criteria of the service is of importance to them. This can be realized with either assigning weights with variance method or by fuzzy weights by asking the user his/her level of importance of each criterion in relation to other. In between these 2 categories we have the main decision maker – the broker who is responsible for employing the decision making algorithm to get us the best service as the result.

Traditional methods involve simple concept of taking an absolute average of all criteria values and applying MCDM to it, however very accepted, this method is not free of flaws when dealing with a huge dataset of criteria values over a considerably long period. Any Cloud provider can vary with its services over a period of time. We may have a trusted cloud server to be performing poorly in the recent past or vice versa. In order to get the most unbiased measure of all the criteria, it is imperative to give more weightage to the recent past. Hence the algorithm is applied on every instance(daily) of the values received and then the overall result is computed. We can dig further and calculate the best provider on quarter hourly or hourly bass, however with the huge amount of data produced, and citing no major upheavals like the stock market, daily basis is sufficient enough.

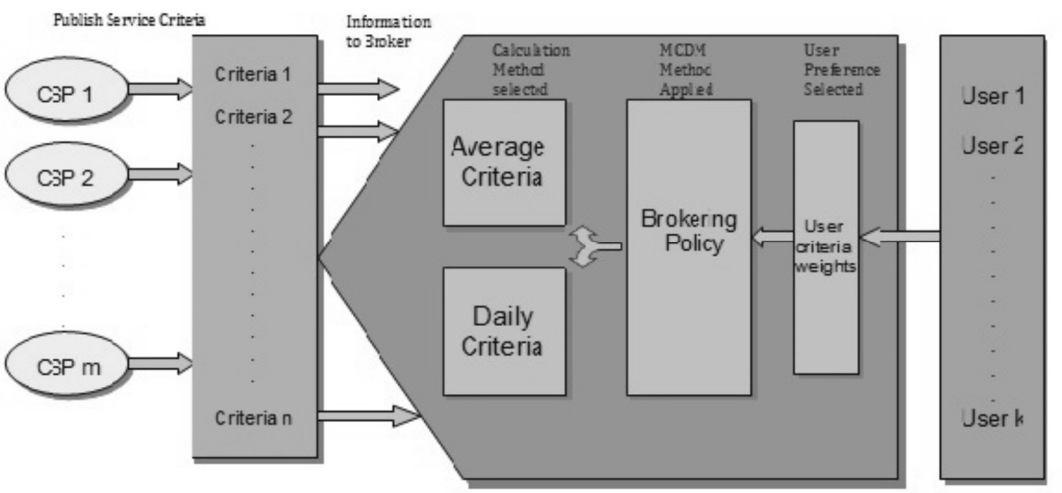


Figure 1 : Architecture of Proposed Model

Phases of Proposed System Setup

Phase 1- Cloud Setup Stage

This stage comprises of setting up an evaluation system which if not for real, atleast mimics the the original cloud environment. For the evaluation of the proposed system, a simulation environment has been devised. This has been realized using tool called Cloudsim[9]. This tool uses classic java based object structure for cloud setup. Major component of this tool is the availability of smaller classes which simulate datacentres,hosts,virtual machines etc. To start with there is a cloud information service. It is a registry which has the resources listed in cloud. After registry of all the distinct objects, individual datacenters are created. Datacenters are synonymous to clouds themselves. Hosts under each datacenter has some hosts, which has some hardware characteristics. Next there will be virtual machines which perform cloud related tasks. A broker is responsible for allocating tasks to datacenter. This is done using the brokering policy which follows any of the several algorithms available. The jobs themselves are called cloudlets, which are assigned to virtual machines/datacenter by broker.

Phase 2 - Average Model Selection

This stage consists of the selection of the main method of approach. As mentioned earlier, there already exists systems with average weighted based decisions. Keeping in mind their limitations, where the relevance of time is not taken into consideration, a new method has been devised. In this method, the time is of essence. The farther away we go into the past, the weightage factor reduces. This phenomenon of weight decay can be found in almost all systems which base their decisions on cumulative time based metrics. This decay can be calculated stochastically or logarithmically as per requirement. In the new method, the metrics are subjected to calculation based on the instance selected. Instance can be any

Flowchart for the Proposed System

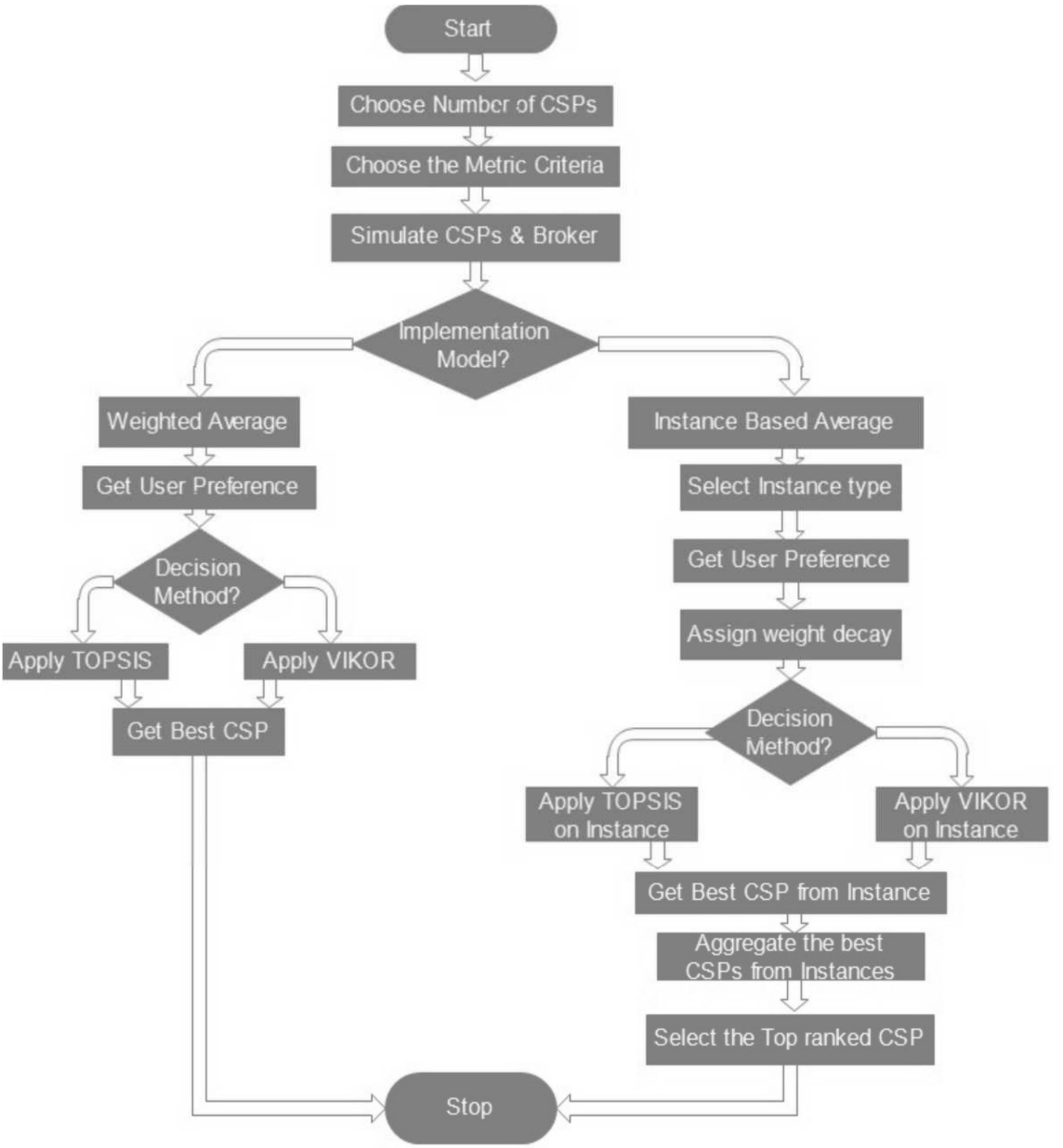


Figure 2: Flowchart for the Proposed System

single point of reference, weeks, days, hours, minutes etc. To depict the way the system works, a user is given access to the option to select any of the model required - either the average weighted or instance weighted.

Phase 3 - Decision Algorithm Selection Stage

After the model is selected, next step is to select the decision algorithm[17]. For the purpose of this evaluation - Topsis and Vikor have been selected. They have both proven their worth in getting the best decision out of the huge datasets of data. Two options have been given for comparison of results and data so that both the algorithms can be applied independently and evaluation be done fairly. Out of the 2 which needs to be applied is again sole decision of the user. To these algorithms we feed the cloud metric data with user preferences, weight decay etc. These perform normalization of the metrics and calculate the ranking of the cloud service providers as outcome.

Phase 4 - Cloud Ranking and Selection

The result derived from the previous step becomes the base for the user decision. The clouds are ranked in the descending manner and the one at the top becomes the best cloud. This is the final step of the entire evaluation process. The outcome may change depending on the model used or depending on the algorithm used. Also the user preferences change every now and then, resulting in variation of results.

Algorithms

TOPSIS - Technique for Order of Preference by Similarity to Ideal Solution

This method as depicted in algorithm Fig.3. works by calculating the geometric distance from the ideal solution. It takes into account the weight of each criteria, normalizing it and then determining the distance to ideal solution. Linear Normalization is performed to the metric data. The Technique for Order Preference by Similarity to Ideal Solution (Hwang and Yoon, 1981) identifies the best alternative among the one having the shortest distance from ideal solution. This method calculates the distance from the ideal alternative and the distance from negative option combining the worst performances of alternatives with respect to the single criterion. This method does not have many checks for acceptability as compared to other methods. The major advantage of this method is the admissibility of various types of criteria and various types of metric in varied ranges. TOPSIS outperforms other methods by being simple and straightforward. It uses euclidean distance method to calculate the distance, hence differentiating between individual and total satisfaction.

VIKOR - VIseKriterijumska Optimizacija I Kompromisno Resenje

The method Algorithm Fig.4. is that it selects the compromised solution among the other alternatives. Here the ranking is done using step by step checking where first best option is compared to second best option. It is also compared to the overall best alternative. Hence it rejects the solution if these checks are not passed. However it may give the solution of compromise if required. This method works excellently with conflicting criteria. Especially where criteria has been long drawn and selecting the best among so many different variable criteria makes this the most useful method. However since it does checking step by step, it becomes highly probable that the end result may be without solution. This method is sometimes extended to be used with other such methods like Analytic Hierarchy Process(AHP) and with fuzzy input values. The algorithm has proven efficient for large datasets where varied datatypes and criteria range are used.

Algorithm 1 TOPSIS algorithm

procedure TOPSIS
  Identify the alternatives $\epsilon \mathbb{N}$
  for $x_{ij} \epsilon D(m,n)$ of $m * n$ **do**,
    obtain x_{ij} where $i = 1, 2, ., m, j = 1, 2, .n$
  end for
  for $x_{ij} \epsilon D(m,n), i = 1$ to $i = m$ and $j = 1$ to $j = n$ **do**,
    Compute $N^*_{ij} = N_{ij} / \sqrt{\sum_{i=1}^{n} N_{ij}^2}$
  end for
  Calculate Variance of weights $V_j = (1/n) \sum_{i=1}^{n} (N^*_{ij} - (N^*_{ij})_{mean})^2$
  for $i = 1$ to $i = m$ **do**
    Obtain weights $W_j = V_j / \sum_{i=1}^{m} V_j$ and $\sum_{j=1}^{m} w_j = 1$
  end for
  for $i = 1$ to $i = m$ and $j = 1$ to $j = n$ **do**
    Compute Weighted normalized matrix $WV_{ij} == W_j * N_i$
  end for
  Determine best and worst ideal solutions as
    $A^+ = \{a_1^+, ., a_m^+\} i.e (maxWV_{ij}) | (minWV_{ij})$
    $A^- = \{a_1^-, ., a_m^-\} i.e (minWV_{ij}) | (maxWV_{ij})$
  Obtain separation of each alternative from A^+ and A^- as below
    $D_i^+ = \sqrt{\sum_{j=1}^{m} (a_{ij} - a_j^+)^2}$
    $D_i^- = \sqrt{\sum_{j=1}^{m} (a_{ij} - a_j^-)^2}$
  Obtain similarity Index $C_i^* = D_i^- / (D_i^+ + D_i^-)$
  Choose best alternatives in increasing order of C_i^*
end procedure

Figure 3: TOPSIS Algorithm

Algorithm 2 VIKOR algorithm

procedure VIKOR
  Identify the alternatives $\epsilon \mathbb{N}$
  for $x_{ij} \epsilon D(m,n)$ of $m * n$ **do**,
    obtain x_{ij} where $i = 1, 2, ., m, j = 1, 2, .n$
  end for
  for $x_{ij} \epsilon D(m,n), i = 1$ to $i = m$ and $j = 1$ to $j = n$ **do**,
    Compute $N^*_{ij} = N_{ij} / \sqrt{\sum_{i=1}^{n} N^2_{ij}}$
  end for
  Calculate Variance of weights $V_j = (1/n) \sum_{i=1}^{n} (N^*_{ij} - (N^*_{ij})_{mean})^2$
  for $i = 1$ to $i = m$ **do**
    Obtain weights $W_j = V_j / \sum_{i=1}^{m} V_j$ and $\sum_{j=1}^{m} w_j = 1$
  end for
  for $i = 1$ to $i = m$ and $j = 1$ to $j = n$ **do**
    Compute Weighted normalized matrix $WV_{ij} == W_j * N_i$
  end for
  Obtain Maximum Criterion Weight and Minimum Criteria Weight as below:
    $F^+_{ij} = \max(N_{ij})$
    $F^-_{ij} = \min(N_{ij})$
  Compute Utility Measure
    $U_i = \sum_{i=1}^{m} W_j (F^+_j - F_{ij}) / (F^+_j - F^-_j)$
  Compute Regret Measure
    $R_i = max[W_j (F^+_j - F_{ij}) / (F^+_j - F^-_j)]$
  Calculate the Vikor Index
    $VI_i = v((U_i - U^-)/(U^+ - U^-)) + (1 - v)((R_i - R^-)/(R^+ - R^-))$
  Choose best ranking alternative in increasing order of VI^*_i
end procedure

Figure 4: VIKOR Algorithm

Results

Discussion

The idea basically deals with 2 different models - one being with the average method and other being Instance Method. The implementation and comparison of the first method has been completed at this juncture. The cloud creation has been completed using cloudsim toolkit. The system developed can create cloud structures with 3 criteria mark values for RAM, Bandwidth and Storage. Number of clouds to be created is kept with the admin of the system. Below referenced Table 1, Table 2, Table 3, Table 4, Table 5, Table 6 show the results of the experiment.[15]. Along with this a broker is also created. Once the clouds are created with the metrics, they are then stored in the sql database for future use. Immediately following this, the traditional method for cloud selection has been established which compares TOPSIS and VIKOR in terms of evaluation time for arriving at the best cloud. \par While evaluating both Topsis and Vikor, it was found that evaluation with a collection of 10 clouds randomly created by the cloudsim, vikor performed better than topsis in terms of evaluation time. Also when it came to memory vikor was better than topsis. As the data is collected over a period of 30 days, the dataset is sufficient to establish the fact that though both the algorithms are excellent choices for cloud service selection, vikor outperforms topsis. Next would be development of the Instance based model using real cloud data value sets and instances of days. This would be to apply averages on individual days rather than entire set, giving a microscopic and accurate result.

Table 1: Average Criteria Values

Cloud	RAM	Bandwidth	Storage
Cloud1	1433	18000	360000
Cloud2	1638	22000	160000
Cloud3	1843	22000	240000
Cloud4	1433	18000	360000
Cloud5	1843	21000	480000
Cloud6	1433	20000	320000
Cloud7	1024	12000	300000
Cloud8	1433	23000	360000
Cloud9	1740	18000	340000
Cloud10	1331	28000	360000

Table 2: Evaluation Matrix

Cloud	RAM	Bandwidth	Storage
Cloud1	0.059	0.083	0.168
Cloud2	0.068	0.101	0.075
Cloud3	0.076	0.101	0.112
Cloud4	0.059	0.083	0.168
Cloud5	0.076	0.097	0.224
Cloud6	0.059	0.092	0.15
Cloud7	0.042	0.055	0.14
Cloud8	0.059	0.106	0.168
Cloud9	0.072	0.083	0.159
Cloud10	0.055	0.129	0.168

Table 3: Similarity Index using TOPSIS

Cloud	Similarity Index
Cloud1	0.446
Cloud2	0.449
Cloud3	0.530
Cloud4	0.446
Cloud5	0.433
Cloud6	0.473
Cloud7	0.432
Cloud8	0.485
Cloud9	0.431
Cloud10	0.535

Table 4: Vikor Index using VIKOR

Cloud	Vikor Index
Cloud1	0.341
Cloud2	0.909
Cloud3	0.605
Cloud4	0.341
Cloud5	0.0
Cloud6	0.445
Cloud7	0.729
Cloud8	0.269
Cloud9	0.35
Cloud10	0.216

Table 5: Proposed Daily Method

Days	TOPSIS	VIKOR
1	cloud6	cloud9
2	cloud5	cloud6
3	cloud2	cloud2
4	cloud9	cloud6
5	cloud8	cloud2
6	cloud2	cloud2
7	cloud1	cloud7
8	cloud2	cloud8
9	cloud6	cloud2
10	cloud2	cloud6
11	cloud1	cloud8
12	cloud9	cloud2
13	cloud2	cloud8
14	cloud5	cloud6
15	cloud2	cloud2
16	cloud5	cloud10
17	cloud2	cloud2
18	cloud9	cloud2
19	cloud7	cloud5
20	cloud3	cloud6
21	cloud2	cloud2
22	cloud8	cloud2
23	cloud2	cloud1
24	cloud7	cloud8
25	cloud1	cloud2
26	cloud10	cloud7
27	cloud2	cloud2
28	cloud9	cloud9
29	cloud2	cloud2
30	cloud2	cloud2

Table 6: Result Comparison

Average Method		Daily Method	
TOSIS	VIKOR	TOPSIS	VIKOR
cloud10	cloud2	cloud2	cloud2

Conclusion

In conclusion , we assert that the method proposed by us is efficient and practical in terms of real time and changing cloud environments. This method keeps into account the instance based calculation of metrics and hence gives a more focused result than the assumptions based on entire average of metrics in whole. As we have shown the results for a small group of data, this method can be tested against a fairly large dataset without issues. Enhancement of this method could be to include fuzzy weights and future predictions of criteria change. Also some important parameters like the cost, vendor lock in, data disruption etc. have not been considered, which can be included in this model.

References

[1] Serafim Opricovic, Gwo-Hshiung Tzeng, “Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS”, European Journal of Operational Research, Elsevier, Volume 156, Issue 2, July 2004, Pages 445–455.

[2] R.Buyya, C.S.Yeo, S. Venugopal,J. Broberg,I. Brandic, “Cloud Computing and Emerging IT Platforms: Vision,Hype,and Reality for Delivering Computing as the 5th Utility” , Future General Computer System, IEEE, Vol. 25, June 2009, pp. 599–616.

[3] Lisa Y. Chen, Tien-Chin Wang “Optimizing partners’ choice in IS/IT outsourcing projects: The strategic decision of fuzzy VIKOR”,International Journal of Production Economics,Volume 120, Issue 1, July 2009, Pages 233–242.

[4] Yager, R.R., “On ordered weighted averaging aggregation operators in multicriteria decision making”, IEEE Transactions on Systems, Man and Cybernetics, Volume:18 ,Issue: 1, 1998, pp .183 - 190.

[5] Toinard, C., Ravier, T., Cerin, C., Ngoko, Y., "The Promethee Method for Cloud Brokering with Trust and Assurance Criteria", IEEE, Parallel and Distributed Processing Symposium Workshop (IPDPSW), 2015, pp 1109 - 1118.
[6] Rajanpreet Kaur Chahal , Sarbjeet Singh, "AHP-Based Ranking of Cloud-Service Providers" , Springer, Advances in Intelligent Systems and Computing, Vol 443, 2016, pp 491-499.
[7] Buyya, R., Ranjan, R., Calheiros, R.N., "Modeling and simulation of scalable Cloud computing environments and the CloudSim toolkit: Challenges and opportunities", IEEE, Conference proceedings on High Performance Computing and Simulation, 2009, pp 1-11.
[8] S. Sindhu, Saswati Mukherjee, "Efficient Task Scheduling Algorithms for Cloud Computing Environment", Springer, High Performance Architecture and Grid Computing, Vol 169, 2011, pp 79-83.
[9] Wang Long, Lan Yuqing , Xia Qingxin, "Using CloudSim to Model and Simulate Cloud Computing Environment", IEEE, Conference proceedings of Computational Intelligence and Security (CIS), 2013, pp 323 - 328.
[10] Treadgold, N.K., Gedeon, T.D. "Simulated annealing and weight decay in adaptive learning: the SARPROP algorithm", IEEE Transactions on Neural Networks, Vol:9, n.4, 1998, pp 662 – 668.
[11] Johan Tordssona, Rubén S. Monterob, Rafael Moreno-Vozmedianob, Ignacio M. Llorenteb, "Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers", Elsevier, Future Generation Computer Systems, Vol 28, n.2, February 2012 pp 358–367.
[12] Wang, T.-C., Lee, H.-D., Chang, M.-S. "A fuzzy TOPSIS approach with entropy measure for decision-making problem", International Conference on Industrial Engineering and Engineering Management, IEEE, 2007, pp. 124-128.
[13] Shangguang Wang , Zhipiao Liu , Qibo Sun , Hua Zou , Fangchun Yang, "Towards an accurate evaluation of quality of cloud service in service-oriented cloud computing", Journal of Intelligent Manufacturing, Vol.25 n.2, April 2014, pp.283-291.
[14] Luis Del Vasto-Terrientesa, Aida Vallsa, Roman Slowinskib c, Piotr Zielniewiczb, "ELECTRE-III-H: An outranking-based decision aiding method for hierarchically structured criteria", Elsevier, Expert Systems with Applications, Vol 42, n. 11, 2015, Pages 4910–4926.
[15] Deepti Rai, Pavan Kumar V, "Instance based multi criteria decision-model for cloud service selection using TOPSIS and VIKOR", IAEME, International Journal of Computer Engineering and Technology , Vol 7 , n.2, 2016, pp 78-87.
[16] Gang Kou , Yong Shi , Shouyang Wang, "Multiple criteria decision making and decision support systems", Decision Support Systems, v.51 n.2, May 2011, p.247-249.
[17] Tao Yu, Kwei-Jay Lin, "Service selection algorithms for Web services with end-to-end QoS constraints", Springer, Information Systems and e-Business Management, Vol 3, n.2, July 2005, pp 103-126.
[18] S Sundareswaran., A.quicciarini, D.Lin, "A Brokerage-Based Approach for Cloud Service Selection",IEEE,Conference proceedings from Cloud Computing , June 2012, pp 558-565.
[19] Saurabh Kumar Garga, Steve Versteegb, Rajkumar Buyyaa, "A framework for ranking of cloud computing services", Future Generation Computer Systems, Vol 29, Issue 4, June 2013, Pages 1012–1023.
[20] Md Whaiduzzaman, Abdullah Gani, Nor Badrul Anuar, Muhammad Shiraz, Mohammad Nazmul Haque, Israat Tanzeena Haque, "Cloud Service Selection Using Multicriteria Decision Analysis", Hindawi, The Scientific World Journal, 2014.

Part of Speech TAGGER for MARATHI Language

Shubhangi Rathod*, Sharvari Govilkar** and Sagar Kulkarni***
*Department of Computer Engg. PIIT, New Panvel, India
rathod.shubhangi30@gmail.com
**Department of Computer Engg. PIIT, New Panvel, India
sgovilkar@mes.ac.in
***Department of Computer Engg. PIIT, New Panvel, India
skulkarni@mes.ac.in

Abstract: A part of speech (POS) tagging is one of the most well studied problems in the field of Natural Language Processing (NLP). Part-of-Speech Tagging (POS) tagging means assigning grammatical classes i.e. appropriate parts of speech tags like noun, adjective, verb, adverb etc to each word in a natural language sentence/word. The main challenge in POS tagging is to resolving the ambiguity in possible POS tags for a word so disambiguation rules and Tagset is vital parts of POS tagger. POS tagging is difficult for Marathi language due to unavailability of corpus for computational processing. In this paper, a POS Tagger for Marathi language using Rule based technique is presented. Our proposed system find root word using morphological analyzer and compare the root word with corpus to assign appropriate tag. If word has assigned more than one tags then by using grammar rules ambiguity is removed. Meaningful rules are provided to improve the performance of the system.

Keywords: Part of Speech (POS), Tagset, Tokenizer, Stemmer, Morphological analyzer, Disambiguation.

Introduction

The work on Part-of-Speech (POS) tagging has begun in the early 1960s [2]. Part-of-Speech Tagging (POS) tagging means assigning grammatical classes i.e. appropriate parts of speech tags to each word in a natural language sentence/word. Assigning a POS tag to each word of an un-annotated text by hand is very time consuming, which results in the existence of various approaches to automate the job [3].The significance of these is the large amount of information they give about a word and its neighbours.

POS tagger is a necessary pre-processing module and extremely powerful as well as accurate tool [1] used in any application that deals with natural language processing. The tagging performance totally depends on tag dictionary. The large numbers of POS tagger available for English language which has got satisfactory performance but cannot be applied to Marathi language. Part-of-speech tagging in Marathi language is a very complex task as Marathi is highly inflectional in nature & morphologically rich language. The main challenge in POS tagging is to resolving the ambiguity in possible POS tags for a word [3].

Taggers can be classified as supervised or unsupervised: Supervised taggers are based on pre-tagged corpora, whereas unsupervised taggers automatically assign tags to words [6]. Furthermore, taggers divide into three types: (i) Rule Base Taggers: The rule based POS tagging approach that uses a set of hand constructed rules. (ii) Stochastic Taggers: A stochastic approach assigns a tag to word using frequency, probability or statistics [6]. It required vast stored contextual information because many high frequency words of POS are ambiguous. (iii) Hybrid Taggers: The hybrid approach, assign tag to the word using statistical approach after that, if wrong tag is found then by applying some rules tagger tries to change it [7].

Part-of-speech tagging is harder than just having a list of words and their parts of speech, because some words can represent more than one part of speech at different times, and because some parts of speech are complex or unspoken. This is not rare in natural languages such as Marathi language that a large percentage of word-forms are ambiguous. For example: In the sentence, **"पुजा देवं पुजा कर."** it is clear that the word **"पुजा"** is occurred two times in a sentence but the meaning of the word is different at both the places. The sentence given here contains ambiguity in the word which must be resolved before assigning tags to it. Designed system recognizes that the word **"पुजा"** has two different tags because of disambiguity rule for noun and verb. Thus the resultant tags assigned to the words in the sentence are

"पुजा::NNP देवं::NN पुजा::VM कर::VM .::RD_PUNC".

In these tags, the word **"पुजा"** has been assigned with two different tags, one act as 'Proper Noun' and other can be 'Verb Main'.

The paper presents, part of speech tagger for Marathi language. In section 2, related work is discussed in detail. Working of system is mentioned in detail in section 3. Section 4 explores accuracy obtained by POS tagger. Finally, paper is concluded in section 5.

Related Work

In this section we cite the relevant past literature that use the various pos tagging techniques. In the last few years the several approaches have been developed for English and other foreign languages. Most of the researchers concentrate on rule base rather than statistical approach for POS tagging. The small set of the meaningful rules of this tagger provides the better improvements over statistical tagger.

Jyoti Singh, et.al. [1] Proposed a Development of Marathi Part of Speech Tagger Using Statistical Approach. They used statistical tagger using Unigram, Bigram, Trigram and HMM Methods. To achieve higher accuracy they use set of Hand coded rules, it include frequency and probability. They use most frequently used tag for a specific word from the annotated training data and use this information to tag that word in the annotated text. They train and test their model by calculating frequency and probability of words of given corpus.

H.B. Patil, et.al. [2] Proposed a Part-of-Speech Tagger for Marathi Language using Limited Training Corpora. It is also a rule based technique. Here sentence taken as an input generated tokens. Once token generated apply the stemming process to remove all possible affix and reduce the word to stem. SRR used to convert stem word to root word. The root-words that are identified are then given to morphological analyzer. The morphological analysis is carried out by dictionary lookup and morpheme analysis rules.

Pallavi Bagul, et.al. [3] Proposed a Rule Based POS Tagger for Marathi Text. Which will assign part of speech to the words in a sentence given as an input and used a corpus which is based on tourism domain. The ambiguous words are those words which can act as a noun and adjective in certain context, or act as an adjective and adverb in certain context. The ambiguity is resolved using Marathi grammar rules.

Jyoti Singh, et.al. [4] Proposed a Part of speech tagging of Marathi text using Trigram method. The main concept of Trigram is to explore the most likely POS for a token based on given information of previous two tags by calculating the transition probabilities between the tags and helps to capture the context of the sentence. The probability of a sequence is just the product of conditional probabilities of its trigrams. Each tag transition probability is computed by calculating the frequency count of two tags which come together in the corpus divided by the frequency count of the previous two tags coming in the corpus.

Nidhi Mishra, et.al. [5] Proposed Part of Speech Tagging for Hindi Corpus. The system scans the Hindi (Unicode) corpus and then extracts the Sentences and words from the given Hindi corpus. Finally Display the tag of each Hindi word like noun tag, adjective tag, number tag, verb tag etc. and search tag pattern from database.

Namrata Tapaswi, Suresh Jain [6] proposed a Treebank Based Deep Grammar Acquisition and Part-Of-Speech Tagging for Sanskrit Sentences. In the Sanskrit morphology meaning of the word is remain same. When affixes are added to the stem, words are differentiated at database level directly. The input is one sentence per line, split the sentence into words called lexeme .read each word to find longest suffix, and eliminated the suffix until the word length is 2. Apply the lexical rules and assign the tag. Remove the disambiguity using context sensitive rules.

Javed Ahmed MAHAR, Ghulam Qadir MEMON [7], proposed a system for “Rule Based Part of Speech Tagging of Sindhi Language”. Take input text, and generate token. Once token generated search and compare selected word from lexicon (SWL) .If word is found one or more times, then store associated tag and if not found add that word into lexicon by generating linguistic rule for new word.

Proposed System

We have designed a rule based part of speech tagger that assigns parts of speech to each word, such as noun, verb, adjective, adverb etc in a sentence. Rule-based part-of-speech tagging is the most powerful approach that uses manually written rules for tagging. Rule based tagger depends on dictionary or lexicon to get possible tags for each word to be tagged. Hand-written rules are used to identify the correct tag when a word has more than one possible tag. The proposed approach consists of following phases:

1. Pre-processing
2. Stemmer
3. Morphological analyzer.
4. Tag Generator
5. Disambiguation.

Preprocessing

Validation of Input document
The input document may contain some words or sentences in other script or language. So, validation of Input document is very important stage because the resultant information is totally depends on the language and nature of query supplied to the system. Here we are analyzing whether the input document is valid in Devanagari script or not. The words which are not valid to Devanagari script are simply removed from further processing. To perform this operation we have used Unicode values called UTF-8 for Devanagari script document. The aim of this phase is to maintain pure Devanagari script document as an input to Morphological Analyzer.

Tokenization
This Tokenization is the process of separating word/tokens from input text. The division of input text into tokens is important for POS tagging. This tokenization task is possible by searching spaces between the words. The words separated from sentence and treat as single token so, we can deal with each word separately.

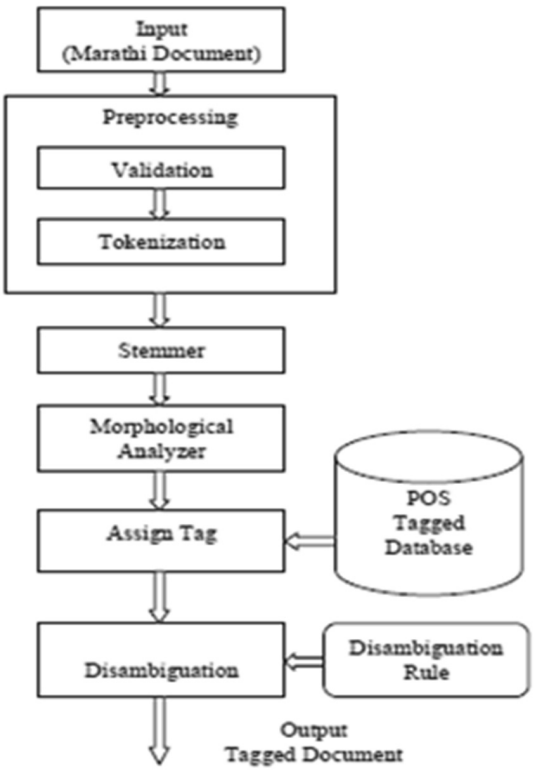


Fig 1. Proposed System

Stemmer
Stemming is important in the system, which uses a suffix list to remove suffixes from words and thus reduces the word to its stem. To remove suffixes from input document the Corpus is used consist of 1059 suffixes which frequently occur in Marathi language. The result of stemming is stem of word that can be given as input to Morphological Analyzer for further processing. The stem word contains inflections. The inflections in the stem word cannot be removed using simple stemming operation.

Morphological analyzer
The aim of morphological analysis is to recognize the inner structure of the word. The words after stemming are analyzed to check whether they are inflected or not. If stem word is inflected then the root word is formed by addition of replacement characters with stem word. A morphological analyzer is expected to produce Root words for a given input document. There is need to design some standard rules called inflection rule which will enable the system to process the stem of words and find the actual Root word.

Tag Generator

Corpus linguistics is the study of language as expressed in samples (corpora) of "real world" text. Corpus is a large collection of texts. It is a body of written or spoken material upon which a linguistic analysis is based [9]. This phase assigns corresponding part of speech tags to the words and we have used tagset developed by IIIT Hyderabad [9] [10]. A well-chosen tagset is important to represents parts of speech. The language tagset represents parts of speech and consist on syntactic classes [8].

Algorithm for POS tagging System:

1) Take input text and generate a token.
2) Use tokens to generate stem of word.
3) Use rule to generate root word using morphological analyzer and stored them.
4) Select each word one by one and compare with corpus.
5) If word is found one or more times, then store associated tag or tags of word and else display "the word is not found" add this new word into corpus.
6) If one tag is stored, then display word with associated tag as an output.
7) Else apply rule to select most appropriate tag for word

Table 1. POS Tag list

No.	Name	Tag	Description	Example
1	NOUN	NN	Common Nouns	मुलगा, साखर, मंडळी, चांगुलपणा
		NNP	Proper Nouns (name of person)	मोहन, राम, सुरेश
		ABN	Abstract noun	गर्व,कौशल्य, क्रोध,चपळाई
2	PRONOUN	PPN	Personal pronoun	मी,आम्ही,तुम्ही
		PPS	Possessive pronoun	माझा,माझी, तुझा,तुझी,त्याचा
		PDM	Demonstrative pronoun	तो, ती, ते, हा, ही
		PRF	Reflexive pronoun	आपण,आम्ही, तुम्ही,तुम्हाला
		PRC	Reciprocal pronoun	एकमेकांचा, एकमेकाला
3	ADJECTIVE	JJ	Modifier of Noun	उत्साही, श्रेष्ठ,बळवान
4	VERB	VM	Verb Main (Finite or infinite)	बसणे, दिसणे, लिहिणे,पडला
		VAUX	Verb Auxiliary	नाही, नको, करणे,हवे, नये
5	ADVERB	RB	(Modifier of Verb)	आता, काल, कधी, नेहमी, लवकर
6	CONJUNCTION	CC	Coordinating and Subordinating	आणि,पण, जर, तर
7	POSTPOSITION	PSP	Postposition	आणि, वर, कडे,जवळ
8	INTERJECTION	INJ	Interjection	आहा, छान, अगो, हाय

9	NUMERAL(NUM)	NUM	Number	१,२,३,४
		NUMCD	Cardinal Numeral	एक, दोन, तीन
		NUMO	Ordinal Numeral	पहिला,दुसरा, तिसरा
10	RESIDUAL	RDS	Symbol residual	$, &, *, (,)
		RD_PUNC	Punctuation	? , ; : !
11	REDUPLICATION	RDP	Reduplications	जवळजवळ-
12	NEGATIVE	NEG	Negative	नाही,नको
13	DETERMINER	QF	Quantifiers	किती,पुष्कळ,खूप,भरपूर, बरेच
14	QUESTION WORDS	WQ	Question Words	काय, कधी, कु ठे
15	INTENSIFIER	INTF	Intensifier	खूप,फार,बराच,अतिशय
16	PARTICLES	RP	Particles	तर,ओहो
17	PHRASE	PHR	Phrase	नमस्कार, अभिनंदन,खेद आहे
18	ECHO	ECH	Echo Word	जेवणबिवण, डोकेबिके
19	QUATATIVE	UT	Quatative word	म्हणजे

Disambiguation

The Natural language has the ambiguity issues as the single word has different tags. To overcome the ambiguity issues and assigning a "correct" tag in particular contexts, disambiguation rules are required. Word Sense Disambiguation (WSD) is the process of identifying the sense of a polysemic word. In modern WSD systems, the senses of a word are typically taken from some specified dictionary. Disambiguation is based on contextual information or word/tag sequences. The ambiguity which is identified in the tagging module is resolved using the Marathi grammar rules.

Following example demonstrates processing of our system:

Input Query: मराठी भाषा हे महाराष्ट्राचे वैभव आहे. It is given in history of Marathi Language

Validation: मराठी भाषा हे महाराष्ट्राचे वैभव आहे.

Tokenization:

मराठी
भाषा
हे
महाराष्ट्राचे
वैभव
आहे
.

Stemmer: मराठी भाषा हे महाराष्ट्रा वैभव आहे.

Morphological Analyzer: मराठी भाषा हे महाराष्ट्र वैभव आहे.

POS Tagging Output: मराठी\\NNP भाषा\\NN हे\\PDM महाराष्ट्र\\NNP वैभव\\JJ आहे\\VAUX . \\RD_PUNC

Performance of System

We have developed our own corpus consisting of 17197 unique words, tagset consist 29 tags and we have developed 141 rules for disambiguation for Marathi languages. The performance of the system is measured for multiple documents as shown in Table. We have used randomly selected Marathi document as input to NLTK and our designed Tagger. While recording correctness of both the system we focused on the strength of both taggers to handle WSD of the words in the sentences. Many times we found that our designed Tagger performs well for both Tagging and handling WSD as compared with NLTK tagger. The designed tagger system is compared with other existing systems such as NLTK and Shallow Parser. The Table 5.1 shows the details of the testing results for ten Marathi language documents. It is found that the efficiency of designed POS tagger to assign correct tags to words in the document is better than that of NLTK and Shallow Parser.

Table 2. Table Performance of the System based on Document analysis

Doc. No.	Document Name	No. of words	Words correctly tagged by NLTK	Performance of NLTK (%)	Words correctly tagged by Shallow Parser	Performance of Shallow Parser (%)	Words correctly tagged by Designed Tagger	Performance of Designed Tagger (%)
1	Marathi Bhasha	106	62	58.49	73	68.87	102	96.23
2	Disambiguity Text	61	18	29.51	36	59.02	59	96.72
3	Agriculture text	130	64	49.23	99	76.15	127	97.69
4	Ramayana	44	20	45.45	34	77.27	42	95.45
5	Shivaji Maharaj	123	74	60.16	94	76.42	114	92.68
6	Panvel info	87	38	43.68	64	73.56	85	97.70
7	Aai text	58	33	56.90	46	79.31	53	91.38
8	Marathi Grammar text	101	38	37.62	70	69.31	95	94.06
9	Modi	101	58	57.43	70	69.31	96	95.05
10	Mumbai info	77	47	61.04	53	68.83	72	93.51

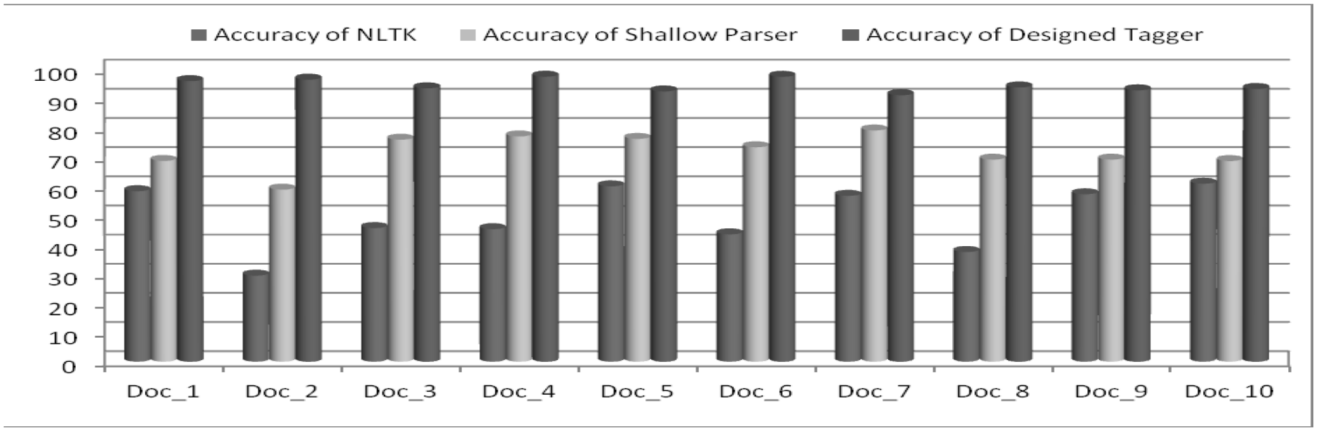


Figure 2. Graphical representation of performance analysis of Designed Tagger

The Figure shows the graphical representation of performance of the designed Tagger, NLTK and Shallow Parser. It is clearly observed that the designed tagger system gives higher performance result to tag the words correctly in the document than NLTK and Shallow Parser.

Table 3. System Testing based on Collection of Documents testing based on Collection of Documents

Overall Analysis on	Total No of Words	NLTK		SHALLOW PARSER		DESIGNED TAGGER SYSTM	
		Correctly tagged words	Accuracy (%)	Correctly tagged words	Accuracy (%)	Correctly tagged words	Accuracy (%)
Collection of 10 Documents	811	347	49.95	516	71.81	677	95.05

We have taken ten randomly selected Marathi documents to analyze the result of our designed tagger and it is compared with the existing systems such as NLTK and Shallow Parser. As shown in Table it can be easily observed that out of total number words (811) the designed POS tagger system gives efficiency of 95.05% i.e. it tags 677 words correctly, whereas NLTK and Shallow Parser gives the efficiency up to 49.95% and 71.81% respectively.

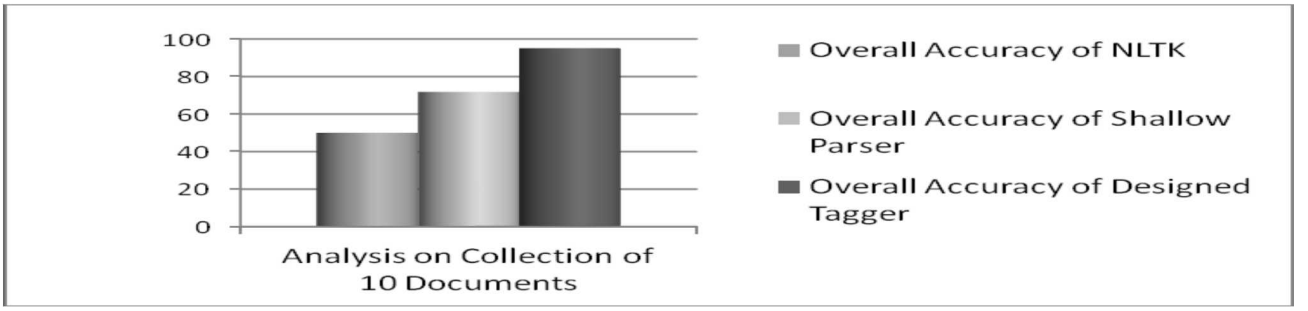


Figure 3. Graphical representation of testing based on collection of documents

The Figure shows that the efficiency of the designed POS tagger is higher than that of NLTK and Shallow Parser. The overall accuracy of the system is 95.05%.

Conclusion

The task of POS tagging is quite complex for Marathi language as the language is morphologically rich in script. There are some issues still present in tagging the words effectively because if stemming and morphology is not performed well then the root form is not generated correctly and thus the tag assigned to such incorrect words are not always correct. The POS tagger we designed for Marathi language uses Rule-based tagging approach which assigns all possible tags to word and WSD uses context rules to disambiguate the tags so that the accuracy is enhanced. It has been proved that the designed POS tagger for Marathi language gives relevant and acceptable performance up to **95.05%**.There is large scope to enhance the rule set of POS tagging as well as WSD context rules to improve the accuracy of the designed system up to maximum extent.

Acknowledgment

I am using this opportunity to express my gratitude to thank all the people who contributed in some way to the work described in this paper. My deepest thanks to my project guide for giving timely inputs and giving me intellectual freedom of work. I express my thanks head of computer department and to the principal of Pillai Institute of Information Technology, New Panvel for extending his support.

References

[1] Jyoti Singh, Nisheeth Joshi, Iti Mathur, “Development of Marathi Part of Speech Tagger Using Statistical Approach”, international conference on Advance in Computing, Communications and Informatics (ICACCI), IEEE DOI 10.1109/ICACCI.2013.6637411,2013,1554-1559.

[2] H.B. Patil, A.S. Patil, B.V. Pawar “Part-of-Speech Tagger for Marathi Language using Limited Training Corpora”, International Journal of Computer Applications (0975 –8887) Recent Advances in Information Technology,2014.

[3] Pallavi Bagul, Archana Mishra, Prachi Mahajan, Medinee Kulkarni, Gauri Dhopavkar, "Rule Based POS Tagger for Marathi Text" 2014 in proceeding of: International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1322-1326.
[4] Jyoti Singh, Nisheeth Joshi, Iti Mathur "PART OF SPEECH TAGGING OF MARATHI TEXT USING TRIGRAM METHOD", 2013 International Journal of Advanced Information Technology (IJAIT) Vol. 3, No.2, DOI: 10.5121/ijait2013.3203.
[5] Nidhi Mishra, Amit Mishra,"Part of Speech Tagging for Hindi Corpus" 2011 in proceeding of : International Conference on Communication Systems and Network Technologies, 978-0-7695-44373/11, 2011 IEEE DOI 10.1109/CSNT.2011.118.
[6] Namrata Tapaswi, Suresh Jain, "Treebank Based Deep Grammar Acquisition and Part- Of-Speech Tagging for Sanskrit Sentences", Software Engineering (CONSEG), 2012 CSI Sixth International Conference on, 971-1-4673-2174-7 IEEE DOI 10.1109/CONSEG.2012.6349476.
[7] Javed Ahmed MAHAR, Ghulam Qadir MEMON, "Rule Based Part of Speech Tagging of Sindhi Language " 2010 proceeding of International Conference on Signal Acquisition and Processing 978-0-7695-3960-7/10,2010 IEEE DOI 10.1109/ICSAP.2010.27.
[8] Sankaran Baskaran , Kalika Bali1, Tanmoy Bhattacharya, Pushpak Bhattacharyya, Monojit Choudhury, Girish Nath Jha, Rajendran S.5, Saravanan K.1, Sobha L.6, and KVS Subbarao, "A Common Parts-of-Speech Tagset Framework for Indian Languages".
[9] Kh Raju Singha Bipul Syam Purkayastha Kh Dhiren Singha "Part of Speech Tagging in Manipuri: A Rule-based Approach "International Journal of Computer Applications, (0975 – 8887) Volume 51– No.14, August 2012
[10] Bharati, A., Sharma, D.M., Bai, L., Sangal, R., "AnnCorra: Annotating Corpora Guidelines for POS and Chunk Annotation for Indian Languages",2006.

Aadhaar based Secure E-Voting System using Cortex-A15 Processor

Prathiba Jonnala*, Joshua Reginald Pullagura** and K. Ashok Kumar Reddy***
*Department of Electronics and communication Engineering, VFSTR University, Vadlamudi, Andhra Pradesh, India
jpvuece@gmail.com
**Department of Electronics and communication Engineering, VFSTR University, Vadlamudi, Andhra Pradesh, India
pjreginald@gmail.com
***Department of Electronics and communication Engineering, VFSTR University, Vadlamudi, Andhra Pradesh, India
ashokreddy41@gmail.com

Abstract: This paper introduces Aadhaar based Secured E-Voting system which will be very much suitable for the most sensitive areas in India where there is a large scope for rigging and other malfunctions. In this study, the prototype is built based on secured and trusted framework for Aadhaar based E- voting. The System allows the voters to participate by using username and password. In the proposed system Cortex-A15 Processor is used for authentication. The Aadhaar number issued by Election Commission of India will be the user name and biometric thumb impression taken at the time of Aadhaar enrolments will be the password. Therefore this system provides the convenience to the voter to cast his/her vote from any location of their choice. Proposed system is cost effective and user friendly system and can be used for the secured voting purpose, thus making the counting process less burdened.

Keywords: Aadhaar, E-Voting Machine, Cortex-A15processor.

Introduction

With the growth of modern communications and Internet, information is very easily available and accessible electronically; The Information technology provides security in many aspects. Usages of information technology in the voting process improve the elections in sensitive areas where there is a chance of misuse and riots. This advanced technology is referred as Aadhaar based secure E-voting system. E-Voting is an election method in which the voters cast their votes electronically. In the past years, information security was used only in military and other important applications. But, now there is a need for secured voting. This E-Voting system helps the voter to select their representatives as per their preference without any unnecessary problems. Therefore this system helps in improving the security of voting process and builds trust to the voters. Voting system has more advantages than any other voting process according to the researchers. Advancements in voting systems include live recording using the cameras, webcams and Internet based recordings and cellular Mobile based recording systems are also used in voting processes. The finger impressions and Unique Identification Numbers details of the voters are stored in the database. The amount of data that is stored is in accordance with the central server. The system which is proposed in this paper has the advantage of reduction of physical effort, fast and accuracy. As the voter cast his/her vote the data is stored in the digital format, therefore more secured.

In section II the present existing system is discussed and in section III the proposed system explanation is given. Section IV concentrates on the components required to implement the proposed system where as Section V explains the Algorithm for the proposed system, followed by conclusion in Section VI.

Existing System

The present Electronic voting machine is the inter connection of two systems one is the control unit and other is Ballet unit. These two systems are interconnected via cable of length around five meters the control unit is operated by the Presiding officer who is the in charge of the Polling booth. The Ballet unit is fixed in the designated voting chamber. The current system allows the voter to cast his or her vote of their choice by pressing the appropriate button on the ballet unit. In EVM Control Unit will have a unique ID Number given by election commission and is painted on the unit with a permanent marker [1]. This ID Number will be allowed to be noted by the Polling Agents of various parties and will be recorded in the Register maintained by the Returning Officer as per the Election commission rules. The address tag attached to the Control Unit will also indicate this ID Number. Therefore, there is no question of replacement of any EVM. In the present Election process the voter needs to provide their Election photo identity card or any other card issued by the govt. recognized by the election commission of India. This verification of the photo identity card is done manually by concerned Polling officer and clearance will be given if found genuine. The present process involves lot of manual work which includes time and human effort.

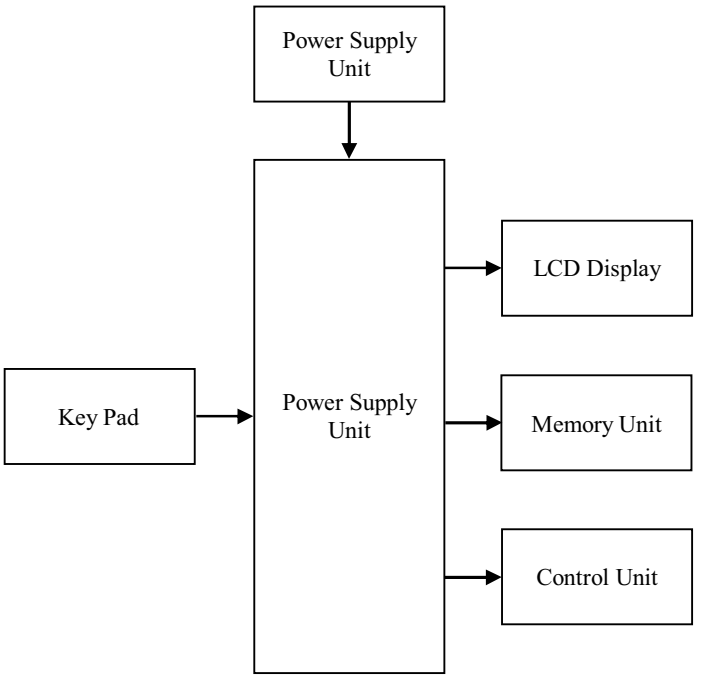


Figure 1. Existing System

Proposed System

In order to reduce the human interference and to overcome the drawbacks of the present system a new system Aadhaar based electronic voting system is proposed. In this system, Aadhaar based authentication is to be used to improve the security levels to the existing EVM. In this system the voter will be able to cast his/her votes with the help of user id and password. Where the user id is the Aadhaar number and password is the biometric finger impressions of the voter. As per the Aadhaar data acquisition system, the demographics and biometrics are captured and packaged at the time of Aadhaar enrollment .This data can be used to check the authenticity, where at the time of voting, the voter places the finger on the finger impression slot [2]. Since this system is linking the EVM with the Aadhaar server it will become more easy and convenient to check and verify whether the vote is casted by the right person and hence duplication of the votes may be avoided. Therefore this system helps in reducing the second time voters and the unauthorized voters to improve the standard of the voting process thus ensuring the security [3]. The Aadhaar identification number which is also known as Unique Identification number is unique for all the citizens of India. The voting process with this system ensures greater security which in turn provides the liberty to the voter to cast his/her vote.

The proposed system consists of a central unit which controls the process and a server unit. The voter can cast his vote by putting his finger in the biometric machine slot and that is encrypted and transmitted to the processor unit. Therefore cortex A-15 processor processes the obtained data and is stored. Then the verification of the data is done with the Aadhaar database via Ethernet.

Components in the Proposed System

Cortex-A15 processor

The Arm Cortex-A15 processor is a high performance engine for high performing flexible devices. Cortex processor carved a niche for itself among the galaxy of many other processors for its reliability and optimum power utilization [6]. Cortex-A15 processor which delivers double the performance is preferred over Cortex-A9 processor which is used in today's smart phones The ARM cortex-A15 processor is a 32-bit processor [7]. Cortex-A15 processor is selected for this project as it is well suited for many biometric applications.

Ballot unit

The ballot unit board is a simple System. It has no processing unit of its own. The processing unit is instead, it uses two electronically programmable logic devices (EPLDs) to interpret signals from the control unit CPU and interface with the

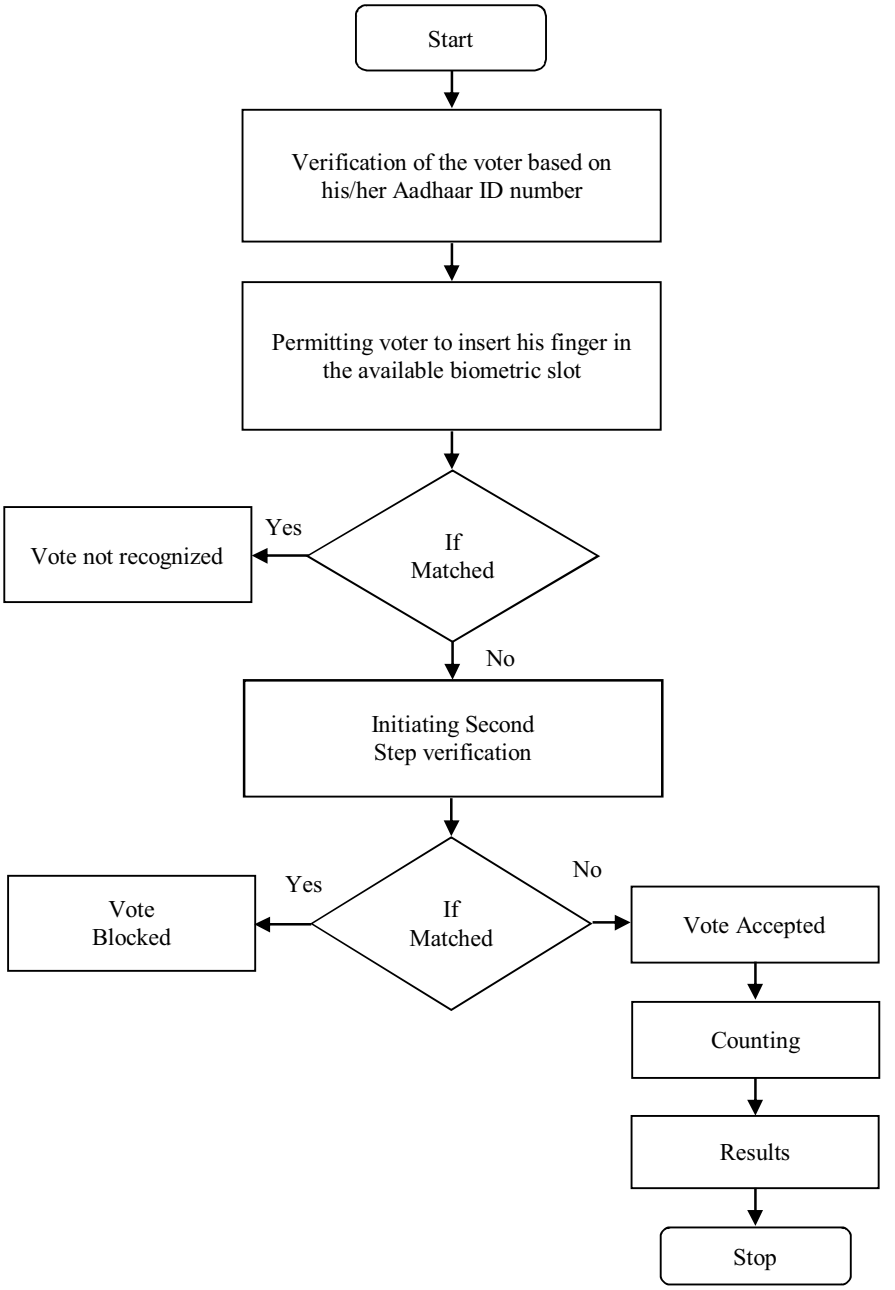


Figure 2. Proposed System

candidate buttons and LEDs on its face [3]. It also contains a four-position switch used to select the ballot unit's position in a multi-unit chain.

Aadhaar Server

The data from citizens is Collected, Organized, Stored and Maintained by Government of India. During the authentication transaction, the voters' record is first selected using the Aadhaar Number and then the demographic/biometric inputs are matched with the database for which the data is provided by the citizens during enrolment/update process [5]. Aadhaar authentication service is exposed as stateless service over hyper text transfer protocol Usage of open data format in XML and widely used protocol such as HTTP allows easy adoption and deployment of Aadhaar authentication.

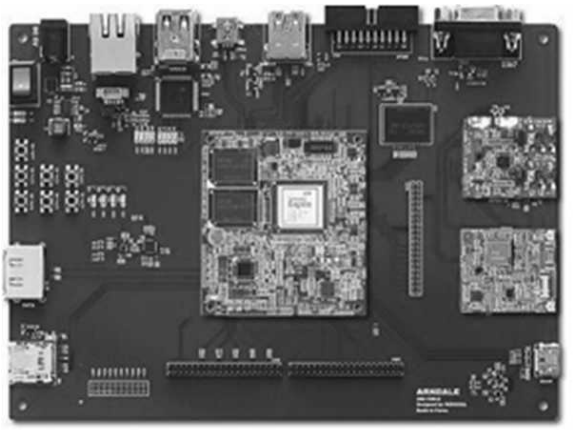


Figure 3. Cortex-A15 processor

Algorithm

Proposed Algorithm:

- Step I: Insert the finger in the provided Biometric slot.
- Step II: If the demographics and biometrics match with the stored data, ballot is released. Voter can cast his/her vote.
- Step III: In case of mismatch of demographics and biometrics, ballot will not be released.
- Step IV: If the Step III is repeated twice, the person will not be allowed to vote.

Conclusion

The inception of the idea Aadhaar based E-voting system is initiated after the current general elections. Keeping in view the bogus voters who exploit the total process, harmony and peace among the society a system which address these problems is developed. As the government of India is making Aadhaar card a mandatory in many sub-systems like driving license, Domestic LPG connections and banking services. Hence people the citizen of the country are motivated towards Aadhaar. The enrollment process of Aadhaar is still under process and expected to achieve 90% enrollment by 2017 positively. Hence this system can be used in the next general elections for the peaceful and secured conduct of polling in Digital India. Based on the design principles and requirement, a prototype of the system for E-voting System has been developed using Cortex-A15 processor. The system has several advantages that had been achieved. The advantages of the system are as follows: i) It gives confidence in voting system, only the genuine voter is allowed to access the voting Machine. ii) The system is user friendly and the user can easily understand the system and can cast his or her vote. The proposed system is fast performing and highly secure there by eradicating the defaulters.

References

[1] Benaloh, J. and Tuinstra, D. “Receipt-free Secret-Ballot Elections,” In proceedings of the 26th ACM Symposium on Theory of Computing (STOC’94), Montreal, Canada, 1994, pp. 544-553.
[2] Benoit, D. K. “Experience with voting overseas,” Appendix 2J to the first report of Ireland’s Commission on Electronic Voting, 2004.
[3] Brennock, M. “Cabinet to press ahead on e-voting in EU and local polls,” The Irish Times, 2004.
[4] California Secretary of State Ad Hoc Touch screen Voting Task Force Report, [Available Online at]: www.ss.ca.gov/elections/taskforce_report.htm.
[5] Caltech-MIT, “Voting: what is, what could be,” Cal Tech-MIT Voting technology Project Report, 2001 [Available Online at]: www.vote.caltech.edu/Reports.
[6] https://en.wikipedia.org/wiki/ARM_Cortex-A15
[7] http://www.arm.com/products/processors/cortex-a/cortex-a15.php

An Secured and Energy Conserved Utilization Path Algorithm using Secret Key and Adaptive Partition Controller in WSN

K.Ramanan* and E.Baburaj**
* Deparment of Computer Science & Engineering, Sathyabama University chennai, India
ramana3483@gmail.com
** Department of Computer Science & Engineering, Sun College of Engineering,Nagercoil, India
alanchybabu@gmail.com

Abstract: In wireless sensor networks (WSNs), secured and energy conserved data gathering are two important performance metrics for critical event monitoring in wireless sensor networks. Current state-of-the-art research is limited to either maximizing security under redundant radix-based approach or security through adversary model while satisfying throughput requirement. Although many prior research efforts resulted in optimized solutions, but how to optimize both objectives simultaneously has to be addressed. Our proposal work aimed on presenting an integrated configuration called Secret Key and Adaptive partition controller (SK-APC) is to maximize the fastness of data gathering of sensed data in the sink. To focus on secure data gathering of sensed data events, the encryption and decryption of sensor data is simplified using Keyed Hash Function. We show that the solution to the integrated configuration problem characterizes the secured energy conserved data gathering through XORing operations. Next, an Adaptive partition controller based on minimum Steiner tree is presented with the objective of reducing the energy consumption of sensed data gathering at the sink. By applying the minimum Steiner tree, minimum hops between source and sink nodes are identified satisfying the energy conservation principal. An extensive simulation analysis to demonstrate the security aspect and energy conservation principal with time taken for data gathering is presented. The obtained results show that SK-APC provides comparatively better performance than the state-of-the-art works in terms of both security and energy efficiency. Experimental analysis shows that SK-APC is able to reduce the time for data gathering by 25.72% and energy consumption by 18.48% compared to the state-of-the-art works.

Keywords: Energy conserved, data gathering, radix-based approach, Keyed Hash Function, Adaptive partition controller, Minimum Steiner tree.

Introduction

Current Improvement in communication have resulted in a significant shift in wireless sensor network research and as a result a secured and energy conserved data gathering configuration is the need of the hour. Many research works have been contributed in this aspect.

Protection Location Privacy in Wireless Sensor Network (PLP-WSN) [1] covered. Redundant Radix-based Approach (RRA) [2] in Wireless Sensor Network (WSN) provided an energy conserved means of communication using hybrid approach called Frequency Shift Keying (FSK) and Amplitude Shift Keying (ASK). Similar approaches to realizing Energy Efficient Routing protocols in WSN are found in [3] [4]. However, all the above mentioned methods lack an integrated effort in ensuring security and energy efficiency of data being collected, which is the core objective of our configuration.

The topological Structure in Layered Configurations (TSLC) [23] is a routing algorithm based on range of WSN which is used to perform the data communication. It is saves the energy of the entire network efficiently. In addition, it also increases the quality of the network service performance, and extends the life cycle of the network. Based on the communication protocol, it fails to realize the network energy protection. It also explains performance of throughput as well as the end-to end delay.

Recently, many data aggregation approaches using privacy homomorphism encryption have been presented and analyzed on wireless sensor networks. The approach in [5] used Recoverable Concealed Data Aggregation (RCDA) in WSN while in [6] a Routing algorithm to increase the lifetime of the network using Data Gathering Scheme (DGS) was presented to ensure data integrity and improving network lifetime. However, both the approaches lack data aggregation rate while exchanging messages between source and sink node. Data aggregation rate is addressed in the configuration SK-APC by applying Keyed Hash Function.

Energy-Efficient and Relay Hop Bounded Mobile Data Gathering Algorithm (BRH-MDG) [24] is established where the data collecting latency is effectively reduced by performing local aggregation through multihop transmissions. Then, the

algorithm uploads the data aggregated to the mobile collector. But, the BRH-MDG algorithm is failed in energy utilization of the entire network optimized. Discrete Cosine Transform (DCT) [26] of compactness attribute is used in sensory data and it is also used to improve the recovery accuracy. The time complexity of the algorithm is evaluated in the DCT method for reducing the computational cost. But, the realistic sensor signal is not accurately sparse. Thus, the low sampling rate causes insufficient measurements and fails in accuracy recovery.

Related works

A wireless sensor network comprises of thousands of small sensors with non rechargeable batteries. For such sensor nodes, the energy consumption during transmission is much more than during computation. Hence such networks employ Energy Efficient and High Accuracy [7] model for data aggregation. Cluster based data gathering [8] to ensure efficient data transmission was presented.

Location-Energy Spectral Cluster Algorithm (LESCA) [21] is determined in many clusters in a network. The spectral classification is providing on location-energy using the residual energy in various network nodes. The Location-Energy Spectral performance is improves the energy efficiency and also increased network lifetime. The clustering algorithms are not handled in an optimal method. So, the entire energy consumed of sensor network per round gets increased rapidly.

Energy Efficient Cluster Based Scheduling Scheme [22] for wireless sensor networks maintains the network lifetime, increases power as well as the high delivery ratio. Ubiquitous wireless connectivity was established in [9] with the objective of improving network lifetime. A resource aware algorithm [10] ensured location privacy while extending network lifetime. Though network lifetime and energy efficient approaches were ensured there observed a tradeoff in terms of time for data gathering which is addressed in the configuration SK-APC using Adaptive partition controller.

Energy-Efficient Adaptive Routing [27] is produced the energy load on multiple routes and it performs the quality of the network lifetime with increasing the end-to-end transmission. However, the energy efficiency is reduced because energy efficiency frequently fetches the additional latency. Secure and reliable routing protocols for WSNs [28] are estimated to handle the application of security requirements and the reliable data transmission using selective encryption method. However, protocol sending more data over the multipath is required in order to recognize a certain number of path failures.

There are several performance hurdles that hamper the design and deployment of WSN during critical monitoring applications. In [11], Latin Squares were adopted to ensure low packet delay and low overhead. However, network lifetime with respect to scalability with which the data packet can be transmitted remained unsolved. To improve network lifetime Eigen Beam forming was applied in [12] ensuring security in fading channels. Energy analysis component for quality of protection- modeling language (QoP-ML) [25] evaluated the inclined of many security stages on the energy utilization of a protocol. However, components fail to consider the trade-off between security and energy efficiency.

While various investigations have been carried out in the direction of energy efficient data gathering in WSN [13] [14] [15] security while data being collected at the sink remains to be explored. In particular, with appropriate encryption and decryption of sensor data, it may be possible to be extended to address security aspect.

Numerous critical monitoring applications like commercial and military require secure operation of sensor networks, and the outcome is affected largely with highly compromised sensor nodes in the network. Similar approaches employing energy efficient and security in WSN are found in [16] [17] [18]. But the rate at which data integrity was provided remained unaddressed, which is addressed in SK-APC by reducing the false positive data aggregate rate using Symmetric Key and Cyclic Redundancy Check. The approach in [19] uses Cyclic Diversionary Routing (CDR) while in [20] Energy Based Connected Dominated Set was used to improve security increasing the lifetime of network.

The movable and deployable resource unit[29] (MDRU)-based network provides communication services in disaster-struck areas where the lack of spectrum and energy resources is intensified due to the high demand from users and the power outages after a disaster. The MDRU-based network attempts to apply spectrum- and energy-efficient methods to provide communications services to users. A novel data collection scheme, called the Maximum Amount Shortest Path[30] (MASP), that increases network throughput as well as conserves energy by optimizing the assignment of sensor nodes. MASP is formulated as an integer linear programming problem and then solved with the help of a genetic algorithm. A two-phase communication protocol based on zone partition is designed to implement the MASP scheme. We also develop a practical distributed approximate algorithm to solve the MASP problem. In addition, the impact of different overlapping time partition methods is studied.

A good energy conserved and secured data gathering configuration should resolve energy imbalance while improving the security with respect to data being collected at the sink in WSN. Motivated by this, we investigate an integrated configuration to realize the objective that secured and energy conserved data gathering is efficiently provided. The main contributions of this paper are as follows. Secured data gathering configuration to reduce the false positive data aggregate rate at the sink node in WSN is proposed, in which a Marvin Keyed Hash Function, used to enhance the fastness of data gathering in the sink node. Construction of Steiner tree to minimize the energy consumption is designed.

The remainder of the paper is organized as follows: Secured Energy Efficient Data gathering configuration is given in Section 2, the Secret Key and Adaptive partition controller (SK-APC). The programming model and implementation with experimental setup is presented in Section 3, the implementation and numeric experimental results with discussions are shown in Section 4. The concluding remark is provided in Section 5.

Secret Key and Adaptive partition controller

In this work, a Bi-criteria optimization problem, to maximize security while minimizing energy consumption in improving the fastness of the data gathering of the sensed data event in the sink is presented. Figure 1 shows the block diagram of the Secret Key and Adaptive partition controller (SK-APC) configuration.

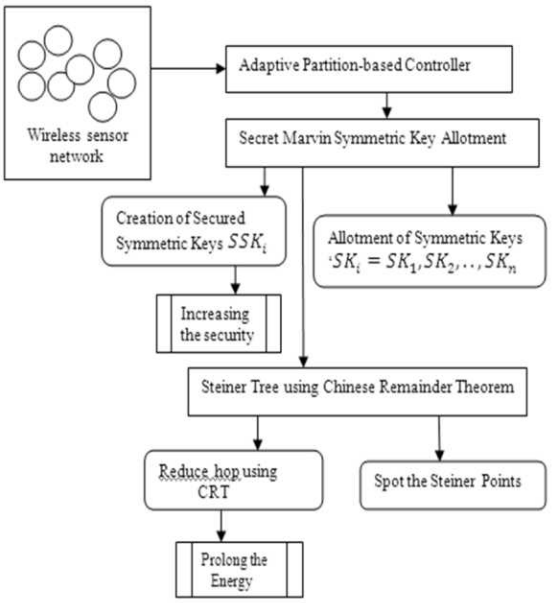


Figure1. Block diagram of Secret Key and Adaptive partition controller (SK-APC)

Fig. 1 shows that, the configuration SK-APC is divided into two parts. With the objective of improving the security, Secret Marvin Symmetric Key Allotment is applied to the sensor nodes in WSN for the sensed data event. Secret Marvin Symmetric Key Allotment model first performs an efficient distribution of Symmetric Key by performing XOR operations and likelihood function. Next, aiming at conserving the energy, a Steiner Tree using Chinese Remainder Theorem is applied that observes the steiner points (i.e., center node or sink node) and reduce the hop by applying CRT. As a result, a secured energy efficient data gathering on wireless sensor network is established.

Representation of Secret Marvin Symmetric Key Allotment

In this module, the representation of Secret Marvin Symmetric Key Allotment is increasing the security is presented. The SK-APC configuration uses a Secret Marvin Symmetric Key Allotment model for data being collected. The Secret Marvin Symmetric Key Allotment model is divided into three steps: Symmetric Key Allotment, discovering Symmetric Key and Route-Symmetric Key generation. The three steps in the design of SSKG model are described briefly as follows.

Let us assume that each sensor node '$SN_i = SN_{1,} SN_{2,.....,} SN_n$' in Wireless Sensor Network (WSN) is assigned with a Symmetric Key '$SK_i = SK_1, SK_2, .., SK_n$' $SK_i = SK_{1,} SK_{2,.....} SK_n$ that is disclosed only to the sink node 'S' and is mathematically formulated as given below.

$$\sum_{i=1}^{n} SN_i \rightarrow SK_i \rightarrow S \qquad (1)$$

From (1), the sensor nodes in WSN include a Symmetric Key and only with the help of this Symmetric Key, the sensor communicates with the sink node. This Symmetric Key 'SK_i' used to encrypt sensor data and generate Keyed Hash Function for each sensor data. In order to achieve security, the SK-APC configuration uses Keyed Hash Function that instead of using

the Symmetric Key 'SK_i' directly, each sensor node obtain a Error Detecting Code (EDC) to generate Secret Symmetric Key 'SSK_i'.

During encryption, for 'j' sensor data, using EDC, source node generates an '$i-j$' Frame Check Sequence and the resulting frame of 'j' sensor data divisible by predetermined threshold value is used. On the other hand during decryption by the sink node, the incoming sensor data frame is divided by the predetermined threshold value.

This Secret Symmetric Key 'SSK_i' is then XORed (i.e., encrypted) with the sensor data in addition to the EDC and is formulated as given below.

$$Encrypted_{SD} \sum_{i=1}^{n} SSK_i \; XOR \; SD_i \tag{2}$$

The Secret Symmetric Key 'SSK_i' are generated by applying a threshold, 'δ' left padded with 'n' sensor data bits and is mathematically formulated as given below.

$$\sum_{i=1}^{n} SSK_i \rightarrow \delta \ll n \tag{3}$$

In this method, a large Symmetric pool of 'n' Keys are first generated using the Symmetric Key Allotment. The second step is the discovering of Symmetric Key where each sensor node in WSN locates which of its nearby sensor node shares the common key with itself by exchanging discovery messages between the sensor nodes.

During the successful accomplishment of discovering common key, a secured link is established between them, ensuring security and reducing false positive data aggregates. Fig.2 shows the diagrammatic representation for discovering symmetric key by the sensor nodes (i.e., 10 sensor nodes). The figure shows two different types of arrows where dotted arrow represents successful discovering of symmetric key between the sensor nodes whereas the dashed arrow specifies the discovering of symmetric keys through multi-hop sensor nodes.

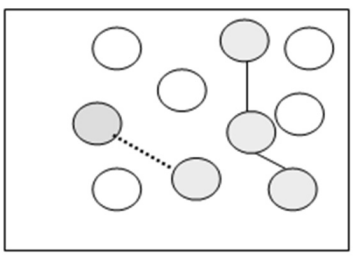


Figure 2. Symmetric key discovering

The Route-Symmetric Key generation step, an end-to-end Route Symmetric Key is assigned to pair neighbor nodes which do not share a common Symmetric Key. These sensor nodes in the SK-APC configuration are then linked by multi-hop secure links at the end of discovering of Symmetric Key. At the end of the Route-Symmetric Key generation, the likelihood that any pair of sensor nodes possesses at least one Symmetric Key is given as below

$$Likelihood\ (SN_i) = \frac{((SSK_i - n)!)^2}{(SSK_i - 2n)! * SSK_i!} \tag{4}$$

The above Secret Marvin Symmetric Key Allotment model, secure routing is performed from a source sensor node to the sink node. The security mechanism for data gathering of the sensed data event is based on the Marvin Keyed Hash Function.

The Marvin Hash Function Symmetric Key (MHF-SK) algorithm for data gathering for sensed data event is given as follows (in Fig. 3). The design of MHF-SK algorithm involves two steps. The two step process involves the encryption and decryption of sensor data to improve the fastness of data Gathering of sensed data event in the sink node, with the aim of increasing the security.

As shown in Fig. 3 ,the design of MHF-SK algorithm, involving efficient encryption and decryption aiming at improving the security of data gathering for sensed data event (i.e., data packet) by the sink node. In order to perform encryption, Symmetric Key and Secret Symmetric Key are obtained for each sensor nodes. Based on these two keys, the source node which wants to send the sensor data senses the presence of the neighbor nodes with same Symmetric Key.

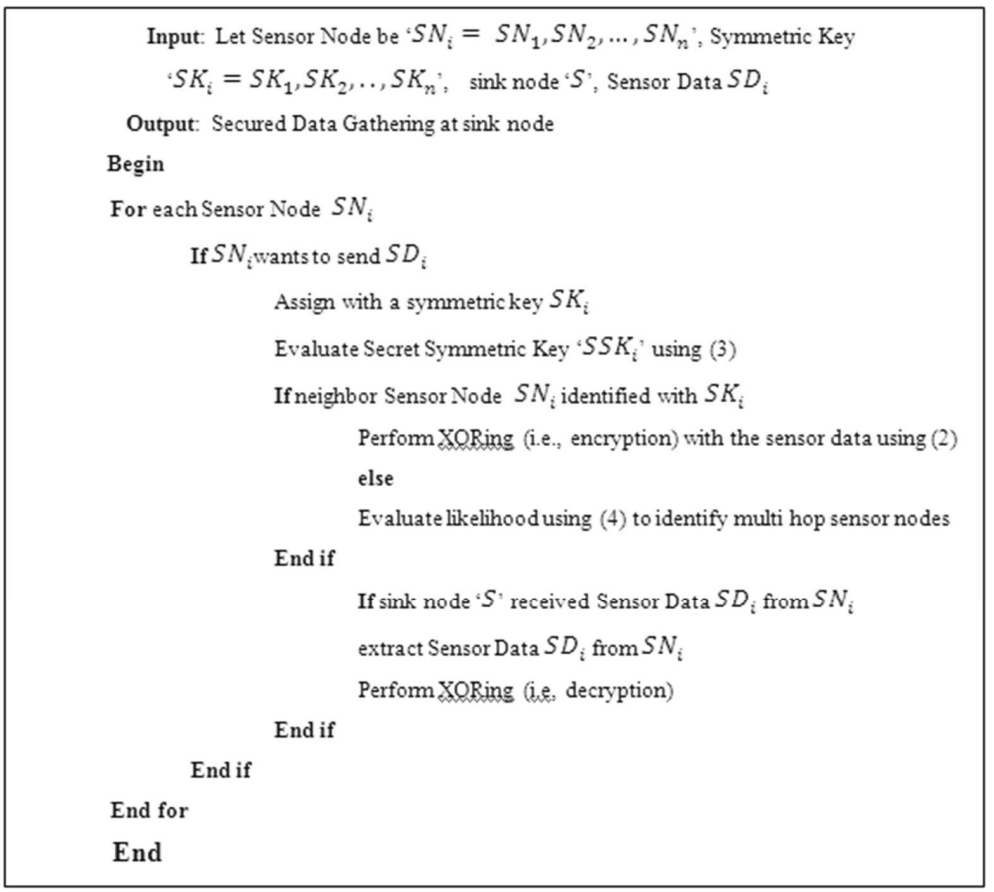


Figure 3. Marvin Hash Function Symmetric Key (MHF-SK) algorithm

On detection, an XOR operation is performed with efficient data gathering at the sink node. On the other hand, if neighboring nodes with same Symmetric Key is not detected, then the likelihood is measured to identify the multi-hop sensor nodes. In this similar manner, data gathering at the sink node is performed in a highly secured manner. One the other hand, decryption process is applied to extract the sensor data using the XORing operation. In this way, security for data being collected at the sink node is ensured. The next section discuss in detail about the energy conservation principal.

Representation of Steiner Tree

This model, we define an Adaptive Partition-based Controller based on Steiner Tree for energy efficient data gathering in WSNs. The aim of this model is to increase the percentage of sensor data successfully transferred, while reducing the energy consumption of sensor nodes. Let us consider a graph '$G = (V, E)$'where 'G' contains all the source nodes SN_i and sink 'S' respectively.

The aim of improving the energy efficiency, the SK-APC configuration uses Adaptive partition controller based on Minimum Steiner Tree that initially pairs up the available source nodes 'SN_i'and then randomly selects a hub node (i.e., center node or the Steiner point) from the node-pair. On the other hand, the loads of the non-hub node are shifted to the hub node, assigning the transmission costs on that specific edge. As a result, the non-hub node is removed and the hub node with aggregated load is grouped as a new set of sources. This process is repeated until the sink node is the only remaining node in the entire network. Fig.4 shows the Steiner Tree representation for four sensor nodes.

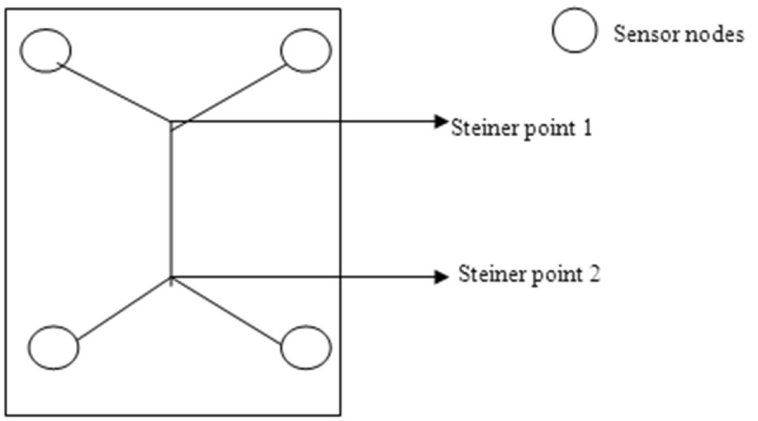


Figure. 4 Representation of Steiner Tree for four sensor nodes with two steiner points

The above-stated energy efficient task definitions, the energy efficient data gathering and successful sensor data transfer is maximized while reducing the energy usage by deciding upon the next hop that can be used as the forwarders (i.e., forwarding sensor nodes). After the formation of Steiner Tree the SK-APC configuration moves forward with the aid of Chinese Remainder Theorem (CRT) for splitting the sensor data packets.
Let us consider 'R' primes where '$r_i > 1$' then consider the product which is given as below

$$M = \Pi_i r_i \tag{5}$$

From (5) 'M' represents the multiplicative sensor data that produces simultaneous congruences and is obtained as below

$$m = \sum_{i=1}^{n} (coeff_i * m_i) \tag{6}$$

According to the CRT, the SK-APC configuration alternatively identifies with the set of numbers 'm_i' provided that 'r_i' are known. In addition, the CRT in the SK-APC configuration applies an Adaptive partition controller that instead of 'm' uses the mathematical formulation as given below

$$m_i = \sum_{i=1}^{n} [m_i * mod(r_i)] \tag{7}$$

By applying the above formulation, the maximum energy consumed by each sensor node for data gathering of the sensed data event in the sink node is reduced substantially.

Experimental Discussion

In this section we present the numerical data obtained as a result of applying SK-APC. Table 1 lists the set of input parameter and evaluates performance of SK-APC via simulation. Our example WSN consists of 100 sensor nodes deployed in a square area of A^2 (1600 m * 1600 m) placed in a random manner in the wireless sensor network that generates traffic for every 10 m/s.
The nodes are distributed in an area using Random Way point model for simulation, whereas the link layer provides the link between two nodes and the design of link is multi direction. The radio ranges are dynamically adjusted between 5m and 40 m to maintain network connectivity. The sink node collects the data packets of range 8 – 56 and forwards the data to the sink node with each data packet size differing from 100 KB to 512 KB. The simulation time varies from 5000 simulation seconds to 1600 simulation seconds.
Omni directional antenna is used for simulation and at any instant of time only single process is performed (i.e., either packet transmission or packet reception). Let us assume that the transmission color purple is selected as a starting point with a moving speed of 30 m/s and ends at anywhere in the network area. The default color for each sensor node is set and also the shape for each node is set by declaring variable m1 and assigned it with three shapes circle, hexagon and square. Communication is performed between the sensor nodes if the frequency matches between them and lies within its communication range.

Table 1. Simulation Parameters

Parameters	Values
Network area	1600 m * 1600 m
Number of sensor nodes	10,20,30,40,50,60,70,80,90,100
Number of data packets i.e., size of data block	9,18,27,36,45,54,63,72
Range of communication	40 M
Speed of node	0 – 30m/s
Simulation time	1600 s
Number of runs	8

Result and Discussion

In this section the result analysis of SK-APC is made and compared with two existing methods, Protection Location Privacy in Wireless Sensor Network (PLP-WSN) [1] and Redundant Radix-based Approach (RRA) [2] in WSN. The nodes in SK-APC configuration are positioned in uniform topology. To evaluate the efficiency of SK-APC, the following metrics like security, energy consumption, execution time for data gathering and data aggregate rate in Wireless Sensor Network is measured.

Impact of Security for data gathering

Security with respect to data being collected is measured on the basis of data packets received at the sink node in WSN. Therefore, security is the difference between the total packets sent to the packets not received at the sink node.

$$S\,(DC) = Packets_s - Packets_{nr} \quad (8)$$

Table 2. Security for data gathering with respect to hop count

Hop Count	Security for data gathering (p/s)		
	SK_APC	PLP_WSN	RRA
2	131	117	99
4	134	119	102
6	138	123	105
8	143	125	107
10	147	128	108

From (8), '$Packets_s$' refers to the data packets sent and '$Packets_{nr}$' refers to the data packets not received at the sink node in WSN. It is measured in terms of packets per second (p/s). The values obtained through (8) is tabulated for different hop count using the proposed SK-APC configuration and compared elaborately with the existing two works PLP-WSN [1] and RRA [2] respectively.

Fig. 5 shows the security for data collected at the sink node with respect to different hop counts. To better perceive the efficacy of the proposed SK-APC configuration, substantial experimental results are illustrated in Figure 5 and compared against the existing PLP-WSN [1] and RRA [2] respectively. The results reported above confirm that with the increase in the number of hop counts in WSN, the data collected at the sink node also increases and comparatively observed to be higher using SK-APC. So, the configuration SK-APC is said to be secured than compared to PLP-WSN and RRA. From the table 2, with a hop count of 2, 130 data packets (i.e., p/s) were efficiently collected at the sink node using the SK-APC configuration, whereas 117 and 99 packets per seconds were collected at the sink using PLP-WSN and RRA. The data packets collected at the sink is improved with the application of Secured Symmetric Key Distribution model. With the objective of improving security, Secured Symmetric Key Distribution model performs an XOR operation with the sensor data packet in addition to the Error Detecting Code (EDC). As a result, the successful packets received using SK-APC is improved and therefore security is improved by 9.77% compared to PLP-WSN and 23.33% compared to RRA.

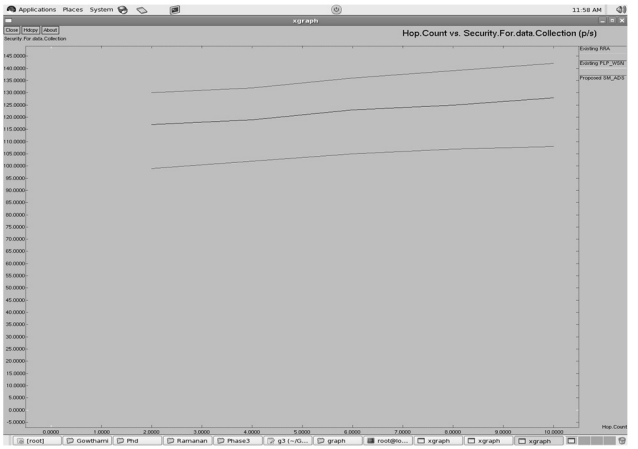


Figure 5. Measure of security for data gathering with respect to hop count

Impact of False Positive Data Aggregate rate

Whenever sensor nodes in the network send data, due to the presence of misbehaving nodes in WSN, possibility of packet drops occur during data packet transmission, that inject the false data into the packet and send to sink node. Then sink node then collects that packet. As a result, false positive data gets aggregated at the sink node. False positive data aggregate rate is the ratio of false data aggregated at the sink node to the overall data packet in WSN. The false positive data aggregate rate is formulated as given below

$$FPDA = \frac{False\ data\ packet}{Overall\ data\ packet} * 100 \quad (9)$$

From (9), ‘’, the false positive data aggregate rate for different number of sensor nodes in the range of 10 to 70 is measured.

In the experimental setup, the number of sensor nodes ranges from 10 to 70. The results of seven simulation runs conducted to measure the false positive data aggregate rate are listed in table 3. As listed in table 3, the SK-APC measure the false positive data aggregate rate which is measured in terms of percentage (%). The false positive data aggregate rate obtained using our configuration SA-ADS offer comparable values than the state-of-the-art methods.

The targeting results of false positive data aggregate rate using SK-APC configuration is compared with two state-of-the-art methods PLP-WSN and RRA and figure 6 is presented for visual comparison based on the relevant information. The figure shows the false positive data aggregate rate with respect to sensor nodes, with each sensor nodes sending different data packets to the sink node. As illustrated in figure 6, when 10 sensor nodes sent data packet to the sink node, the false positive data aggregate rate using SK-APC configuration was 21% compared to PLP-WSN and RRA that showed 27% and 25% respectively. Our configuration SK-APC differ from the ACSDTP differs from the PLP-WSN and RRA in that we have incorporated Marvin Keyed Hash Function. The advantage of applying using Keyed Hash Function in SK-APC configuration is that instead of using the Symmetric Key directly, each sensor node obtains a Error Detecting Code (EDC) to generate Secret Symmetric Key. This Secret Symmetric Key are then used for the communication between sensor and sink node which reduces the false positive data rate by 29.16% compared to PLP-WSN and 21.15% compared to RRA.

Table 3. False Positive data aggregate rate with respect to sensor nodes

Sensor Nodes (N)	False Positive data aggregate rate (%)		
	SK_APC	PLP_WSN	RRA
10	23	27	25
20	27	32	29
30	29	34	31
40	30	35	32
50	31	37	34
60	32	38	36
70	36	40	38

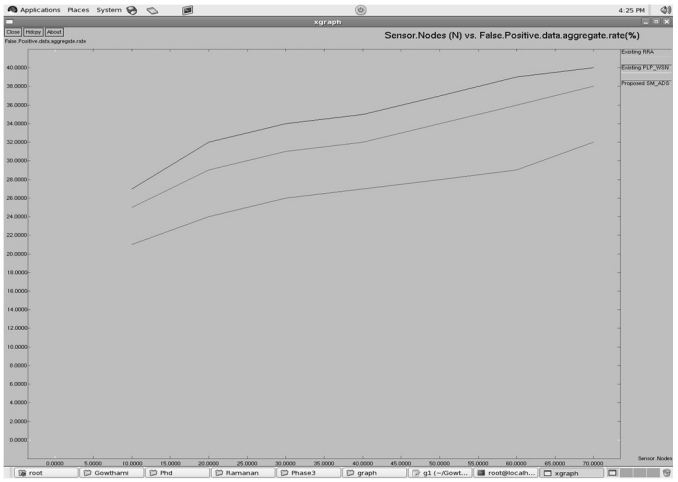


Figure 6. Measure of false positive data aggregate rate

Impact of Energy consumption for data gathering

Energy consumption for data being collected at the sink node is the product of energy consumed by a single sensor node and the total sensor nodes in WSN.

$$EC = Energy_{sn} * Total_{sn} \quad (10)$$

From (10), 'EC' is the energy consumption for data gathering at the sink node whereas 'sn' represents the sensor nodes. The consumption of energy is measured in terms of Joules.

Table 4. Energy consumption with respect to sensor nodes

Sensor Nodes (N)	Energy Consumption (Joules)		
	SK_APC	PLP_WSN	RRA
10	47	58	64
20	51	60	66
30	53	61	69
40	54	64	70
50	56	65	72
60	60	67	73
70	62	69	75

In Table 4 we further compare the energy consumed by different number of sensor nodes for data gathering at the sink in WSN. The experiments were conducted using seventy sensor nodes and the energy consumed is measured in terms of Joules (J).

Fig. 7 given above shows the energy consumption rate for SK-APC configuration, PLP-WSN [1] and RRA [2] versus seventy different sensor nodes. The energy consumption returned over SK-APC configuration increases gradually though not linear for differing sensor nodes when compared to the two other methods. From figure 7, it is illustrative that the energy consumption for the data being collected at the sink node is reduced using the proposed SK-APC configuration This is because with the application of Steiner tree based on the minimum number of hops being selected between source and sink nodes, the energy consumption is reduced. The Steiner tree with the aid of Chinese Remainder Theorem split the sensor nodes data packets that decide the next hop to be selected as the forwarding nodes. Only after the selected forwarding nodes sensor nodes data packets are sent that reduces the energy consumption of data being collected at the sink node by 13.08% compared to PLP-WSN and 23.89% compared to RRA respectively.

Impact of time for data gathering

Time taken for data gathering is the difference between the end time and start time for data gathering by the sink node in WSN. It is measured in terms of milliseconds and is formulated as given below.

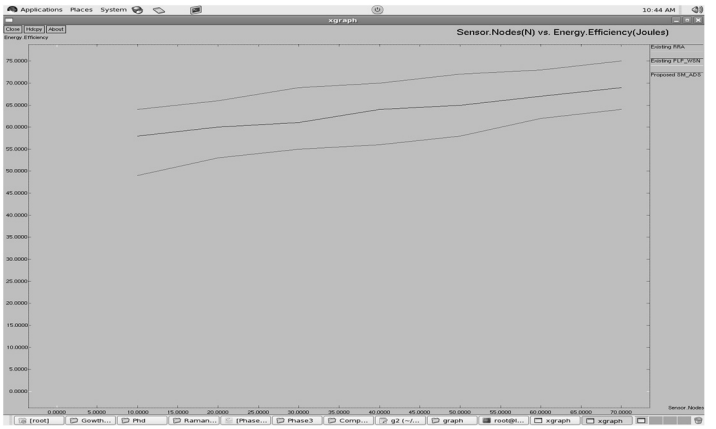


Figure 7. Measure of energy consumption

$$DC_t = (Endtime_{DC} - Starttime_{DC}) \qquad (11)$$

Table 5. Time Taken Data Gathering With Respect To Sensor Nodes

Sensor Nodes (N)	Time taken for Data gathering (ms)		
	SK_APC	PLP_WSN	RRA
10	124	173	185
20	142	177	197
30	154	182	201
40	158	191	216
50	171	198	221
60	173	209	224
70	192	215	227

From (11), the time for data gathering is measured using 'DC_t', whereas '$Endtime_{DC}$' represents the end time for data gathering by the sink node and '$Starttime_{DC}$' represents the start time for data gathering by the sink node in WSN. Table 5 shows the time for data gathering with respect to 70 sensor nodes with a moving speed of 25 m/s. To better perceive the efficacy of the proposed SK-APC configuration, substantial experimental results are illustrated in Figure 7 and compared against the existing PLP-WSN [1] and RRA [2] respectively.

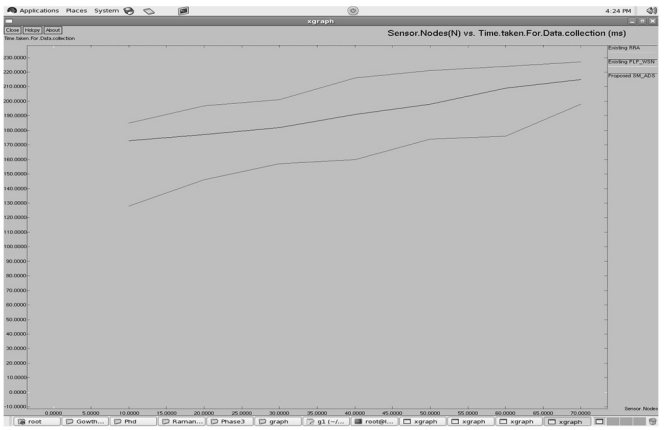


Figure 8. Impact of Time taken for data gathering

Fig. 8 shows the impact of time taken for data gathering with respect to varying sensor nodes in the range of 10 to 70 and the time for data gathering using three methods differs according to the size of sensor nodes. The results reported above confirm that with the increase in the number of sensor nodes being sent to the sink node for data gathering, the time for data gathering also increases. From figure 8, the time for data gathering using three methods differs according to the size of sensor nodes. As illustrated in Figure, the proposed SK-APC configuration performs relatively well when compared to two other methods PLP-WSN [1] and RRA [2]. This is because of the application of Adaptive Partition-based Controller randomly selects a hub node from the node pair. As a result, the loads of the non-hub node are shifted to the hub node. The non-hub node is removed and the hub node with aggregated load is grouped as a new set of sources. Through this, the time taken for data gathering using the SK-APC configuration is reduced by 21.86% compared to PLP-WSN and 29.58% compared to RRA respectively.

Conclusion

This article presents a novel configuration Secured Marvin and Adaptive Partition-based Controller (SK-APC) using the Marvin Keyed Hash Function. The performance of the proposed configuration is compared with secured data gathering and energy efficient data gathering methods (namely, PLP-WSN and RRA). The proposed configuration has the following advantages. (i) Improves security for data being collected at the sink node, (ii) provides low false positive data aggregate rate, (iii) representation of Steiner tree for achieving the energy conservation principle. The security in SK-APC configuration is improved using Secured Marvin Symmetric Key Distribution that applies Symmetric Key and Secret Symmetric Key during encryption and decryption that is available only to the source and sink node. By applying Marvin Keyed Hash Function and CRS, the false positive data aggregate rate is reduced significantly. Finally with the construction of steiner tree, the energy consumption for data gathering at the sink node is reduced considerably. Simulations were conducted to measure the performance of SK-APC configuration and evaluated the performance in terms of different metrics, such as security, false positive data aggregate rate, energy consumption and time to perform data gathering at the sink node in WSN. The results show that SK-APC configuration offers better performance with an improvement of security by 16.55% and reducing the false positive data aggregate rate by 25.15% compared to PLP-WSN and RRA respectively.

References

[1] Kiran Mehta, Donggang Liu, and Matthew Wright,," Protecting Location Privacy in Sensor Networks against a Global Eavesdropper", IEEE Transactions on Mobile Computing, Vol. 11, No. 2, February 2012.
[2] Koushik Sinha, Bhabani P. Sinha, and Debasish Datta," An Energy-Efficient Communication Scheme for Wireless Networks: A Redundant Radix-Based Approach", IEEE Transactions On Wireless Communications, Vol. 10, NO. 2, February 2011.
[3] Manish Kumar Jhaa, Atul Kumar Pandey, Dipankar Pala, Anand Mohan," An energy-efficient multi-layer MAC (ML-MAC) protocol for wireless sensor networks", International Journal of Electronics and Communications (AEÜ), Elsevier, Mar 2010.
[4] Jalel Ben-Othman, Bashir Yahya," Energy efficient and QoS based routing protocol for wireless sensor networks, J. Parallel Distrib. Comput, Elsevier, Mar 2010.
[5] Chien-Ming Chen, Yue-Hsun Lin, Ya-Ching Lin, and Hung-Min Sun," RCDA: Recoverable Concealed Data Aggregation for Data Integrity in Wireless Sensor Networks", IEEE Transactions On Parallel And Distributed Systems, Vol. 23, No. 4, April 2012.
[6] Yi-hua Zhu, Wan-deng Wu, Jian Pan, Yi-ping Tang," An energy-efficient data gathering algorithm to prolong lifetime of wireless sensor networks", Computer Communications, Elsevier, Jan 2010.
[7] Hongjuan Li, Kai Lin, Keqiu Li," Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks", Computer Communications, Elsevier, Mar 2010.
[8] Hongbo Jiang, Shudong Jin, and Chonggang Wang," Prediction or Not? An Energy-Efficient Framework for Clustering-based Data Collection in Wireless Sensor Networks", IEEE Transactions On Parallel And Distributed Systems, Volume:22 , Issue: 6, Apr 2011.
[9] Ahmad Rahmati, and Lin Zhong," Context-Based Network Estimation for Energy-Efficient Ubiquitous Wireless Connectivity",IEEE Transactions On Mobile Computing, Vol. 10, No. 1, January 2011.
[10] Chi-Yin Chow, Mohamed F. Mokbel and Tian He," A Privacy-Preserving Location Monitoring System for Wireless Sensor Networks",IEEE Transactions On Mobile Computing, Vol. 10, No. 1, January 2011.
[11] Chih-Kuang Lin, Vladimir I. Zadorozhny, Prashant V. Krishnamurthy, Ho-Hyun Park, and Chan-Gun Lee," A Distributed and Scalable Time Slot Allocation Protocol for Wireless Sensor Networks", IEEE Transactions On Mobile Computing, Vol. 10, No. 5, April 2011.
[12] Xiangyun Zhou, Radha Krishna Ganti and Jeffrey G. Andrews," Secure Wireless Network Connectivity with Multi-Antenna Transmission",IEEE Transactions On Wireless Communications, Vol. 10, No. 2, February 2011.
[13] Anfeng Liu, Laurence T. Yang, Motoki Sakai, andMianxiong Dong," Secure and Energy-Efficient Data Collection in Wireless Sensor Networks", Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2013.
[14] Anfeng Liuyz , Zhongming Zhengz , Chao Zhangy, Zhigang Cheny, and Xuemin (Sherman) Shenz, " Secure and Energy-Efficient Disjoint Multi-Path Routing for WSNs", Volume:61 , Issue: 7, Sep 2012.
[15] Kamlesh A. Waghmare, Dr. P.N. Chat," Energy Efficient Data Collection and Routing Algorithm in Wireless Sensor Network: A Survey", International Journal For Research In Emerging Science And Technology, Volume-1,Issue-2, July-2014.
[16] Sung Jin Choi, Kyung Tae Kim, and Hee Yong Youn," An Energy-Efficient Key Predistribution Scheme for Secure Wireless Sensor Networks Using Eigenvector", Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2013.

[17] Jiliang Zhou," Efficient and Secure Routing Protocol Based on Encryption and Authentication for Wireless Sensor Networks", Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2013.
[18] Chaoran Li and Yun Liu," ESMART: Energy-Efficient Slice-Mix-Aggregate for Wireless Sensor Network", Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2013.
[19] Ju Ren, Yaoxue Zhang, and Kang L," An Energy-Efficient Cyclic Diversionary Routing Strategy against Global Eavesdroppers in Wireless Sensor Netwo", Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2013.
[20] Xiaoyan Kui, Yu Sheng, Huakun Du, and Junbin Liang," Constructing a CDS-Based Network Backbone for Data Collection in Wireless Sensor Networks", Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2013.
[21] Ali Jorio, Sanaa El Fkihi, Brahim Elbhiri, and Driss Aboutajdine "An Energy-Efficient Clustering Routing Algorithm Based on Geographic Position and Residual Energy for Wireless Sensor Network" Hindawi Publishing Corporation Journal of Computer Networks and Communications Volume 2015, Article ID 170138, 11 pages.
[22] E. Srie Vidhya Janani1 and P. Ganesh Kumar "Energy Efficient Cluster Based Scheduling Scheme for Wireless Sensor Networks" Hindawi Publishing Corporation The Scientific World Journal, Volume 2015, Article ID 185198, 9 pages.
[23] Jun Yu and Xueying Zhang "A Cross-Layer Wireless Sensor Network Energy-Efficient Communication Protocol for Real-Time Monitoring of the Long-Distance Electric Transmission Lines" Hindawi Publishing Corporation Journal of Sensors Volume 2015, 13 pages.
[24] Ling Chen, JianxinWang, Xiaoqing Peng, and Xiaoyan Kui, "An Energy-Efficient and Relay Hop Bounded Mobile Data Gathering Algorithm in Wireless Sensor Networks" Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2015, 9 pages.
[25] Damian Rusinek, Bogdan Ksiezopolski, and Adam Wierzbicki "Security Trade-Off and Energy Efficiency Analysis in Wireless Sensor Networks" Hindawi Publishing Corporation, International Journal of Distributed Sensor Networks, Volume 2015, 17 pages.
[26] Kefu Yi, Jiangwen Wan, Tianyue Bao, and Lei Yao "A DCT Regularized Matrix Completion Algorithm for Energy Efficient Data Gathering in Wireless Sensor Networks" Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2015, 11 pages.
[27] Shaohua Wan "Energy-Efficient Adaptive Routing and Context-Aware Lifetime Maximization in Wireless Sensor Networks" Hindawi Publishing Corporation, International Journal of Distributed Sensor Networks, Volume 2014,16 pages.
[28] Hind Alwan and Anjali Agarwal "A Multipath Routing Approach for Secure and Reliable Data Delivery in Wireless Sensor Networks" International Journal of Distributed Sensor Networks,Volume 2013 (2013),10pages.
[29] T. Ngo; H. Nishiyama; N. Kato; T. Sakano; A. Takahara," A Spectrum Energy- Scheme for Improving the Utilization of MDRU-Based Disaster Resilient Networks",IEEE Transactions on Vehicular Technology,Year:2014,Volume:63,Issue:5,Pages:2027-2037.
[30] S.Gao;H.Zhang;S.K.Das,"Efficient Data Collection in Wireless Sensor Networks with Path Constrained Mobile Sinks",IEEE Transactions on Mobile Computing,Year:2011,Volume:10,Issue:4,Pages:592-608.

Smart Watchmen with Home Automation System based on Raspberry Pi

Vanita Jain*, Ashwani Sinhal** and Saksham Jain***
*Bharati Vidyapeeth's College, of Engineering, New Delhi
vanita.jain@bharatividyapeeth.edu
** Bharati Vidyapeeth's College
ashwani@crystalpbx.com
***Bharati Vidyapeeth's College
sakshamjain2612@gmail.com

Abstract: The Smart Home is an advancement of the current existing home system. Home automation is an essential point of accomplishing brilliant home system and is regularly energizing field that has blown in the course of recent years. Launching latest advances have made homes more advantageous, effective and considerably more secure. Home security framework, equipped for movement and unsettling influence identification at passage focuses and making a caution notification with email notice alarms having a picture, was actualized to permit continuous observing for the house. In this paper, we have developed a smart home prototype which is composed of a two-way communication where electricity and data are triggered by the user and its utility. We have used Raspberry Pi which is an open source hardware which gives various customizations to transform a standard home into a high-class home. We have used different sensors which can be controlled and observed by the Raspberry Pi by means of the Internet of Things (IoT). The Raspberry Pi is interfaced with a camera, any movement detected in front of the camera will automatically shoot a mail to the admin and he can access the security lock from any remote area. Python codes have been used for interfacing every sensor; a model of the smart home is created. The smart home was completely tried and execution was discovered palatable.

Keywords: Raspberry Pi, Smart Home, Home Automation, Home Security, Internet of things.

Introduction

The Smart Home is a propelled stage to the way we get power today. In prior times, the interest for power was considerably contrasted with that right away. Since the interest for power has enormously expanded, an update of the present Home framework is greatly required. With the innovation accessible in these current times, the savvy Home could be planned in such a way, that it utilizes computerized correspondences innovation to recognize and respond to neighborhood changes in utilization. The framework will highlight a two-way dialog where power and data can be traded between the buyer and utility. This can expand or diminish the measure of vitality a buyer needs by breaking down the input of the two-way dialog. The exchange of power and data in the middle of customer and utility would expand productivity, unwavering quality and security. The shrewd Home likewise empowers renewable vitality innovation to be incorporated into the framework for a greener, all the more naturally inviting technique for acquiring vitality, in this way diminishing a rate of reliance on fossil fuel [1-3]

Home Automation can be considered as a demonstration of utilizing electronic frameworks/gadgets and programming them to supplant various human connections for the control of fundamental home capacities. This works on the base of associating sensors and gadgets to the IoT. IoT can be considered as a system of physical articles which can be gotten to through the Internet. For articles to considered IoT-based, systems should be changed over to an IP-based system for exclusive conventions. The article being joined with the web can speak to itself, digitally consequently being controlled from anyplace there is a web association. These articles have the capacity to transmit and get information over a system without human-to-human or human-to-PC association. This also means that more data can be assembled from these articles, even at a number of places with on-going information being presented from the objects. This super errand can increment effectiveness, well-being and security [4].

The Smart home should be in normal association with its inside and outer situations. The outer environment comprises of the considerable number of substances having a place to the Smart Home and the inward environment comprises of all machines and gadgets having a place with the keen home, which is halfway overseen by an element in it [5]. A smart home having mechanized framework can be made to build up control of specific parts of a home. This upgrades security and effectiveness of a home, which likewise gives constant observing by means of the web. Aside from controlling the sensors and parts by means of a gadget associated with the web, a midway controlled board having an LCD screen with keypad or an LCD

touchscreen can be utilized to control a significant number of the utilizations of home mechanization [6-7]. A few utilizations of home robotization incorporate computerized lights, which can be programmed to turn ON when somebody goes into a room, or relying upon the season of day or turn ON by means of a manual order over the web. Security framework, which can be controlled with constant checking of the house, where diverse sensors utilized as a part of the framework can be programmed to do specific things, for example, take photographs, turn on a siren, send email alarms and a great deal more. The aggregate business sector esteem for European brilliant homes business sector will be worth $13.81 Billion by 2020 at an expected compound yearly development rate CAGR (Compound Annual Growth Rate) of 18.17% [8].

The optimization problem lies within the constraints of minimum, forecasted, and learned energy requirements for consumers, thermal and renewable generation capacity, weather forecasts, and storage levels, as well as within physical and stochastic models to provide accurate forecasting of energy demand on the grid. Distinct hierarchical structures of these constraints based on increased penetration of renewable generation are necessarily considered. The optimization problem and its solutions become the basis for a pervasive communication framework that leads to a “just-in-time” model of energy distribution while utilizing energy generators at their maximum efficiency. [9].

The Raspberry Pi and the subject of home mechanization is a striking one. The Raspberry Pi can be utilized to computerize a home at a generally minimal effort. It works on the idea as the IoT. There are various things that make the Raspberry Pi crucial for home robotization yet the one that emerges the most must be the surprisingly reasonable expense. The incomprehensible measure of sensors at to a great degree eases makes it's wonderful for home computerization. A security framework to be introduced into a house can be expensive from security organizations, however with a few aptitudes and time the Raspberry Pi can be programmed to be a security framework with as you require, for a small amount of the expense [10-11]. Figure 1 demonstrates the Raspberry Pi B model with its details.

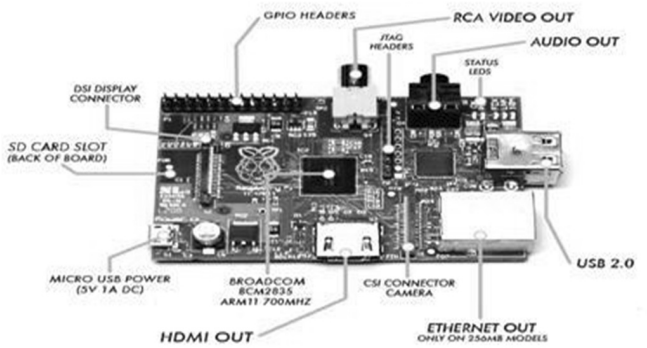


Figure 1. Raspberry Pi B Model [9]

In the present based work, we have written an IP-based system was built up, Python codes were composed of the sensors, which were associated with the Pi, an order then did from the system site, which was prepared by the Raspberry Pi and responded with the joined sensors. The reason for the present study is to manufacture an arrangement of interconnected gadgets and sensors, which permit the client to control and screen certain electrical/electronic gadgets in their home by means of the web from the Raspberry Pi. This implies executing a framework which permits ongoing checking of the home, likewise sending messages to the client when certain gadgets are activated.

Experimental Details

The Raspberry Pi works on a UNIX-based open source working framework called Raspbian OS [12]. This permits more control and adaptability in the product accordingly making it simple to program the Pi. The Raspberry Pi speaks with the appended gadgets and sensors through PYTHON codes to control their capacities. The Raspbian working framework was introduced onto Raspberry pi, which was acquired by downloading NOOBS onto the SD card from the maker's site [13].

Raspberry Pi Face Rack is a development board with 4 set of 26 GPIO pins [14]. We have interfaced piface advanced expander, Char LCD plate with the Raspberry Pi. Piface computerized expander was utilized to build a number of advanced I/Os and LCD Pi plate was utilized to show the status of Raspberry Pi.

Each sensor for home mechanization and security applications e.g. PIR infrared movement sensor, an attractive contact switch (Reed switch), siren, speaker, LEDs, push catches, DHT 11 sensor and so on were joined with Raspberry Pi through Pi Face advanced board. A proper block diagram which gives the detailed view of the Home Automation has been shown in Figure 2.

In order to have low energy consumption in the smart grid, electrical equipment optimized controllers are required [15-16]. The prototype build in this work, the principle controller was additionally joined with the Wi-Fi module to acquire the

entrance of web and HTML page utilizing advanced mobile phone/PC to control/access sensors and gadgets of the home at anytime, anyplace and from a remote place.
Table 1, demonstrates the rundown of significant segments utilized amid this work. A model of the brilliant home was likewise created, which appears in Figure 3 alongside controller unit.

Table 1. List of Major Components

Component	Picture
Raspberry Pi B	
Piface Rack	
Piface Digital I/O Expander	
Adafruit RGB LCD Plate with Keypad	
PIR Motion Sensor	
DHT 11 Sensor	
Reed Switch	
Wi-Fi Module	
USB Camera	

Results and Discussion

Usage points of interest and results identified with the two parts of keen home i.e. home security and home robotization are talked about here.

Home Security

The Raspberry Pi-based home security framework was fabricated utilizing Adafruit RGB positive 16x2 LCD Pi plate, Wi-Fi module, PIR sensor, USB camera (webcam), Reed switch and web access point. To enact the security framework, the keypad of the LCD Pi plate was utilized. Security caution can be equipped or handicapped by squeezing up or left keys of keypads for 3 seconds.
Figure 4 shows the status of caution on LCD Pi plate. After initiating the caution, a sound yield is sent to 3.5mm sound jack of Raspberry pi and after that "Framework is Armed" sound can be heard through the self-fuelled sound speaker. Every one of the gadgets mentioned above will also become active.

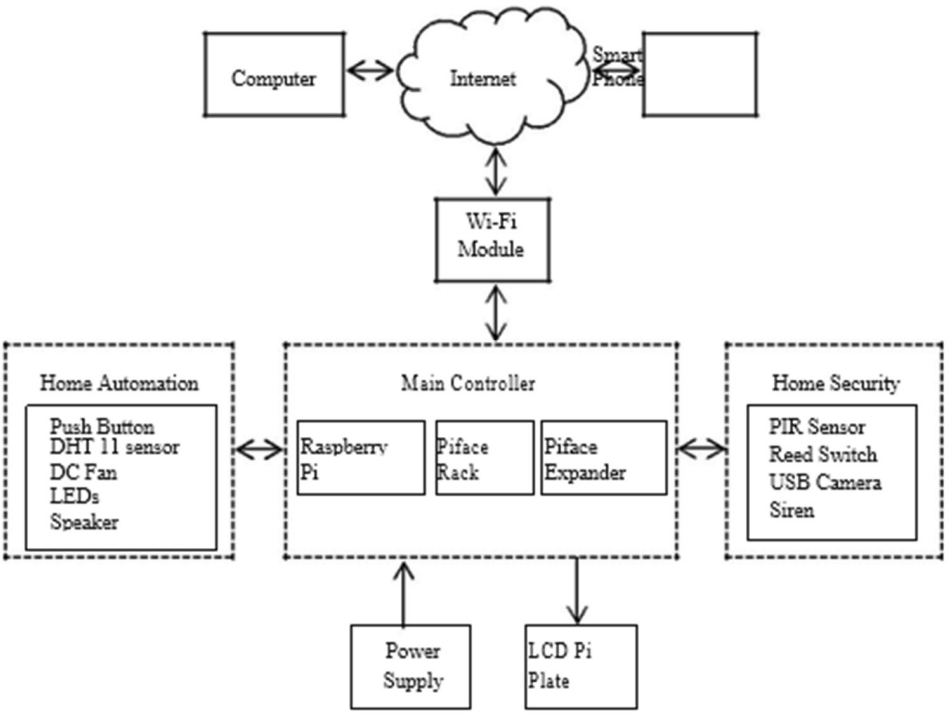


Figure 2. Block Diagram of Raspberry Pi-Based Smart Home

Figure 3. (Left)Prototype of Smart Home (Right) Main Controller

At the point when an article moves inside of the scope of PIR sensor, a sign is sent to the controller, which starts the webcam. Webcam snaps a photograph which is put away onto memory card of Raspberry Pi. The putaway photograph is then sent through email to the proprietor with the title "Movement notice" as appeared in Figure 5.

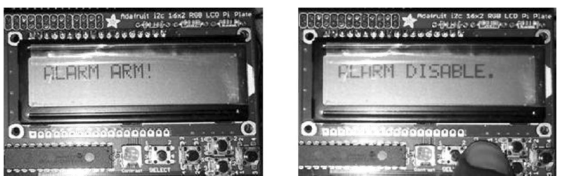


Figure 4. Home Security alarm system status

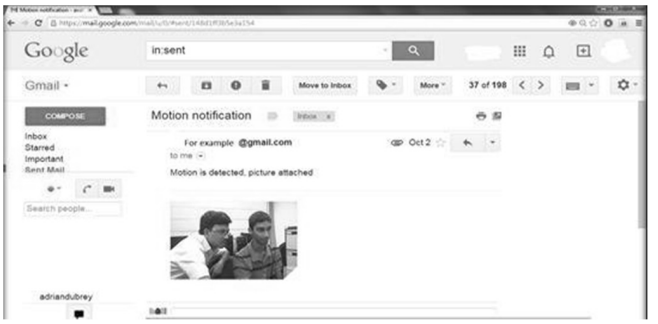


Figure 5. Email Notification upon motion detection

At the point when the entryway of home outfitted with reed switch is opened as appeared in Figure 6, the alert framework identifies it as a break, which turns ON the siren and a message "Framework Breached" is heard on the speaker. In the meantime, an email is sent to the proprietor with the title "Entryway Open" and picture around the entryway. Messages, notice titles, beneficiary email ID and so on were gone into the framework through python scripts.

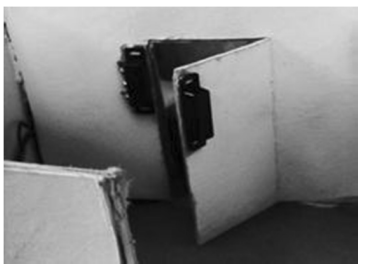


Figure 6: Entry detection using Reed Switch

Home Automation

The home automation system was also implemented around the same Raspberry Pi, which incorporates a brilliant doorbell, an electronic mechanized lighting framework and a temperature and moistness controller that turns an aerate and cool unit or fan on/off naturally under given conditions.

Figure 7 shows brilliant entryway ringer, which was manufactured utilizing a push catch, LEDs and a speaker. At the point when the push catch is squeezed, LED 1 will turn ON for 15 seconds and voice message

"Somebody will go to you in no time" will be heard through the speaker. After LED 1 cut off, LED 2 will be ON for 10 sec. furthermore, a voice message "Sorry, nobody is at home, please come back again later" will be listened.

Figure 7: Smart Door Bell

Home robotized light framework was composed utilizing HTML and LEDs. A site page was produced to control the lights in the home naturally from a web utilizing a PC, tablet or even a wireless. This framework utilizes a server location facilitated by the Raspberry Pi itself, which is shown on the LCD Pi plate amid start up. Figure 8 demonstrates the website page, which was made to control/check the status of lights and different sensors examined before.

Every button on the website page is clickable. Catches in the information area detect signals from the controller. As indicated by which info channel/sensor is being used, the particular catch will appear ON and its shading will get changed from dim to sky blue. At the point when any catch of the yield segment is squeezed, the separate yield would appear ON in red.

Figure 8. HTML Website to control smart house

Conclusion

Soon in not so distant future, the conventional Homes of today will advance into a powerful, successful, environment inviting and vitality proficient framework known as the Smart Home.

The present paper tried to plan normal home utilizing different sensors to be controlled and checked by the Raspberry Pi by means of the IoT. It is centred around two parts of keen home i.e. home security and home mechanization, for home security, the Raspberry Pi is programmed to work as a caution framework in which it distinguishes interruption at passage focuses alongside movement inside of the home and email alarms will be sent with pictures to permit ongoing changes of the home. This framework is additionally furnished with robotized lights and virtual switches for controlling lights in the home remotely utilizing outside and/or inside systems for administration with the Raspberry Pi by means of an HTML page.

References

[1] X. Fang et. Al. , " Smart grid- The new and improved power grid: A Survey", IEEE Communication Surveys & Tutorials, , 2012, Vol. 14, No.4, 944-980.

[2] H. Gharavi and R. Ghafurian, "Smart grid: The electric energy system of the future", Proceedings of the IEEE, 2011, Vol. 99, No. 6, 917 – 921.

[3] A. Ipakchi and F. Albuyeh, "Grid of the future", IEEE Power & Energy Magazine, 2009, Vol. 7, No. 2, 52–62.

[4] Bassi, Alessandro. Enabling Things to Talk: Designing IoT Solutions with the IoT Architectural Reference Model, 2013, Vol. 1 No. 1, 1-12.

[5] N. Komninos, E. Philippou and A. Pitsillides, "Survey in smart grid and smart home security: issues, challenges and countermeasures", IEEE Communication Surveys & Tutorials, 2014, Vol. 16, No. 4, 1933-1954.

[6] Dennis, A. K. 2013. Raspberry Pi Home Automation with Arduino. Packt Publishing.

[7] Wei, L. Wei, L. Xin, L. 2013. Design and implement on the smart home system. In proceedings of the Fourth International Conference on Intelligent Systems Design and Engineering Applications.

[8] European Smart Homes Market by Products (Security, Access, Lighting, Entertainment, Energy Management Systems, HVAC, and Ballast & Battery Pack), Services (Installation & Repair, Renovation & Customization) & Country - Global Forecasts and Analysis 2013 – 2020.

[9] Vaz, Eduardo; Shakshuki, Elhadi "Storage-based smart grid communication framework", *Sustainable Technologies (WCST), 2012 World Congress on,* On page(s): 37 - 43, Volume: Issue: 19-22 Nov. 2012. Upton, E. Halfacree, G. 2014 Raspberry Pi User Guide, 3rd. Ed. Wiley.

[10] Richardson, M. Wallace, S. 2014 Getting Started with Raspberry Pi, 2nd. Ed. Maker Media Inc.

[11] Bradbury, A. Everard, B. 2014 Learning Python with Raspberry Pi. Wiley.

[12] FOUNDATION, RASPBERRY PI. 2014. NOOBS SETUP. Accessed 11 21, 2014. http://www.raspberrypi.org/help/noobs-setup/.

[13] Mohsenian-Rad, A.-H.; Leon-Garcia, A. "Optimal Residential Load Control With Price Prediction in Real-Time Electricity Pricing Environments", *Smart Grid, IEEE Transactions on,* On page(s): 120 - 133 Volume: 1, Issue: 2, Sept. 2010

[14] Paudyal, S.; Canizares, C.A.; Bhattacharya, K. "Optimal Operation of Distribution Feeders in Smart Grids", *Industrial Electronics, IEEE Transactions on,* On page(s): 4495 - 4503 Volume: 58, Issue: 10, Oct. 2011.

[15] Mets, K.; Verschueren, T.; De Turck, F.; Develder, C. "Exploiting V2G to optimize residential energy consumption with electrical vehicle (dis)charging", *Smart Grid Modeling and Simulation (SGMS), 2011 IEEE First International Workshop on,* On page(s): 7 – 12

[16] Rieken, D.W.; Walker, M.R. "Ultra Low Frequency Power-Line Communications Using a Resonator Circuit", *Smart Grid, IEEE Transactions on,* On page(s): 41 - 50 Volume: 2, Issue: 1, March 2011.

Big Data Analytics using Hadoop Collaborative Approach on Android

Altaf Shah*, Amol Bhagat** and Sadique Ali***
*Department of Computer Science and Engineering,
Prof Ram Meghe College of Engineering and Management, Badnera, Amravati, India
altafshah7@gmail.com
**Innovation and Entrepreneurship Development Center,
Prof Ram Meghe College of Engineering and Management, Badnera, Amravati, India
amol.bhagat84@gmail.com
***Department of Computer Science and Engineering,
Prof Ram Meghe College of Engineering and Management, Badnera, Amravati, India
softalis@gmail.com

Abstract: Big data also known as data sets which are so large and complex that they are not easy to understand or handle using traditional versions of data processing or database management systems. The Data is always important for searching, transferring, visualizing, storing this data the Data Storage technologies are responsible. Now as the business grows the data pertaining to the business organization grows drastically - hence storage technologies are emerging as critical IT system on which the business depends. At the same time, businesses across the board are also increasing their investment in various mobility solutions. Areas those are getting popular like 'the integration of videos, images, personal information and streaming, displaying, them in cross platform mobile devices – and streaming it in cross platform mobile devices. This paper presents how to mine data on a mobile from a big-data on the remote server.

Keywords: Android, Big Data, Data Analytics, Hadoop, Mapreduce, Mobile Data Analytics.

Introduction

Recent development of various areas of Information and Communication Technology (ICT) has contributed to an explosive growth in the volume of data. According to a report published by IBM in 2012 [1], 90 present of the data in the world was generated in the previous two years. As a consequence, the concept of the big data has emerged as a widely recognized trend, which is currently attracting much attention from government, industry, and academia. Figure 1 show the various by which big data can be gathered. Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it.

The hot IT buzzword of 2012, big data has become viable as cost-effective approaches have emerged to tame the volume, velocity and variability of massive data. Within this data lie valuable patterns and information, previously hidden because of the amount of work required to extract them. To leading corporations, such as Wal-Mart or Google, this power has been in reach for some time, but at fantastic cost. Today's commodity hardware, cloud architectures and open source software bring big data processing into the reach of the less well-resourced. Big data processing is eminently feasible for even the small garage start-ups, who can cheaply rent server time in the cloud.

The value of big data to an organization falls into two categories: analytical use, and enabling new products. Big data analytics can reveal insights hidden previously by data too costly to process, such as peer influence among customers, revealed by analysing shoppers' transactions, social and geographical data. Being able to process every item of data in reasonable time removes the troublesome need for sampling and promotes an investigative approach to data, in contrast to the somewhat static nature of running predetermined reports.

The past decade's successful web start-ups are prime examples of big data used as an enabler of new products and services. For example, by combining a large number of signals from a user's actions and those of their friends, Facebook has been able to craft a highly personalized user experience and create a new kind of advertising business. It's no coincidence that the lion's share of ideas and tools underpinning big data has emerged from Google, Yahoo, Amazon and Facebook. The emergence of big data into the enterprise brings with it a necessary counterpart: agility. Successfully exploiting the value in big data requires experimentation and exploration. Whether creating new products or looking for ways to gain competitive advantage, the job calls for curiosity and an entrepreneurial outlook.

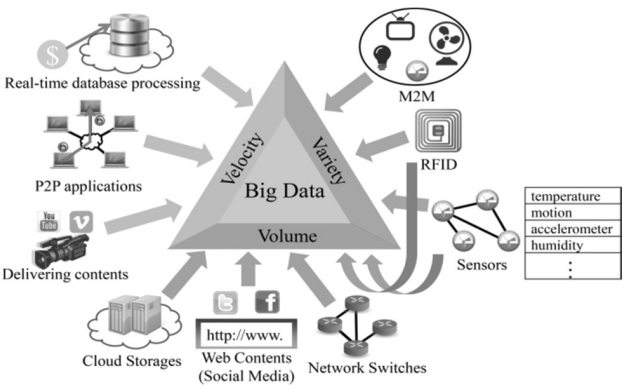


Figure 1. Major trends of big data gathering

The primary goal of big data analytics is to help companies make more informed business decisions by enabling data scientists, predictive modellers and other analytics professionals to analyse large volumes of transaction data, as well as other forms of data that may be untapped by conventional business intelligence (BI) programs. That could include Web server logs and Internet clickstream data, social media content and social network activity reports, text from customer emails and survey responses, mobile-phone call detail records and machine data captured by sensors connected to the Internet of Things. Some people exclusively associate big data with semi-structured and unstructured data of that sort, but consulting firms like Gartner Inc. and Forrester Research Inc. also consider transactions and other structured data to be valid components of big data analytics applications.

Big data can be analysed with the software tools commonly used as part of advanced analytics disciplines such as predictive analytics, data mining, text analytics and statistical analysis. Mainstream BI software and data visualization tools can also play a role in the analysis process. But the semi-structured and unstructured data may not fit well in traditional data warehouses based on relational databases. Furthermore, data warehouses may not be able to handle the processing demands posed by sets of big data that need to be updated frequently or even continually -- for example, real-time data on the performance of mobile applications or of oil and gas pipelines. As a result, many organizations looking to collect, process and analyse big data have turned to a newer class of technologies that includes Hadoop and related tools such as YARN, MapReduce, Spark, Hive and Pig as well as NoSQL databases. Those technologies form the core of an open source software framework that supports the processing of large and diverse data sets across clustered systems [17].

For years SAS customers have evolved their analytics methods from a reactive view into a proactive approach using predictive and prescriptive analytics. Both reactive and proactive approaches are used by organizations, but let's look closely at what is best for your organization and task at hand. Enterprises are increasingly looking to find actionable insights into their data. Many big data projects originate from the need to answer specific business questions. With the right big data analytics platforms in place, an enterprise can boost sales, increase efficiency, and improve operations, customer service and risk management.

Webopedia parent company, QuinStreet, surveyed 540 enterprise decision-makers involved in big data purchases to learn which business areas companies plan to use Big Data analytics to improve operations. About half of all respondents said they were applying big data analytics to improve customer retention, help with product development and gain a competitive advantage. Notably, the business area getting the most attention relates to increasing efficiencies and optimizing operations. Specifically, 62 % of respondents said that they use big data analytics to improve speed and reduce complexity.

Related Work

The different aspects of hadoop distributed file system are described in [1]. It presents the working of hadoop components. It describes comparison of hadoop technique with other system technique and concludes that Hadoop is possibly one of the best solutions to maintain the Big Data. Paper provide study of other techniques such as Grid Computing tools, Volunteering Computing and RDBMS techniques, paper presents that Hadoop is capable enough to handle such amount of data and analyse such data. Big data analytics define the analysis of large amount of data to get the useful information and uncover the hidden patterns [2]. Big data analytics refers to the Mapreduce Framework which is developed by the Google. Apache Hadoop is the open source platform which is used for the purpose of implementation of Google's Mapreduce Model. In this the performance of SF-CFS is compared with the HDFS using the SWIM by the Facebook job traces. SWIM contains the workloads of thousands of jobs with complex data arrival and computation patterns.

To explore Big Data, [3] have analysed several challenges at the data, model, and system levels. To support Big Data mining, high-performance computing platforms are required, which impose systematic designs to unleash the full power of the Big Data. At the data level, the autonomous information sources and the variety of the data collection environments, often result in data with complicated conditions, such as missing/uncertain values. In other situations, privacy concerns, noise, and errors can be introduced into the data, to produce altered data copies. Developing a safe and sound information sharing protocol is a major challenge. At the model level, the key challenge is to generate global models by combining locally discovered patterns to form a unifying view. The report [4] is an attempt to assess the potential value of data equity to the companies that gather and store big data, and who combine this with high-performance analytics to harness the big data's full potential. they undertook this at the sector level by reviewing and developing an understanding from the available literature of how and by what means big data could be expected to impact on each of twelve sectors: Retail Banking, Insurance, Investment Banking, Retail, Central Government, Healthcare, Transport & Logistics, Telecommunications, Energy & Utilities, Manufacturing, Professional Services and Other Activities.

In [5], authors provide an overview of state-of-the-art research issues and achievements in the field of analytics over big data, and they extend the discussion to analytics over big multidimensional data as well, by highlighting open problems and actual research trends. Their analytical contribution is finally completed by several novel research directions arising in this field, which plays a leading role in next-generation Data Warehousing and OLAP research. [6] have characterized four requirements for the data placement structure: (1) fast data loading, (2) fast query processing, (3) highly efficient storage space utilization, and (4) strong adaptively to highly dynamic workload patterns. they have examined three commonly accepted data placement structures in conventional databases, namely row- stores, column-stores, and hybrid-stores in the context of large data analysis using MapReduce. In this paper, they present a big data placement structure called RCFile (Record Columnar File) and its implementation in the Hadoop system.

Starfish, a self-tuning system for big data analytics is introduced in [7]. Starfish builds on Hadoop while adapting to user needs and system workloads to provide good performance automatically, without any need for users to understand and manipulate the many tuning knobs in Hadoop. While Starfish's system architecture is guided by work on self-tuning database systems, they discussed how new analysis practices over big data pose new challenges; leading us to different design choices in Starfish. The approach in the paper enables Starfish to handle the significant interactions arising among choices made at different levels. In [8] authors found Facebook daily operation results, certain types of queries are executed at an unacceptable low speed by Hive (a production SQL-to-MapReduce translator). In this paper, they demonstrate that existing SQL-to-MapReduce translators that operate in a one-operation-to-one-job mode and do not consider query correlations cannot generate high-performance MapReduce programs for certain queries, due to the mismatch between complex SQL structures and simple MapReduce framework. They propose and develop a system called YSmart, a correlation aware SQL-to-MapReduce translator. YSmart applies a set of rules to use the minimal number of MapReduce jobs to execute multiple correlated operations in a complex query.

The article [9] discussed big data techniques and technologies, the transformative potential of big data in five domains. They focussed on data have swept into every industry and business function and are now an important factor of production, big data creates value in several ways. The book [10] is the culmination of five years' worth of in-memory research presenting: overview of their vision of how in-memory technology will change enterprise applications, technical foundations of in-memory data management, in-depth description of how we intend to realize our vision, resulting implications on the development and capabilities of enterprise applications. In [11] authors shown from an application perspective, many websites dedicated to social media are among the most popular Wikipedia (collective knowledge generation), MySpace and Facebook (social networking), YouTube (social networking and multi- media content sharing), Digg and Delicious (social browsing, news ranking, and bookmarking), Second Life (virtual reality), and Twitter (social networking and microblogging), becoming the source of Big-Data and how users, customers, volunteers getting interacted with these sites to make businesses. In [12], author report their survey on a selection of state-of-the- art VA systems as a basis for analysing current market and trend, discussing space for improvement and identifying future research directions. They evaluate the functionality and performance of each system by surveying the vendor with a structured questionnaire as well as testing with real world data and detailed findings and outline the main characteristic of each system. Their survey provides a comparative review of ten products on the market. We also investigate a larger number of systems, including Cognos, SQL Server BI, Business Objects, Teradata, PowerPivot, Panopticon, KNIME, Oculus, Palentir and in-Spire to gain a better overview of the VA software market.

The paper [13] is about how the SP theory of intelligence and its realization in the SP machine may, with advantage, be applied to the management and analysis of big data. The SP system introduced in this paper and fully described, may help to overcome the problem of variety in big data; it has potential as a universal framework for the representation and processing of diverse kinds of knowledge, helping to reduce the diversity of formalisms and formats for knowledge, and the different ways in which they are processed. It has strengths in the unsupervised learning or discovery of structure in data, in pattern recognition, in the parsing and production of natural language, in several kinds of reasoning, and more. In [14], authors have investigated the privacy challenges in the big data era by first identifying big data privacy requirements and then discussing

whether existing privacy-preserving techniques are sufficient for big data processing. They have also introduced an efficient and privacy-preserving cosine similarity computing protocol in response to the efficiency and privacy requirements of data mining in the big data era. Although they have analysed the privacy and efficiency challenges in general big data analytics to shed light on the privacy research in big data, significant research efforts should be further put into addressing unique privacy issues in some specific big data analytics.

In [15] described Smart Data that is realized by extracting value from Big Data, to benefit not just large companies but each individual. He considered If his child is an asthma patient, for all the data relevant to his child with the four V-challenges, what he care about is simply, "How is her current health, and what are the risk of having an asthma attack in her current situation (now and today), especially if that risk has changed?" As he shown, Smart Data that gives such personalized and actionable information will need to utilize metadata, use domain specific knowledge, employ semantics and intelligent processing, and go beyond traditional reliance on ML and NLP.

In [16], authors proposed architecture for real-time Big Data analysis for remote sensing application. The proposed architecture efficiently processed and analysed real-time and offline remote sensing Big Data for decision-making. The proposed architecture is composed of three major units, such as 1) RSDU; 2) DPU; and 3) DADU. These units implement algorithms for each level of the architecture depending on the required analysis. The architecture of real-time Big is generic (application independent) that is used for any type of remote sensing Big Data analysis. Furthermore, the capabilities of filtering, dividing, and parallel processing of only useful information are performed by discarding all other extra data. The proposed architecture welcomes researchers and organizations for any type of remote sensory Big Data analysis by developing algorithms for each level of the architecture depending on their analysis requirement.

Big Data Hadoop Android a Collaborative Approach

In this stage an in-depth analysis is performed to obtain a detailed understanding of the business needs as defined in the business case and scope documents. The challenges, features, requirements of big data platform and how important Big-Data analytics are? analysed in this paper. For the ease of use and for ease of business, analytics should be available of smart phone (Android). For solving this issue we have proposed logic. Contemporary mobile users show an increasing trend in consuming information as data analytics, with bigger screens and smarter visualization on mobile devices. Instead of showing the information on computers now it can be available on mobile phone. Applications running on a mobile device (Android, iOS) want to access file systems stored in the Hadoop server. For that, applications use the HTTP requests.

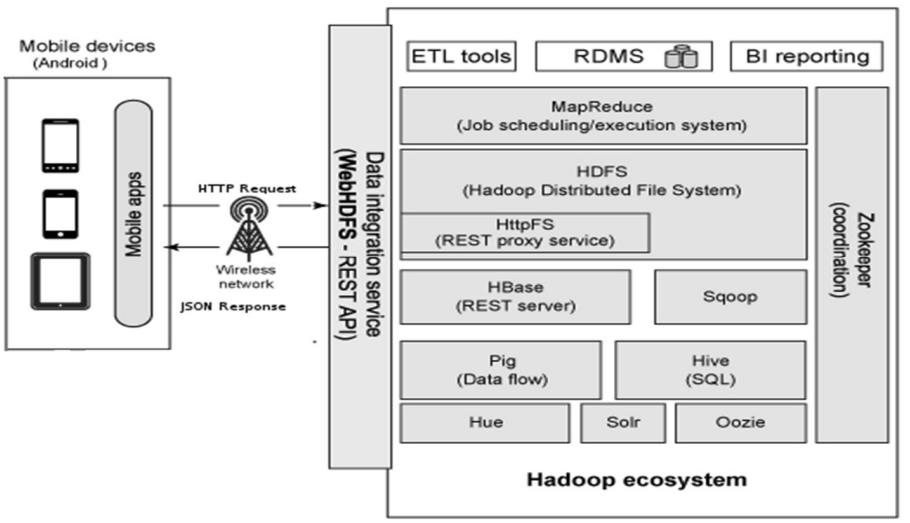


Fig 2. Big data analytics on Android platform

The procedure then fires a REST (JSON/HTTP) request to the web service running inside the Hadoop container system as MapReduce service. Upon receiving the REST request processing starts, server then returns data back in JSON format. Once the application retrieves data in JSON format then Application running on mobile device (Android, iOS) parse it and represent it in desired format. The approach utilized for analysing Big data on android platform is depicted in figure 2. In this paper an approach for showing Big-Data analytics on mobile devices running android OS is developed. The main objectives of this paper is addressing challenges, requirements, and importance of Big-data, application of hadoop for data analytics, utilization of JSON and web-services for Big-data analytics on Android mobile phone. The proposed work is implemented by using following steps:

Step 1: Copy input file- copy input file from local file system to Hadoop Distributed File System (HDFS) using the command *$ hdfs dfs -put /home/altaf/Desktop/MyHadoop/ finalin*

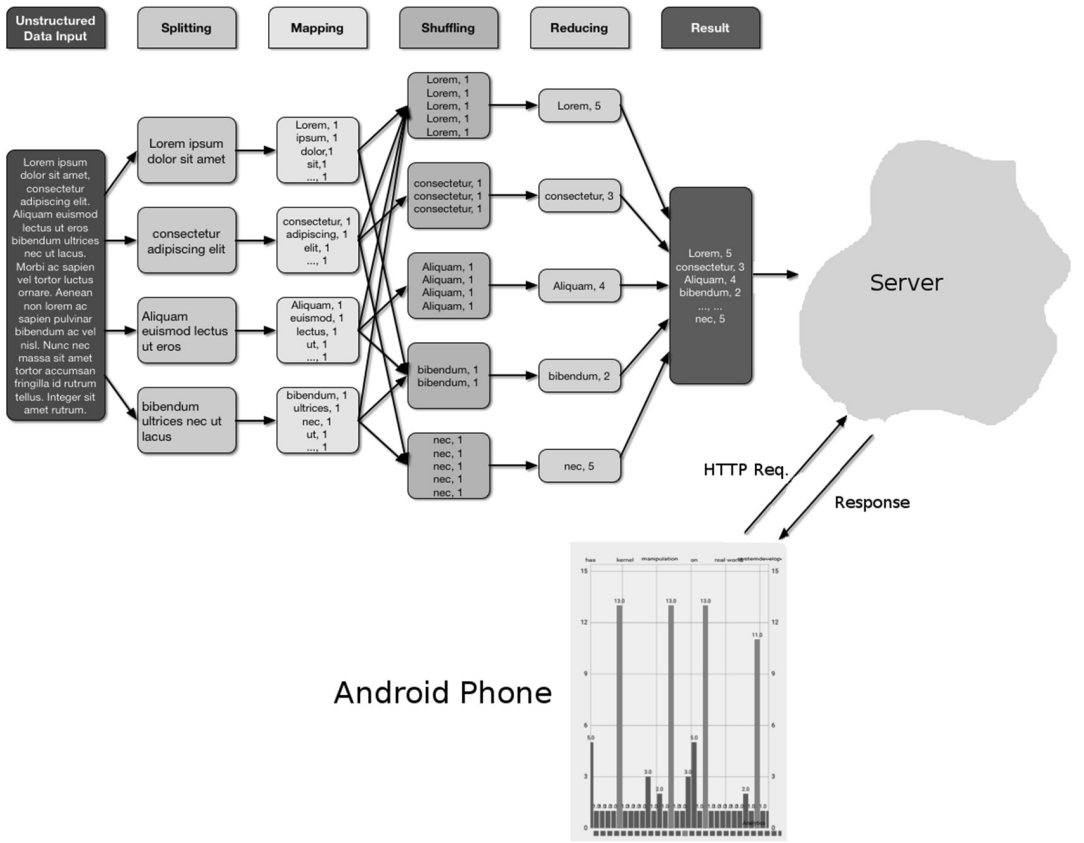


Fig 3. Proposed Big-Data Hadoop Android collaborative approach

Step 2: Execute the job- execute the job and obtain desired output which is going to be store in directory of HDFS. This is done by command *$ hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0.jar wordcount finalin/final final*

Step 3: Display output file- The output file can be displayed using command *$ hdfs dfs -cat finalop/**

Step 4: Copy output to local file system- After getting output in output file copy it from HDFS to Local file System using (Local Server) command $ hdfs dfs -copyToLocal finalop/* /var/www/html/xyz/

Step 5: Android User- Android makes the HTTP connection to local server and reads the copied file. The analytics are shown in the form of bar chart. If output file contain word whose count is greater than threshold value (11) then represent that word with red bar.

The overall implementation is carried out using Android framework, XML, PHP, and Hadoop. In the experimental controlled environment we have created a Mobile (Android) and Hadoop Application. The developed system contains two maor components: Hadoop Application and Android Application (for End User). Hadoop application which operates on data or dataset and perform number of operations like mapping, reducing, counting words, creating output file and so on. Android application shows the number of time word occurrence of word in file of text or in dataset. The complete process followed for the implementation of the proposed collaborative approach is shown in the figure 3.

Experimental Results

The word count operation takes place in two stages a mapper phase and a reducer phase. In mapper phase first the test is tokenized into words then we form a key value pair with these words where the key being the word itself and value '1'. In the reduce phase the keys are grouped together and the values for similar keys are added. So here, there is only one pair of similar keys 'tring' the values for these keys would be added.

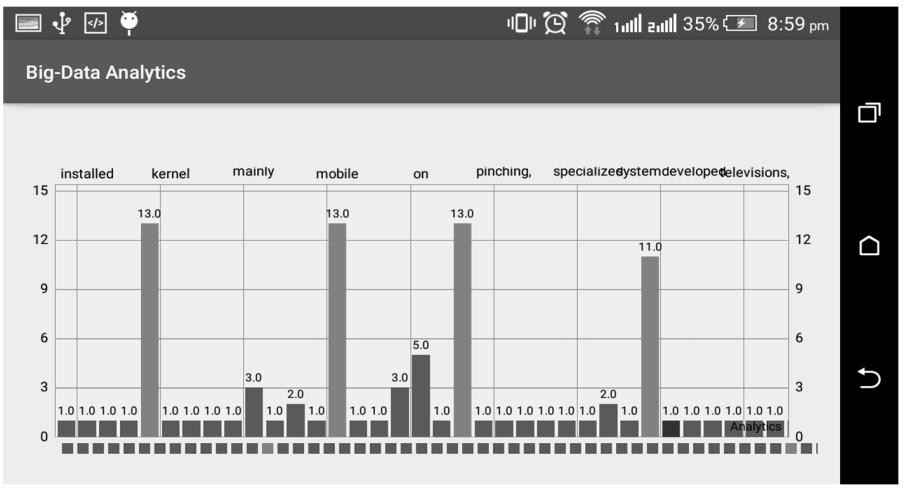


Fig 4. Big-data analytics obtained on Android 5.1.1

The point to be noted here is that first the mapper class executes completely on the entire data set splitting the words and forming the initial key value pairs. Only after this entire process is completed the reducer starts. Say if we have a total of 10 lines in our input files combined together, first the 10 lines are tokenized and key value pairs are formed in parallel, only after this the aggregation/ reducer would start its operation. Now coming to the practical side of implementation we need our input file and map reduce program jar to do the process job. In a common map reduce process two methods do the key job namely the map and reduce, the main method would trigger the map and reduce methods. After getting the number of occurrences of a word then output is uploaded to the server. The end user make request and get the analytics in the form of bar chart. Figure 4 shows the analytics obtained on the Android 5.1.1 in landscape view.

The analytics are easily available on users smartphone, because Android phones are very common. Enables access anywhere with a Mobile connection i.e. globalised the work. Faster, better decision making by observing chart. It enables access anywhere with a web connection i.e. globalised the work. Device and location independence: Enables users to access systems using a Mobile Phone regarding of their location. No need to buy updates or newer versions of software and hardware. Updating and managing software or applications i.e. cost can be reduce by spending on technology. No need to worry about your lot of data and files to store, this provides more data to save the files in server. Here depending upon the data and usage you can choose the plans. Everything is online, store your entire data in cloud and can access at any time in browser.

Conclusion and Future Scope

The proposed work strives to load the Big-Data analytics on Mobile phone (Android) which is done by using Hadoop and analytics are easily available to user with high throughput. In experimental platform we have implemented Mobile application (Android) and hadoop application and analytics are shown on a Android Application. Mobile has become one of the most active topics in today's world. In future work, we can use the Mobile as commodity hardware by enhancing the configuration of mobile like, RAM, secondary storage, etc. for Hadoop configuration and also we can extend the applications for machine learning.

References

[1] Rathod, H.; Chauhan, T.: A Survey on Big Data Analysis Techniques. 9. (2013).
[2] Mukherjee A., Datta J., Jorapur R., Singhvi R., Haloi S., Akram W.: Shared disk big data analytics with Apache Hadoop. Pp. 18-22, (2012).
[3] Xindong W., Xingquan Z., Gong-Qing W., Wei D.: Data Mining with Big Data. (2014).
[4] Cebr: Data equity Unlocking the value of big data. SAS Reports, pp. 1–44, (2012).
[5] Cuzzocrea A., Song I., Davis K.C.: Analytics over Large-Scale Multidimensional Data: The Big Data Revolution! in Proceedings of the ACM International Workshop on Data Warehousing and OLAP, pp. 101–104 (2011).
[6] He Y., Lee R., Huai Y., Shao Z., Jain N., Zhang X., Xu Z.: RCFile: A Fast and Space-efficient Data Placement Structure in MapReduce-based Warehouse Systems. IEEE International Conference on Data Engineering (ICDE), pp. 1199–1208, (2011)
[7] Herodotou H., Lim H., Luo G., Borisov N., Dong L., Cetin F.B., Babu, S.: Starfish: A Self-tuning System for Big Data Analytics. In Proceedings of the Conference on Innovative Data Systems Research, pp. 261–272 (2011).
[8] Lee R., Luo T., Huai Y., Wang F., He Y., Zhang X.: Ysmart: Yet Another SQL-to-MapReduce Translator. In IEEE International Conference on Distributed Computing Systems (ICDCS), pp. 25–36 (2011).

[9] Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Byers A. H.: Big Data: The Next Frontier for Innovation, Competition, and Productivity. In: McKinsey Global Institute Reports, pp. 1–156 (2011)
[10] Plattner, H., Zeier, A.: In-Memory Data Management: An Inflection Point for Enterprise Applications. Springer, Heidelberg (2011)
[11] Zeng D., Hsinchun C., Lusch R., Li S.H.: Social Media Analytics and Intelligence. IEEE Intelligent Systems 25(6), 13–16 (2010).
[12] Zhang L., Stoffel A., Behrisch M., Mittelstadt S., Schreck T., Pompl R., Weber S., Last H., Keim D.: Visual Analytics for the Big Data Era—A Comparative Review of State-of-the-Art Commercial Systems. in IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 173–182 (2012).
[13] Gerard J.W.: Big Data and the SP Theory of Intelligence. IEEE Translations on Content Mining, (2014)
[14] Rongxing Lu, Hui Zhu, Ximeng Liu, Joseph K. Liu, and Jun Shao: Toward Efficient and Privacy-Preserving Computing in Big Data Era, IEEE Network, (2014)
[15] Amit Sheth, Knoesis, : Smart Data - How you and I will exploit Big Data for personalized digital health and many other activities. 2014 IEEE International Conference on Big Data.
[16] Muhammad Mazhar, Paul Anand, Ahmad Awais, Chen Bo-Wei, Huang Bormin, Ji Wen: Real-Time Big Data Analytical Architecture for Remote Sensing Application. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. (2014).
[17] Badase P. S., Deshbhratar G. P., Bhagat A. P.: Classification and analysis of clustering algorithms for large datasets. in Proceedings International Conference on Innovations in Information, Embedded and Communication Systems. (2015).

Adaptive Dynamic Surface Control of Chaotic Micro-Electro-Mechanical System with Unknown System Parameters and Dead-Zone Input

Shaohua Luo*, Suqun Cao**, Zhong Chen*** and K. Sujatha****
*The Jiangsu Key Laboratory of Advanced Manufacturing Technology, Huaiyin Institute of Technology, Huai'an 223003, China; School of Automation, Chongqing University, Chongqing 400044, China
hua66com@163.com
** The affiliated Huai'an Hospital of Xuzhou Medical College, Huai'an 223003, China ; School of Computer Science and Engineering, Southeast University, Nanjing 211189, China; Faculty of Mechanical and Material Engineering, Huaiyin Institute of Technology, Huai'an 223003, China
caosuqun@126.com
***The Jiangsu Key Laboratory of Advanced Manufacturing Technology, Huaiyin Institute of Technology, Huai'an 223003, China
chenzhong2006@126.com
****EEE Department, Center for Electronics, Automation and Industrial Research (CEAIR), Dr. M.G.R. Educational and Research Institute, Maduravoyal, Chennai - 600095, Tamil Nadu, India
drksujatha23@gmail.com

Abstract: This paper focuses on chaos control of the micro-electro-mechanical system with unknown system parameters and dead-zone input existing in the engineering application. The phase diagrams, corresponding time histories and bifurcation diagram are employed to reveal the chaotic dynamics performance of the micro-electro-mechanical system. For eliminating chaos and vibration, an adaptive neural-network-based dynamic surface control is proposed to convert the chaos motion into regular motion without imposing any condition on parameters of system model and the boundedness of control gain. Meanwhile, to achieve high accuracy and quick response, a neural network is employed to approximate unknown nonlinear item of model and an adaptive law is designed to estimate unknown control gain in the framework of dynamic surface control. Finally, some simulations are executed and corresponding results show effectiveness and robustness of the proposed scheme.

Keywords: Micro-electro-mechanical system; Adaptive dynamic surface control; Chaos control; Neural network; Dead-zone input.

Introduction

The research aimed at micro-electro-mechanical system has garnered significant attention recently because of its advantages. A lot of researches have been carried out on the behaviors and modeling. Though some achievements about micro-electro-mechanical system have been made, there still exists some challenges associated with it [1-3]. A key problem is how to effectively control chaotic behavior of micro-electro-mechanical system with unknown system parameters and dead-zone input. Chaos phenomenon with geometrical strangeness usually leads to deterioration of the system performance.

To suppress chaos in the motion process of micro-electro- mechanical system, various methods and analysis results have been reported recently [4, 5]. The OGY is a basic method for suppressing chaos behavior [6]. But it is difficult to choose a reasonable parameter in real practice. Chaos control using the time-delay feedback control method is introduced to the engineering applications [7]. But it is applied under strict restriction that the control objective must be the equilibrium. For suppressing the chaos motion, the robust fuzzy sliding mode controller is designed [8]. This approach uses fuzzy logic linguistic rules to generate a suitable chatter-free control signal for driving the error dynamic system and ensures the track error to converge asymptotically to zero. The drawback of this approach needs to obtain precision model in the process of design. However, the parameters of system and actuator dead-zone input are hard to do a high-precious measurement because parameters are influenced by temperature and material wear.

Sliding mode control (SMC) is recognized as a useful and effective approach, which deals with time varying properties, uncertainties and bounded external disturbances [9, 10]. However, the chattering associated with SMC is a serious impediment for engineering application. In view of this, a novel second-order fast terminal sliding mode control is proposed to suppress the random motion of the micro-electro-mechanical system with system uncertainty and external disturbance[11]. The dynamic surface control (DSC) is proposed by introducing a first-order filtering of the synthetic input at each step of the backstepping,

so the differentiation items on the virtual function can be thoroughly eliminated [12]. An adaptive dynamic surface control combined with SMC to compensate for friction and backlash nonlinearities in a motion system is designed, wherein the updated laws of the recurrent wavelet neural networks and friction estimation are derived to approximate and compensate for the backlash and friction nonlinearities [13]. However, the explicit consideration of input constraint within the framework of DSC have not received attention. Furthermore, the control gain which is usually unknown in real environment is supposed to a bounded constant in previous works. To solve it, the Nussbaum gain is adopted to deal with unknown sign on the problem of input uncertainty [14, 15]. Then, the unknown gain can be effectively estimated in finite time. But the drawback is the complicated calculation procedures and time consuming.

Dead-zone input usually exists in the control input because of the actuator's physical limitations [16, 17]. This nonlinearity can cause oscillation and then deteriorate the micro-electro- mechanical system performance. An adaptive control method based on neural network is presented to control a direct current motor system with dead-zone characteristics, wherein neural networks are adopted to accomplish traditional identification [18]. Unfortunately, this research result is only limited to symmetric dead-zone input.

Many researchers have achieved considerable progress in chaos of the micro-electro-mechanical system, but chaos control remains to face new challenges. In this paper, an adaptive dynamic surface control combining with neural network is designed to apply on tracking control for the micro-electro-mechanical system with unknown system parameters and dead-zone input. First, the scheme is designed by Lyapunov stability theory, which can guarantee stabilization of the closed-loop error system. Second, in the recursive process of DSC, a neural network is employed to approximate unknown nonlinear item of math model which reduces the requirement about precise parameters. It not only improves tracking accuracy but also obtains a smooth control input without high-frequent chattering phenomena. Moreover, stability analysis is carried out so that the error converges to a small neighborhood of the origin. Finally, numerical simulation results show a satisfactory performance.

System description

System model

The schematic diagram of the micro-electro-mechanical system under the combined DC and AC actuation voltages is depicted in Fig.1. An external driving force on the resonator is implemented by using an electrical driving voltage that leads to electrostatic excitation between electrodes and resonator.

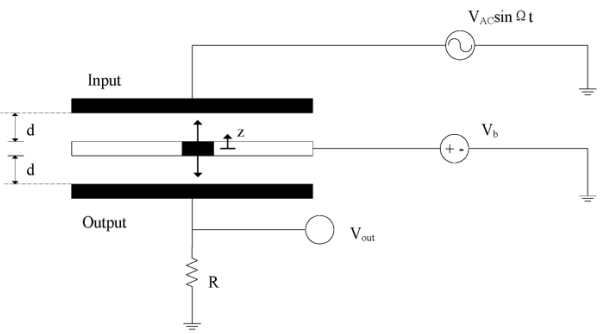


Fig 1.Schematic diagram of the micro-electro-mechanical system

For establishing dynamic model, it makes assumptions that the amplitude of the AC driving voltage is much lower than the bias voltage. Then, the math model of chaotic micro-electro- mechanical system can be defined as follow [5].

$$\ddot{x}+\mu\dot{x}+\alpha x+\beta x^3=\gamma\left(\frac{1}{\left(1-x\right)^2}-\frac{1}{\left(1+x\right)^2}\right)+\frac{A}{\left(1-x\right)^2}\sin\left(\omega\tau\right) \tag{1}$$

where the non-dimensional variables x and ω are defined as

$$x=\frac{z}{d},\omega=\frac{\Omega}{\omega_0},\ A=2\gamma\frac{V_{AC}}{V_b}$$

where d is the initial width of the gap, z is the vertical displacement of the beam midpoint, Ω is the frequency of AC voltage, ω_0 is the natural frequency, V_{AC} is the AC amplitude, V_b is the bias voltage.

For simplicity, the following notations are employed:

$$x_1 = x, x_2 = \dot{x}, G(x) = \gamma\left(\frac{1}{(1-x)^2} - \frac{1}{(1+x)^2}\right) \tag{2}$$

Substituting the notations into (1) yields the following nominal form with non-symmetric dead-zone input.

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\mu x_2 - \alpha x_1 - \beta x_1^3 + G + \dfrac{A}{(1-x_1)^2}\sin(\omega\tau) + \Gamma(u) \end{cases} \tag{3}$$

The micro-electro-mechanical system has been studies for V_{AC} in $(0, 0.47)$ and constant values of $\alpha = 1$, $\beta = 12$, $\gamma = 0.338$, $\mu = 0.01$, $V_b = 3.8$ and $\omega = 0.5$. The phase diagrams and corresponding time histories are shown in Fig. 2 in given initial conditions like $(x_1, x_2) = (0, 0)$ and the fixed bias voltage. Beginning at the neighboring value of zero, the transient chaos and regular motion around the center points are shown in Fig. 2(a). Obviously, the more increase in AC voltage gives rise to longer transient and random vibration. As can be seen from the Fig. 2(c), after the transient chaotic response, regular motion can come into being in homoclinic orbit and the amplitude of harmonic oscillation is much larger comparing to the case in Fig. 2(a)-(b). Fig. 3 shows the bifurcation diagram. In the case, the qualitative behavior of the micro-electro-mechanical system is concluded against a varying voltage from 0 to 0.4. Along with increasement of AC voltage, regular motion appears around one of the center points.

Actuator dead-zone input

Actuator dead-zone input which is considered as a complex nonlinearity phenomenon occurs in the micro-electro-mechanical system. The phenomenon usually leads to oscillatory activity and deteriorates the system performance. Thus, it is necessary to eliminate it.

The non-symmetric dead-zone of actuator can be written as a combination of a line and a disturbance-like term [19]

$$\Gamma(u) = m(t)u + d_1(t) \tag{4}$$

where $m(t) = \begin{cases} m_l, u \le 0 \\ m_r, u > 0 \end{cases}, d_1(t) = \begin{cases} -m_r b_r, u \ge b_r \\ -m(t)u, \ -b_l < u < b_r \\ m_l b_l, \ u \le -b_l \end{cases}$, m_r and m_l stand for the right and left slopes of the dead-zone characteristic, respectively, b_r and b_l denote the breakpoint of the actuator dead-zone input, respectively.

Chaos controller

In this section, an adaptive neural-network-based dynamic surface control method is used to stabilize the micro-electro-mechanical system with non-symmetric dead-zone input in a high amplitude oscillation state. The proposed scheme can easily accommodate change and has strong robustness in the face of dynamic uncertainties.

The boundary layer error is given in the first place as follow

$$y_2 = \alpha_{2f} - \alpha_2 \tag{5}$$

where α_{2f} is the output of the first-order filter, α_2 is the virtual control input.

Then, for any given x_{1d}, the dynamic surfaces are generally taken to be

$$S_1 = x_1 - x_{1d}, S_2 = x_2 - \alpha_{2f} \tag{6}$$

RBF neural network

The RBF neural network is universal approximator [20]. It approximates any smooth function $f_n(X): R^n \to R$.

$$f_n(X) = \theta'^T \xi(X) \tag{7}$$

where $X \in D \subset R^n$ is the input vector, $\theta' = \left[\theta'_1, \theta'_2, \cdots, \theta'_l\right]^T \in R^l$ is the weight vector, $l > 1$ is the node number of neuron, and $\xi(X) = [\xi_1(X), \xi_2(X), \cdots, \xi_l(X)]^T \in R^l$ is a basic function vector, with $\xi_i(X)$ being chosen as the commonly used Gaussian functions, which have the following form:

$$\xi_i\left(X\right)=\exp\left(-\left(X-\mu_i\right)^T\left(X-\mu_i\right)/2\sigma_i^2\right),i=1,2,\cdots,l \tag{8}$$

where $\mu_i=\left[\mu_{i1},\mu_{i2},\cdots,\mu_{in}\right]^T$ is the center of the receptive field and σ_i is the width of the Gaussian function.

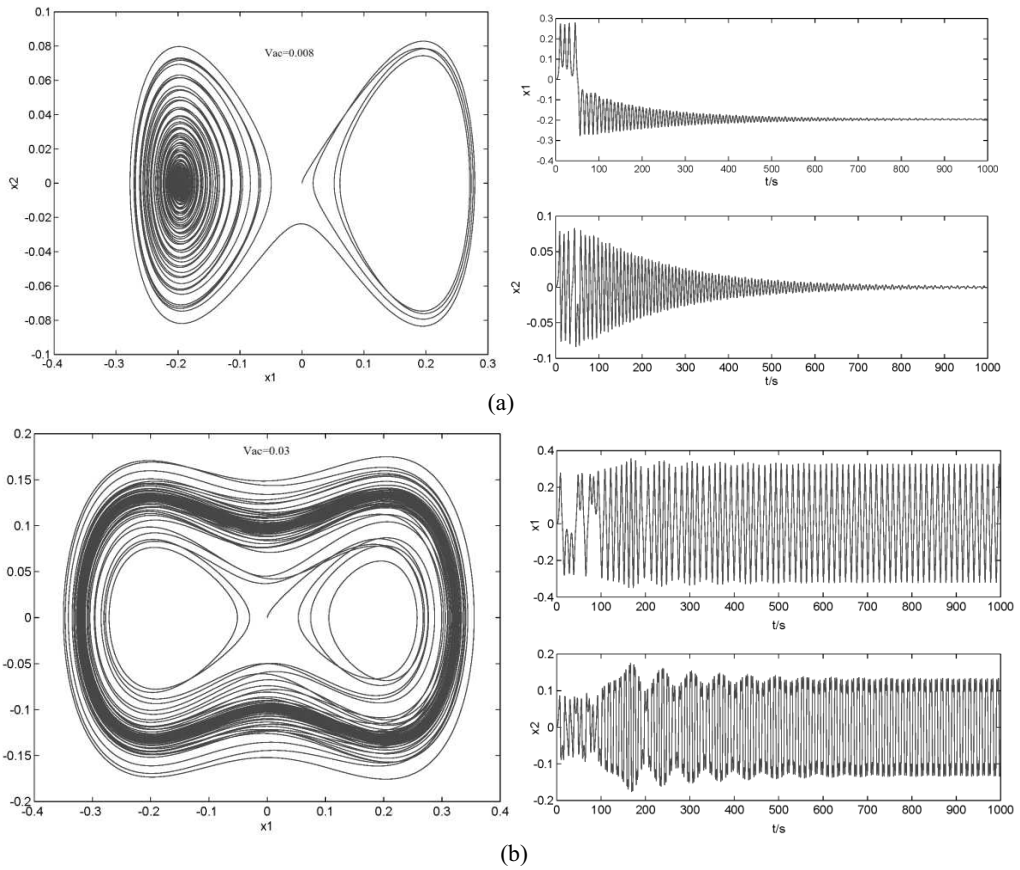


Fig 2.Phase diagrams and corresponding time histories

Due to the approximation capability, the nonlinear term can be approximated as

$$f\left(X\right)=\theta^{*T}\xi\left(X\right)+\varepsilon \tag{9}$$

where ε is the approximation error, the optimal parameter vector θ^* is bounded and defined as

$$\theta^*=\arg\min_{\theta\in\Omega}\left\{\sup_{X\in D}\left|f\left(X\right)-\theta^{'T}\xi\left(X\right)\right|\right\} \tag{10}$$

where Ω is the compact region for $\theta^{'}$ .There exists known constants ε_0 such that $0<\left|\varepsilon\right|\le\varepsilon_0$.

Controller design

Step 1: Let the Lyapunov function of system be defined as

$$V_1=\frac{1}{2}S_1^2 \tag{11}$$

Then, taking the time derivative of V_1 along the trajectory (6), it follows that

$$\dot{V}_1=S_1\left(S_2+y_2+\alpha_2-\dot{x}_{1d}\right) \tag{12}$$

Then, the virtual control input is defined by the following form:

$$\alpha_2=-k_1S_1+\dot{x}_{1d} \tag{13}$$

where k_1 is the design constant.

Substituting (13) into (12), $\dot{V}_1$ is rewritten as

$$\dot{V}_1 \le \left(1-k_1\right)S_1^2 + S_1S_2 + \frac{1}{4}y_2^2 \tag{14}$$

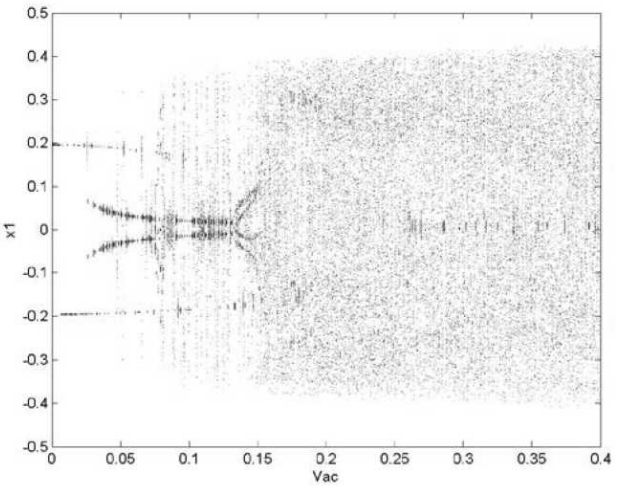


Fig 3.The bifurcation diagram

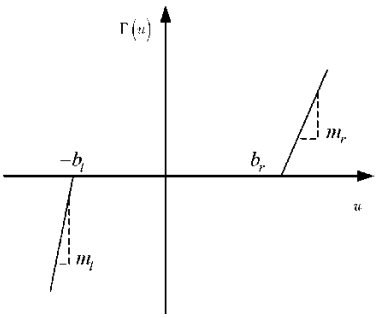


Fig 4. Non-symmetric dead-zone characteristic

Step 2: Filtering α_2 through the first-order filter yields

$$\varpi_2\dot{\alpha}_{2f} + \alpha_{2f} = \alpha_2,\ \alpha_{2f}(0) = \alpha_2(0) \tag{15}$$

where ϖ_2 is a time constant.

From (5) and (15), the derivative of α_{2f} is given by

$$\dot{\alpha}_{2f} = -\frac{y_2}{\varpi_2} \tag{16}$$

Then, the derivative of y_2 can be obtained

$$\left|\dot{y}_2 + \frac{y_2}{\varpi_2}\right| \le B_2\left(\cdot\right) \tag{17}$$

According to Young's inequality, there exists

$$y_2\dot{y}_2 \le -\frac{y_2^2}{\varpi_2} + y_2^2 + \frac{1}{4}B_2^2 \tag{18}$$

where B_2 is a continuous function.

Introduce the variables as

$$\tilde{\lambda}_2 = \hat{\lambda}_2 - \lambda_2,\ \tilde{g}_2 = \hat{g}_2 - g_2 \tag{19}$$

where $\hat{\lambda}_2$ and $\hat{g}_2$ mean the estimation of λ_2 and g_2.

Choose the Lyapunov function candidate

$$V_2 = V_1 + \frac{1}{2}S_2^2 + \frac{1}{2}y_2^2 + \frac{1}{2\gamma_2}\tilde{\lambda}_2^2 + \frac{1}{2\Gamma_2}\tilde{g}_2^2 \tag{20}$$

where γ_2 and Γ_2 are the design constant of controller.

Then, it is easy to obtain

$$\dot{V}_2 \le \dot{V}_1 + S_2\left(f_2(\cdot) + g_2u - \dot{a}_{2f}\right) - \frac{y_2^2}{\varpi_2} + y_2^2 + \frac{1}{4}B_2^2 + \frac{1}{\gamma_2}\tilde{\lambda}_2\dot{\hat{\lambda}}_2 + \frac{1}{\Gamma_2}\tilde{g}_2\dot{\hat{g}}_2 \tag{21}$$

where $f_2(\cdot) = -\mu x_2 - \alpha x_1 - \beta x_1^3 + G(x_1) + \dfrac{A}{(1-x_1)^2}\sin(\omega\tau) + d_1,\ g_2 = m$.

In the engineering application, precise measuring for the parameters $(\mu, \alpha, \beta, \gamma, \omega, m,\ d_1)$ of system and actuator dead-zone input becomes difficult because of effect of temperature and material wear, etc. To solve the problem, employ the neural network to approximate the nonlinear coupling function $f_2(\cdot)$.

Therefore, for any given $\varepsilon_2 > 0$, there exists a neural network $\theta_2^{*T}\xi_2$ such that

$$f_2(\cdot) = \theta_2^T\xi_2 + \varepsilon_2 \tag{22}$$

where $\theta_2 = \theta_2^*$.

According to Young's inequality, substituting (14) and (22) into (21) yields

$$\begin{aligned}\dot{V}_2 \le{}& S_2\left(\frac{1}{2a_2^2}\lambda_2S_2\xi_2^T\xi_2 + g_2u - \dot{a}_{2f} + S_1 + \frac{1}{2}S_2\right) - \frac{y_2^2}{\varpi_2} + \frac{5}{4}y_2^2 + \frac{1}{4}B_2^2 + \frac{1}{\gamma_2}\tilde{\lambda}_2\dot{\hat{\lambda}}_2 + \frac{1}{\Gamma_2}\tilde{g}_2\dot{\hat{g}}_2 \\ &+ \frac{a_2^2}{2} + (1-k_1)S_1^2 + \frac{1}{2}\varepsilon_{20}^2\end{aligned} \tag{23}$$

where a_2 is the design constant.

The actual control law and adaptive laws are given by

$$u = \frac{\hat{g}_2}{\hat{g}_2^2 + \eta_2}\left(-\left(\frac{1}{2} + k_2\right)S_2 - S_1 - \frac{1}{2a_2^2}\hat{\lambda}_2S_2\xi_2^T\xi_2 + \dot{\alpha}_{2f}\right) \tag{24}$$

$$\dot{\hat{\lambda}}_2 = \frac{1}{2a_2^2}\gamma_2\xi_2^T\xi_2S_2^2 - m_2\hat{\lambda}_2 \tag{25}$$

$$\dot{\hat{g}}_2 = \Gamma_2\left(S_2u - c_2\hat{g}_2\right) \tag{26}$$

where k_2, m_2 and c_2 are the design constant, η_2 is a small positive constant, and $\hat{\lambda}_2 = \left\|\hat{\theta}_2\right\|^2$.

In addition, the inequalities $-\dfrac{m_2}{\gamma_2}\hat{\lambda}_2\tilde{\lambda}_2 \le -\dfrac{m_2}{2\gamma_2}\left|\tilde{\lambda}_2\right|^2 + \dfrac{m_2}{2\gamma_2}\left|\lambda_2\right|^2$ and $-c_2\tilde{g}_2\hat{g}_2 \le -\dfrac{c_2}{2}\left|\tilde{g}_2\right|^2 + \dfrac{c_2}{2}\left|g_2\right|^2$ are used here.

Therefore, using (24)-(26), it has

$$\begin{aligned}\dot{V}_2 &\le \frac{-\eta_2}{\hat{g}_2^2+\eta_2}\left(-\left(\frac{1}{2}+k_2\right)S_2 - S_1 - \frac{1}{2a_2^2}\hat{\lambda}_2 S_2\xi_2^T\xi_2 + \dot{\alpha}_{2f}\right)S_2 - \frac{1}{2a_2^2}\tilde{\lambda}_2 S_2^2\xi_2^T\xi_2 - k_2 S_2^2 - \tilde{g}_2 u S_2 - \frac{y_2^2}{\varpi_2} \\ &\quad + \frac{5}{4}y_2^2 + \frac{1}{4}B_2^2 + \frac{1}{\gamma_2}\tilde{\lambda}_2\dot{\hat{\lambda}}_2 + \frac{1}{\Gamma_2}\tilde{g}_2\dot{\hat{g}}_2 + \left(1-k_1\right)S_1^2 + \frac{a_2^2}{2} + \frac{1}{2}\varepsilon_{20}^2 \\ &\le \left(1-k_1\right)S_1^2 + \left(\frac{1}{2}-k_2\right)S_2^2 + \left(\frac{5}{4}-\frac{1}{\varpi_2}\right)y_2^2 + \frac{1}{4}B_2^2 - \frac{m_2}{2\gamma_2}\left|\tilde{\lambda}_2\right|^2 - \frac{c_2}{2}\left|\tilde{g}_2\right|^2 + \frac{m_2}{2\gamma_2}\left|\lambda_2\right|^2 + \frac{c_2}{2}\left|g_2\right|^2 \\ &\quad + \frac{a_2^2}{2} + \frac{1}{2}\left(\varepsilon_{20}^2+\delta_2^2\right)\end{aligned} \tag{27}$$

where δ_2 is the continuous function and satisfies $\frac{-\eta_2}{\hat{g}_2^2+\eta_2}\left(-S_1 - \left(\frac{1}{2}+k_2\right)S_2 - \frac{1}{2a_2^2}\hat{\lambda}_2 S_2\xi_2^T\xi_2 + \dot{\alpha}_{2f}\right)$ $\le \delta_2\left(S_1, S_2, y_2, \hat{\theta}_2, \hat{g}_2, x_{1d}, \dot{x}_{1d}\right)$.

In order to illustrate the proposed scheme clearly, the schematic block diagram of the micro-electro-mechanical system with unknown dead-zone input is shown in Fig.5.

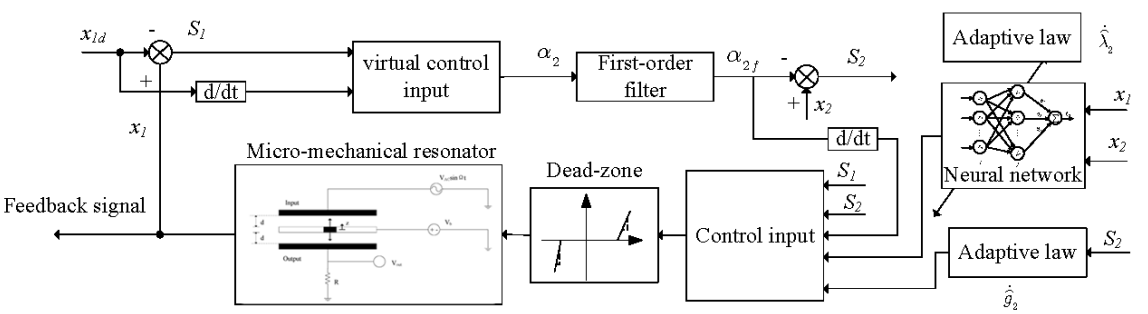


Fig 5. Schematic block diagram of the micro-electro-mechanical system

Stability analysis

Theorem 1: Suppose that chaos controller (24) with updated laws (25) and (26) is used to reduce the trajectory tracking error of the micro-electro-mechanical system with unknown dead-zone input described by (4), by selecting the reasonable parameters as $k_1, k_2, a_2, \gamma_2, m_2, \Gamma_2, \eta_2, \varpi_2, c_2$, then the closed-loop control system is uniformly ultimately bounded, and S_1 converges to a vicinity of zero when the initial condition satisfies $\sum_{i=1}^{2} S_i^2 + \frac{1}{\gamma_2}\tilde{\lambda}_2^2 + \frac{1}{\Gamma_1}\tilde{g}_2^2 + y_2^2 \le 2p$ for any given $p > 0$.

Proof: The derivative of this system with respect to time can be written as

$$\begin{aligned}\dot{V} = \dot{V}_2 &\le \left(1-k_1\right)S_1^2 + \left(\frac{1}{2}-k_2\right)S_2^2 + \left(\frac{5}{4}-\frac{1}{\varpi_2}\right)y_2^2 - \frac{m_2}{2\gamma_2}\tilde{\lambda}_2^2 - \frac{c_2}{2}\tilde{g}_2^2 + b_0 \\ &\le -a_0 V + b_0\end{aligned} \tag{28}$$

where $b_0 = \frac{1}{4}B_2^2 + \frac{m_2}{2\gamma_2}\left|\lambda_2\right|^2 + \frac{c_2}{2}\left|g_2\right|^2 + \frac{a_2^2}{2} + \frac{1}{2}\left(\varepsilon_{20}^2+\delta_2^2\right), a_0 > \frac{b_0}{p}$.

Finally, (28) implies that

$$0 \le V(t) \le \frac{b_0}{a_0} + \left(V(t_0) - \frac{b_0}{a_0}\right)e^{-a_0\left(t-t_0\right)} \le \frac{b_0}{a_0} + V(t_0) \tag{29}$$

Numerical Simulation

In this section, the numerical simulation demonstrates the effectiveness of the proposed scheme in suppressing the oscillatory and chaos motion of the micro-electro-mechanical system.

The initial conditions are set as $x_1(0) = 0.3$, $x_2(0) = 0.1$. The given reference signal is given as $x_{1d} = 0.07\sin(3t) + 0.08\cos(2t)$. The parameters of the controller are selected as $k_1 = 15$, $k_2 = 35$, $\gamma_2 = 0.5$, $m_2 = 0.3$, $c_2 = 15$, $a_2 = 30$, $\Gamma_2 = 0.04$ and $\eta_2 = 0.001$. The initial values of estimates are set to be $\hat{g}_2(0) = 0.1$, $\hat{\lambda}_2(0) = 0.3$, and the first-order filter constant is selected as $\varpi_2 = 0.01$. In addition, the RBF neural network are chosen in this way. The center of neural network μ_i is uniformly distributed in the field of [-5,5], and its width σ_i is equal to 2.

The unknown dead-zone input-output characteristic which appears at the sixth second is defined as

$$\Gamma(u) = \begin{cases} 0.9(u + 0.6) & \text{if } u \le -0.6 \\ 0 & \text{if } -0.6 < u < 0.4 \\ 1.1(u - 0.4) & \text{if } u \ge 0.4 \end{cases}.$$

The simulation results of the state response are displayed in Fig.6. Four kinds of curves basically overlap with deadzone on the whole time. Fig.7 shows that the tracking errors between actual signal and desired signal are equal to $\pm 10^{-3}$ for different value of AC voltage. Obviously, the system state x_1 can track the given smooth reference signal x_{1d} precisely.

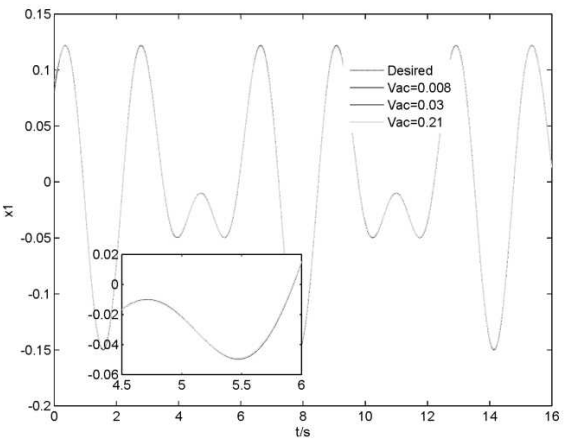


Fig 6. Tracking performance with varying AC voltage

Comparing with phase trajectory in Fig. 2(a-c), the effectiveness and feasibility of the proposed control scheme in suppress the chaos motion can be demonstrated in Fig. 8.

The curves of the actual control are shown in Fig.9. It should be noticed that a key role of designing controller is to try to avoid vibration of control input effectively within the threshold of the non-symmetric dead-zone. This means that the micro-electro- mechanical system possesses excellent tracking performance, and the chattering phenomenon of the controller is greatly weakened.

To further illustrate the changes of the non-symmetric input model parameters which do not influence the performance of the micro-electro-mechanical system, another simulation by changing the model parameters is executed at the sixth second as

$$\Gamma(u) = \begin{cases} 1.1(u + 0.7) & \text{if } u \le -0.7 \\ 0 & \text{if } -0.7 < u < 0.3 \\ 0.9(u - 0.3) & \text{if } u \ge 0.3 \end{cases}$$

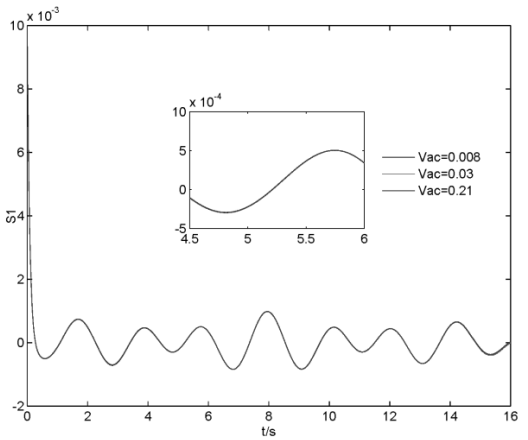


Fig 7. Tracking error with varying AC voltage

The simulation results are shown in Fig.10. Obviously, the closed-loop system still has excellent robustness. It can be known that vibration phenomenon does not appear in the actual control within the threshold of the non-symmetric dead-zone. Additionally, the control input is chatter-free even if the overall system confronts with uncertainty and varying AC voltage.

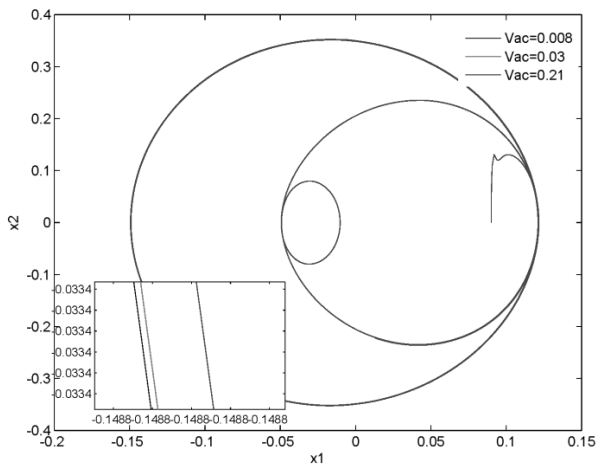


Fig 8.Phase trajectory with varying AC voltage

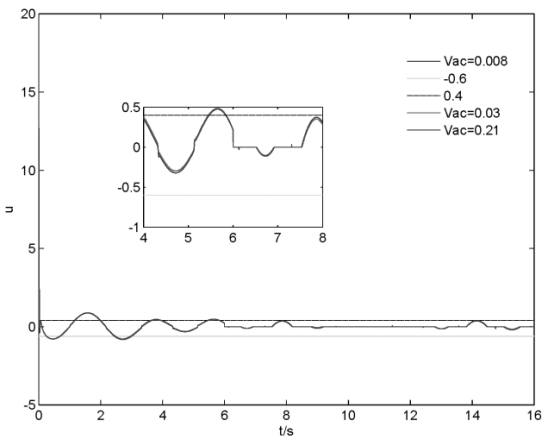


Fig 9. The curve of the actual control with varying AC voltage

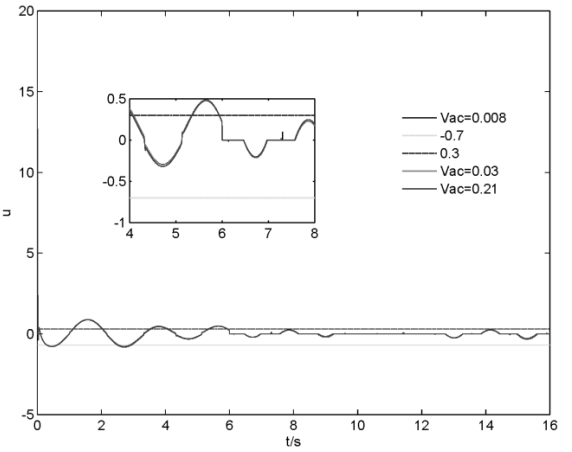


Fig 10. The curve of the actual control with varying AC voltage

Conclusion

This paper discusses adaptive dynamic surface control of chaotic micro-electro-mechanical system because of nonlinear coupling between the electrostatic force and resonator deflection. The transient chaos behavior accompanying with increase of AC voltage can destroy system stability disastrously. An adaptive dynamic surface approach via neural network for eliminating and stabilizing the chaotic motion is developed to oblige system state to approximate a reference signal with small error and compensate parameters variation in the presence of non-symmetric dead-zone input. The presented scheme can guarantee the closed-loop error system is stable in the sense of uniform ultimate boundedness. The more precise parameters of system model and the boundedness hypothesis of control gain which are available in advance in previous works can be cancelled automatically. Finally, numerical simulations for the micro-electro-mechanical system are given to demonstrate the effectiveness of the presented scheme.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 51505170), China Postdoctoral Science Foundation funded project (No. 2015M572445), Chongqing Postdoctoral Science Foundation funded project (No. Xm2015017) and Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No.14KJB460002). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers and editors, which have greatly improved the presentation.

References

[1] M. S. Siewe and U. H. Hegazy, "Homoclinic bifurcation and chaos control in MEMS resonators," *Applied Mathematical Modelling,* **35**, 5533–5552 (2011).

[2] E. M. Miandoab, A. Yousefi-Koma, H. N. Pishkenari, and F. Tajaddodianfar, "Study of nonlinear dynamics and chaos in MEMS/NEMS resonators," *Communications in Nonlinear Science & Numerical Simulation,* **22**, 611–622 (2015).

[3] J. Sharma and D. Grover, "Thermoelastic vibrations in micro-/nano-scale beam resonators with voids," *Journal of Sound and Vibration,* **330**, 2964-2977 (2011).

[4] E. M. Miandoab, H. N. Pishkenari, A. Yousefi-Koma, and F. Tajaddodianfar, "Chaos prediction in MEMS-NEMS resonators," *International Journal of Engineering Science,* **82**, 74–83 (2014).

[5] H. S. Haghighi and A. H. D. Markazi, "Chaos prediction and control in MEMS resonators," *Communications in Nonlinear Science & Numerical Simulation,* **15**, 3091–3099 (2010).

[6] M. F. Danca, "Random parameter-switching synthesis of a class of hyperbolic attractors," *Chaos,* **18**, 033111 Sep (2008).

[7] H. Ma, V. Deshmukh, E. Butcher, and V. Averina, "Delayed state feedback and chaos control for time-periodic systems via a symbolic approach," *Communications in Nonlinear Science & Numerical Simulation,* **10**, 479-497 (2005).

[8] H. T. Yau, C. C. Wang, C. T. Hsieh, and C. C. Cho, "Nonlinear analysis and control of the uncertain micro-electro-mechanical system by using a fuzzy sliding mode control design," *Computers & Mathematics with Applications,* **61**, 1912–1916 (2011).

[9] M. Roopaei and M. Z. Jahromi, "Synchronization of two different chaotic systems using novel adaptive fuzzy sliding mode control," *Chaos,* **18**, 033133 (2008).

[10] Bagheri and J. J. Moghaddam, "Simulation and tracking control based on neural-network strategy and sliding-mode control for underwater remotely operated vehicle," *Neurocomputing,* **72**, 1934–1950 (2009).

[11] S. Zhankui and K. Sun, "Nonlinear and chaos control of a micro-electro-mechanical system by using second-order fast terminal sliding mode control," *Communications in Nonlinear Science & Numerical Simulation,* **18**, 2540–2548 (2013).

[12] D. Swaroop, J. K. Hedrick, P. P. Yip, and J. C. Gerdes, "Dynamic surface control for a class of nonlinear systems," *IEEE Transactions on Automatic Control,* **45**, 1893-1899 (2000).
[13] S. I. Han and J. M. Lee, "Adaptive dynamic surface control with sliding mode control and RWNN for robust positioning of a linear motion stage," *Mechatronics,* **22**, 222-238 (2012).
[14] S. Luo, "Adaptive fuzzy dynamic surface control for the chaotic permanent magnet synchronous motor using Nussbaum gain," *Chaos,* **24**, 5880-5885 (2014).
[15] Y.-J. Liu and Z.-F. Wang, "Adaptive fuzzy controller design of nonlinear systems with unknown gain sign," *Nonlinear Dynamics,* **58**, 687-695 (2009).
[16] S. Ibrir and C.-Y. Su, "Simultaneous state and dead-zone parameter estimation for a class of bounded-state nonlinear systems," *IEEE Transactions on Control Systems Technology,* **19**, 911 - 919 (2011).
[17] D. J. Li, "Neural network control for a class of continuous stirred tank reactor process with dead-zone input," *Neurocomputing,* **131**, 453–459 (2014).
[18] P. J and D. R, "Identification and adaptive neural network control of a DC motor system with dead-zone characteristics," *ISA Transactions,* **50**, 588–598 (2011).
[19] X. Wu, X. Wu, X. Luo, Q. Zhu, and X. Guan, "Neural network-based adaptive tracking control for nonlinearly parameterized systems with unknown input nonlinearities," *Neurocomputing,* **82**, 127-142 (2012).
[20] R. J. Wai, J. X. Yao, and J. D. Lee, "Backstepping fuzzy-neural-network control design for hybrid maglev transportation system," *IEEE Transactions on Neural Networks & Learning Systems,* **26**, 302-317 (2015).

Five Level Inverter Fed Squirrel Cage Induction Motor Drive with Reduced Number of Power Elements

R Harikrishnan, C N Ravi**, S Vimala***
* ** ***Faculty of Electrical and Electronics Engineering, Sathyabama University, Chennai-600 119
*Rhareish@Gmail.Com,**Dr.Ravicn@Gmail.Com,***Vimalaeee2010@Gmail.Com

Abstract: A topology for four switch five level inverter fed squirrel cage induction motor drive was presented in this paper. The three phase five level inverters is realized by cascading a three single phase inverter with output voltage phase sequence 120 degree lagging each other. Multi carrier sinusoidal pulse width modulation was used as a gate drive for the proposed inverter topology. The frequency of the reference wave determines switching time of MOSFET Switches. The proposed five level inverter configuration entail lesser number of switching devices as compared to the conventional five level inverter. Similarly, power supply requirement also less when compared to the conventional inverter drives.

Keywords: Multi carrier pulse width modulation (PWM); Four Switch Five level Inverter(FSFLI), Three Phase Squirrel Cage Induction Motor, Total Harmonic Distortion.

Introduction

Paper [1] proposes an improved single phase five level PWM inverter. In this paper firing signal for proposed inverter can be generated by comparing the reference signal with two carrier waves. Both having the same frequency and in phase with each other but two different offset levels. Paper [2] suggest the design of control scheme for a Nine Level Neutral Point C/H bridge inverter interfaced with the grid through an LCL filter. This Inverter topology produces the Nine level output waveform with low harmonic contents because of LCL filter. In five-level neutral-point clamped (NPC) H-bridge PWM inverter, a five level output voltage can be obtained by combining the three level outputs from both legs of the inverter. By choosing the phase of reference and carrier signal, a three-level output can be obtained. This inverter show signs of superior harmonics suppression property and leaving the multiples of the fourth order harmonics. Gating signal can be generated by comparing the triangular carrier wave with the sinusoidal reference wave[3]. In Paper [4], recommend the Model Predictive Direct Torque Control (MPDTC) which is used to reduce the converter's switching losses and improves the torque's Total Harmonic Distortion (THD). They introduced the MPDTC and applied to a five-level converter for driving the high frequency induction machine. By using MPDTC technique, they reduce the switching loss and Total harmonic distortion by fifty percentages. In Paper [5], they offer the PWM techniques with multi carrier for a 5-level Neutral Point Clamped (NPC) inverter. In that paper, they are using a new inverter topology for a 3-phase coupled inductor. This suggested inverter performance is mainly dependent on the PWM strategies. The use of the PWM techniques is reduction of current ripple and losses in the inductor side. In general, the conventional inverter fed induction machines rotor current and voltage quality is low due to the presents of harmonics in inverter output voltage. So there is a significant amount of energy losses due to the harmonics. This problem was overcome in Paper [6] by implementing the nine level inverter as a energy supply. While increasing the number of level in multi level inverter produce the high quality of output profile for induction motor drive. As previously discussed, harmonic losses are the important things in high power applications like AC drive. It should be reduced before applying to induction motor drive. Paper [7], they characterized the harmonic content by loss factor. They introduced the three PWM strategies for the three level inverters. Paper [8] proposes the five levels and seven level inverter fed induction motor drive system. Both the levels can be obtained by same two H Bridge circuit. The harmonic analysis was done for both the topologies and comparative analysis was shown. It clearly depict when number of levels was increased gradually the harmonic content also reduced. Paper [9] proposed that three phase three level inverters reduces the medium voltage drives especially used to reduce the harmonic content. They are verifying their model for 7.5 HP, 400 V squirrel cage induction motor drive system. Its used to increase no of steps of a induction motor step voltage its leads to reduce the dv/dt applied to the machine terminal. The Proposed [10] three phase five level topology having the 18 MOSFET switches. It reduces the convention Neutral Point Clamped inverter which is having the 18 clamping diodes and also it reduces 5 flying capacitor compared to the conventional flying capacitor which is using 6 capacitor. Hence proposed inverter reduces the circuit complexity.

Five Level Inverter

Proposed Five Leve Inverter Cicuit Model

The Fig. 1 shows the proposed model of single phase five level inverter which comprises of four MOSFET switches and single DC Source. The advantage of proposed inverter is reduced number of power switches over conventional 5 level inverters. Generally Multi Level Inverters are widely used because of its lowest harmonic contents in its output voltages. Due to the reduced harmonic contents requirement of output filter also reduced.

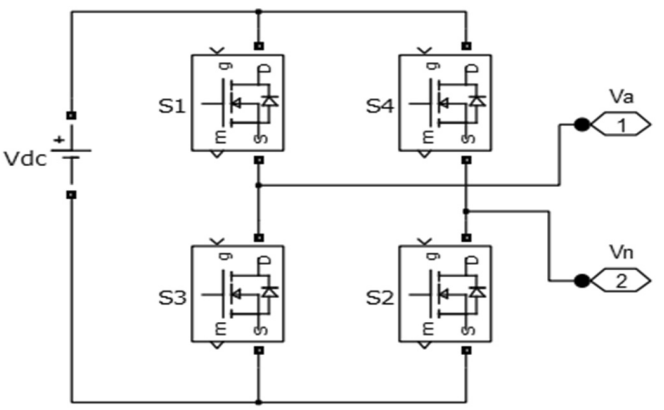
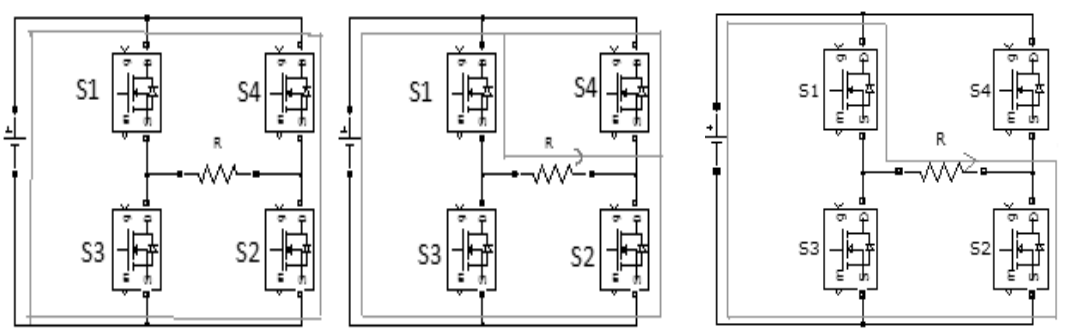


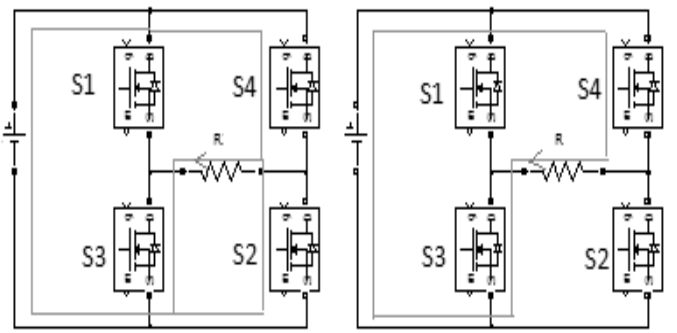
Figure 1. Proposed Four Switch 1 Phase Five Level Inverter Circuit Simulink Model

Operation of Proposed Inverter

The below shown Table 1depict the switch status of various levels of output voltage. For producing the Zero output voltage, the switches S1 and S3 are in OFF status and S2 and S4 are in ON status. So the current is circulating through the DC source, Switches S2 and S4. There is no current is flowing through the Load until the changes of the next triggering patterns. The next level of output $V_{DC}/2$ can be obtaining by giving the firing pulse to the switch S2.The Current flow direction is shown in the Fig. 2(b). The maximum voltage level V_{DC} can be obtained by giving the firing pulses to the switches S1 and S2. The current flow direction is shown in the Fig. 2(c). Voltage level $-V_{DC}$ can be obtained by making the switch S2 in ON status. The current flow direction is shown in the Fig. 2(d). The next level of voltage $-V_{DC}$ can be obtained by making switches S3 and S4 in ON status. The Current flow direction is shown the Fig. 2(e). The Fig. 3 shows the five level inverter simulation output voltage waveform. The harmonic analysis was done for this inverter output voltage. The THD value for this output wave form is 16.83%. The Fig. 4 shows the hardware image of the proposed single phase four switch five level inverter and Fig. 5 shows the hardware output stored by DSO.



(a) (b) (c)



(d) (e)

Figure 2. Five Level Inverter Operation Modes (a) 0 V (b) $V_{DC}/2$ V (c) V_{DC}V (d) $-V_{DC}/2$ V (e) $-V_{DC}$V

Table 1. Switching Table

V_{AC}	SWITCHES STATE			
	S1	**S2**	**S3**	**S4**
V_{DC}	ON	ON	OFF	OFF
$V_{DC}/2$	ON	ON	OFF	ON
0	OFF	ON	OFF	ON
$-V_{DC}/2$	OFF	ON	ON	ON
$-V_{DC}$	OFF	OFF	ON	ON

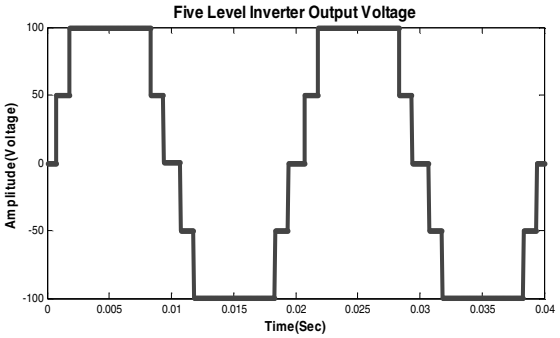


Figure 3. Proposed FLI Simulation Output Voltage Waveform

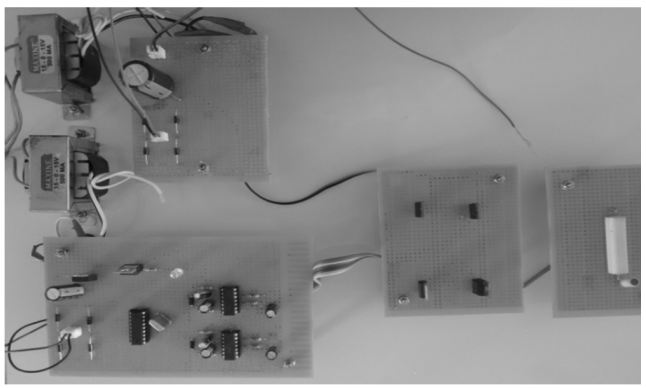


Figure 4. Hardware Image of Proposed FLI .

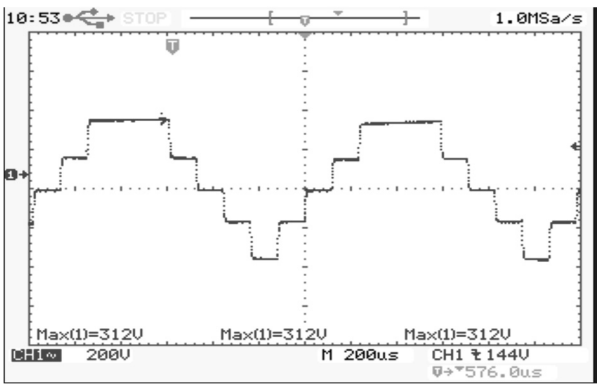


Figure 5. Proposed FLI Hardware Output Voltage Waveform

PWM Switching Strategy

The Switching pattern of the proposed inverter is generated by comparing the modulating wave which is a pure sinusoidal wave having the frequency of 50Hz with the four triangular carrier waves. The four triangular carrier waves having the same frequency and amplitude. The first two carrier waves having the same phase sequence but different offset levels. Similarly the next carrier waves having the same phase sequence but different offset levels.

The PWM pulse generation circuit consists of four carrier waves and one reference wave. Each carrier wave is shifted with the amplitude of AC with the other carrier waves. The firing pulses for the switches S1 and S2 can be generated when the amplitude of the reference wave is greater than the amplitude of the second and fourth carrier waves. In the same way the firing pulses for the Switches S3 and S4 can be generated when the amplitude of the reference wave is less than the third and first carrier waves respectively. The firing patterns for the five level inverter power switches shown in the Fig. 6.

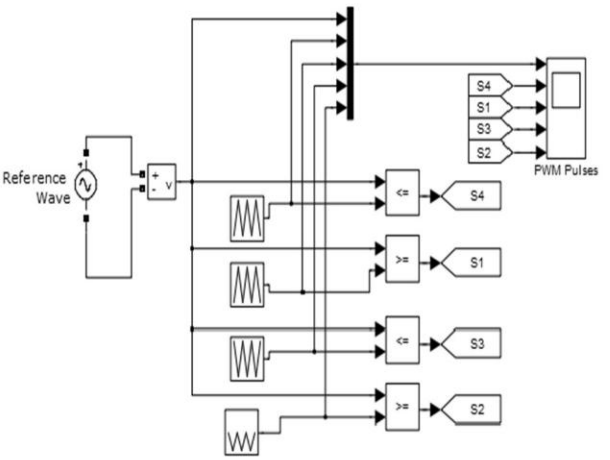


Figure 6. PWM Generation for 1 phase Five Level Inverter

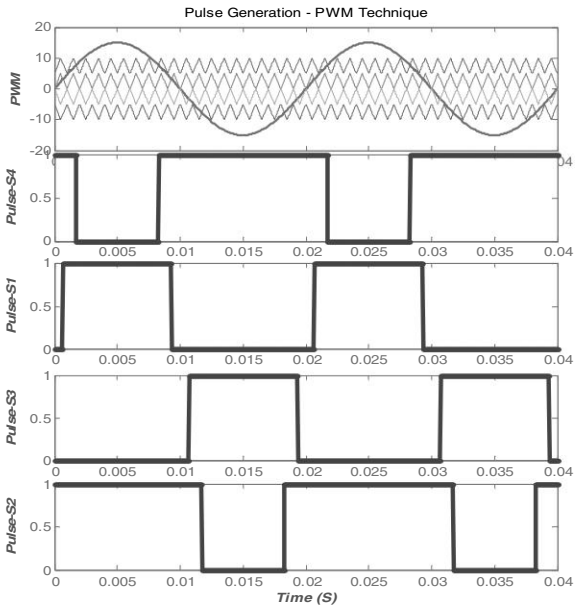


Figure 7. PWM Switching Pattern for Proposed Single Phase FLI

Three Phase Five Level Inverter Fed Induction Motor Derive

The proposed five level inverter was implemented with MOSFET as a switching elements and was tested with a 5.4 HP(4KW), 400 V, 1430 RPM, 50 HZ squirrel cage induction motor. The simulation was done using MATLAB 7.10.0 (R2010a)/ Simulink and corresponding simulation output was plotted. The Fig. 10 shows the simulink model for three phase five level inverter fed squirrel cage induction motor drive. The DC supply used for each single phase five level inverter is 200 V. The Fig. 8 shows the phase voltage of induction motor and Fig. 3 shows the single phase five level inverter pole voltage. FFT analysis was done for stator voltage, stator current and rotor speed of the squirrel cage induction motor drive. The gate drive system used for three phase five level inverter uses the below given reference wave as modulating signal for each phase of a power switches.

$$V_{aref} = V_m \sin \omega t$$
$$V_{bref} = V_m \sin (\omega t - 120)$$
$$V_{aref} = V_m \sin (\omega t - 240)$$

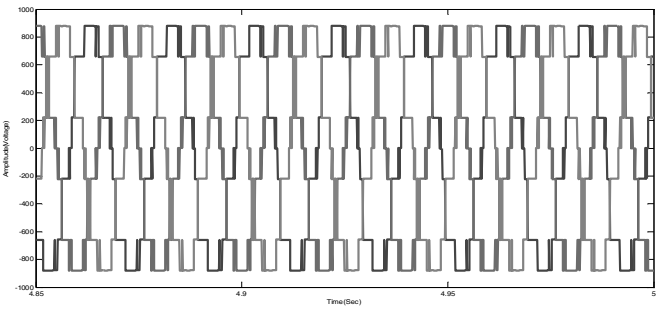


Figure 8. Three Phase Five Level Inverter Output Voltage Waveform

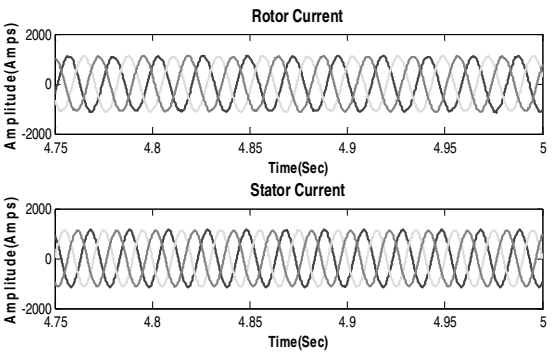


Figure 9. Stator and Rotor Current Profile of Multi Level Inverter fed three phases Induction Motor

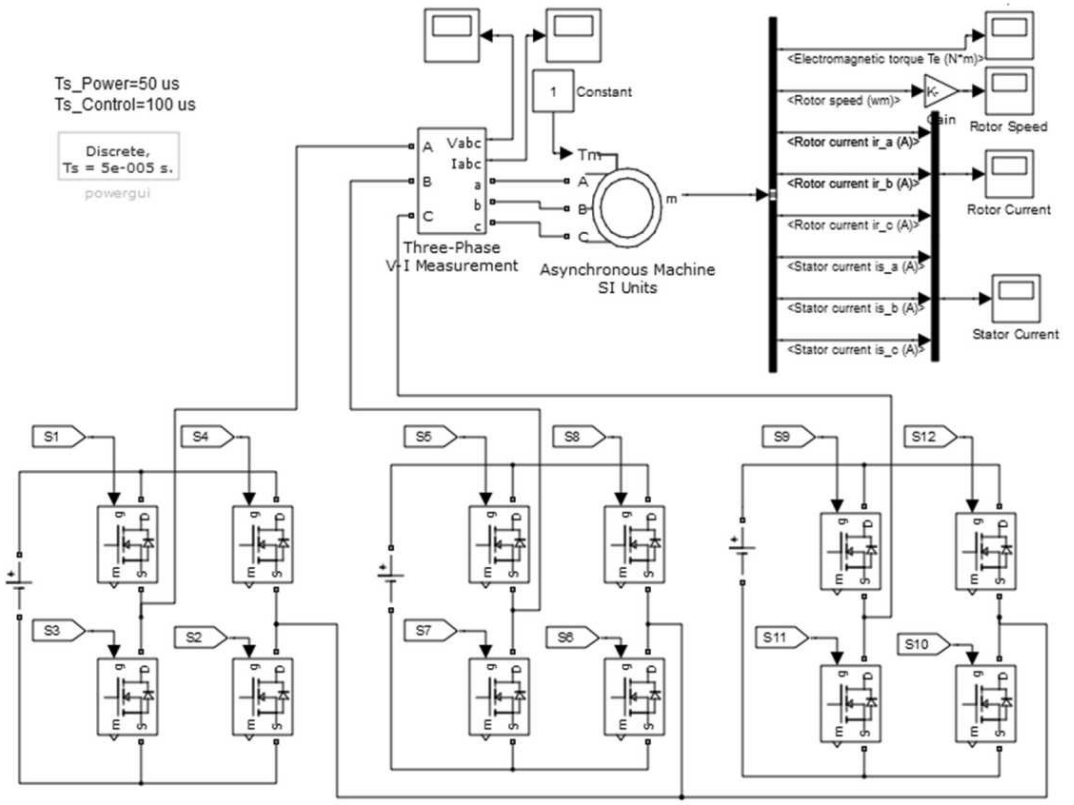


Figure 10. Simulink Model for Three Phase Five Level Inverter fed Squirrel Cage Induction Motor

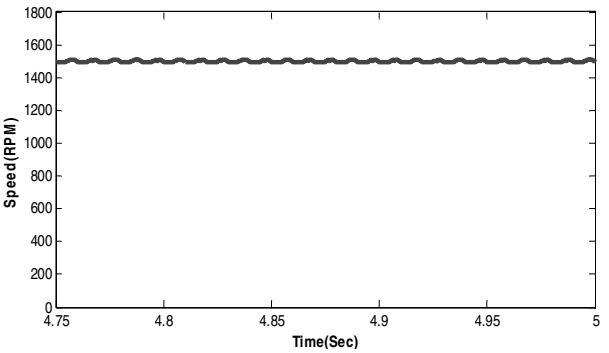


Figure 11. Rotor Speed of Multi Level Inverter fed three phases Induction Motor

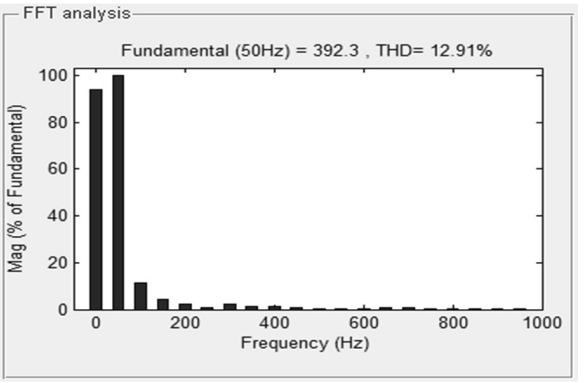


Figure 12. FFT analysis for Torque

Conclusion

The Proposed four switch five level inverter drive system can be used in high voltage industrial drive with minimum amount of harmonic content than the conventional speed drives. In this paper we have discussed about design and simulation of the 3φ five level inverter fed squirrel cage induction motor drive system. FFT analysis was done for the inverter output voltage, stator current, stator voltage and rotor speed of the induction motor .Four Carrier signal was used to generate the firing signal for MOSFET switches of proposed single leg of the 3φ five level inverter. The proposed inverter cost is very less compared to other inverter topology due to the less number of power elements.

References

[1] T. Wanjekeche , D.V.Nicolae, A.A.Jimoh, "Modeling and Control of a Cascaded NPC/H-Bridge Inverter with LCL Filter in PV- Grid Application", IEEE Conference on IPEC2010.Page(s):334-339

[2] Maha G. Elsheikh, Mahrous E. Ahmed, EmadAbdelkarem, Mohamed Orabi "Single-Phase Five-Level Inverter with Less Number of Power Elements" IEEE 33rdInternational Conference on Telecommunication Energy Conference(INTELEC) 2011.

[3] C.M. Wu, W.H. Lau and H. Chung and Tat Chee Ave., "A Five-Level Neutral-Point-Clamped H-Bridge PWM Inverter With Superior Harmonics Suppression: A Theoretical Analysis" Published in IEEE Conference on ISCAS 99 Proceeding of the 1999,Vol 5, Pages:198-201

[4] Tobias Geyer and GeorgiosPapafotio, " Model Predictive Direct Torque Control of a Variable Speed Drive with a Five-Level Inverter" Published in 35thIEEE Annual Conference on Industrial Electronics 2009. Pages:1203-1208

[5] BehzadVafakhah ,Jeffrey Ewanchuk and John Salmon, "Multi-Carrier Interleaved PWM Strategies for a New Five-Level NPC Inverter Using a 3-Phase Coupled Inductor" Published in IEEE Conference on Energy Conversion Congress and Exposition (ECCE) in 2010. Paes:2312-2319.

[6] Budi, S.R., Sharma, S "A novel method of nine level inverter fed induction motor drive" Published in International Conference on Advances in Engineering and Technology Research (ICAETR) on 1-2 Aug. 2014.

[7] Hálasz, S, Hungary ; Zakharov, A. "PWM strategies of three-level inverter-fed AC drives" presented in International conference on Industry Applications in 13-18 October 2002.

[8] Manasa S, Balaji Ramakrishna S, Madhura SAndmohan H M "Design And Simulation Of Three Phase Five Level And Seven Level Inverter Fed Induction Motor Drive With Two Cascaded H-Bridge Configuration" published in International Journal of Electrical and Electronics Engineering (IJEEE), ISSN (PRINT): 2231 – 5284 Vol-1 Iss-4, 2012.

[9] Suvajit Mukherjee, GautamPoddar "A Series Connected Three-Level Inverter Topology For Medium Voltage Squirrel Cage Motor Drive Applications" published in IEEE on Industry Applications in 2010. Volume:46, Issue: 1, Pages(s):179-186.

[10] K. Sivakumar, Anandarup Das, RijilamchandChintan Patel, K Gopakumar "A Simple Five-Level Inverter Topology for Induction Motor Drive Using ConventionalTwo-Level Inverters and Flying Capacitor Technique" Published in 35th Annual Conference of IEEE on Industrial Electronics, 2009(IECON '09) Pages:1009-1013.

Intelligent Web based Home Automation and Security System using Raspberry PI

R Harikrishnan*, C N Ravi**, Liji S Job*** and Rudra J Kumar****
* ** *** ****Faculty of Electrical and Electronics, SathyabamaUniversity, Chennai, India
*rhareish@gmail.com,**dr.ravicn@gmail.com,***sjobliji@gmail.com,****rudra.jkumar@gmail.com

Abstract: Home automation is a system used to control the home appliances without going to them physically. This paper emphasis on controlling home appliances through web. A security system is also included which provides images of the area under vision. The whole system is designed with Raspberry Pi which is a mini computer. Raspberry Pi has proved to be most suitable platform for home automation. Suitable sensors like motion detector are used for energy conservation.

Keywords: Raspberry Pi, Home Automation, GPIO, Python, Linux.

Introduction

Home Automation(HA) include control of lighting, HVAC appliances, security locks of gates and other system to provide convenience to the user and security. Home automation has become popular in recent years due to its higher affordability and simplicity through connectivity of smart phones and tablet. Devices are connected through a home network which is controlled by a PC where internet has made remote access possible. This results in convenience, energy efficiency and safety benefits. There has been significant research in the field of home automation.

Home automation has to overcome a number of obstacles to make it easier and affordable for the users. Existing system can be classified on the basis of communication protocol like blue tooth, hand gestures, DTFM and others [1].A Java based home automation system incorporates an inbuilt security features but it requires high end computers and wired installation .This makes Java bases HA system expensive and complex. A Bluetooth based HA system has two controllers, primary and bluetooth sub-controllers to which all the peripherals are connected. This leads to access delay since all the peripherals are connected a single Bluetooth module.

Another HA [2] system is a phone based system where the control operation is done through fixed telephone lines. It suffers from the problem of flexibility as the user has to remember the access codes. The controller in a control network using hand gestures uses a glove to relay hand gestures to control the system. This system lacks security[3]. Sougata Das et.al [4] proposed a system that control Home appliances through mobile phones using GSM technology.SMS is given as control commands to the central control system which has Atmega8 as the main controller. The system lacks of the network coverage area and the user has to be aware of AT commands for controlling the devices.

M.G.Golzar et.al [5] proposed a system whose design and implementation was based on an embedded controller which is connected to a home web server that enables communication through USB ports. The system was able to save energy. However there was some limitation such as remote connection delay and actuator status feedback information. J.A.Nazabal et.al [6] designed a system for monitoring the behavior of elderly people at home by developing a low cost home automation based on sensor system. According to the data acquired by the sensors and based on certain predefined rules for a particular user, an abnormal behavior can be detected and action can be taken. The sensors used are wired which would need some kind of house reformation.

Md.Syaduset.al [7]developed a security system with ARM processor as CPU and communication system is designed by GSM technology. The system counts the number of people in the room. When the presence of human is detected an SMS is sent to the user and the video is recorded for further inspection whereas switches off the load when no movement is detected.

N.Sriskanthanet.al [8] proposed a Bluetooth wireless technology to control the home appliances. The system developed, demonstrated the control of the room temperature. In this paper they use I^2C to interface between the DC modules and AD's where one module of Bluetooth for each DC.

This paper proposes a HA system using Raspberry Pi, it is a mini-computer developed by the Raspberry Pi foundation in UK for stimulating the teaching computer science in schools. The system developed demonstrates an LED which glows when a movement is detected by the motion detector connected to the GPIO pins of the Rpi. It also has a Webcam that is able to take a snapshot of the area under surveillance. The snap shot can be viewed by using the Web browser in desktop PC or mobile. Figure 1 shows a RPi board. There are presently four models in the market. Model A, Model B version 1, Model B version 2 and the latest model, Model B+. Model A has RAM memory of-256 MB , one USB Port and no Ethernet Port with 26 GPIO pins. Model B v1 is same as Model A except it has 2 USB port. Model B v2 has RAM memory of 512 MB and similar

features of the previous version. Lastly the latest model, Model B+ has 512MB RAM, 4 USB ports and 40 GPIO pins. The model used in this paper is Model B +.

Fig 1. Raspberry Pi board

Figure 2 shows the pin description of RPi board.The micro USB power port is used to power the Raspberry Pi device. The Raspberry Pi has ARM1176JZF-S 700 MHz processor. The SD card slot is used for inserting the SD card which stores the operating system, programs and the data needed to run the Raspberry Pi[9].The GPIO headers are used to connect the Raspberry Pi to other hardware devises. The HDMI output is used to plug into modern television or monitor. The video output is used to connect to an older type television. The audio output can be used to plug into an external amplifier or an audio docking station. The ethernet port is used to connect the Raspberry Pi to the internet or a local network. USB 2.0 ports are used to plug in a keyboard, mouse, web cam, external hubs etc. The Camera Serial Interface (CSI) is a serial interface between digital camera module and RPi. The *Display* Serial Interface (DSI) is a serial interface between *LCD Display monitor* and RPi. Status LEDs will show the status of power, booting and network.

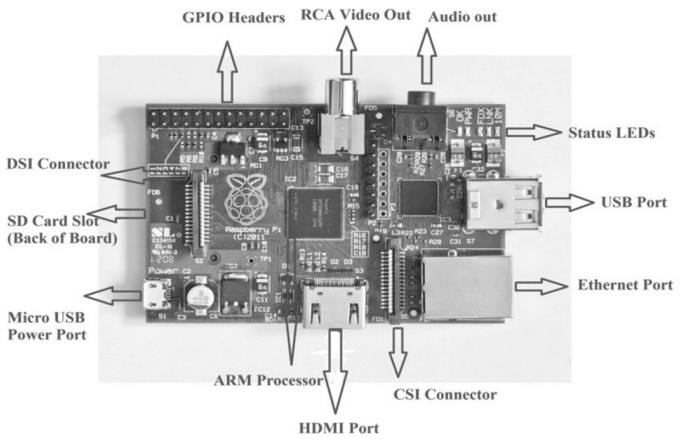


Fig 2. Pin description of RPi

Design of the Proposed Work

Figure 3 shows the block diagram of the proposed system. The whole system is designed with Raspberry Pi which acts as the CPU and the programming language is written in python which is already incorporated in Rpi by default.

The user can access the Raspberry Pi board to control and view the status of the device through internet from PC, desktop, mobile or tablet. This is done by enabling SSH in the Rpi configuration. The RPi has a total of three connections. A 5V, 700Ma-2A rating adapter which provide the supply to the board, an HDMI connection, (the other end is connected to the monitor) and a USB hub. The USB hub has a Wi-Fi dongle (optional in case of LAN connections), a power supply since the

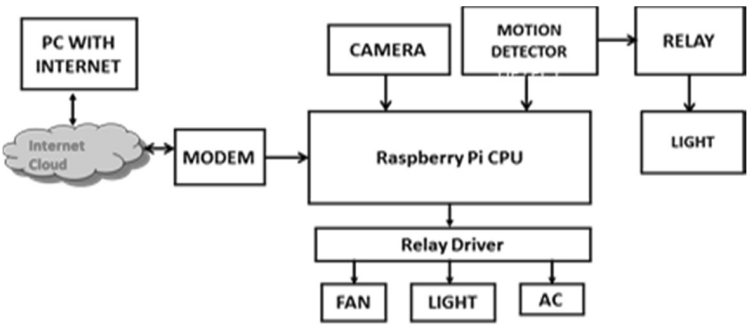


Fig 3. Block Diagram of Proposed system

Wi-Fi dongle draws more current than the board, keyboard and a mouse. The monitor, keyboard and the mouse are used only during the development of the system. The figure 4 shows the pin configuration of the Rpi GPIO used in the system.

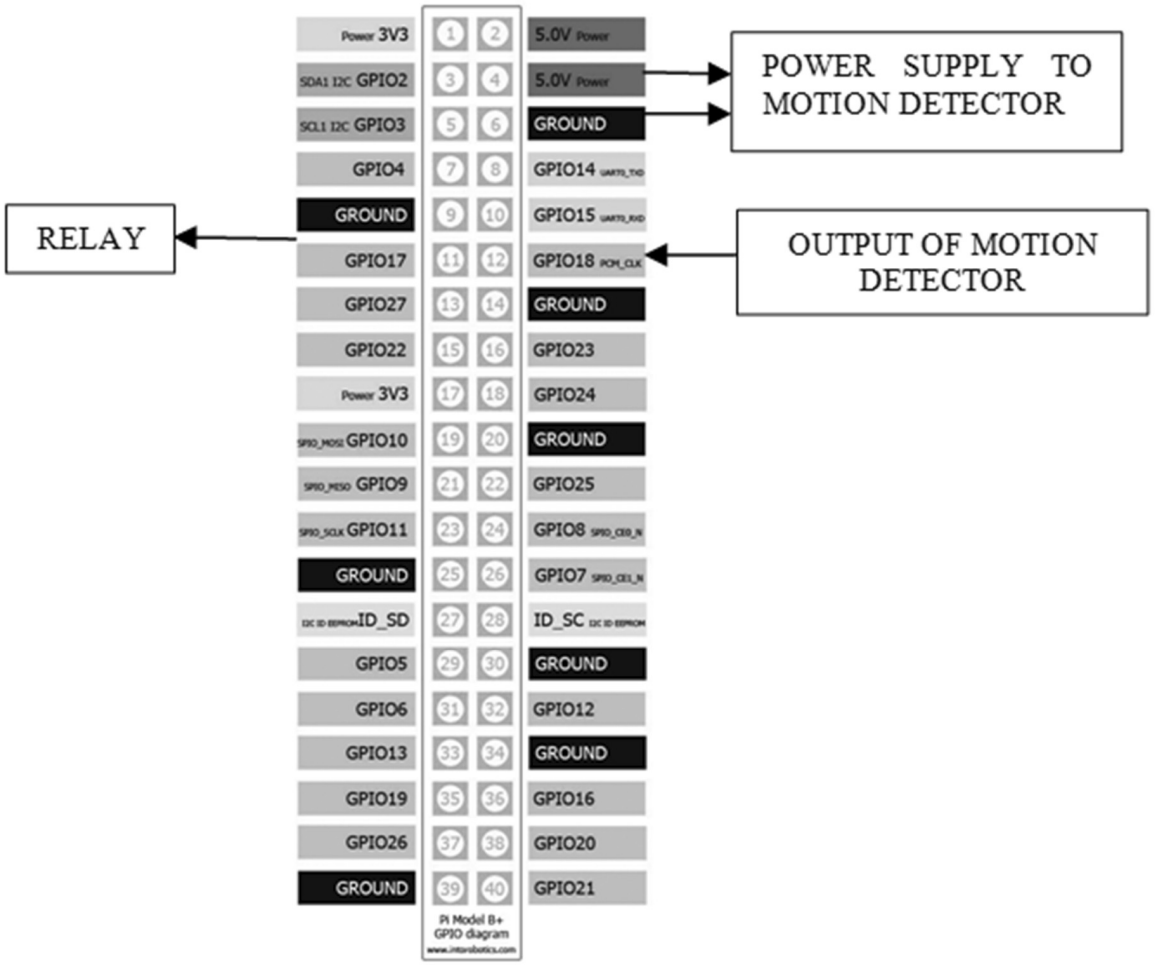


Fig 4. Pin configuration of RPi GPIO

Camera Interfacing

A web cam is used for capturing images for the area under observation. It is connected to RPi through the USB port. The RPi does not come up with the software for the web cam so firstly it has to be installed by using the Linux command.

Sudo apt-get installfswebcam
The python program is written in such a way it takes images after every one second and the previous image is refreshed after every two seconds. To use the file having the program at startup is done by using the Linux command

Sudo update-rc.dpiWebCamserve defaults
piWebCam is the file in which the python script for the operation is saved. An HTML script is written to view the images taken from the web through PC, desktop, etc

Motion Detector Interfacing
A PIR sensor is used for detecting motion in the vicinity. It is a 5v passive infrared sensor. Since the rating of IR sensor is as same as that of Rpi, it can be directly interfaced with the GPIO pins. The PIR is a 3 pin module consisting of a 5Vsupply pin, ground pin and an output pin. The Figure 5 shows the interfacing of PIR module with the RPi. When a motion is detected the output pin goes high with +3V for a duration ranging from few second to 200 seconds. The default setting is only for few seconds. When there is no motion detected the out pin goes low. Depending on the status of the output pin a relay is operated to switch ON/OFF the device connected to it by using suitable level shifters.

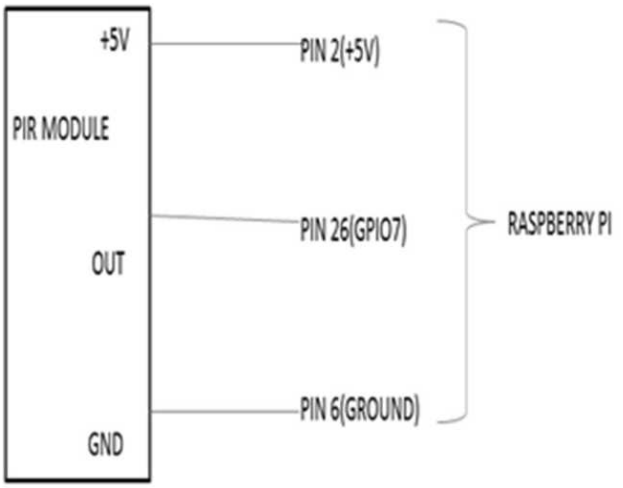


Fig 5. Interfacing of PIR module

Interfacing Home Appliances To Be Controlled
Home appliances such as lighting, HVAC devices are connected to RPi GPIO pins directly via relays. Home appliances are switched ON/OFF by the user depending upon the status. The user controls the appliances through the web page created for HA. The web page is accessed by using the unique IP address for the system. The webpage shows the status of the device and provide controlling operation. The controlling signal is send to the board by the internet cloud. The relay operates which in turn controls appliances.
The proposed setup has the raspberry pi, motion detector and the webcam. The control of home appliances is shown with the help of an LED which is turned ON/OFF depending upon whether there is any human motion or not.

Result

The Raspberry pi is first booted up and updated using the command
Sudo apt-get update
The IP address of the Raspberry pi is got through the command:
ifconfig
The required operation program and the HTML script for the web page is been saved to a folder called home_monitoring. Therefore the directory should be changed to home_monitoring from the LXT terminal. This can be done by using the command:
cdhome_monitoring
With the command ***ls***all the files in home_monitoring is listed. The required program to be executed can be done using the command:
Sudo ./monitoring

This command allows the execution of the file ***monitoring*** where the program is saved. The Figure 6 shows the LXT terminal of the RPi.

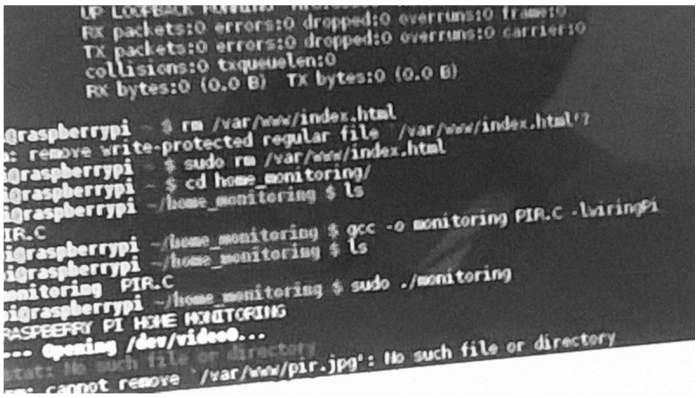


Fig 6. LXT terminal of RPi

Figure 7 shows the led output which glows when a motion is detected. The LED is turned ON for a few seconds and is turned OFF when no motion is detected.

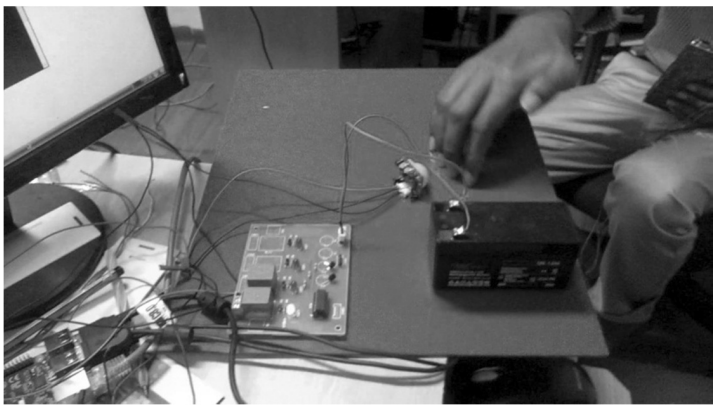


Fig 7. LED Output for motion detection

The Figure 8 shows the webpage from where the image can be viewed which is captured by the webcam. The webpage is accessed using the IP address of the Raspberry pi.

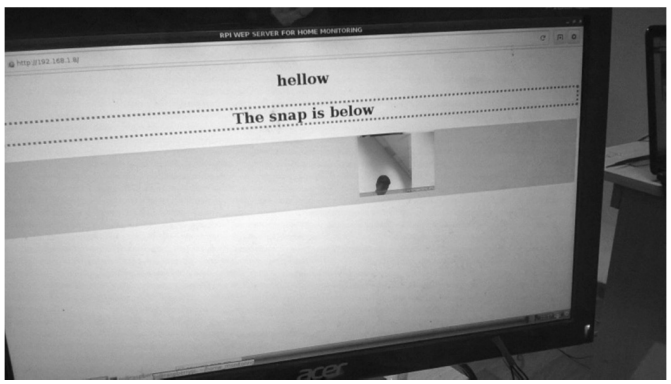


Fig 8. Webpage for webcam image

Conclusion

The proposed system has proved to be an efficient Home Automation system in terms of flexibility, reliability and preventing wastage of energy. Additional operations can be included without changing the core features. Experimental results show how the system is user friendly and acting to its purpose. This system can allow the users to control home appliances from anywhere anytime thus it is ubiquitous. Home appliances were controlled by the user from the web page created specifically for home automation.

References

[1] S.Jain,AnantVaibav,LovelyGoyal, "Raspberry Pi based Interactive Home Automation System Through Email",InternationalConferanceonRelaibility,Optimization and Information Technology,2014,pp 277-281

[2] A.Gurek,C.Gur,C.Gurakari,M.A.Kdeniz,S.K.Metin,I.Korkmaz,,"An Android based Home Automation System",10th International IEEE Conference,2013,pp 121-125

[3] N. Hasan, A.Al M. Khna, NisamUddin, Abu FarzanMitul,"Design and implementation of touchscreen and remote control based home automation system",2nd International Conference on Advances in Electrical Engineering,2013,pp347-352

[4] S.Das,N.Debabhuthi,R.Das,S.Dutta,A.Ghosh,,,"Embedded System For Home Automation Using SMS",IEEE Conference Publications 2014, pp 1-6,

[5] M.G.Golzar, H.R.Tajozzakerin, "A New intelligent remote control system for home automation and energy consumption",4th Asia International conference On Mathematical/Analytical modeling and Computer Simulation,2010,pp174-179

[6] J.A.Nazabal,I.R.Matrias,C.FernandezValdivielso,F.Falcone,P.Branchi,S.C.Mukhopadhyay,,"Home Automation Based Sensor System For Monitoring Elderly People Safety",6th International Conference on Sensing Technology,2012,pp 142-145,

[7] Md.SyadusSefat,A .Al Mamum Khan,Md.Shahjahan,"Implementation of Vision Based Intelligent Home Automation And Security System",3rd International Conference on Information Electronics and vision, 2014,pp 1-6

[8] N. Sriskanthan, F. Tan, A. Karande," Bluetooth based home automation system", Microprocessors and Microsystems 26,2002,PP 281-289

[9] M.S.BinBahrudin,R.Abu Kasem,N.Buniyamin,,"Development Of Fire Alarm System Using Raspberry Pi And Arduinouno",International Conference on Electrical, Electronics And System Enginering,2013,pp 43-48.

Enhancement of Smart Grid Performance Through Logic based Fault Tolerant MPSoC

D.Vijayakumar* and V.Malathi**
*Assistant Professor, Dept. of Electronics and Communication Engineering, Latha Mathavan Engineering College, Alagarkoil, TamilNadu, India. Email: vkkumarin@gmail.com
**Professor, Dept. of Electrical and Electronics Engineering, Anna University Regional campus Madurai, TamilNadu, India

Abstract: Sustainable energy is the energy production without compromising the energy production for the future generations. This paper presents the picture of today's power system structure. It also portrays a reasonable picture about the different challenges that are confronted by the present day smart grid archi tectures used in transmission framework. This paper presents a unique vision for the development of smart power grids. This paper addresses major issues in smart grid. A Multiprocessor System on Chip (MPSoC) is designed to specifically meet the niche requirements of modern power system processing elements. The existing power grid model does not provide real-time information of transmission devices during emergency events. In this paper the significance of restructuring the existing smart grid architecture using MPSoC with power system components. An embedded intelligence is inserted into the power-electronics to facilitate the reconfiguration of the system, and thereby ensuring security. As the system is designed with MPSoC modified smart grid architecture, the computational complexity of the proposed system architecture significantly improving the performance of the smart grid. This paper reveals the fault tolerant methodology using MPSoC with self-diagnosis, which is essential for enhancing the proposed architecture for smart grid functionalities. Using application-specific instructions for Heterogeneous MPSoC allows finding a good performance/energy tradeoff. The functions will enhance the general execution regarding execution to accomplish the framework a reconfigurable and thusly the execution may be enhanced by upgrading which means edges in gadget system readiness and practicality and at last, worth reduction. This leads to reduce computational complexity of the existing architecture, latency and improves performance tradeoff.

Keywords: Smart grid, MPSoC, Embedded system, Reconfigurable Smart grid architecture, Fault tolerance.

Introduction

A Multiprocessor System on Chip (MPSoC) is a complete computational system integrated in a single chip, combining multiple Processing Elements (PEs) as main components. In addition, the MPSoC is built by interconnecting Intellectual Property (IP) cores with local ports of Network on Chip (NoC) routers. Messages are transmitted between IP cores wrapped in NoC packets. The growing interest of MPSoCs lies in its ability to combine high performance, flexibility and its reconfigurable feature. In this work, MPSoC uses multiple Remote Terminal Unit (RTUs) along with other Distributed State Estimator (DSEs) to implement a system. The wide ranges of MPSoC architectures have been developed over the conventional methods to progress real time management system.

Literature Survey

The literature survey has been conducted on architectures of the power grid by using different means and approaches. This section reviews the relevant literature of the existing power grid model and infrastructure basics. Based on the different literature, the smart architecture is proposed and an low data latency and low energy consumption for communication topology are implemented for smart grid. From [8] it is proposed that a modern electrical grid infrastructure for higher efficiency and reliability via automated control, modern communications, sensing, high power converters, metering technologies, and energy management schemes. In this approach, smart grid infrastructure is highly complex. From [18] an approach about the recent smart grid management and protection systems. This approach focusses the performance of protection in smart grid. From [20] an efficient and privacy-preserving aggregation methodology for securing the smart grid communications is proposed. The performance of this method is more securable and flexible in terms of computation for consumer communication overhead. A distributed interval integration algorithm [9] is explained. In this method, the generator was chosen to collect the device output to develop an overlap function. This proposed algorithm guaranteed that the mere variations in the input intervals resulted only in mere variations of the integrated result. From [6] a method is proposed for comparative analysis for fault detection and classification using Functional Analysis and Computational Intelligence (FACI).This method lags with the computation time to detect the fault. A smart grid controller [1] for optimizing energy consumption is proposed. A framework for high level power estimation of MPSoCs architecture on FPGA is proposed in

[19]. The approach uses different interconnects method and memory hierarchies. This method does not discuss the power model of real time system-level design space exploration. From [21] proposes a algorithm of task scheduling issue to avoid the inter-core communication overhead in on-chip MPSoC architecture. This algorithm is obtained with minimum optimal solution for memory usage with reduced schedule length.
From [13] it describes the changing requirements due to privatization and the deregulations have created needs for analyzing information from different sources within DW. These needs require new high performance solutions represented by the new data warehouse of Supervision Control And Data Acquisition (SCADA). The system security and memory protection service in a wide range of NoC-based MPSoC platforms is discussed in [7]. The protection mechanism is to prevent the malicious data throughout system with the help of self-contained NoC at Network Interface (NI) layer.
A technique to eliminate inter core communication overhead on Multiprocessor System-on-Chips (MPSoCs) with combined computation and communication task scheduling of delivering virtualized applications is proposed in [12]. From [14] it is exposed that the emergence of MPSoC processor to efficiently exploit the low latency and high bandwidth of modules in the hierarchy in terms of programming framework. An MPSoC architecture which has thermal effect was presented by [5]. It leads to cause violation in timing constraints. Therefore, MPSoC have high power density and temperature which will degrade the reliability and increase the cost. The MILP solver provides the improvement of performance and reduces peak power. Hence, scheduling and assignment technique are used to overcome peak temperature.
State Estimator (SE) in the electric power grids makes technological alterations and is being proposed as a part of the smart grid development. SE turns out to be a key operation in supervisory manipulation and planning of electric power grids. It serves to observe the state of the grid and enables Energy Management Systems (EMS) to participate in quite a lot of foremost control and planning tasks corresponding to open near real-time community units for the grid, optimizing power flows, and bad information detection/evaluation as acknowledged by way [22] and [16].
In the past decade, Multiprocessor system on-chips (MPSoCs) have been developed as a significant class of Very Large Scale Integration (VLSI) systems. In VLSI system, an MPSoC is a system on-chip that includes most or the entire components essential for an application and it utilizes the multiple programmable processors as system components. MPSoCs are widely used in networking, communications, signal processing, multimedia and other applications. An method used to [15] analyze the design challenges faced by MPSoC designers at all levels. At the application level, there is an necessity for programming models and communications APIs that permit applications to be simply re-configured for much different possible architecture without rewriting, while at the same time guaranteeing efficient production code. From [13] it discusses about the communication architecture analysis for multi-processor Systems-on-Chips (MPSoCs) and it leverages a SystemC-based platform to simulate a whole multi-processor system at the cycle-accurate and accurate signal level.
The packet data flow affects the MPSoC performance and power consumption that are flowing in the network. It was [25] discussed about the packetized on-chip communication power model and quantified the effect of packet size deviation on the performance and energy consumption. For assisting MPSoC programmers with C application parallelization, MPSoC Application Programming Studio (MAPS) is developed by [4] which is an integrated framework. To extract coarse coarse-grained parallelism in C, a novel granularity level has been proposed on MAPS partitioning tools. A temperature-aware task allocation and scheduling algorithm are proposed by [23] for MPSoC embedded systems. In that method, both power-aware and thermal-aware schemes are investigated to the task allocation and scheduling.
To raise the lifetime reliability of the platform-based MPSoC embedded systems, new solutions have been proposed by [11]. An analytical model is presented to evaluate the lifetime reliability of platform-based MPSoC, when executing periodical tasks. From [24] it proposed a multi-task mapping/ scheduling heuristic based on the QEA technique considering data and temporal parallelisms as well as task parallelism for MPSoC. Compared with an ILP (Integer Linear Programming) approach, experiments with real-life examples show the feasibility and the efficiency of the proposed technique. From [3] a methodology was introduced for power modeling of global framework for power/energy evaluation and optimization of different MPSoC. In order to attain more accurate power evaluation and some improvement of power model, it must be realized by focusing on more complex heterogeneous platforms. A innovative [11] technique for task share and scheduling algorithm that are added to the processors with aging effects, based on the annealing technique. This technique exploits the life time of MPSoC design and limits the various parameter performances. A method [2] proposes a system level flexible methodology for Heterogeneous multi-processor system-on-chip (HtMPSoCs), which addresses the fault tolerance and performance trade-off such, that an economic utilization of resources and it will minimize the overall cost of the application.

Requirements of a Smart Power Grid

The power grid must be designed securely at system level and device level. The following are the requirements for a secure power grid.

Real-Time Behavior

The high frequency (per 0.1 – 10s) of event occurrence in the power grid makes the control devices incapable for real-time control and reconfiguration. The conventional data collection and contingency analysis are performed only for every few

seconds and the State Estimator (SE) executes only every 5 minutes. It cannot fulfill the "real-time condition". The factors which hinder the implementation of a real-time management scheme in smart grid are communication delay, and large amount of raw data. Real-time scheduling are required to guarantee end-to-end real-time behavior. The fault detection and mitigation processes can be accelerated via distributed processing. And the other characteristics to be considered are

- *Grid Sensitivity*
- *Fault Resiliency*
- *Communication technologies and networks*

Smart hybrid architectures for Smart Grid

Data Processing Architectures

The current topology of the smart grid network is arranged according to the demand of the consumers. The distribution network is organized into multiple subgrids and consequently, forms a hierarchical topology. The key elements of architectural model are: (i) State Estimator (SE) - the state estimator is an integral part of the overall monitoring and control systems of transmission networks. It is used for the local measurements in a substation by the local estimator to estimate the states and then, it is transmitted to a central coordinator for further estimation. Remote Terminal Unit (RTU) acts as an intermediate node between substations and power grid network. RTU receives information from the substation within the neighborhood and multiple RTUs are utilized to cover several substations (iii) Grid network represents the central power grid to control the entity of a utility provider. This node is responsible for data processing and maintenance of all substations.

Decentralized Architecture

In this architecture, only RTUs have data processing and storage capabilities which are explained. SEs transmits data periodically to the respective RTU, but instead of forwarding the data, the RTU stores and processes this data locally. In the decentralized architectures, since complete data are available at the RTUs, data aggregation is possible. For instance, RTUs can aggregate data while reporting to the grid network. The grid generates queries to retrieve information from RTUs only when required. Thus, RTUs act as central entities in this architecture.

Distributed Architecture

In distributed architectures, all SEs have data processing and storage capabilities. SEs periodically sense and store the data values locally. Also, they process the data locally by the processor. Grid initiates a query to fetch the state of the substation, which is forwarded to the RTU and in turn to the SEs. RTUs process the query and send the reply to the grid. Thus, making the architecture distributed is made as Distributed State Estimations (DSE).

Signal Processing and the Smart Grid

While some researchers could consider the power grid SE research quite mature, new tactics for SE must be developed as the power grid turns into more intricate, extra interconnected, and more shrewd. Any development in the Signal Processing (SP) can largely facilitate and benefit the progress of the smart grid. In return, study of SE inside the framework of one of the most problematic man-made systems can invigorate the signal processing research community. In distributed estimation, several nodes estimate a usual parameter vector through neighborhood collaborations. Within the case of Multi Area State Estimation (MASE), the measurement of each area relates to a small part of the whole state vector. Hence, ensuing computational and communication expenditures of an allotted estimation procedure rely upon whether local knowledge of the entire state vector is required or not.

The vector equation as

$$[\sum_{p-1}^{P} H_p^T(j) W_p^{-1} H_p(\mathrm{j})]\Delta w(j) = \sum_{p-1}^{P} H_p^T(j)[G_p - h_p(w_p(j))]$$

$$\widehat{W}(j+1) = \widehat{W}(j) + \Delta w(j) \qquad \text{.....(1)}$$

Where H_p (j) is the measurement Jacobian of area p obtained with the local state estimate $\widehat{Wp}$ (j). The signal processing community can contribute to the research on MASE for the future grid by building upon recent advances in distributed estimation.

Remote Terminal Unit (RTU) in Smart Grid

RTU is an acronym for Remote Terminal Unit. An RTU is an electronic device that is controlled by a microprocessor. The device interfaces with physical objects to a Distributed Control System (DCS) or Supervisory Control and Data Acquisition (SCADA) system by transmitting telemetry data to the system. A remote terminal unit is a control device that helps the master device (SCADA), monitors and communicates with grid devices placed across different substation locations. They can be used in a wide variety of electrical and process automation devices where SCADA is required to monitor and control equipment. RTUs are used to collect data from the equipment in the field and to transmit commands back from the master

device. This can be done through wired connections (telephone lines, cables, Ethernet) or wireless connections. The data are sent to the operator in real time so that, they are aware of the situation and can carry out any necessary changes to the grid.

Restructuring of architecture for Smart Grid

MPSoC architecture for smart grid

A Multiprocessor systems-on-chip (MPSoC) is a single chip to meet strict requirements of embedded applications, such as real time, high performance and high reliability. Technology scaling has enabled the integration of an increasing number of Processor Elements (PEs) on a single chip (MPSoC). The increasing computational power in turn supports the ability of permitting higher number of applications to execute simultaneously on the chip. The increasing complexity and the demanding workload of these applications, coupled with short time-to-design and time-to-market of portable devices has given rise to the application specific Multiprocessor System on-Chip (MPSoC) platforms for multimedia. The increase in demand of the NoC-based MPSoCs in terms of processing elements (PEs), new applications can run simultaneously on such systems, which require management techniques to meet the application constraints.

In this architecture configuration, it is possible to design either Symmetric Ht-MPSoC or a AHt-MPSoC. In the case of SHt-MPSoC, the number of private processors and hardware accelerators, which are shared, are equal in number. Whereas in the case of AHt-MPSoC, the hardware accelerators are interfaced with the different processors and each processor differs with one another. The proposed Ht-MPSoC architecture shares the hardware accelerators among processors. The result is based on the MIP formulation, which is used to explore the very huge space of feasible configurations in reasonable time. In the AHt-MPSoC architecture, for improving the performance of the processors, application-specific instructions are effectively used. In these processors, the run time of the critical computations is degraded by the usage of newly added instructions implemented in HW accelerators. This HW accelerator utilization of each processor can be either connected by the system bus or memory controller to the instruction pipeline. The MILP model space explorations used to connect the computational patterns are represented by existing method on the various applications, which degrade the overall area utilization with respect to application-performance restraints. The necessary limit to be taken for producing the execution time of each processor is maintained. Thus, the MIP model is able to determine an ideal AHt-MPSoC configuration that attains desired area and performance. The proposed model, in order to prove the increasing performance and energy consumption of the system, is reduced compared to the conventional method. The modified MILP architecture is shown in Fig.1.

The conventional work focuses only on area reduction without considering the performance improvement. Whereas, the proposed work, on examining the entire feasible sharing configuration reduces the area usage and the execution time is desired based on the performance constraints. In all possible ways of sharing configuration, the usage of area diminishes and satisfies the performance in terms of its execution time.

Modified Integer Linear Programming for MPSoC configuration

An MPSoC is a complete computational system integrated in a single chip, combining multiple processing elements (PEs) as main components. In addition, the MPSoC is built by interconnecting Intellectual Property (IP) cores to local ports of NoC routers. Messages are transmitted between IP cores wrapped in NoC packets. The growing interest in MPSoCs lies in its ability to combine high performance and flexibility and its reconfigurable feature. In this work, the MPSoC uses multiple Remote Terminal Unit (RTUs) along with other Distributed State Estimator (DSEs) to implement a system. By sharing hardware accelerators among the cores, the conventional FPGA-based MPSoC architectures has been extended. In these architectures, cores on the FPGA may have different resources to share in different manners.

The application needs a fast and accurate exploration tool for exploring the enormous space of possible configurations of MPSoC on FPGA. Taking this reason, a Mixed Integer Linear Programming (MILP) model is proposed further in order to determine the MPSoC configuration. It consumes the least HW resources and also values the application execution time constraints. With the help of this MIP model, the design space of several hundreds of private and shared HW accelerators can be explored in a sensible time with high accuracy. Figure 1.1 shows the proposed MILP MPSoC architecture. As the MPSoC is designed with MILP, the computational complexity of the proposed system is significantly reduced.

Fault Indication and Propagation Infrastructure (FIPI)

An essential part of the demonstrator is the Fault Indication and Propagation Infrastructure (FIPI). When the faults are detected in the instruments, the operation of the MPSoC is threatened, unless proper action is taken. The FIPI employs an interrupt-like function that propagates fault indications into fault detection instrument and a system-level fault indication instrument, which controls a global fault indication. The fault origin is identified by the fault detection instrument, but this can take a particular time to access. The FIPI applies a noticeable track back to the instrument for easy indication of the fault origin.

Fault Tolerant MPSoC

The fault tolerance enables the correct operation of a device even in the presence of faults (errors) and it has been established along with the rise of the very first devices used in critical applications. To enable correct operation in the presence of errors, fault tolerance provides techniques that are capable of error-detection, i.e. to detect the presence of errors, and error-recovery, i.e. recover the system from errors. Usually, this is achieved by introducing a hardware and time redundancy. The methodology accepted in the scheme is defined as follows. In a NoC based homogeneous MPSoC, three different redundancy schemes are induced and the obtained results are compared. The first scheme is Dual Modular Redundancy (DMR), which is a hardware redundant technique. In DMR, two processes are performed simultaneously. By using comparator, the outputs are compared to check whether the fault has occurred or not.

To execute critical tasks, the hardware is needed additionally in this scheme. The Temporal Redundancy (TR) is the second technique in which two processes are performed sequentially and then, the comparator is used to compare the fault detection. The third technique is parity Error Correcting Code (ECC) which compares the input and the produced output to detect faults. It seems to have intermediate hardware compared to conventional schemes.

Logic Based Real time Multiprocessor Architecture (LRMA)

The Logic based Real time Multiprocessor Architecture (LRMA) is designed with the combination of DSE and RTU based MPSoC. The power consumption of this system is significantly minimized. An MPSoC is a system on-chip that includes most of the entire components essential for an application and it utilizes the multiple programmable processors as system components. The transmission parameters are transmitted to the corresponding RTU, which performs as an intermediate between all substations. In the proposed method, to minimize the execution time, the DSE is used. The proposed LRMA is attained with multiple factors that are low power consumption and low execution time.

Logic Based Fault Tolerance MPSoC (LFT-MPSoC)

The proposed Logic based Fault Tolerance MPSoC (LFT-MPSoC) is designed with the combination of DSE and RTU based FT-MPSoC method. The power consumption and fault tolerance of the proposed method are significantly achieved. Fault tolerant MPSoC has a good performance and low hardware overhead. It provides the reliability of the system under real-time constraints. In the event of the failure, Fault tolerance property enables a system to continue operating properly. The RTU acts as an intermediate between all substations. The DSE is utilized to reduce the execution time of the substation and to deliver the feedback on local faults. The numerous factors achieved by the proposed LFT-MPSoC are fault tolerant, reliability, low power consumption and low execution time.

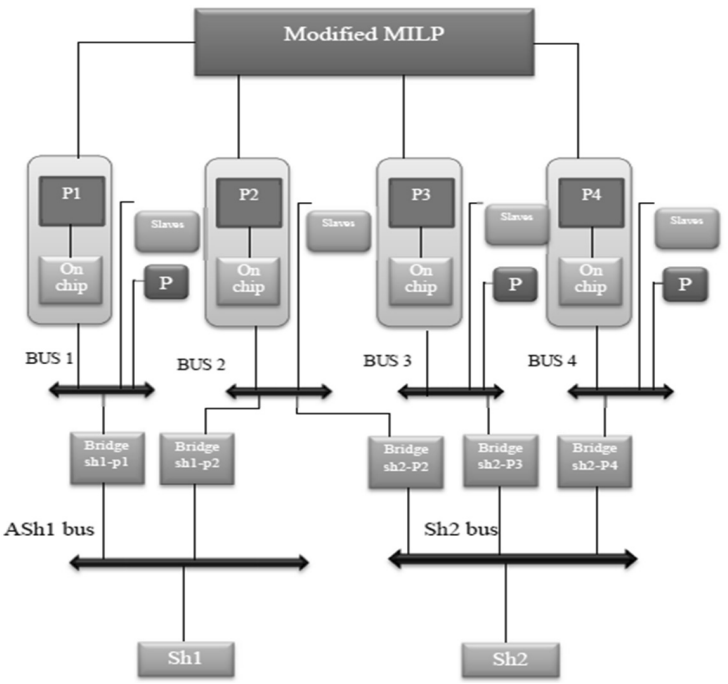


Figure 1. MILP MPSoC architecture

Testing Analysis Strategies

When there is increasing needs of performance and scalability, the commercial reconfigurable architectures such as available COTS FPGAs are used. Additionally, the usage of dynamically reconfigurable architectures can deliver a higher degree of flexibility, scalability and simultaneously maintains the computing power. The major challenge in MpSoC design is system integration. Complex hardware and software component interactions pose a serious threat to all kinds of performance pitfalls, including transient overloads, data loss and missed deadlines.

Static State Estimation (SSE)

For the last four decades, more researches on SE have been conducted about SSE. Peculiarly, for the reason that the ordinary monitoring technologies, comparable to those carried out in the SCADA method, can be best taken non synchronized measurements in each two to four seconds. Additionally, to reduce the computational complexity required in implementing SE, also it estimates and large up-to-date for simplest once in every few minutes. Thus, the usefulness of SSE as a means to furnish actual-time monitoring of the power grid is quite restrained to apply.

As specified by Yih-Fang Huang et al (2012), in an N-bus system, the (2N-1) x 1 state vector has the form $w=[\Phi_2, \Phi_{3,\ldots\ldots}\Phi_{N,} |E_N|]^T$ where Φ_i denotes the phase angles and |Ei| refers the magnitudes of the voltages at the *i*th bus. The phase angle Φ_1 at the reference bus is assumed and is normally set to zero radians. The measurements are typically obtained within SCADA systems, and are specifically related to the state vector by an over determined system of nonlinear equation given in equation (2).

$$G=h(w)+n \qquad \ldots..(2)$$

where h(w) is a set of M nonlinear functions of the state vector (determined by Kirchhoff's laws and the power network admittance matrix) and n is a zero-mean Gaussian measurement noise vector with covariance matrix $b_n \epsilon R^{MxM}$.

In the traditional SSE approach, the state vector is estimated from the measurement Equation in (2) using the Weighted Least Squares (WLS) method as mentioned by schweppe et al (1990). In particular, the SSE problem is solved by finding

$$\widehat{W}= \arg\min[G-h(w)]^T D^{-1}[G-h(w)] \qquad \ldots.. (3)$$

where weighting matrix D is commonly taken as diagonal with elements related to background noise covariance as D=Cn. The solution for $\widehat{W}$ is obtained in an iterative fashion by linearizing equation (2) around the available estimation (at iteration j), and applying the Gauss-Newton algorithm to improve the estimate, and by using the following equations:

$$N(j)\Delta w(j)=H^T(j)D^{-1}[G-h(w(j))] \qquad \ldots..(4)$$

$$\widehat{W}(j+1) = \widehat{W}(j)+ \Delta w(j) \qquad \ldots..(5)$$

where $N(j)=H^T(j)\, D^{-1}\, H(j)$ is the gain matrix at iteration j. Equation (4) is usually referred to as the normal equation. A more recent process for reducing the computational price is to use a nested, or multilevel, system of the nonlinear dimension model as mentioned by Gomez et al 2011. This method can sustain the development in size, complexity, and data information. It is designed to operate at extraordinary phases of the modelling hierarchy to accomplish very big-scale interconnection-vast monitoring. This method uses the same over determined set of measurement equations as in (2). The equations are then unfolded into L sequential WLS problems by introducing a set of intermediate variables $X=\{x_1,x_2,\ldots x_L\}$ with the following nested structure,

$$G=f_1(x_1)+n$$
$$x_{1=} f_2(x_2)+ n_1$$
$$\ldots\ldots\ldots\ldots$$
$$x_{L-1=} f_L(x_L)+ n_{L-1}$$
$$x_{L=} f_{L+1}(w)+ n_L \qquad .. (6)$$

The set X is chosen such that the solution of the nested system of Equations (6) offers some desired advantage over solving (2), e.g., reduction of the computational complexity or the amount of information exchanged between different levels. This is particularly an appealing solution, when the measurement model can be factorized into separate linear and nonlinear parts.

Latency

Latency is distinct at the time that intervenes the injection start of the its data into the network of the transmitter section and its arrival at the receiver section. The latency is used as the performance metric. It is assumed that the packet latency time is the instant when the packet is created, to the time when the packet is delivered to the receiver section. Also,it is assumed that the packets are consumed immediately once they reach their receiver section. For a flit to reach the receiver section, it must travel through a path consisting of a set of links and other components.

Using network calculus, the latency D_{nocQi} in each node Q_i is expressed as follows.

$$D_{nocQi} = \frac{B_{wij}}{U_{ii}} + Llat \qquad \ldots..(7)$$

where B_{wij} is the service bandwidth and L_{lat} is the latency. Thus, the latency can be intended based on the equation. The latency for the application can also be intended after evaluating the latency D_{bi} of each switch b_i. In the simulator, each process is linked with a traffic generator that injects flits according to the real time traffic model at a deterministic rate which

is varied from low traffic (data) to heavy traffic (video). In this evaluation, data flows are considered dependent on each other and the interaction causes congestion in some switches. It increases the latency at a high injection rate.

Power

Link power

From the power models, for a NoC router and the power model considering the cross-coupling effect for N-wire interconnect, we may determine the total power for an N-wire link per unit length as follows: gate leak wire bias short.

$$P_{link} = \frac{1}{2} N_{wire} V_{sv}^2 (C_{self}\alpha_{saw} + (C_{ou}\alpha C_{ou})f + N\tau\alpha_{saw} VI_{short}.f + N.(VI_{bias\text{-}wire} + VI_{leak,gate}) \quad \text{.....(8)}$$

where N_{wire} is the total number of wires in the link, C_{self} and C_{coupl} are the self and coupling capacitance of a wire and neighbouring nodes, respectively. α_{saw} is the switching activity on a wire and α_{Cou} is the switching activity with respect to the adjacent wires, τ is the short circuit period, V_{sv} is the supply voltage and $I_{short,bias}$, wire and $I_{leak,gate}$ are currents

Static power consumption

Static power is the power immoral by a gate or a wire, after it is inactive or in an active state. The static power is mostly inclined by the structure of the circuit. The static power dissipation can be more precise by the equation

$$E_{Static} = VI_{bias,wire} + VI_{leak,gate} \quad \text{.....(9)}$$

Energy

At upper levels, energy spent due to the scattering of one bit of data from one router (R_1) to another (R_2) via the links is a utility of the number of routers and the number of links. The full energy can be talked into the energy spent on the replacements and energy spent per wire or link coldness traveled. The total energy(t) can be intended as follows

$$\text{Energy(t)} = \sum_{i=1}^{Ni} \Phi_{consumed@time}(t) + \sum_{i=1}^{Ni} \Phi_{consumed@switch,link}(t) \quad \text{.....(10)}$$

where $\Phi_{consumed@time}(t)$ is the energy spent, at time t, on the link l_i, $\Phi_{consumed@consumed@time)}(t)$ is the energy consumed inside the switch sw and N_{link} and N_{switch} are the number of links and switches, respectively involved in transporting the application flows. Using network calculus arrival curves, the total energy consumption can be calculated.

Results and Discussion

From the proposed modified MPSoC architecture, the performance parameter model has attained the optimal results for the following parameters. They are delay, energy consumption and power consumption. In the development section of this proposed scheme, the work is done under the combination of packages. This computes further improvement, which is done in the previous conventional works to improve the performance of the smart grid architectures. As mentioned above, the performance measurements are compared with the existing system.

Table 1. Energy Consumption comparison of Proposed AHt-MPSoC Architecture

Energy (j)					
Method	Barnes	Ocean	Radiosity	Raytrace	Avg
SHT	0.275	0.312	0.298	0.278	0.29
AHT	0.215	0.278	0.266	0.233	0.248
VRMA	0.118	0.185	0.166	0.185	0.163

Table 2. Delay comparison of Proposed AHt-MPSoC Architecture

Delay(Ps)					
Method	Barnes	Ocean	Radiosity	Raytrace	Avg
SHT	78	96	86	77	84
AHT	66	74	66	55	65
VRMA	45	48	47	49	47

Table 1 and Figure 2 depicts that the comparison of energy consumption of different architectures. The proposed AHt-MPSoC Architecture is achieved to reduce the energy consumption of different benchmark which gives better performance compared with SHT and AHT-MPSoC architecture. In Barnes, the energy consumptions of the SHT, AHT and the proposed

architecture are 0.275, 0.215 and 0.118 respectively. It shows that the proposed AHt-MPSoC architecture gives better improvement than the conventional architecture.

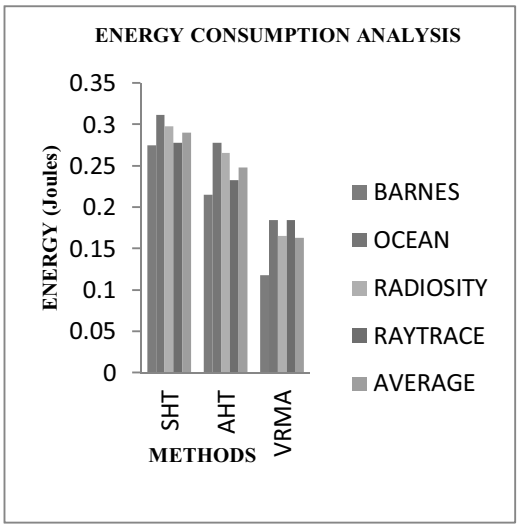


Figure 2. Energy Consumption comparison of Proposed AHt-MPSoC Architecture

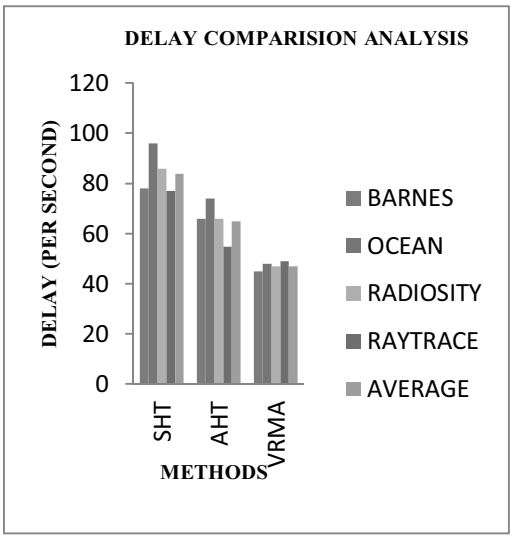


Figure 3. Delay Comparison of Proposed AHt-MPSoC Architecture

The proposed AHt-MPSoC Architecture is attained to minimize the delay of different benchmark which gives better enhancement compared with SHT and AHT-MPSoC architecture. Delay comparison of different architecture is shown in Table 2.In Barnes, the delay of the SHT, AHT and the proposed architecture are 78, 66 and 45 respectively, which shows that the proposed AHt-MPSoC architecture is better than the conventional architecture and it is shown in Figure 3.

Table 3, 4 and Figure 4,5 presents the comparison of static power of different architectures with FT-MPSoC. Static power minimization of different benchmark is achieved by the proposed FT-MPSoC Architecture and it gives better attainment compared with the base MPSoC architecture. It depicts that the proposed FT-MPSOC architecture provides better improvement than the conventional architecture.

Table 3. Static Power Comparison of Proposed AHt-MPSoC Architecture

Static Power(mw)					
Method	Barnes	Ocean	Radiosity	Raytrace	Avg
SHT	220	240	255	233	237
AHT	180	190	196	210	194
VRMA	120	110	121	123	118

Table 4. Static Power Comparison of Proposed MPSoC with FT unit Architecture

Static Power(mw)					
Method	Barnes	Ocean	Radiosity	Raytrace	Avg
MPSoC	180	190	196	210	194
FT-MPSoC	150	140	135	138	140

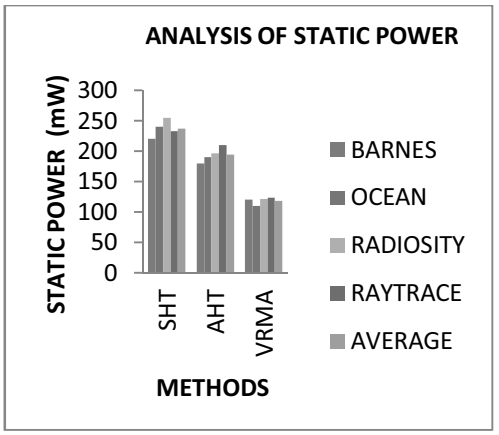


Figure 4. Static Power Comparison of Proposed AHt-MPSoC Architecture

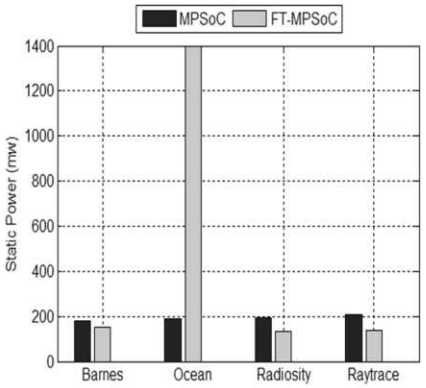


Figure 5.Static power Comparision of MPSoC with proposed Fault tolerant MPSoC

Conclusion

In this paper, the significance of restructuring the existing smart grid architecture using MPSoC is designed. The significance of the fault tolerant unit FT-MPSoC with existing smart grid architecture is designed The MPSoC is designed with modified MILP algorithm and the computational complexity of the proposed system is significantly reduced. Hence, the MPSoC with the fault tolerant unit is designed and the architecture performs better than the existing architectures. As the proposed MPSoC

is designed with modified architecture the delay, energy and Power consumption are calculated and they compared with the existing architecture which gives better performance than the existing architectures.

References

[1] Al-Ali AR , "Smart grid controller for optimizing HVAC energy consumption", International conference on Computer Systems and Industrial Informatics (ICCSII), 2012, pp.1-4.

[2] Amin, Mohsin, Mihkel Tagel, Gert Jervan & Thomas Hollstein , "Design methodology for fault-tolerant heterogeneous MPSoC under real-time constraints", IEEE 7th International Workshop on Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC), 2012, pp.1-6.

[3] Ben Atitallah, Rabie, Eric Senn, Daniel Chillet, MickaelLanoe & Dominique Blouin , "An efficient framework for power-aware design of heterogeneous MPSoC", IEEE Transactions on Industrial Informatics, vol 9, no.1, 2013, pp.487-501.

[4] Ceng, Jianjiang, Jeronimo Castrillon, Weihua Sheng, Hanno Scharwachter, Rainer Leupers, Gerd Ascheid, Heinrich Meyr, Tsuyoshi Isshiki & Hiroaki Kunieda , "MAPS: an integrated framework for MPSoC application parallelization", ACM Proceedings of the Design Automation Conference, 2008, pp. 754-759.

[5] Chantem, Thidapat, Sharon Hu & Robert P Dick , "Temperature-aware scheduling and assignment for hard real-time applications on MPSoCs", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol 19, no.10, 2011, pp. 1884-1897.

[6] De Souza Gomes A , "Detection and classification of faults in power transmission lines using functional analysis and computational intelligence", IEEE Transactions on Power Delivery, vol.28, 2013, pp.1402-1413.

[7] Grammatakakis, Miltos, Kyprianos Papadimitriou, Polydoros Petrakis, AntonisPapagrigoriou, George Kornaros, IoannisChristoforakis, OthonTomoutzoglou, George Tsamis & Marcello Coppola , "Security in MPSoCs: A NoC Firewall and an Evaluation Framework", IEEE Transcations on Computer Aided Design of Integrated Circuits and Systems,vol.34, 2015, pp.1344-1357.

[8] Gungor, VC , "Smart grid technologies: communication technologies and Standards", IEEE Transactions on Industrial Infrastructure, vol.7, 2011, pp.529-539.

[9] Hairong, Q , "A resilient real-time system design for a secure and reconfigurable power grid", IEEE Transactions on Smart Grid, vol.2, 2011, pp.770-781.

[10] Huang, Lin, Feng Yuan & Qiang Xu , "Lifetime reliability-aware task allocation and scheduling for MPSoC platforms", Conference on Design, Automation and Test, 2009, pp. 51-56.

[11] Huang, Lin, Feng Yuan & Qiang Xu , "On task allocation and scheduling for lifetime extension of platform-based MPSoC designs", IEEE Transactions on Parallel and Distributed Systems, vol.22, no.12, 2011, pp. 2088-2099.

[12] Liu Weichen, Yu Wang, Xuan Wang, Jiang Xu & Huazhong Yang, "On-Chip Sensor Network for Efficient Management of Power Gating-Induced Power/Ground Noise in Multiprocessor System on Chip", IEEE Transactions on Parallel and Distributed Systems, vol.24, no.4, 2013, pp.767-777.

[13] Loghi, Mirko, Federico Angiolini, Davide Bertozzi, Luca Benini & Roberto Zafalon, "Analyzing on-chip communication in a MPSoC environment", IEEE Proceedings on Design, Automation and Test, vol.2, 2004, pp.752-757.

[14] Marongiu, Andrea & Luca Benini, "An OpenMP compiler for efficient use of distributed scratchpad memory in MPSoCs", IEEE Transactions on Computers, vol.61, no.2, 2012, pp.222-236.

[15] Martin & Grant, "Overview of the MPSoC design challenge", ACM Proceedings of the 43rd Annual Design Automation Conference, 2006, pp. 274-279.

[16] Monticelli, "Electric power system state estimation", IEEE Proceedings, vol.88, no.2, 2000, pp.262–282.

[17] Motra, AP, Bora & Chakrabarti,I, "An efficient hardware implementation of DWT and IDWT", International conference on Convergent Technologies for Asia-Pacific Region TENCON, vol.1, 2003, pp.95–99.

[18] Pagani, GA & Aiello, M, "The power grid as a complex network: a survey", International Journal on Physica A:Statistical Mechanics and its Applications, vol.392, 2013, pp.2688-2700.

[19] Roberta Piscitelli & Andy D Pimentel, "A High-Level Power Model for MPSoC on FPGA", IEEE Transactions on Parallel and Distributed Process, vol.11, no.1, 2012, pp.16-20

[20] Rongxing,L, "EPPA: An efficient and privacy-preserving aggregation scheme for secure smart grid communications", IEEE Transactions on Parallel Distributed Syststem,vol.23, 2012, pp.1621-1631.

[21] Wang, Xuan, Jiang Xu, Wei Zhang, Xiaowen Wu, Yaoyao Ye, Zhehui Wang, Mahdi Nikdast & Zhe Wang5, "Actively Alleviate Power Gating-Induced Power/Ground Noise Using Parasitic Capacitance of On-Chip Memories in MPSoC", IEEE Transactions on Very Large Scale Integration Systems, vol.23, 2015, pp. 266-279.

[22] Wu, FF, "Power system state estimation: A survey", International Journal of Electrical Power Energy System, vol.12, no. 2, 1990, pp.80–87.

[23] Xie, Yuan & Wei-Lun Hung, "Temperature-aware task allocation and scheduling for embedded multiprocessor systems-on-chip (MPSoC) design", Journal of VLSI signal processing systems for Signal, Image and Video Technology, vol.45, no.3, 2006, pp.177-189.

[24] Yang, Hoeseok & Soonhoi Ha, "Pipelined data parallel task mapping/scheduling technique for MPSoC", IEEE International Conference on Design, Automation and Test, 2009, pp.69-74.

[25] Ye, Terry Tao, Luca Benini & Giovanni De Micheli, "Packetized on-chip interconnect communication analysis for MPSoC", IEEE Conference on Design, Automation and Test, 2003, pp.344-349.

The ASVI - Assistance System for Visually Impaired

Arvind Prasad*, H Govindsingh**, R Hari Krishnan*** and C N Ravi****
* ** *** **** Sathyabama University, Chennai 600 019
arvindprasad123@gmail.com, govindpurohit321@gmail.com, rhareish@gmail.com,dr.ravicn@gmail.com

Abstract: The ASVI- Assistance system for visually impaired incorporates the use of a mobile robot obstacle avoidance system as an assisting device for the visually impaired. During operation the mobile robot moves in front of the user at a prefixed distance and guides the user through auditory signals. Just as electronic signals are sent to a mobile robot's motion control system, auditory signals can guide the visually impaired around the obstacle. An Android phone is used to provide the direction to the destination from the current location of the user. The user can steer the mobile robot to the desired direction as per the navigation from the phone manually with the help of a hand held joystick. The robot's obstacle avoidance system will steer the robot avoiding obstacles safely.

Keywords: Visually Impaired, Ultrasonic Sensor, Infrared Sensor, Assistance System for visually impaired (ASVI), Arduino.

Introduction

The term blindness is used for complete or nearly complete vision loss. Visual impairment may cause people difficulties with normal daily activities such as driving, reading, socializing, and walking. People with complete blindness or low vision often have a difficult time self-navigating outside well-known environments. In fact, physical movement is one of the biggest challenges for the visually impaired. Because of this, many people with low vision bring a sighted friend or a trained guide dog to help navigate unknown environments.

There are various tools that have been developed for the visually impaired. IDog [1], the Guide dogs provide the impaired person with the highest degree of mobility and independence, but require expensive training and selective breeding. Navbelt [2], provides information not only about obstacles along the traveled path, but also assists the user in selecting the preferred travel path. The Navbelt consists of a belt, a portable computer, and ultrasonic sensors.

Guidecane [3], the GuideCane during operation, the user pushes the lightweight GuideCane forward. When the GuideCane's ultrasonic sensors detect an obstacle, the embedded computer determines a suitable direction of motion that steers the GuideCane and the user around it. The steering action results in a very noticeable force felt in the handle, which easily guides the user without any conscious effort on his/her part. RGB-D camera [4], instead of using several sensors, a consumer RGBD camera is used to take advantage of range and visual information. In particular, the main contribution is the combination of. It constitutes of a camera that hangs from the user's neck, and a laptop carried in a backpack. ROVI-[5], in this system, the white cane is attached to a mobile platform which will has an array of sensors used to detect obstacles. The mobile robot will be given a predefined target to move. The sensors will always look for any obstacles along the path. It will avoid by an angle once the obstacle is detected. The deflection from original path due to presence of obstacle will be controlled by microcontroller. It will also provide instruction to the wheels to return to the original path once the obstacle has been avoided. The feedback signals from the optical encoders attached to the wheels are important to determine the distance travelled by the wheels.

Almost all the existing systems incorporate high performance microprocessors or mini computers or even laptops to process the input. This increases the cost involved in the project. Also most of the existing systems can only navigate the user to a prefixed target.

Proposed System

The proposed system works on the similar principle as the guide stick [6] concept. But instead of the user holding the stick connected to the robot, the user will have a joystick in his hand, to steer the robot moving in front of him at a particular distance. The proposed system can navigate the user to any location and also detect potholes and prevent the user from tripping over. When an obstacle is encountered, the robot will move around the obstacle and emit sound signals that will guide the user around the obstacle [7].

It is observed that for visually impaired people to navigate the shape of the obstacle does not matter, instead the user just wants to move around the obstacle. As the robot moves by its own, unlike the guide cane concept the user will also be warned about potholes, thereby increasing the safety of the user.

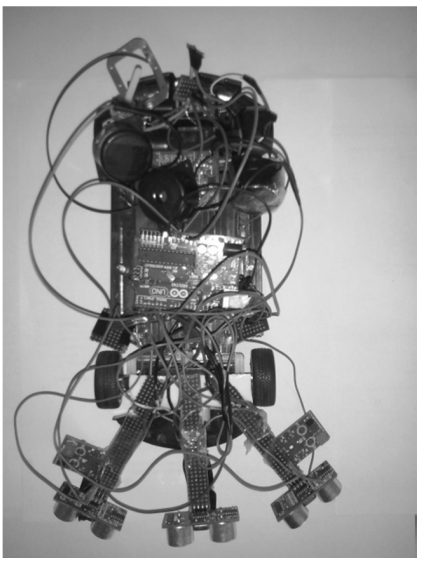


Fig 1. ASVI Prototype

Since audio signals are used to steer the user around the obstacle, the navigation is more precise than the various existing systems in which there are chances that the user brushes past the obstacle [8]. As the robot moves at a distance from the user, there is always a minimum distance between the user and the obstacle. As the robot can be controlled by a joystick, the user can steer the robot to any particular direction depending on the navigation from an android phone if necessary. The proposed system consists of an Arduino board, ultrasonic sensors, infrared sensors, potentiometer, motor driver IC, DC Motors, buzzers.

The Arduino board has a processing speed of 16 MHz. This processing speed is enough for processing the data received from the ultrasonic as well as infrared sensors. The Arduino needs a 5V supply to work.

The ultrasonic sensor used is HCSR04. It has 4 pins +VCC, GND, Trig, Echo. The trigger pin of the ultrasonic sensor must be made high for 10 microseconds for the ultrasonic sensor to be triggered. The transmitted ultrasonic signal reflects back after striking an object if present. The frequency of the pulses reflected by the object and detected by the receiver is recorded. This is used to calculate the distance of the sensor from the obstacle. The number of pulses detected by the detector when divided by 58.2 gives the distance of the object from the sensor in centimeters. In our proposed system the ultrasonic sensors are used to detect the obstacles in front of the user. These ultrasonic sensors will give the distance data to be processed, to the Arduino. The Arduino calculates the distance in cm. The program in the Arduino is such that when the distance is below a particular threshold, the necessary commands to steer the robot away from the obstacle is executed.

The infrared sensor also has a transmitter and receiver. It has 3 pins, +VCC, GND, Signal pin. If the transmitted beam is received, then signal pin is high, else low. The infrared sensor is generally used in a line follower robot. But the infra red rays are also reflected by a smooth non black floor surface. So this property of the infrared sensor is used to detect the presence of potholes. In the absence of the surface ie: in the presence of a pothole, the transmitted ray will not be reflected back, thereby setting the signal pin low. The user must be allowed to move only if the surface if the surface is flat for the reason of safety. So the checking of the presence of the pothole is set as the first condition. Only if the infrared sensor is high then the rest of the program is executed. Else the stop command is executed.

The system uses 2 DC Motors. One for the forward and backward movement and another for the left, right movement of the robotic system. The motors are driven by a motor driver IC L293D.The robotic system can manually steered depending on the users need, and also is automatically steered in the presence of an obstacle.

Triggering the sensors

The ultrasonic sensor send a sound wave only if the trigger pin is made high for a duration of 10 micro second. So in order to trigger the ultrasonic sensor, the triggerport should be made high accurately for a duration of 10 micro seconds. The infrared sensors do not need any trigger.

Obtaining data from sensor

Both the ultrasonic sensors as well as the infrared sensors work by the echo principle. The sent wave is reflected back from an obstacle. The infrared sensor receives the sent signal in the presence of an obstacle and the range can be adjusted by a

potentiometer. So the input from an infrared module is either Logic 0 in the absence of the obstacle or a Logic 1 in the presence of an obstacle.
The ultrasonic sensor's echo pin gets the reflected wave in pulses. So the frequency of the pulse varies with the distance. If the distance is far, then the frequency of echo pulses is less and vice versa.To calculate the distance in cm, the duration of the pulses is measured and divided by 58.2 .

Working with the input from Sensors
The infrared sensors produce only 2 types of outputs either logic 1 or 0. But the ultrasonic sensors distance has range of upto 400cm. So ultrasonic sensors distance must be calculated and a prefixed distance must be set, below which the necessary action must be taken.
This project uses ultrasonic sensors and infrared sensors instead of an RGB-D camera or stereo camera to navigate the visually impaired person. This change in design reduces the cost of the system and also does not need a high performance processor.
The robot can be manually steered by a joystick shown in Fig 3 to enable the user to be navigated to a desired location using the help of voice navigation from an android phone. The proposed system uses an Arduino board, L293D motor driver IC, DC motors, logic gates, HC-SR04 ultrasonic sensors and infrared sensors.

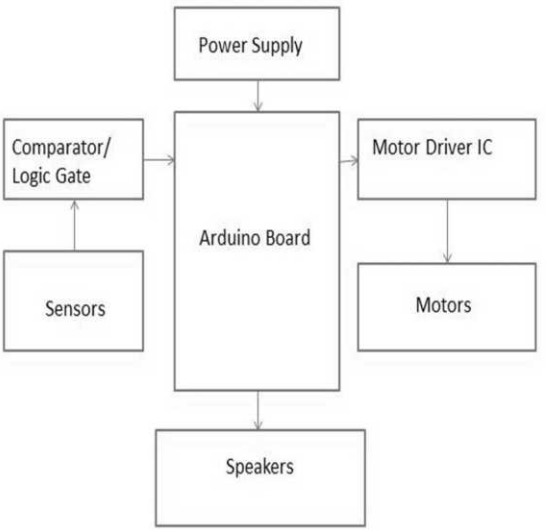


Fig 2. Block Diagram of the proposed system

Fig 3. Joystick module

Performance Analysis

A performance analysis was conducted and it was found that, since auditory guidance is used, the precision with which the user navigates around the obstacle is much higher than in the existing systems using physical guidance. Also since the robot moves at a distance from the user and the ultrasonic sensor's range is high, the user always maintains a safe distance from the obstacle. It was found that the robot cannot detect vehicular movement, therefore it is not to be used in environments with

vehicular traffic. It can be used in environments where there is a special walk way or pavement for walking. Since the robot has infrared sensors, any pothole can also be detected and the user can be steered around it. The robot cannot tackle staircases. The training time required is minimal. The user just needs to become familiar with the sounds emitted and the amount he should turn.

If there was a walkway and a predefined map of the environment then the robot can be programmed more efficiently taking into consideration the road crossings etc. The area covered by the sensors is shown in Fig 4. Even though the area covered is lesser compared to the existing designs, it is ensured that the path the user will take behind the robot will be obstacle free. Fig 5 is a screen shot of the serial monitor feature in the Arduino software. This feature helps us to observe calculated distance value by the ultrasonic sensor data. This also helps us in calculating the error with the actual distance value.

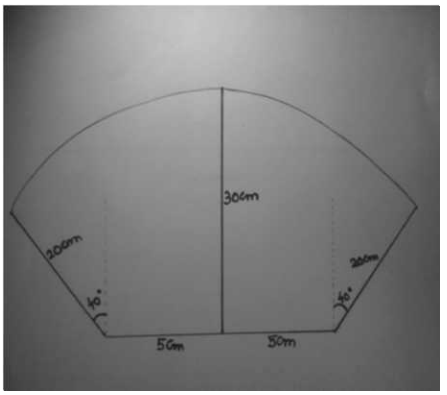


Fig 4. Area covered by sensors

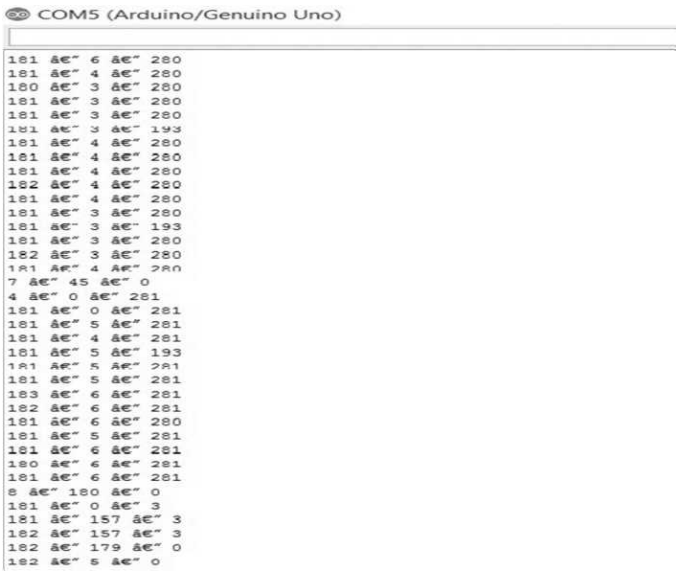


Fig 5. Sample Output in Serial Monitor

Conclusion

The ASVI works using the similar hardware as compared to the existing systems, but the brain of the system is just an Arduino. The ASVI offers solution to a few shortcomings

- ASVI can be steered manually so, the flexibility of usage is increased, which is absent in the pre-existing systems.
- ASVI can also detect potholes thereby increasing the safety of the user.
- ASVI does not need a laptop or a high specification microprocessor for processing the inputs, thereby reducing the cost
- ASVI uses sound guidance which is more accurate than other forms of physical guidance.

ASVI can be further developed by incorporating an inbuilt GPS navigation system, so the user can concentrate more on the robots voice assistance rather than the robot's and the phones navigation. ASVI can also be designed to detect staircases.

References

[1] Georgios Galatas, Christopher McMurrough, Gian Luca Mariottini*, FilliaMakedon “iDog : An Assistive-Guide Robot for the Visually Impaired” 2011
[2] ShragaShoval, Johann Borenstein, and YoramKoren “Auditory Guidance with the Navbelt—A Computerized Travel Aid for the Blind” IEEE transactions on systems , manand cybernetics—Part C: Applications and review VOL. 28, NO. 3, August 1998
[3] Iwan Ulrich and Johann Borenstein “The GuideCane—Applying Mobile Robot Technologies to Assist the Visually Impaired” IEEE transactions on systems ,manand cybernetics—Part C: Applications and review VOL. 31, NO. 2, March 2001
[4] A. Aladrén, G. López-Nicolás, Member, IEEE, Luis Puig, and Josechu J. Guerrero, Member, IEEE “Navigation Assistance for the Visually Impaired Using RGB-D Sensor With Range Expansion” IEEE systems journal 2014
[5] Navigation A. Allan Melvin, B. Prabu, R. Nagarajan , Bukhari Illias “ROVI: A Robot for Visually Impaired for Collision- Free” Proceedings of the International Conference on Man-Machine Systems 2009
[6] Sung Jae Kang', Young Ho, Kim', In HyukMoon ”Development of an Intelligent Guide-Stick for the Blind”Proceedings of the 2001 IEEE Seoul, Korea. May 21-26, 2001 International Conference on Robotics 8, Automation
[7] Edwige PISSALOUX “A Characterization of Vision Systems for Blind People Mobility”. IEEE SMC 2002
[8] Jack M loomis, Reginald G Golledge, Roberta L Klatzky ”Navigation System for the Blind: Auditory Display Modes and Guidance” Presence, Vol. 7, No. 2, April 1998, 193–203 MIT

Biometric Recognition System (Algorithm)

Rahul Kumar Jaiswal* and Gaurav Saxena**
*The LNM Institute of Information Technology
rahulkumarjaiswal@gmail.com
**Jaypee University of Engineering and Technology
gaurav.saxena@juet.ac.in

Abstract: Fingerprints are the most widely deployed form of biometric identification. No two individuals share the same fingerprint because they have unique biometric identifiers. This paper presents an efficient fingerprint verification algorithm which improves matching accuracy. Fingerprint images get degraded and corrupted due to variations in skin and impression conditions. Thus, image enhancement techniques are employed prior to singular point detection and minutiae extraction. Singular point is the point of maximum curvature. It is determined by the normal's of each fingerprint ridge, and then following them inward towards the centre. The local ridge features known as minutiae is extracted using cross-number method to find ridge endings and ridge bifurcations. The proposed algorithm chooses a radius and draws a circle with core point as centre, making fingerprint images rotationally invariant and uniform. The radius can be varied according to the accuracy depending on the particular application. Morphological techniques such as clean, spur and H-break is employed to remove noise, followed by removing spurious minutiae. Templates are created based on feature vector extraction and databases are made for verification and identification for the fingerprint images taken from Fingerprint Verification Competition (FVC2002). Minimum Euclidean distance is calculated between saved template and the test fingerprint image template and compared with the set threshold for matching decision. For the performance evaluation of the proposed algorithm various measures, equal error rate (EER), D_{min} at EER, accuracy and threshold are evaluated and plotted. The measures demonstrate that the proposed algorithm is more effective and robust.

Keywords: Fingerprint images, Singular point, Minutiae extraction, ridge endings, ridge bifurcations, Morphological technique, Euclidean distance, Fingerprint matching.

Introduction

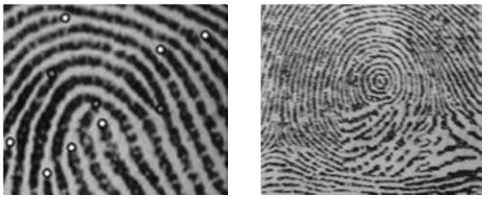
Biometric is the science of automatically and uniquely recognizing individuals based upon one or more instinctive physical or behavioral characteristics [12]. Present day biometric systems are becoming an essential component of effective personal identification. Fingerprints are the most widely adopted tool for personal identification among all biometrics due to easily acceptance, distinctiveness and permanence. Despite the ingenious methods improvised to increase the efficiency of the manual approach to fingerprint indexing and matching, the ever growing demands on fingerprint recognition quickly became overwhelming. Therefore, fingerprint identification is one of the main motives of various research endeavors in the field of Pattern recognition over the last few decades. The principle area of fingerprint identification includes law enforcement, criminal identification, fraud prevention, and computer access etc [15].

The main purpose of biometric recognition is to minimize the degree of forgery of the desired system and to improve one or more perceptual aspects of identification, such as the quality and/or matching accuracy so that the required service will be accessed by the legitimate user only [20]. A fingerprint is a unique pattern of ridges and valleys on the surface of a finger of an individual. A ridge is defined as a single curved segment, and a valley is the region between two adjacent ridges. The classification of fingerprint is based on micro and macro features [13]. Macro features include ridge patterns like loop, whorl and arc, along with core point, delta point and ridge count. Core point is found at the center of finger image and it is the uppermost of the innermost curving ridge whereas delta point is the meeting point of two or three ridges. Micro features include minutiae, it's orientation and position. Minutiae points are the local ridge discontinuities. These are of two types namely; terminations, which is the immediate ending of a ridge and bifurcations, which is the point on the ridge from which two branches derive [1]. A good quality image has around 40 to 100 minutiae [5]. Minutiae points are used for determining uniqueness of a fingerprint. Orientation refers to the direction in which a minutiae feature appears to be moving and position implies the relative location of the minutiae. The micro and macro features are shown in Fig. 1 and Fig. 2.

Fingerprint recognition system is categorized as enrollment, verification, and identification process system [14]. Enrollment includes capturing image, extracting feature, creating templates and making database. Verification authenticates a person's identity by comparing captured biometric traits with previously enrolled biometric reference template, pre-stored in the system. It conducts one-to-one comparison. Identification or Authentication recognizes an individual by searching the entire

(a) (b) (c)

Fig. 1. Macro features: ridge pattern (a) arc (b) loop (c) whorl



(a) (b)

Fig. 2. Micro features: minutiae points (a) terminations (b) bifurcations

enrolled template database for a match. It conducts one-to-many comparisons. Authentication can be understood as a combination of four RIGHTS [20].

Fig. 3. Components of authentication

In this paper, Fingerprint matching system based on minutiae extraction is proposed to enhance the accuracy of the biometric system degraded by elements of noise and large variability in different impressions of same finger [4]. The minutiae points are determined by cross number [8, 21]. The cross number is 1 for termination points and for bifurcation points, it is 3. Our first part of the problem is to identify fingerprints based on feature extraction. After getting minutiae as feature vector in region of interest (ROI), templates of fingerprint database are created. The minimum Euclidean distance is determined and based on certain threshold, matching will be decided with the database saved templates [7]. The proposed recognition algorithm allows in finding the best tradeoff between the radius of circle, matching accuracy and the space requirements in selection of radius in a perceptive view.

The paper is systematized as follows. In Section 2, the review of Pattern based fingerprint recognition method is described with its connection in fingerprint matching. In Section 3, the Minutiae based finger print recognition approach is described. In Section 4, the simulation results, performance measures and analysis of proposed algorithm is detailed. Finally, paper is concluded in Section 5.

Pattern based Fingerprint Recognition Method:

Pattern based fingerprint recognition method compares the basic fingerprint patterns (arch, loop, and whorl) between a previously stored template and a test fingerprint. The necessity of this method is that the images should be aligned in the same orientation. To do this, the algorithm detects a central point in the fingerprint image and centers on that point. In this method, the template contains the type, size, and orientation of patterns within the aligned fingerprint image. The test fingerprint image is graphically compared with the previously stored template to determine the degree of match [3]. The major disadvantage of pattern based fingerprint recognition method is that they are sensitive to proper placement of finger and need large storage for templates [7]. Several variations of pattern based fingerprint recognition method have been proposed to overcome this problem. The pattern of fingerprint [16-18] is shown in Fig. 1.

Minutiae based Fingerprint Recognition Algorithm

The minutiae based fingerprint recognition algorithm is relatively stable, robust to contrast, image resolutions, and global deformation as compared to pattern based fingerprint recognition method [6]. This approach is the backbone of the current available fingerprint recognition system. Fingerprint identification with minutiae extraction is mainly based on the minutiae points i.e. the direction and location of the ridge endings and bifurcations along the ridge path [11]. It reduces the complex fingerprint recognition problem to a point pattern matching problem. The proposed algorithm includes pre-processing of fingerprint image, feature extraction, post-processing and finally matching decision. Essentially, the matching consists of finding the minimum difference of distance (D_{min}) between the saved template and the test minutiae sets having maximum number of minutiae pairings. The test fingerprint matches if (D_{min}) is lower than the set threshold.

Step-by-step description of Algorithm:

Step 1: Input the fingerprint image, f (x, y).

Step 2: Conversion of image f (x, y) into grayscale image, $f_g(x, y)$.

Step 3: Resizing image $f_g(x, y)$ to 400 x 400, new image $f_r(x, y)$.

Step 4: Enhancing image using histogram equalization and wiener filter to improve quality, degraded by noise like smudgy area, break in ridge, wounds and sweat. The histogram of a digital image with gray levels in the range [0, L-1] is a discrete function, defined as [15]

$$h(r_k) = n_k \quad (1)$$

r_k - k^{th} Gray level,
n_k - Number of pixels in the image.

Step 5: Finding the core point of the fingerprint image $f_r(x, y)$.

The image is divided into a non-overlapping blocks of size 'w' 10 x 10. The horizontal gradient '$G_x(x, y)$' and vertical gradient '$G_y(x, y)$' at each pixel (x, y) is computed using Sobel mask of size 3 x 3 and the ridge orientation '$\theta(x, y)$' of each pixel is given by [2],

$$G_{xx} = \sum_{(x,y)\epsilon w} G_x^2(x, y) \quad (2)$$

$$G_{yy} = \sum_{(x,y)\epsilon w} G_y^2(x, y) \quad (3)$$

$$G_{xy} = \sum_{(x,y)\epsilon w} G_x(x, y).G_y(x, y) \quad (4)$$

$$\theta(x, y) = \frac{1}{2}\tan^{-1}\left(\frac{2G_{xy}}{G_{xx}-G_{yy}}\right) \quad (5)$$

Now, ridge orientation is smoothed using Gaussian low pass filter. As singular point has the maximum curvature. so, it is located by measuring strength of the peak. Further, applying thinning followed by Morphological closing and opening to locate singular point in original fingerprint image.

Step 6: Extraction of a circle of radius 'R' with core point as centre of the fingerprint image $f_r(x, y)$ to get new image $f_c(x, y)$ in the region of interest (ROI) because area near singular point contains correct and efficient information about fingerprint.

Step 7: Conversion of image $f_c(x, y)$ into binary image $f_b(x, y)$ by thresholding [10]. Pixel value above the threshold is assigned to 1 and below to 0. Here threshold = 160.

Step 8: Applying thinning operation on the image $f_b(x, y)$ to get thinned image $f_t(x, y)$. Thin operation reduces width of ridges to one pixel wide.

Step 9: Extracting minutiae points (terminations and bifurcations) of $f_t(x, y)$ using Cross-number (CN) concept [8]. It is computationally efficient and inherently simple. The minutiae points are extracted by scanning the local neighborhood of each pixel in the ridge thinned image, using a 3 x 3 window (Fig. 4).

a)	P_4	P_3	P_2	b)	1	**0**	1	c)	1	**0**	1
	P_5	**(x,y)**	P_1		1	**(x,y)**	1		1	**(x,y)**	**0**
	P_6	P_7	P_8		1	1	1		1	**0**	1

Fig 4. a) 3 × 3 window b) Ridge ending and c) bifurcation

The CN value [8] is defined as half the sum of the differences between pairs of adjacent pixels, P_i and P_{i+1} in the eight neighborhood and computed by,

$$CN_{(x,y)} = \frac{1}{2}\sum_{i=1}^{8} |P_i - P_{i+1}| \ , \ P_9 = P_1 \quad (6)$$

The ridge pixel is classified as a ridge ending and bifurcation having cross-number 1 and 3 respectively. Data matrix is generated to get the position, orientation and type of minutiae.

Step 10: Post-processing to remove spurious minutiae, observed due to undesired spikes, breaks, and holes. Morphological operation [9] namely clean, spur and H-break is employed on thinned image $f_t(x,y)$ to get image $f_m(x,y)$ as described,

	0	0	0		0	0	0	
Clean:	0	1	0	becomes	0	0	0	after clean operation
	0	0	0		0	0	0	
	0	0	0		0	0	0	
Spur:	0	1	0	becomes	0	0	0	after spur operation
	1	0	0		1	0	0	
	1	1	1		1	1	1	
H-Break:	0	1	0	becomes	0	0	0	after H-Break operation
	1	1	1		1	1	1	

Step 11: Finding true minutiae points in ROI of $f_m(x,y)$ to get final image $f_{final}(x,y)$ after removing spurious minutiae [14] in the cases, if

i) distance between a termination and a bifurcation is smaller than D

ii) distance between two bifurcations is smaller than D

iii) distance between two terminations is smaller than D

'D' is the average distance between minutiae points. Here D = 6.

Step 12: Representation of linear distance and angle of each minutia in ROI with respect to core point in polar form. The linear distance and angle between core point (x_1, y_1) and minutia (x_2, y_2) is given by,

$$r = D(x,y) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (7)$$

$$\theta = \tan^{-1}\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \quad (8)$$

$$z = re^{i\theta} \quad (9)$$

Step 13: Taking Fourier transform of (9) and saving the Fourier coefficients in '.dat' file.

Step 14: Creation of template of fingerprint database.

Step 15: Calculation of the parameter Euclidean distance (D_{min}) between saved template and the test fingerprint template using (7).

Step 16: Minimum (D_{min}) is compared with the set threshold to get result whether 'match' or 'not match'.

The block diagram of the minutiae based fingerprint recognition algorithm is shown in Fig. 5.

Simulation and Results Analysis:

This section presents the simulation results and performance evaluation of the proposed algorithm, minutiae based fingerprint recognition. For simulations, we have employed MATLab software as the simulation platform. For the experimental purpose, the fingerprint samples have been taken from Fingerprint Verification Competition (FVC2002) [22], which is a publicly available fingerprint database and are usually used for benchmark experiments. FVC2002 comprises of four different image sized fingerprint database, collected by four different sensors or technologies. We have employed first database (DB1) due to good quality and image size of fingerprint samples. The resolution of fingerprint image is set to 500 dpi. We have taken gray scale fingerprint image samples for processing.

Each fingerprint images have different and unique characteristics and therefore present a different impact on recognition system. Thus, to achieve a good number of minutiae pairings resulting in maximum accuracy, an appropriate radius of circle is drawn and processed as seen in Table 1. Series of experiments have been performed at different radius value for the test fingerprint samples and the templates are tested with the pre-stored saved template database.

In our experiments, the image size of fingerprint samples is chosen to be 400 x 400. The samples have been contrasted and enhanced with Histogram equalization and Wiener filter. After getting singular point and minutiae extraction, morphological operations have been used to remove false minutiae to get high accuracy.

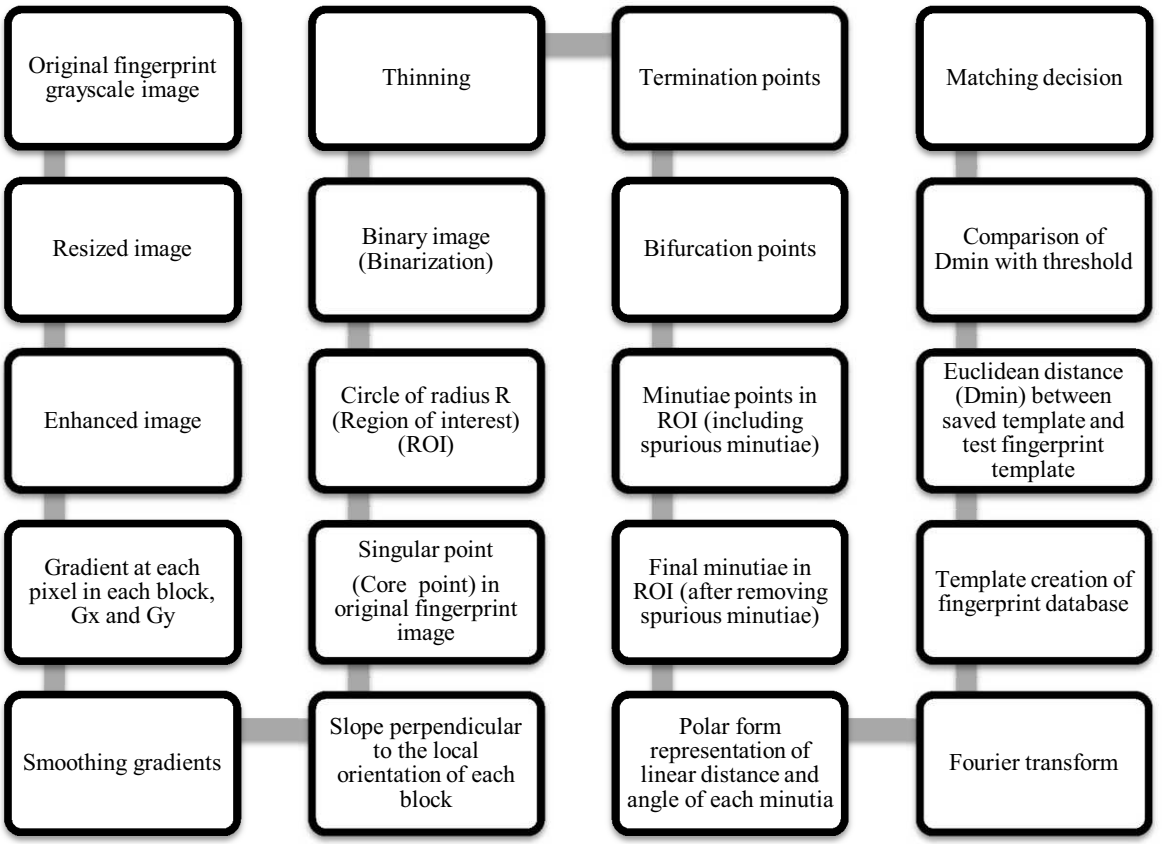


Fig. 5. Block diagram of proposed algorithm

To test the performance of our proposed fingerprint recognition algorithm, the objective measurements like FAR, FRR, EER, D_{min} at EER, and Accuracy have been used. A false acceptance occurs when two images from different fingers are matched, and a false rejection occurs when two images from the same finger are not matched [19]. At equal error rate, both acceptance and rejection errors are equal and it is used to compare the accuracy of biometric system. The EER is calculated using the graph between FAR and FRR versus threshold, which lies in the expected range. The calculated values of these measuring parameters have been tabulated in Table 2.

Further, the step-by-step simulation results of the minutiae based fingerprint recognition algorithm is shown in fig. 6. The FAR, FRR, EER and accuracy with probability and threshold at different radius and same Fourier descriptors are shown in fig. 7.

Table 1. Radius of circle 'R' vs. minutiae points

S. No.	Radius of circle 'R'	Number of minutiae points in circle of radius 'R'	
		Fingerprint image sample 1	Fingerprint image sample 2
1.	90	75	95
2.	100	84	120
3.	150	186	256

From the extensive study, it is observed that for every increase in radius of circle (discussed in this paper), the value of input feature vectors or minutiae points of fingerprint image increases and the number of minutiae points is better for fingerprint image samples to identify legitimate user.

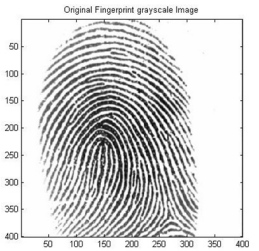
The comparative results Equal error rate (EER), D_{min} at EER, Accuracy and Threshold, shown in Table 2, is for different values of radius of circle 'R' and different values of Fourier descriptors. The accuracy gradually increases as the radius of circle 'R' increases, at the same Fourier descriptors level. This means, we can varry radius 'R' according to the accuracy as

Table 2. Variations of various deciding parameters

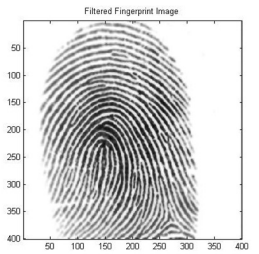
S. No.	Radius of circle 'R'	Fourier descriptors (Real)	Equal error rate (EER)	D_{min} at EER	Accuracy
1.	90	80	3.00 %	420	94.00 %
2.	90	120	3.19 %	1290	93.61 %
3.	100	80	2.90 %	350	94.29 %
4.	100	120	3.19 %	1570	93.86 %
5.	150	80	3.09 %	290	94.81 %
6.	150	120	3.18 %	450	93.97%

required in particular application. The EER is minimum and has slight variation for the increasing value of radius of circle. The distance, D_{min} at EER has different values for different values of radius. There is no clear relation between the radius of circle and value of D_{min} due to non coherence but the minimum possible distance is considered for comparison. The overall accuracy is better for our proposed algorithm.

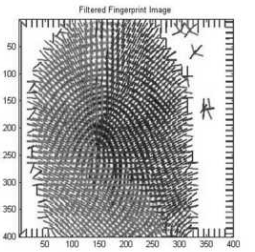
The simulation results of the minutiae based fingerprint recognition algorithm is shown in Fig. 6.



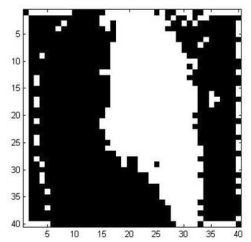
a) Original fingerprint gray scale image from FVC2002



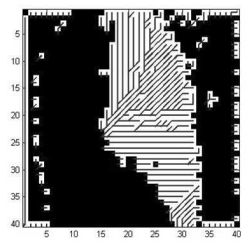
b) Enhanced fingerprint image



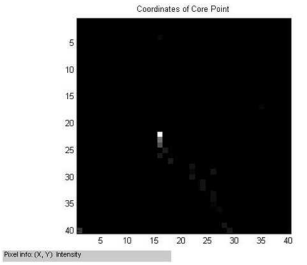
c) Gradient at each pixel in each block, G_x and G_y



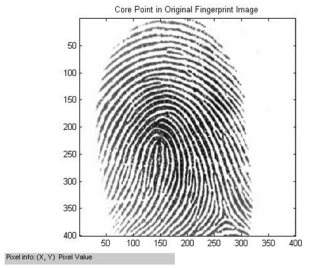
d) Smoothing gradients



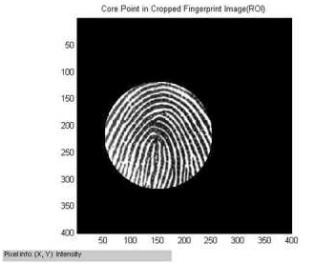
e) Slope perpendicular to local orientation of each block



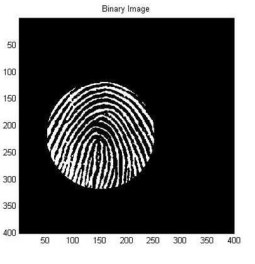
f) Core point in fingerprint image



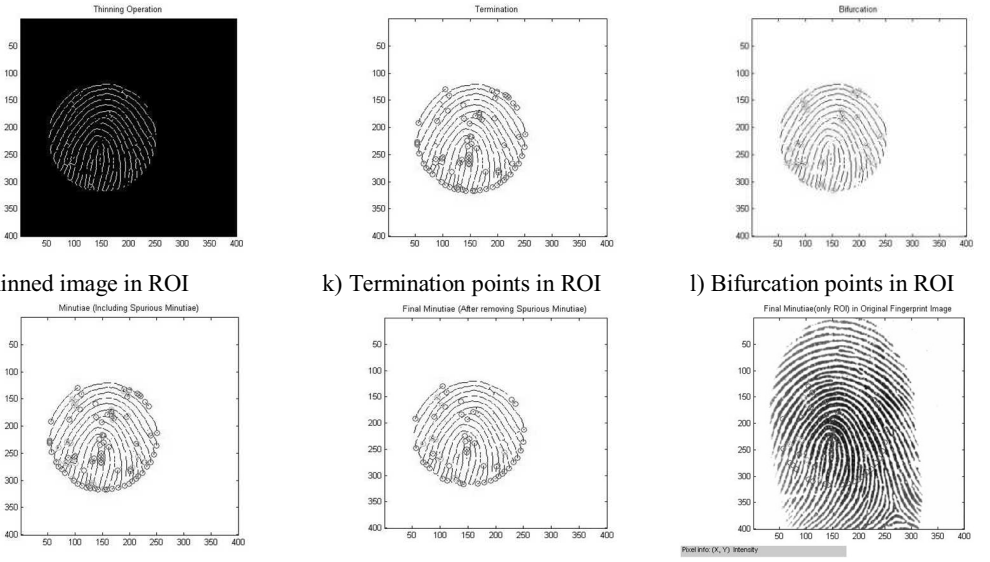
g) Core point in original fingerprint image



h) Core point in cropped fingerprint image (ROI)



i) Binary image in region of interest (ROI)



j) Thinned image in ROI

k) Termination points in ROI

l) Bifurcation points in ROI

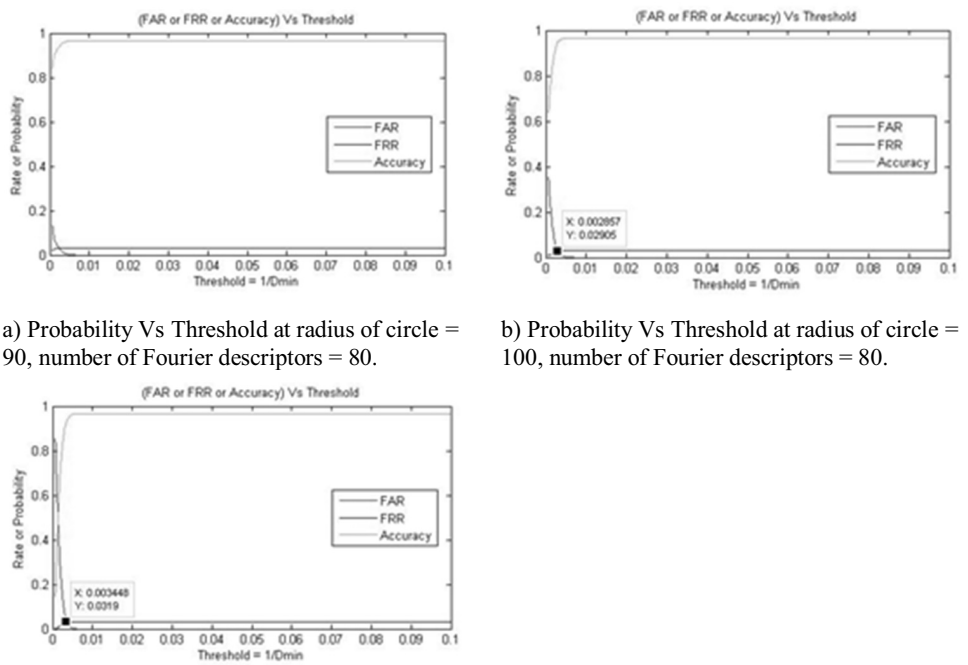
m) Minutiae points in ROI (including spurious minutiae)

n) Final minutiae in ROI (after removing spurious minutiae)

o) Final minutiae (ROI) in original fingerprint image

Fig 6. Simulation results of the proposed minutiae based fingerprint recognition algorithm

The graph between Probability and Threshold showing FAR, FRR, index of EER and Accuracy at same Fourier descriptors but different Radius is depicted in Fig. 7.



a) Probability Vs Threshold at radius of circle = 90, number of Fourier descriptors = 80.

b) Probability Vs Threshold at radius of circle = 100, number of Fourier descriptors = 80.

c) Probability Vs Threshold at radius of circle = 150, number of Fourier descriptors = 80.

Fig 7. Graph between probability and threshold showing FAR, FRR, index of EER and Accuracy

Conclusion

In this paper, the Minutiae based Fingerprint Recognition algorithm was implemented to deal the accuracy of fingerprint matching. For the performance evaluation of the algorithm, different thresholds are set to obtain different level of accuracy.

The objective measures were conducted at different values of radius for the fingerprint images. Results show that the quality and accuracy of matched fingerprint enhanced by our proposed algorithm is good while the EER of fingerprint image remains acceptable. Also, the experimental results reveal the effectiveness and robustness of our algorithm but still very few spurious minutiae are left uneliminated. So, there is a scope of developing new techniques which could remove them completely.

References

[1] Jianjiang Feng and Anil K. Jain, "Fingerprint Reconstruction: From Minutiae to Phase", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 33, Issue: 2, February 2011, Pages: 209-223.

[2] Mohammed S. Khalil, Dzulkifli Mohamad, Muhammad Khurram Khan, Qais Al-Nuzaili, "Fingerprint verification using statistical descriptors", Digital Signal Processing 20, Issue: 4, 2010, Pages: 1264-1273.

[3] T. Divan and V. Gulhane, "A fingerprint matching technique using minutiae based algorithm for voting system: A survey," Electrical, Computer and Communication Technologies (ICECCT), IEEE International Conference, Coimbatore, 2015, Pages: 1-5.

[4] W. Wen, Z. Qi, Z. Li, J. Zhang, Y. Gong and P. Cao, "A Robust and Efficient Minutia-Based Fingerprint Matching Algorithm," Pattern Recognition (ACPR), 2nd IAPR Asian Conference, Naha, 2013, Pages: 201-205.

[5] Bansal Roli, Sehgal Priti, and Bedi Punam, "Minutiae extraction from fingerprint images- A Review." IJCSI International Journal of Computer Science Issue, Volume: 8, Issue: 5, September 2011, Page: 74.

[6] Boashash B., Deriche M., Kasaei S., "Fingerprint feature extraction using block-direction on reconstructed images", IEEE region TEN Conf., Digital Signal Processing applications, TENCON, Volume: 1, 1997, Pages: 303–306.

[7] A. K. Jain, Y. Chen, and M. Demirkus, "A fingerprint recognition algorithm combining phase-based image matching and feature-based matching", in Proc. of International Conference on Biometrics (ICB), 2005, Pages: 316-325.

[8] S. A. Sudiro, M. Paindavoine and T. M. Kusuma, "Simple Fingerprint Minutiae Extraction Algorithm Using Crossing Number On Valley Structure," Automatic Identification Advanced Technologies, IEEE Workshop, Alghero, 2007, Pages: 41-44.

[9] V. Humbe, S. S. Gornale, R. Manza and K. V. Kale, "Mathematical Morphology approach for Genuine Fingerprint Feature Extraction", Int. Journal of Computer Science and Security (IJCSS), Volume: 1, 2007, Pages: 53-59.

[10] Yuheng Zhang and Qinghan Xiao, "An Optimized Approach for Fingerprint Binarization," Neural Networks, (IJCNN), International Joint Conference, 2006, Pages: 391-395.

[11] W. Chen and Y. Gao, "A Minutiae-Based Fingerprint Matching Algorithm Using Phase Correlation," Digital Image Computing Techniques and Applications, 9th Biennial Conference of the Australian Pattern Recognition Society, Glenelg, 2007, Pages: 233-238.

[12] Anil K. Jain, Arun Ross, Karthik Nandakumar, "An Introduction to Biometrics Recognition", IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image and Video Based Biometrics, Volume: 14, Issue: 1, 2004, Pages: 4-20.

[13] R.O. Duda, P.E. Hart and D.G. Stork, Pattern Classification, 2nd edition, Wiley-Interscience, New York, 2000.

[14] Jain, Maio and Prabhakar, Handbook of Fingerprint Recognition, Springer-Verlag London Limited, 2009.

[15] R.C. Gonzalez and R.E. Woods, Digital Image Processing, 3rd edition, Prentice-Hall, Englewood Cliffs, NJ, 2007.

[16] Chirillo and Blaul, "Implementing Biometric Security", Wiley publications, 2009.

[17] J. Parker, Algorithms for Image Processing and Computer Vision, Wiley, New York, 1996.

[18] A.K. Jain, Fundamentals of Digital Image Processing, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[19] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman and A. K. Jain, "Performance evaluation of fingerprint verification systems," IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 28, Issue: 1, 2006, Pages: 3-18.

[20] A{ XE "**A**" }. Li. Stan, A. K. Jain, "Encyclopedia of Biometrics", Springer, New York, 2009.

[21] K. Castleman, "Digital Image Processing", Prentice Hall, 1996.

[22] Fingerprint database, FVC2002, "http://bias.csr.unibo.it/fvc2002"

A Comprehensive Measurement Placement Method for Power System State Estimation

Rakesh J Motiyani* and Ajitsinh R Chudasama**
*Principal, Pacific School of Engineering (Diploma Studies), Research Scholar-Rai University, Ahmedabad
motiyani_me@rediffmail.com
**Academic Director, Neotech Campus, Vadodara
jyotiajit@yahoo.com

Abstract: Power System State Estimators requires a set of redundant measurements. The meter placement problem involves selection of number, type and place of meters. The main objective in designing a metering scheme is to satisfy cost, accuracy, observability and bad data processing requirements for power system state estimators. This paper presents a comprehensive method using graph theoretic approach to solve the problem of measurement placement for power system state estimators. The measurement configuration on IEEE 30 bus system is presented. The suggested technique ensures reliable and accurate estimation of state variables at lower cost.

Keywords: Bad data detection, Graph theory, Measurement placement, Network observability, State estimation.

Introduction

Within the energy management system, state estimation is a key function for building network real time model that is static mathematical representation of current condition of an interconnected power system. The state of electrical power system is defined by the vector of voltage magnitude and angle at all network buses. The static state estimator is the data processing algorithm for converting redundant and not so reliable meter readings and other available information about the network connectivity in to an estimate of the static state vector [1]-[4].

The real-time modelling of a power network usually follows following procedure involving [5]:

- Data gathering
- Network topology processing
- Observability analysis
- State estimation (SE)
- Processing of bad data and
- Identification of network model

Network topology processor identifies energized and de-energized electrical islands and is performed before state estimation and other related functions such as observability analysis and bad data processing [5], [6]. The network observability can be tested for P-δ and Q-V models by triangular factorization of the gain matrix. The obsevability algorithm checks that the information available in the form of measurements is sufficient to fulfil the computational needs of state estimator or not [7].

Power system state estimators requires a set of redundant measurements, which are appropriately chosen according to the type, number and location of the measurement points in the supervised electric network [8]. The main objective in designing metering scheme is to satisfy the requirements like cost, accuracy, reliability, and bad data processing for power system state estimator [9]. In late 90's, researchers started giving importance to the development of cost effective and reliable metering scheme which can yield accurate state estimation. The measurement data are usually power flows, power injections and bus voltages [10]. In 1996, a meter placement method developed by Mesut E Baran et al. The developed method was utilized for designing the measurement system configuration on IEEE 14 bus system. The measurement placement method using Simulated Annealing (SA) aims at attending to the requirements such as observability and reliability- taking in to account the associated monetary costs was developed by Alessandra B Antonio et al. in 2001. The developed method was utilized for designing the measurement system configuration on IEEE 30 bus system.

In 2006, the optimal meter placement method using hybrid Genetic Algorithm and Simulated Annealing (GA/SA) was developed by Thawatch Kerdchuen et al. [11]. The metering configuration on 10-bus and IEEE 14 bus system was presented. In 2011 the metering schemes for IEEE-10, 14, 30 and 57 bus systems was presented by K. Jamuna and K.S.

Swarup using Biography Based Optimization [12]. In 2013, the metering schemes for IEEE-10 and 14 bus systems were presented using Key Cutting Algorithm (KCA) by Yuttana Kongjeen et al. [13].

State Estimation

The state estimation is a mathematical procedure by which the state of electric power system is extracted from a set of measurement. In standard SE, in order to relate measurements and non linear equations, the following model is used:

$$z = h(x) + e$$

where, z is the (m×1) measurement vector, h(x) is the (m ×1) vector of non linear functions, x is the (2n ×1) state vector, e is the (m×1) measurement error vector, n is the total number of buses in the power system network and m is the total number of measurements.

The state estimator is a mathematical algorithm formulated to minimize the error between a real time measurement and a calculated value of the measurement. The minimization criterion often selected is the weighted sum of error squares of all the measurements. The estimator favors accurate measurements over the less accurate ones by weighing the errors with the measurement standard deviation (σ_j) [14].

$$\min J(x) = \sum_{j=1}^{m} \left(\frac{e_j}{\sigma_j} \right)^2$$

The condition for optimality is obtained at a point when the gradient of J(x) is zero. From weighted least square method, the iterative equation can be obtained as follows:

$$x^{k+1} = x^k + \Delta x$$

$$\Delta x = \left(H^T R^{-1} H \right)^{-1} H^T R^{-1} \left(z - h\left(x^k \right) \right)$$

$$H = \frac{\partial h(x)}{\partial x} = \begin{bmatrix} \frac{\partial h_1(x)}{\partial x_1} & \frac{\partial h_1(x)}{\partial x_2} & \cdots & \frac{\partial h_1(x)}{\partial x_{Ns}} \\ \frac{\partial h_2(x)}{\partial x_1} & \frac{\partial h_2(x)}{\partial x_2} & \cdots & \frac{\partial h_2(x)}{\partial x_{Ns}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial h_m(x)}{\partial x_1} & \frac{\partial h_m(x)}{\partial x_2} & \cdots & \frac{\partial h_m(x)}{\partial x_{Ns}} \end{bmatrix}$$

Where, $N_s = 2n-1$

$$W = R^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & & \\ & \frac{1}{\sigma_2^2} & & \\ & & \cdots & \\ & & & \frac{1}{\sigma_m^2} \end{bmatrix}$$

The gain matrix is defined as

$$G = H^T R^{-1} H$$

While the power system not only has Supervisory Control and Data Acquisition System (SCADA), but also has Phasor Measurement Units (PMUs) placement, the sub problem is formed by PMU placement and SCADA measurements. The state variables measured by PMU are assumed true value and the known state variables are x_1. The unknown state variables are required to be estimated by reduced power system state estimation model [15].

Hence, equations can be rewritten as follows:

$$z = h\left(x_2 \right) + e$$

$$\Delta x_2 = \left(H_2^T R^{-1} H_2 \right)^{-1} H_2^T R^{-1} \left(z - h\left(x_2^k \right) \right)$$

$$H_2 = \frac{\partial h(x_2)}{\partial x_2}$$

$$G_2 = H_2^T R^{-1} H_2$$

Measurement Placement

The state estimator uses a set of measurement consisting of bus injections, branch flows and bus voltages. If all the quantities are measured as shown in the Fig. 1, the measurement set will become full and Jacobian (H) has 3n + 4b rows. Where, n is the total number of network buses and b is the total number of network branches.

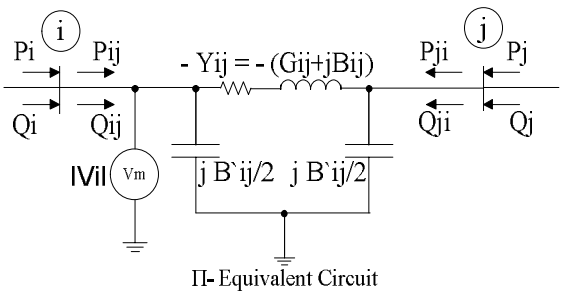


Figure 1. The per phase representation of transmission line- showing possible measurements

Factors Affecting Measurement Placement

1. Cost: The total investment cost for the metering scheme – meter, Remote Terminal Units (RTUs) and communication system should be kept to minimum.
2. Accuracy: The measurements obtained through metering scheme should yield a state estimation with desired accuracy.
3. Reliability: The power system network should be observable with the measurements obtained from the metering scheme and there should be enough redundancy so that the state estimation can be carried out even after loss of measurements or meter failure.
4. Bad Data Processing: The measurement system should allow state estimator to detect and eliminate bad data from the measurement set.

Considering the cost of a voltage measurement, a power measurement and an RTU as 1, 4.5 and 100 Money Units (MU) respectively, Mesut E Baran et al. designed a metering scheme for IEEE 14 bus system. The total cost of metering scheme is 1112 MU. The metering scheme acquires redundancy of 1.92. The metering scheme designed for IEEE 30 bus using SA acquires redundancy of 1.9 and considering meters cost as mentioned above, the total cost of metering scheme is 3052 MU. The metering scheme designed using hybrid GA/SA and KCA gives redundancy of 1.04. The metering scheme obtained through Biography Based Optimization acquires redundancy of 1.02.

The measurement system obtained through hybrid GA/SA, KCA and Biography Based Optimization provides poor redundancy. Hence, the cost of measurement system configuration will become less. But, the metering scheme will fail to satisfy requirement of execution of bad data finding and abolition because; after loss of one measurement pair the redundancy will become less than 1. Always measured quantities should be more than the necessary minimum number to make provision for eliminating erroneous quantities from the measurement set to ensure accurate state of the power system.

Proposed Measurement Placement Method

The proposed measurement placement method is based on network graph theory. The metering scheme assures that each branch of power system network is incidental by power injection measurements at either ends or a flow measurement and an injection measurement at one of its terminal node. Selection of meter locations also assures least requisite of RTUs.

Preliminarily, to reduce meter requirements for observability, the proposed method gives the greatest priority to power injection measurements at the buses of maximum adjacency.

The redundancy can be considered as a function of the accuracy requirement in the state estimation. Always measuring quantities should be more than the necessary minimum number to make provision for eliminating erroneous quantities from the measurement set to ensure accurate state of the power system.
Hence, in the second stage, branches consisting power injection measurement at one of it terminal node are identified and power flows are measured through such branches to increase redundancy. Further, to enhance redundancy, voltages at all additional RTU locations are measured.
The proposed meter placement method proceeds as follows:

- Read bus data, initialize measurement set of interest by injection measurements at all the zero injection buses in the power system network.
- For n bus power system network, read line data and prepare n x n adjacency matrix A= [a_{ij}] where; a_{ij} =1, if i^{th} bus is incident to j^{th} bus and a_{ij} =0, if otherwise. Modify adjacency matrix by making all a_{ii} =0, as these elements of matrix represent the bus itself.
- Compute total ones for each row of modified adjacency matrix. Identify buses of maximum (p) and minimum (q) adjacency. Place RTUs and measure power injections at the buses of adjacency p, p-1, p-2 ……., till p, p-1, p-2 ……., = q+2.
- Identify branches contain no power injection measurement at one of its end, place RTU and measure injection at any end. Add power injection measurements at the buses of q+1 adjacency and voltage measurements at all RTU locations till redundancy becomes ≥ 1.
- Update line data file by removing all the lines comprising of injection measurements at both ends. Measure power flows through the remnant lines such that minimum requisite of additional RTUs. Measure voltages at additional RTU locations.

When SCADA measurements are combined with PMU, the direct measurement of state variables with enough accuracy will be available through PMU data. Hence, excluding the buses of PMU locations, the proposed measurement placement method can be utilized to configure SCADA measurements on modified power system network. Making the use of proposed measurement system configuration, the state variables can be estimated using reduced state estimation model.

Test Results

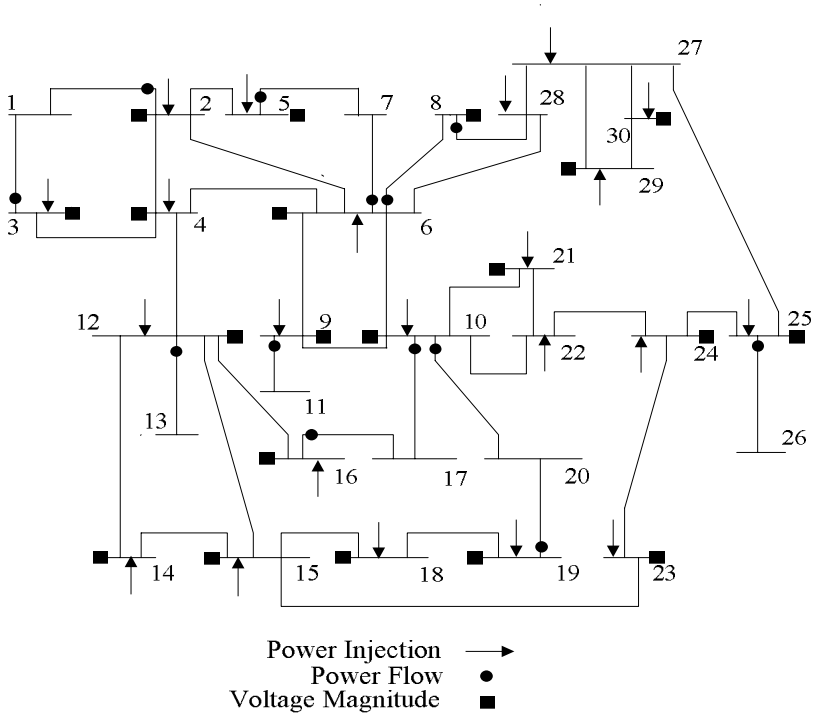


Figure 2. Meter placement on IEEE 30 bus system using proposed method

The simulation study is performed on IEEE 30 bus system. Fig. 2 shows measurement system configuration obtained through the proposed method of measurement placement. Out of 254 possible measurements, metering scheme utilizes 90 measurements and redundancy becomes 1.52.

The metering scheme consists: 20 RTUs, 22 power injections, 20 voltage magnitudes and 13 power flows. Estimated state obtained by utilizing the proposed metering scheme, SA and biography based optimization is shown in the table 1.

Table 1. Estimated state

Bus No.	---------- Estimated State ----------					
	Proposed Metering Scheme		Metering Scheme of SA Alessendra B Antonio et al.		K Jamuna, K S Swarup et al.	
	Voltage Magnitude |V| pu	Bus Angle δ (Degree)	Voltage Magnitude |V| pu	Bus Angle δ (Degree)	Voltage Magnitude |V| pu	Bus Angle δ (Degree)
1	1.027	0.000	3.338	0.000	0.178	0.000
2	1.010	-2.679	0.578	-0.972	1.175	-11.489
3	1.001	-4.073	1.142	-6.801	0.787	-11.818
4	0.995	-4.890	0.949	-8.907	0.906	-13.143
5	0.977	-7.026	0.850	-10.654	1.014	-16.231
6	0.993	-5.778	1.149	-10.103	0.966	-14.072
7	0.983	-6.598	0.963	-10.801	0.945	-15.851
8	0.985	-6.064	0.692	-10.360	0.979	-14.390
9	1.033	-7.362	1.006	-15.851	0.974	-14.941
10	1.031	-8.219	0.994	-19.315	0.940	-15.336
11	1.052	-7.305	0.809	-15.928	1.023	-15.067
12	1.039	-7.844	0.996	-24.142	1.011	-14.247
13	1.036	-7.843	1.011	-26.010	1.021	-14.319
14	1.038	-8.400	0.978	-26.969	1.006	-14.120
15	1.034	-8.376	0.994	-25.920	1.025	-14.302
16	1.024	-8.156	1.003	-23.050	1.019	-14.937
17	1.024	-8.330	1.013	-20.236	0.985	-15.378
18	1.023	-8.774	0.999	-25.290	1.206	-7.558
19	1.016	-8.851	0.983	-24.512	1.055	-6.694
20	1.015	-8.747	0.643	-23.742	1.056	-6.923
21	1.030	-8.484	0.997	-19.819	0.940	-15.431
22	1.031	-8.481	1.000	-19.918	0.944	-15.373
23	1.038	-8.561	1.006	-25.467	1.036	-14.531
24	1.033	-8.611	1.021	-22.426	0.995	-15.357
25	1.037	-8.437	0.993	-18.359	1.045	-15.242
26	1.031	-8.754	0.966	-17.758	1.038	-15.462
27	1.040	-8.064	1.007	-16.754	1.082	-14.974
28	0.995	-6.170	1.031	-10.720	0.982	-14.435
29	1.039	-8.796	0.966	-17.628	1.133	-15.519
30	1.036	-9.207	0.970	-17.976	1.209	-16.738

Fig. 3 and 4 shows errors in estimated state variables computed with the use of proposed measurement system configuration and the metering scheme obtained through SA and Biography Based Optimization.

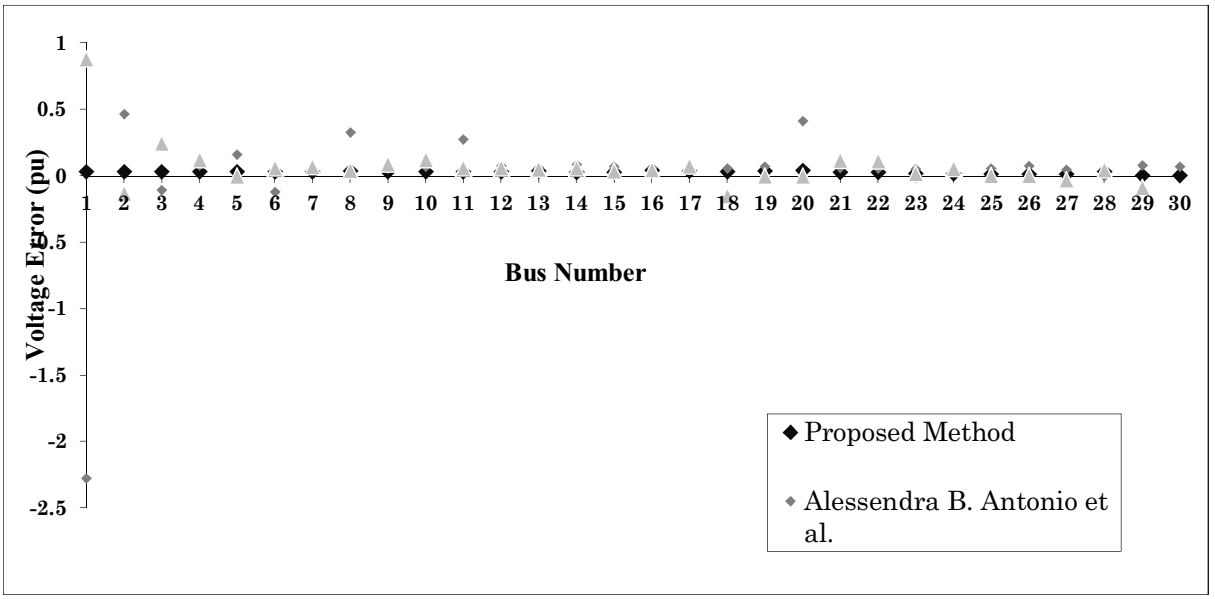


Figure 3. Voltage Magnitude Errors

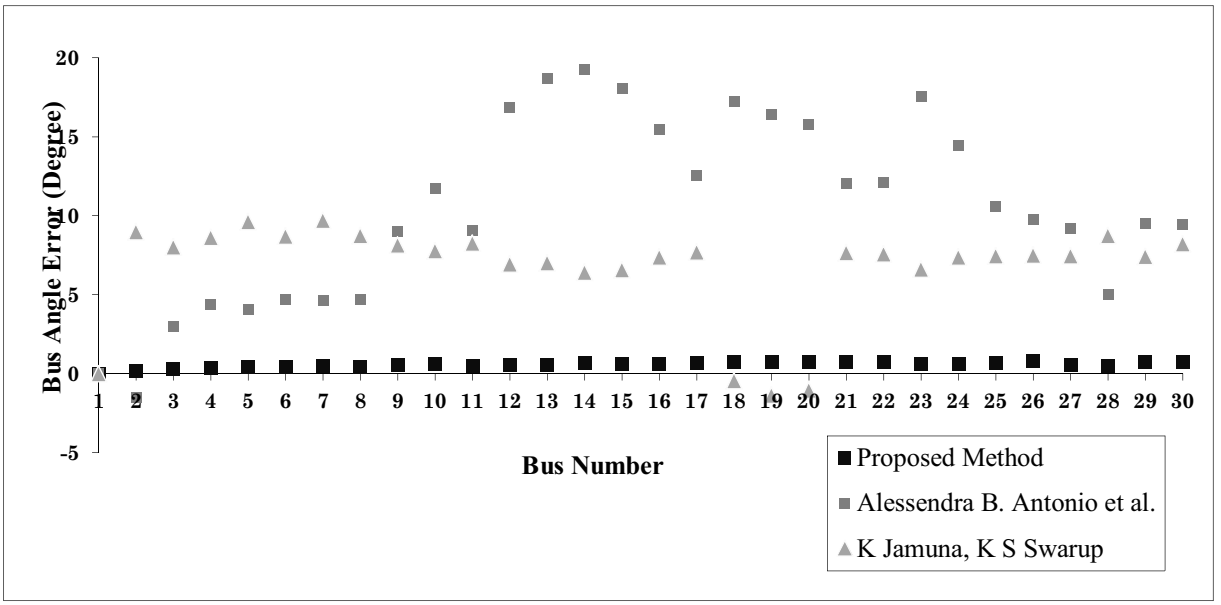


Figure 4. Bus Angle Errors

Conclusion

The suggested measurement system configuration accomplishes better bad data processing and observability requirements for state estimator compared with the metering scheme obtained through the Biography Based Optimization method. The proposed measurement configuration yields much accurate state of power system than the metering schemes obtained through SA and Biography Based Optimization method. Also, the suggested method acquires cost effective measurement system configuration compared with SA.

In presence of PMU, excluding the buses of PMU locations, the proposed measurement placement method can be utilized to configure SCADA measurements on modified power system network. Making the use of proposed method of measurement system configuration, the state variables can be estimated using reduced state estimation model.

The proposed method of measurement placement can be implemented in existing state estimators as an off- line measurement systems planning tool.

References

[1] Robert E. Larson, William F. Tinney, John Peschon, "State Estimation in Power Systems – Part I: Theory and Feasibility", IEEE Transaction on Power Apparatus and System, Volume 80, Issue 3, 1970, 345-352.
[2] Fred C. Schweppe and J. Wildes, "Power System Static State Estimation Part I": Exact Model", IEEE Transaction on Power Apparatus and System, Volume 89, Issue 1, 1970, 120-125
[3] Fred C. Schweppe and Douglas B. Rom, "Power System Static State Estimation Part II: Approximate model", IEEE Transaction on Power Apparatus and System, Volume 89, Issue 1, 1970, 125-130.
[4] Fred C. Schweppe, "Power System Static State Estimation Part III: Implementation", IEEE Transaction on Power Apparatus and System, Volume 89, Issue 1, 1970, 130-135.
[5] A. Monticelli, "Electrical Power System State Estimation", Proceeding of IEEE, Volume 88, Issue 2, 2000, 262-282.
[6] Alsac O, Vempati N, Stott B and A. Monticelli, "Generalized State Estimation", IEEE Transaction on Power Systems, Volume 13, Issue 3, 1998, 1069-1075.
[7] Sandhya Sankar, Noel N. Schulz, "Intelligent Placement of Meters/Sensors for Shipboard power System Analysis", IEEE Proceeding, 2007, 504-509.
[8] Alessandra B. A., Jose R. A. and Milton B. D, "Meter Placement for Power System State Estimation Using Simulated Annealing", IEEE Porto Power Tech Conference, Volume 3, 2001, 1-5.
[9] Mesut E. Baran, Jinxiang Zhu, Hongbo Zhu, Kenneth E. Garren, "A Meter Placement Method for State Estimation", IEEE Transaction on Power Systems, Volume 10, Issue 3, 1995, 1704-1710.
[10] A. Shafiu, N. Jenkins and G. Strbac, "Measurement Location for State Estimation of Distribution Networks with Generation", IEEE Proceeding Volume 152, Issue 2, 2005, 240-246.
[11] Thawatch Kerdchuen, Weerakorn Ongsakul, "Optimal Measurement Placement for Power System State Estimation Using Hybrid Genetic Algorithm and Simulated Annealing", Proceeding of International Conference on Power System Technology, 2006, 1-5.
[12] K. Jamuna and K. S. Swarup, "Contingency Constraint Optimal Meter Placement for Power System Observability using Biography Based Optimization", International Journal of Machine Learning and Computing, Volume 1, Issue 1, 2013, 86-92.
[13] Yuttana Kongjeen, Prajuab Inrawong, Kittavit Buayai and Thawatch Kerdchuen, "Key Cutting Algorithm Application to Measurement Placement for Power System State Estimation", Elsevier Energy Procedia, 2013, 142-147.
[14] L.Mili, Th. Van Cutsem and M. Ribbens-Pavella, "Bad Data Identification Methods in Power System State Estimation – A Comparative Study", IEEE Transaction on Power Apparatus and System, Volume 104, Issue 11, 1985, 3037-3049.
[15] Fang C, Xueshan H, Zhiyuan P and Li H., "State Estimation Model and Algorithm Including PMU", IEEE Electric Utility Deregulation and Restructuring and Power Technologies Conference, 2008, 1097-1102.

Voltage Control of DC to DC Converter in Integrated Renewable Energy System through Fuzzy based GA

K.Harinadha Reddy*, T.Raghu** and K.P.Prasad Rao***
*Professor, EEE Department, K L University, A.P., India
kadapa.hari@gmail.com
** Research Scholar, EEE Department, K L University, A.P., India
thumuraghu@gmail.com
***Research Scholar, EEE Department, K L University, A.P., India
prasadraokondenti@gmail.com

Abstract: A control plan based on improved dc bus reference voltage parameter is proposed for a photovoltaic (PV) generation system with battery backup energy system. Genetic Algorithm (GA) based MPPT technique was proposed and proposed work consists of design model and control scheme of converters for photovoltaic cells. The pattern with different ways like two grid-connected converters, and one dc/dc converter for battery charging and discharging operation and other is considered for power control. Genetic Algorithm Optimization (GAO) is used to get MPPT signal of solar cell. Fuzzy system is producing a signal for converter module. The available of either grid-connected operation or islanding operation, using the proposed control strategy, the operations of a modular PV generation system are categorized into islanding with battery operation, grid-connected conversion, and islanding with constant voltage invention.

Keywords: Photovoltaic array, Grid connection, Islanding, MPPT algorithm, Genetic Algorithm Optimization.

Introduction

Recently in Power network with solar energy systems are having considerable method are developed for improvement of poor voltage profile due to convertor losses and they effect in power control and charging and discharging operation of battery banks and so these energy systems are most want of voltage profile improvement. PV system has become an important source for generating electricity due in part to the development that has happen in the semiconductor region, which has made it possible to increase the energy output to meet the required load power. Due to environmental pollution and the immanent exhaustion of fossil fuel, distributed generation systems using renewable energy sources, including wind power, microhydro, solar photovoltaic, and landfill gas, have become one of the main power generation points of interest. The main distributed generator (DG) requirements are high power quality, highly efficient operation, and safety. Therefore, the main issue in renewable energy research is to reduce the cost and increase the efficiency of production. In recent years, wind power and PV energy have been the two main areas of the research and development.

Renewable energy sources have many advantages over conventional energy sources, as they are green, do not emit carbon dioxide. Power flow in each stage i.e. with conventional and proposed of power plant is clearly depicted in schematic diagram of Figure.2. Also general load at the stage of generator is considered for test case. Each control area as possible supply its own load demand [4, 5] and power transfer through tie line should be on mutual agreement. All control areas [11] should controllable to the specific range of frequencies control. In totaling, the cost of renewable energy sources [8, 9] is higher than the conventional energy sources when generating large volumes of energy. As the intensity of light falling on the panel varies, its voltage as well as its internal resistance varies. Internal resistance of the panel due to variation in irradiation and temperature causes the mismatch between the source [2] and the load. Hence panel is not able to generate power which it is capable of producing, resulting in lower generation of power. This results in generation of less power than what the panel is actually capable of generating. However, much work needs to be done in this field in order to make renewable as efficient and reliable as possible. Since this method likes a constant, its accuracy cannot be guaranteed. Consequently, the tracked power would most likely be below the MPP, resulting in significant power loss.

The Photovoltaic (PV) generating systems need maximum power point tracker because the output power of the PV panel depends on the operating terminal voltage and current. To extract maximum power from the PV array, the load impedance should be equal to the internal resistance of the panel. However, the use of renewable energy still has a many limitations, as most renewable energy sources stands on the weather conditions, such as wind in wind power generation, rain in hydropower and clear skies in photovoltaic systems.

Modelling of Solar PV Cell

Energy systems are deadly need of voltage profile improvement. Various MPP tracking methods have been proposed. These techniques differ in complexity, accuracy, and speed. Various type of the control variables and they are voltage, current or duty cycle. For the voltage and current-based techniques, approaches are used. The first one is the observation of MPP voltage V_{PV} or current I_{PV} with respect to the open circuit voltage and short circuit current, respectively.

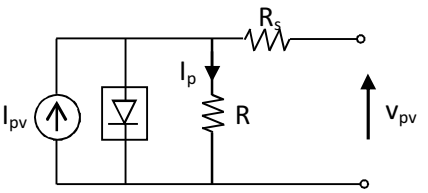


Figure 1. PV Cell Representation

The current source is the photocurrent generated by irradiation on solar cell can be represented by a current source connected in parallel with a PN junction diode

$$I_{PV} = I_{sr}\left(e^{qv/cT} - 1\right) - I_{sc} \qquad (1)$$

The series resistance is the sum of structural resistance of PV panel and it has strong influence when PV panel act as voltage source. The parallel resistance R_{sh} has great influence when PV panel act as current source. The light generated current of the photovoltaic cell depends linearly on the solar irradiation and is influenced by the temperature according to the following equation. The output power of a solar cell is given the product of voltage and current According to the solar cell equation, the output power as a function of the output voltage 'V' is

$$I_{PV} = IV = I_{sc}V - I_{sr}\left(e^{qv/cT} - 1\right)V \qquad (2)$$

$$I_{PV} = I_{sc} - \left[\exp\left(\frac{V_{pv} + I_P R_s}{V_d}\right) - 1\right] - I_p\frac{R_s}{R} \qquad (3)$$

Test system with conventional control gain k_d at voltage measuring system and V_{PV}, I_{PV} are used to obtain the V_{MPPT} signal.
A PV array is the combination of multiple PV panels connected in series and parallel, hence any fault in one PV panel will affect the recital of the overall PV array. This is why some of the case where a small number of PV panels have their proper controller [4]. In this work we consider that the PV array is composed of large number of PV panels and have a sole federal controller with present proposed solution.

Test System with Proposed Fuzzy based GA-MPPT Method

Inputs to a fuzzy logic controller are usually error 'e' and deviation of error. The two inputs for fuzzy logic controller. Linguistic terms used for the membership functions are such that, the terms will be NL(Negative Large), NM(Negative Medium), NS(Negative Small), ZE(Zero), PS(Positive Small), PM(Positive Medium) and PL(Positive Large). Membership Grade NL NM NS ZE. Rules are formed with 49 rule base [9] and expressed in linguistic variables relating input signals to the control signal and shown in Table 1.

Table 1. Knowledge base table with 49 rules

e	Δe						
	NL	NM	NS	ZE	PS	PM	PL
NL	ZE	PS	PM	PL	PL	PL	PL
NM	NS	ZE	PS	PM	PM	PL	PL
NS	NM	NS	ZE	PS	PS	PM	PL
ZE	NM	NM	NS	ZE	PS	PM	PM
PS	NL	NM	NS	NS	ZE	PS	PM
PM	NL	NL	NM	NM	NS	ZE	PS
PL	NL	NL	NL	NL	NM	NS	ZE

Genetic algorithm is popular optimization technique that maps input characteristics to output characteristics with fulfilling of user require constraints. Proposed genetic algorithm in step by step procedure is represented as follows
Step 1: Initialize the output control vector of Fuzzy logic system.
Step 2: With defined fitness function, evaluate the inputs to for mutation of GA process.
Step 3: Individuals of control vector can be determined with selecting the cross over data points,
Step 4: Generate the new set of data for control vector.
Step 5: Calculate the function values for each data of n^{th} data set
Step 6: Check for stopping condition.

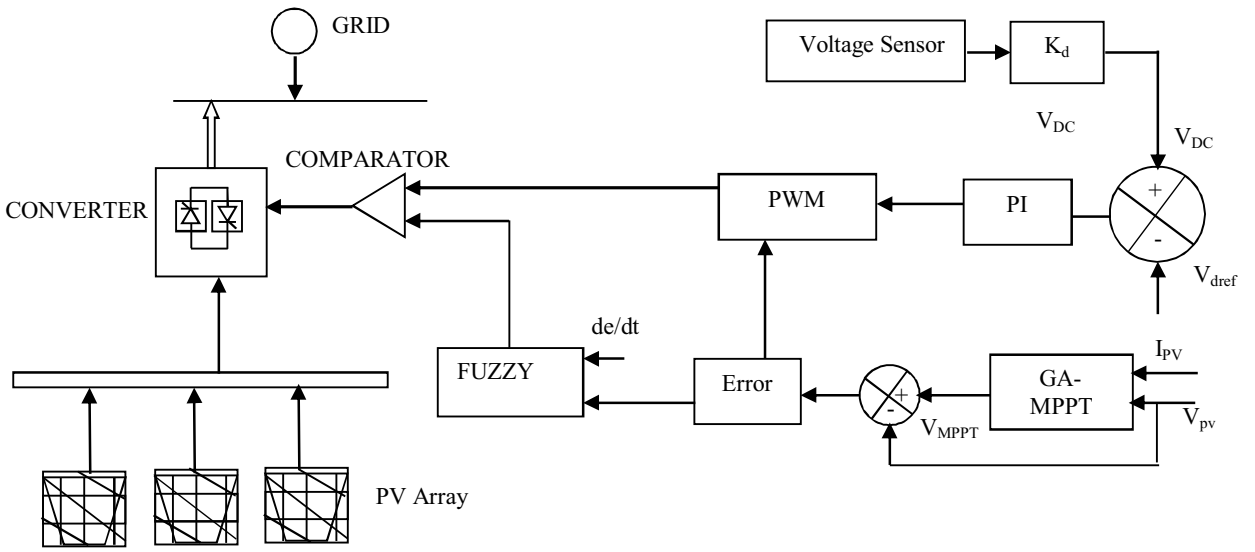


Figure 2. PV Array Module with GA-MPPT Method

The output is the MPPT based PI method is used and compared with the GA method. Test system shown in Figure 2 illustrates the controller configuration. The GA is a swarm intelligence-based algorithm used to find the global optimal solutions. GA-MPPT techniques mostly rely on perturb and observe steps and use the hill-climbing concept in subsequent iterations. While doing so, these methods constantly compare present and previous power values, and when they reach the first local maximum, the algorithm stops progressing [3] in the forward. The DC to DC boost converter is used to regulate a chosen level of the solar photovoltaic module output voltage and to keep the system at the maximum power point.
The proposed method is mainly useful for PV maximum power tracking purposes, where the objective is to draw maximum possible power from solar panels at all times, regardless of the load. It can able to regulate the perturbed voltage by increasing or decreasing the voltage reference of the PWM (Pulse width modulation) signal.A Fibonacci search-based MPPT [5] realization for PV sources has been reported. It compares the values of measured power at two operating points and then determines the operating point movement. The GA algorithms are developed for two separate dataset points and swarm region is obtained from the output of them. The data set is $(n_1\ p_i)$ □ and the second one is $(n_1\ p_i\ P_{bi})\beta$. It is similar to the hill-climbing method with variable step size; the only difference here is that the step size is determined by the Fibonacci sequence. The MPPT search performance, however, is almost identical to that of the conventional hill-climbing algorithm, and hence this scheme, Fibonacci based-search, also settles to local MPP under certain operating conditions.

Results and Analysis

The PV module current and voltage are fed to the converter and the MPPT controller simultaneously. The presentation of the fuzzy based MPPT technique is compare with the conventional P&O MPPT. It shows that the conventional MPPT tracks the maximum power point at 0.1s and also it does not have the ability to reduce the anxious voltage.
The voltages at converter terminal during constant voltage control are shown in Figure 3. Voltage at converter of solar array during maximum power tracking is observed and plot shown in Figure 4. Power curve with MPPT-GA is improved to as that of conventional PI controller. The active and reactive powers are also plotted by exporting data to MATLAB workspace from successful execution of simulation.

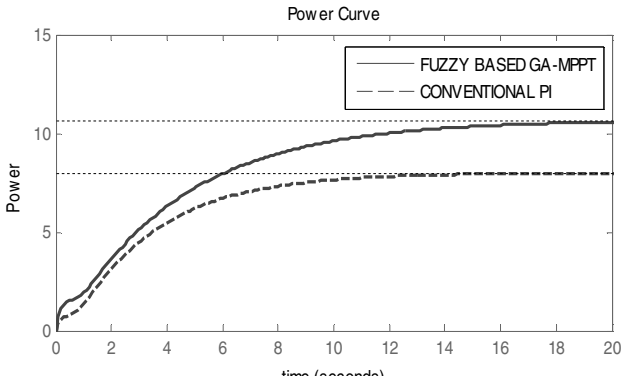


Figure 3. Coverter Power with Fuzzy based and GA-MPPT Technique

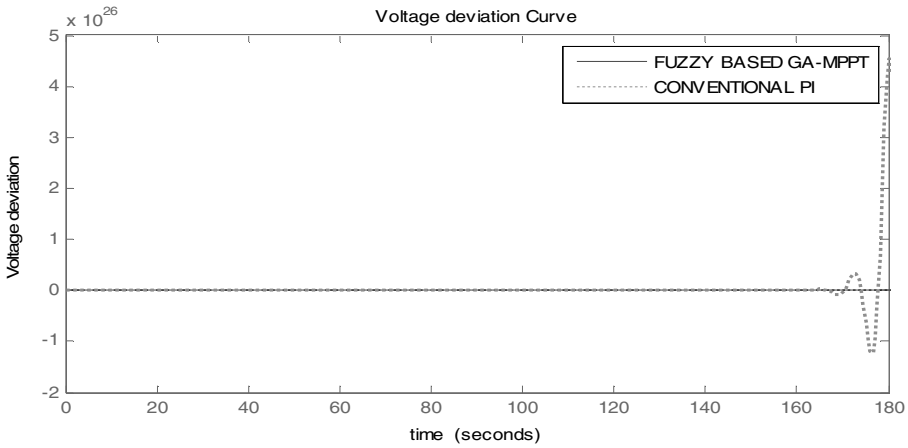


Figure 4. Control of voltage with proposed Fuzzy based and GA-MPPT Technique

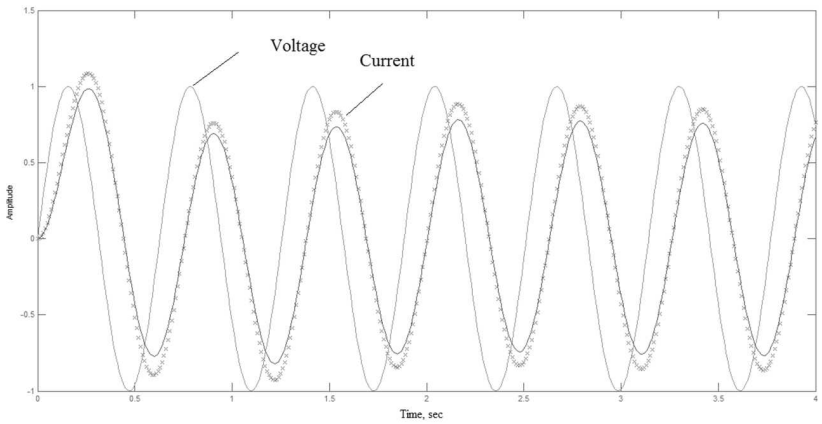


Figure 5. Wave forms of voltages and currents at converter terminals

Power at proposed method at converter terminals is shown in Figure 5. Is observed in simulation and with proposed GA-MPPT method. Active and reactive powers and output voltages are shown in Figure 6 and Figure 7 shows the Control region for active and reactive powers with proposed GA-MPPT method.

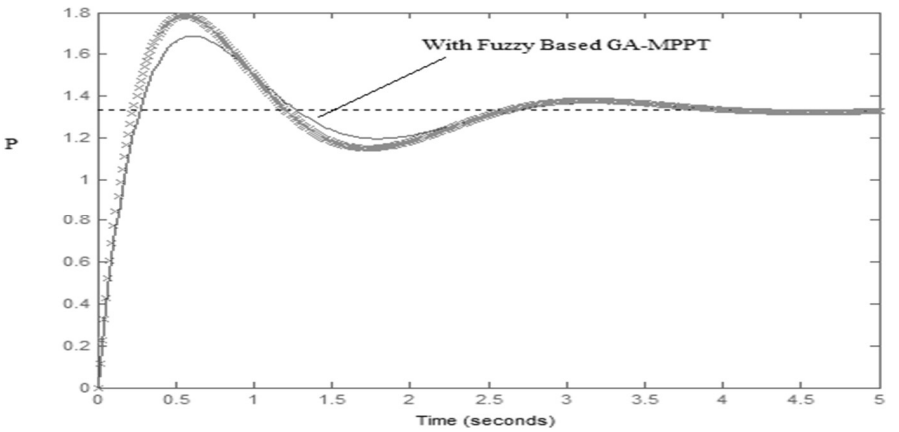


Figure 6. Power with 2% increase in load

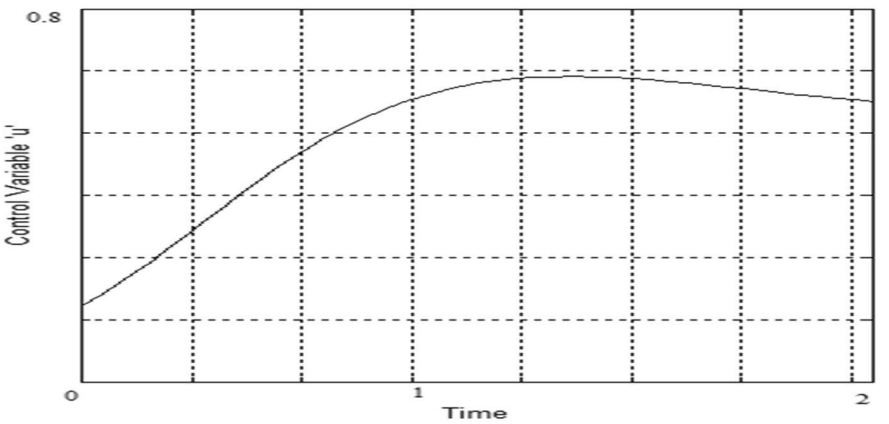


Figure 7. Control vector region for power control

Conclusion

Solar PV array with battery bank is tested during charging and discharging operation. The GA-MPPT is better than the conventional PI method. The power with increase in load is also tested for proposed fuzzy based GA-MPPT technique. The simulation results show the effectiveness of the MPPT control strategy in the case of the attendance of the data set with GA tuned vector. Converter voltage during the test condition is maintaining enhanced profile with proposed optimized voltage profile based on proposed method.

References

[1] Lasseter R.H. and Paigi P., "Microgrid: A conceptual solution", in Proc. IEEE Power Electron. Spec. Conf. (PESC'04), Vol. 6, June 2004, PP.4285– 4290.

[2] Esram T. and Chapman P., "Comparison of Photovoltaic Array Maximum Power Point Tracking Techniques", IEEE Trans. on Energy Conversion, Vol.22, No.2, 2007, PP.439-449.

[3] Femia N. and al., "Optimization of perturb and observe maximum power point tracking method" IEEE Trans. on Power Electronics, Vol.20, 2005, PP.963-973.

[4] Aldeen M. and Marah J.F., "Decentralized Proportional-Plus-Integral Design Method for Interconnected Power Systems", IEE Proceedings C - Generation, Transmission and Distribution, Vol. 138, No. 4, 1991, PP.263-274.

[5] Li Y.W. and Kao C.N., "An accurate power control strategy for power electronics- interfaced distributed generation units operating in a low voltage multibus microgrid", IEEE Trans. Power Electronics, vol. 24, No. 12, 2009, PP.2977–2988.

[6] Kakigano H., Miura Y., and Ise T., "Configuration and control of a DC microgrid for residential houses," in Proc. IEEE Transmission & Distribution Conference & Exposition: Asia and Pacific, Oct. 2009, PP.1–4.

[7] Kakigano H., Miura Y., Ise T. and Uchida R., "DC micro-grid for super high quality distribution-system configuration and control of distributed generations and energy storage devices", in Proc. IEEE Power Electron. Spec. Conf. (PESC'06), June 2006, PP. 1–7.

[8] Spooner J.T., Maggiore M., Ordonez R., Passino K.M., "Stable adaptative control and estimation for nonlinear system, Neural and fuzzy approximator techniques", Willey-Interscience, May 2002.

[9] Dr.K.Harinadha Reddy., "Power Control using Fuzzy Logic Controller in DFIG Wind Farm", IJST, Vol. 2, No. 11, 2014, PP.47-51.
[10] Shayeghi H., Shayanfar H.A., "Power system load frequency control using RBF neural network based on μ-Synthesis theory", Proceeding of IEEE conference on cybernetics and intelligent systems , Vol. 1, Dec 2004, PP.93-98.
[11] Hua C. and Shen C., "Comparative study of peak power tracking techniques for solar storage system", in Proc. APEC, Vol.2, Feb. 1998, PP.679–685.
[12] Hohm D.P. and Ropp M.E., "Comparative study of maximum power point tracking algorithms using an experimental, programmable, maximum power point tracking test bed", in Proc. Photovoltaic Specialist Conference, Sep 2000, PP.1699–1702.
[13] Hussein K.H., Muta I., Hoshino T. and Osakada M., "Maximum power point tracking: an algorithm for rapidly changing atmospheric conditions", IEE Proc.-Gener. Transm. Distrib., Vol. 142, 1995, PP.59–64.
[14] Sun X., Wu W., Li X. and Zhao Q., "A research on photovoltaic energy controlling system with maximum power point tracking", in Power Conversion Conference, Vol. 2, 2002, PP.822–826.
[15] Quaschning V. and Hanitsch R., "Numerical simulation of current-voltage characteristics of photovoltaic systems with shaded solar cells", Solar Energy, Vol. 56, No. 6, 1996, PP.513-520.
[16] Villalva M. G., Gazoli J. R. and Filho E. R., "Comprehensive Approach to Modelling and Simulation of Photovoltaic Arrays", IEEE Trans. on Power Electronics, Vol. 24, 2009, PP.1198-1208.
[17] Nakayama H., Hiraki E., Tanaka T., Koda N., Takahashi N. Noda S., "Stand-alone photovoltaic generation system with combined storage system using EDLC", in Proc. IEEE Ind. Electron. Ann. Conf. Soc., 2009, PP.883–888.

Placement of Synchronized Measurements in Power Networks for Redundant Observability

Satyendra Pratap Singh* and S. P. Singh**
Department of Electrical Engineering, IIT (BHU) Varanasi, Varanasi, 221005, India
*spsingh.rs.eee@iitbhu.ac.in, **spsingh.eee@iitbhu.ac.in

Abstract: The rapid expansion of bulk power grids need full monitoring, operation, protection and control. To tackle these challenges, Phasor Measurement Unit (PMU) provides the real time monitoring which improves the reliability of power system and prevent blackouts. Due to higher cost of installation, PMU cannot install at every power system buses. In this paper, Gravitational Search Methodology is introduces to optimize the placement of PMUs and provide the higher observability of power system buses. This new Gravitational Search Algorithm (GSA) based method has been applied to the IEEE 14-bus, IEEE 30-bus, IEEE 118-bus and Indian Northern Regional Power Grid (NRPG) 246-bus test systems for normal and contingencies conditions both. The effectiveness of proposed methodology reveals optimal number of PMUs with redundant observability by proposed method.

Keywords: Phasor Measurement Unit, Law of Gravity, Gravitational Search Algorithm, Observability (Obs.), Optimal Placement.

Introduction

Today, synchrophasor technology is the first priority in power industries because of their reliable supports and prevents blackouts. Wide Area Measurement System (WAMS) has grown from a handful of PMUs in the power grid. Introduction of phasor measurement units have facilitated producing time synchronized measurements of voltage and current signals facilitating real time synchronized measurements [1]. Day by day, the time frame of synchronized information has been persistently reduced from minutes to microseconds. In 1980s, the first prototypes of PMUs were developed at Virginia Tech and first commercial manufacture of PMUs with Virginia Tech collaboration was started by Macrodyne in 1991. PMU data are time-stamped with high precision at the source and used for wide area measurement systems (WAMS) applications by power engineers and system operators as a time synchronized tool. PMUs are highly accurate and advanced time synchronized technology which provides the voltage and current phasor and frequency information with Global Positioning System (GPS) receivers that allow the synchronization of the several readings taken at distance points. The PMUs at different buses collects voltage and current phasors and send that data to the Phasor Data Concentrator (PDC) and PDC send that to the super PDC where many other PDCs and directly PMUs are connected. Then super PDC send that data to the control center for advance applications of the power system. Selection of power system buses to install the PMU and get desired measurements is a challenging job. To see the dynamic view of bulk power grid, PMU is the most important and reliable real time measuring device.

There are many algorithms and approaches have been published in the literature for optimal placement of PMUs in power system. Initial work on PMU development and utilization has been reported by Phadke et al. [1-2]. In [3, 4], author proposed an algorithm which finds the minimal set of PMU placement needed for power system where the graph theory and simulated annealing method have been used. In [5], a strategic PMU placement algorithm is developed to improve the bad data processing capability of state estimation. Providing selected buses with PMUs can make the entire system observable. This will only be possible by proper placement of PMUs among the system buses. The authors in [6-10] developed an optimal placement algorithm for PMUs by using integer linear programming. In [12-14], a genetic algorithm is used to find out the optimal locations of PMUs. Reference [14], is a combination of immunity algorithm and genetic algorithm. In [15], a Tabu Search (TS) method is proposed for the OPP problem in which augment incidence matrix is used for the observability analysis of PMU. A recursive TS method is suggested in [16] which is more superior than multiple Tabu Search.

The final results of OPP are in binary form hence researchers directly utilize the binary objective function. In [17, 18], a binary search method is proposed to find the optimal PMU placement. In [19], authors presented a binary particle swarm optimization (PSO) methodology for optimal placement of PMUs when using a mixed measurement set. A modified binary PSO algorithm is suggested in [20] for optimal placement of PMUs for state estimation. In [21], author used an iterated local search method to minimize the size of the PMU configuration needed to observe the power system network. In [22], author proposed the multi-objective problem of PMU placement and used non-dominated sorting differential evolution algorithm

based on pareto non-dominated sorting. An unconstrained nonlinear weighted least square algorithm is developed in [23] for optimal placement of PMUs.

This paper has been proposed Gravitational Search Algorithm (GSA) to minimize the number of PMUs with redundant observability (Obs.) for normal and contingencies both the conditions. The GSA is based on the Newton's Law of Gravity and mass interactions [24]. This methodology has been found high quality performance in solving many optimization problems in the literatures [11, 25-26]. In GSA, agents are considered as objects and their performance is measured by their masses. In GSA, heavy masses correspond to good solutions and move more slowly and conversely light masses correspond to weak solutions and fast move toward heavy masses. Finally heavy masses correspond to best solution. The objective of this paper is optimal placement of PMUs in order to achieve redundant observability.

The paper contains following discussion: First discussed the problem of minimum PMU placement in the power system then explain brief discussion of the GSA and presented the optimal PMU placement (OPP) by using GSA. Finally test results are given in this paper concludes the paper.

PMU Formulation

This section formulates the objective of optimal placement of PMUs with maximum observability. Objective function in this paper contains two objectives, first is to find the minimum number of PMUs for full observability and second is to find the best location for maximum observability. Reference [11], has been used single objective function with the same Gravitational Search Methodology but probability to get maximum observability at same number of PMUs in next trial is comparatively low due to absence of maximum observability term in the objective function. This problem overcomes in this paper which is formulated as follows [7]:

$$Min\left[\sum_{i=1}^{N} w_i\, z_i + \delta \sum_{i=1}^{N} (u_i - f_i)\right] \tag{1}$$

$$\text{Subject to, } f = AZ \geq b \tag{2}$$

$$Z = [z_1 z_2 \ldots\ldots\ldots z_N]^T$$

where, N is total number of system buses, w_i is weight factor accounting to the cost of installed PMU at bus i, A is binary connectivity matrix of the system, Z is a binary variable vector having elements z_i define possibility of PMUs on a bus i i.e x_i=1, if a PMU is needed at bus i, otherwise 0. AZ is a vector such that its entries are non-zero if the corresponding bus voltage is observable using the given measurement set and according to observability rules mentioned above. It ensures full network observability while minimizing the total installation cost of the PMUs, otherwise its entries are zero. b is a vector whose entries are all ones, it means each bus should be observed at least one time. If the entries in vector b is modified from one to two it means each bus should be observed at least two times then it provide the optimal location of PMUs in the case of single PMU outage or a single branch outage. δ is normalizing coefficient for the observability redundancy. A suitable value of δ should be chosen as equ. (3) because higher the value of δ increases the number of PMUs.

$$\delta = \frac{1}{\sum_{i=1}^{N} u_i} \tag{3}$$

u_i is the summation of i^{th} row in connectivity matrix (A). f is the observability times of bus i. The entries in A are defined as follows:

$$a_{ij} = \begin{cases} 1 & if\ i = j \\ 1 & if\ i\ and\ j\ are\ connected \\ 0 & otherwise \end{cases} \tag{4}$$

After getting the optimal number of PMUs, we can easily check the observability of each bus of the system and expression for total observability (O_{total}) is given as:

$$O_{total} = \sum_{k=1}^{P} A_{L(k)} \tag{5}$$

where, P is the total optimal number of PMUs, A is the connectivity matrix, and L is the location of PMUs at the power system buses.

Gravitational Search Algorithm (GSA)

Gravitational search algorithm is a heuristic algorithm that uses the concept of gravity and laws of motion to provide a suitable solution for an optimization problem. It is a well known fact that any two particles in the universe attract each other by a gravitational force directly proportional to the product of their masses and inversely proportional to the distance between them [24]. This concept is utilized in GSA along with the laws of motion, where agents are treated as objects and performance of agents is considered by their masses with gravitational force acting as a mode of communication between them.

The basic idea of GSA has been discussed in [11]. Now, consider a system with Q agents or masses. We define the position of the i_{th} agent by:

$$X_i = \left(x_i^1, \ldots, x_i^d, \ldots, x_i^n\right) \text{ for } i=1,2,3..,Q \tag{6}$$

where, x_i^d represents the position of i_{th} agent in the d_{th} dimension. At a specific time't', the force acting on mass 'i' from mass 'j' can be defined as following:

$$F_{ij}^d = G(t)\frac{M_{pi}(t)\times M_{aj}(t)}{R_{ij}(t)+\varepsilon}\left(x_j^d(t) + x_i^d(t)\right) \tag{7}$$

where M_{aj} is the active gravitational mass of agent j, M_{pi} the passive gravitational mass of agent i, $G(t)$ gravitational constant at time t, ε a small constant, and $R_{ij}(t)$ is the Euclidean distance between two agents i and j. The total force acting on the i_{th} agent ($F_i^d(t)$) is calculated as follows:

$$F_i^d(t) = \sum_{j\in Kbest, j\neq i} rand_j\, F_{ij}^d(t) \tag{8}$$

where $rand_j$ is a random number in the interval [0, 1] and k_{best} is the set of first K agents with the best fitness value and biggest mass. k_{best} is a function of time, with the initial value of K_0 at the beginning and decreasing with time. In such a way, at the beginning, all agents apply the force, and as time passes, k_{best} is decreased linearly and at the end there will be just one agent applying force to the others.

By the law of motion, the acceleration of agent i at time t, in direction d_{th} is given by:

$$Ac_i^d(t) = \frac{F_i^d(t)}{M_{ii}(t)} \tag{9}$$

where M_{ii} is the inertial mass of i_{th} agent.

Furthermore, the next velocity of an agent is a function of its current velocity added to its current acceleration. Therefore, the next position and the next velocity of an agent can be calculated as follows:

$$v_i^d(t+1) = rand_i \times v_i^d(t) + Ac_i^d(t) \tag{10}$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{11}$$

where $rand_i$ is a uniform random variable in the interval [0, 1]. The gravitational constant G, is initialized at the beginning of problem and will be decreased with time to control the search accuracy [24].

$$G(t) = G_o e^{-\alpha\frac{t}{T}} \tag{12}$$

where α is a user specified constant, t is the current iteration and T is the total iterations. Gravitational and inertial masses are simply calculated by the fitness evaluation. Both of the masses namely gravitational and inertial are assumed to be equal and their values are calculated using the map of fitness. The gravitational and inertial masses are updated by using following equations:

$$M_{ai} = M_{pi} = M_{ii} = M_i \tag{13}$$

where, i=1, 2, 3....,Q

$$m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \tag{14}$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{Q} m_j(t)} \quad (15)$$

where $fit_i(t)$ represents the fitness value of the agent *i* at time *t*, and the *best(t)* and *worst(t)* in the population of agents respectively indicate the strongest and the weakest agent according to their fitness and can be defined as follows:
For a minimization problem:

$$best(t) = \min_{j\in\{1,..,Q\}} fit_j(t) \quad (16)$$

$$worst(t) = \max_{j\in\{1,...,Q\}} fit_j(t) \quad (17)$$

The procedure to apply the GSA for solving the OPP problem is as follow:

Step 1. Read bus data and line data of the test system.
Step 2. Find the connectivity matrix (*A*) from line data.
Step 3. Initialize GSA parameters *T*, *Q*, G_o and α.
Step 4. Identify the search space.
Step 5. Generate initial population between minimum and maximum values of the control variables.
Step 6. The fitness values of each agent in the population are calculated for the OPP problem.
Step 7. Based on fitness value, update *G*(*t*), best(t), worst(t) and M_i(t) for i=1, 2, ….., *Q*
Step 8. Calculation of total force in different directions using Eq. 8.
Step 9. Acceleration of each agent is modified using Eq. 9.
Step 10. The velocity and position of each agent is updated using Eq. 10 and Eq. 11 respectively.
Step 11. Repeat steps 6-10 until the stop criterion is reached.
Step 12. Stop

Case Study

This paper, proposed a simple case of power system which has been applied to check the effectiveness of proposed methodology for both the simple and contingency conditions. Single PMU outage and single line outage have been considered as a contingency in this paper. Finally the objective function of proposed method has been compiled in Newton method based GSA which is the new heuristic optimization technique. Fig. 1 and Fig. 2 show the single line diagrams of IEEE 14-bus and IEEE 30-bus test systems obtained from DIgSILENT software respectively. All the experiments are executed on a computer having the following configuration: Intel core i3 CPU @ 3.40 GHz, 2 GB RAM.

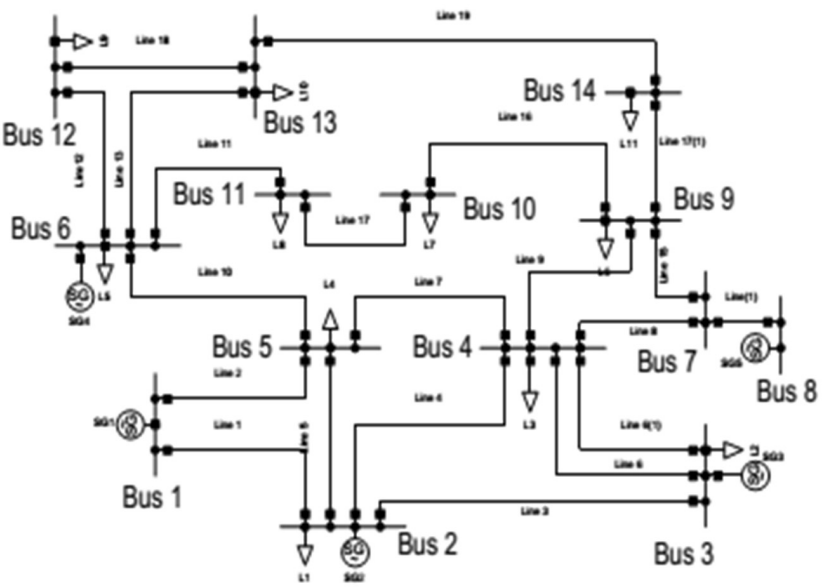


Figure 1. IEEE 14-bus test system

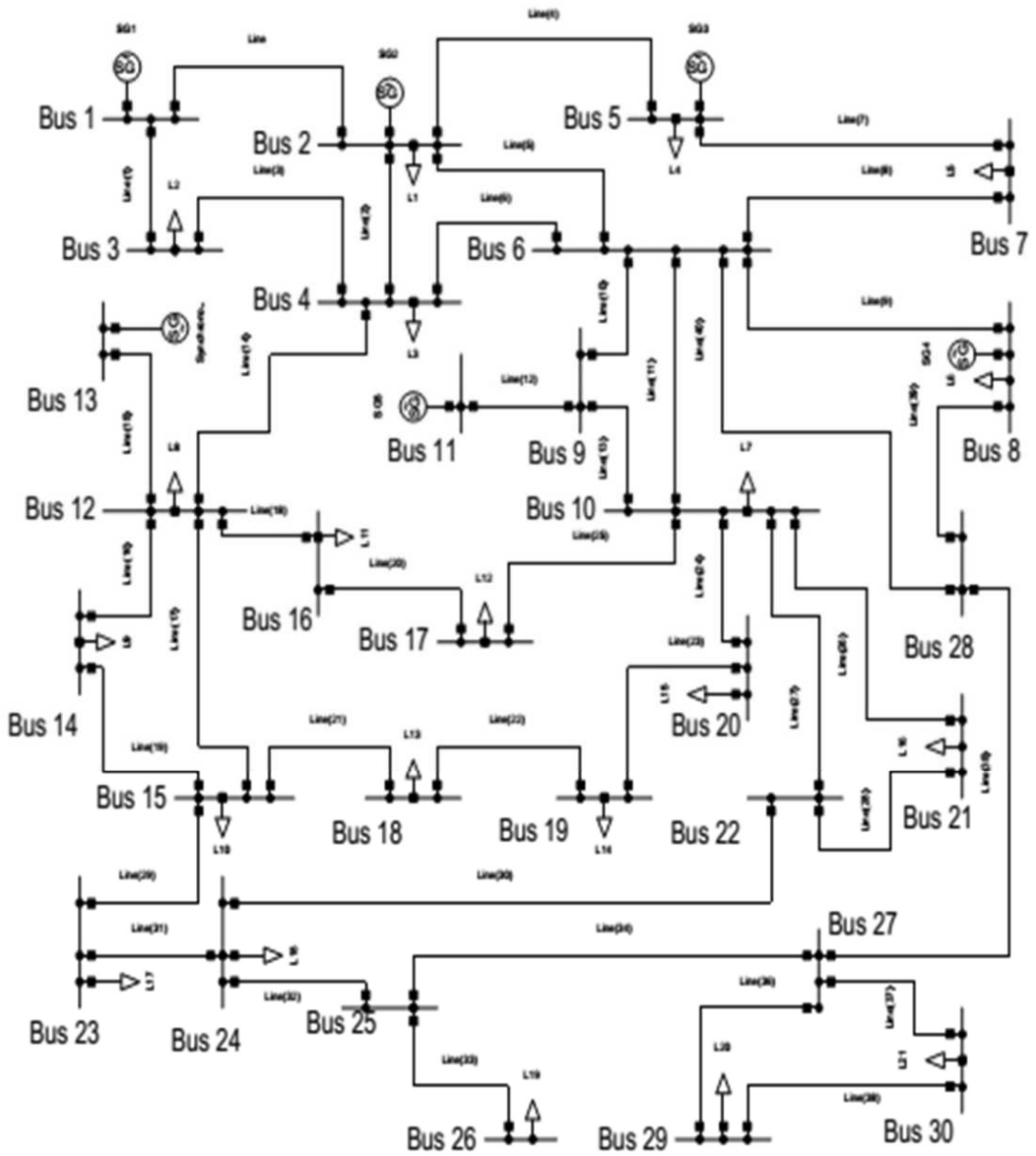


Figure 2. IEEE 30-bus test system

Table 1 shows the chosen values of the parameters for the GSA used for the OPP problem solution. To find out the best result, selections of GSA parameters play an important role. These values have been arrived at by various trials of OPP solutions on all the system tested. The values listed in Table 1 produced best performance in terms of finding the optimal solution and computational time.

Table 2 shows the optimal number of PMUs and their locations which provide the full observability of power network. Fig. 3 shows the comparison of convergence curve of GSA for the IEEE 118-bus system. A steep decline in objective function value is observed in Fig. 3. The computation times of proposed OPP method for IEEE 14-bus, IEEE 30-bus, IEEE 118-bus and NRPG 246-bus [27] test systems are 0.60 sec, 0.76 sec, 3.32 sec and 15.3 sec respectively. It is observed from Fig. 3, the GSA converged in seventy three iterations for ref. [11] and suggested thirty two PMUs for full observability of the power system. But same problem with proposed methodology suggested same number of PMUs in less convergence of GSA as compared to [11] and probability to get same results (maximum Obs.) is fixed with proposed methodology which is the main advantage over the ref. [11].

The results of proposed algorithm have been compared with the previously reported methods in the literature and shown in Table 3. For all the test systems, minimum numbers of PMUs are same as the previously reported methods but the locations of PMUs by proposed method are different which provide the redundant observability of each bus as shown in Table 4.

Table 1. GSA Parameter

G_0	α	$Iter_{max}$	Q
1	20	150	50

Table 2. Optimal Location of PMUs for Test Systems

System configuration	Minimum no. of PMUs	Optimal PMU Locations	CPU Time (s)
IEEE 14-bus	4	2, 6, 7, 9	0.60
IEEE 30-bus	10	2, 4, 6, 9, 10, 12, 15, 20, 25, 27	0.76
IEEE 118-bus	32	3, 5, 9, 11, 12, 17, 21, 25, 28, 34, 37, 41, 45, 49, 52, 56, 62, 63, 68, 70, 71, 76, 79, 85, 86, 89, 92, 96, 100, 105, 110, 114	6.32
NRPG 246-bus	70	6,21,23,24,29,34,40,45,48,54,55,57,60,61,62,63,65,69,73,74,75,78,80, 88,93,95,98,100,101,102,103,106,109,116,117,121,122,125,126,129,1 32,134,140,141,142,147,157,158,160,163,168,173,181,183,185,187,19 0,191,194,199,201,202,203,215,216,219,234,235,243,245	15.3

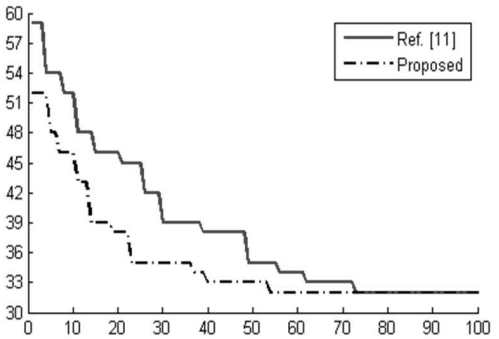


Figure 3. Convergence of GSA for IEEE 118-bus test system

Table 3. Comparison of Obtained Results by Several Methods

Test System	IEEE 14-Bus	IEEE 30-Bus	IEEE 118-Bus	NRPG 246-Bus
Proposed method	4	10	32	70
Generalized ILP [8]	4	10	NA	NA
Ref. [11]	4	10	32	NA
Binary Search Algorithm [17]	4	10	--	NA
WLS [23]	4	10	32	NA

Table 4. Comparison of obtained results on the basis of each bus observability for IEEE 30-bus

	Observability (Obs.) of each bus	Total observability
Proposed method	1 3 1 4 1 5 1 1 3 **4** 1 3 1 2 2 1 1 **1** 1 2 1 1 1 1 2 1 2 2 1 1	52
Ref. [11]	1 3 1 4 1 5 1 1 3 **3** 1 3 1 2 2 1 1 **2** 1 2 1 1 1 1 2 1 2 2 1 1	52
Ref. [17]	2 3 1 3 1 4 1 1 3 3 1 2 1 2 2 1 1 2 1 2 1 1 1 1 2 1 2 2 1 1	50
Ref. [23]	1 1 1 1 1 3 1 1 2 1 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	35

After the modification of entries in vector *b* of Equ. (2), proposed method provides the optimal PMU placement in the case of single PMU outage. If the system is observable in case of single PMU outage it means system will also be observable in case of single branch outage, because single branch outage case is the subset of single PMU outage case. Table 5 shows the results in case of single PMU outage or single branch outage with comparison to other method [6]. For all the IEEE test systems, and Indian NRPG 246-bus system, proposed method provides the maximum observability as compared to [6] in Table 5. Also the

Table 5. OPP in case of single PMU outage or single branch outage case with comparison

Test System	IEEE 14-Bus		IEEE 30-Bus		IEEE 118-Bus		NRPG 246-bus	
	No. of PMUs	Total Obs.	No. of PMUs	Total Obs.	No. of PMUs	Total Obs.	No. of PMUs	Total Obs.
Proposed method	9	39	21	85	68	305	153	614
Ref. [6]	9	34	22	85	72	316	NA	NA

number of PMUs in IEEE 30-bus and 118-bus systems are minimum with maximum observability. In this case, computation time for IEEE 14-bus and IEEE 30-bus is around same as previous case but for IEEE 118-bus system and for Indian NRPG 246-bus system it is 7.8 sec and 17.9 sec respectively.

Conclusion

This paper presented a Gravitational Search method for solving the multiobjective OPP problem, which minimizes the total number of PMUs and provides the maximum observability of the power system for both the normal and contingencies conditions. The method was applied on IEEE 14-bus, IEEE 30-bus, IEEE 118-bus and Indian Northern Regional Power Grid (NRPG) 246-bus test systems and its results are compared with other methods reported in the literatures. The simulation results and fast convergence time indicate that the proposed algorithm satisfactorily provides maximum observable system measurements with minimum number of PMUs. The maximum observability of proposed methodology was found to be better than other methods.

References

[1] A. G. Phadke, J. S. Thorp and K. J. Karimi, "State Estimation with Phasor Measurements", IEEE Trans. on Power Systems, Vol. 1, No. 1, February 1986, pp. 233- 241.

[2] EPRI Final Rep., 1997 "Assessment of Applications and Benefits of Phasor Measurement Technology in Power Systems," GE Power Syst. Eng.

[3] L. Mili, T. Baldwin and R. Adapa, "Phasor Measurement Placement for Voltage Stability Analysis of Power Systems." Proceedings of the 29th Conference on Decision and Control, Honolulu, Hawaii, Dec. 1990.

[4] T. L. Baldwin, L. Mili, M. B. Boisen, and R. Adapa, "Power System Observability With Minimal Phasor Measurement Placement", IEEE Transactions on Power Systems, Vol. 8, No. 2, May 1993, pp. 707-715.

[5] J. Chen and A. Abur, "Placement of PMUs to Enable Bad Data Detection in State Estimation." IEEE Trans. on Power Systems, Vol. 21, No. 4, Nov. 2006, pp. 1608-1615.

[6] B. Xu, A. Abur, "Optimal placement of phasor measurement units for state estimation", Final project report, PSERC, 2005.

[7] M. Esmaili, K. Gharani and H. A. Shayanfar, "Redundant Observability PMU Placement in the Presence of Flow Measurements Considering Contingencies", IEEE Trans. on Power Systems, Vol. 28, No. 4, 2013, pp. 3765-3773.

[8] B. Gou, "Generalized integer linear programming formulation for optimal PMU placement" IEEE Trans. Power Syst., Vol. 23, No. 3, Aug. 2008, pp. 1099-1104.

[9] S. P. Singh, S. P. Singh, "Optimal PMU Placement in Power System Considering the Measurement Redundancy", Int. J. of Advances in Electronic and Electric Engineering, Vol. 4, No. 6, Jan. 2014, pp. 593-598.

[10] S. Kesharvani, S.P. Singh and S.P. Singh, "Voltage Stability Assessment using Phasor Measurement Units in Power Network with Full System Observability", ICPCES-2012, Dep. of Electrical Engg., MNNIT, Allahabad, Dec. 17-19, 2012

[11] Satyendra. P. Singh, S. P. Singh, "Optimal Placement of Phasor Measurement Units Using Gravitational Search Method", Int. J. of ECECE, Vol. 9, No. 3, 2015, pp. 268-272

[12] B. Milosevic and M. Begovic, "Nondominated Sorting Genetic Algorithm for Optimal Phasor Measurement Placement." IEEE Trans. On Power Systems, Vol. 18, No. 1, Feb. 2003, pp. 69-75.

[13] F. J. Marın, F. Garcıa-Lagos, G. Joya, and F. Sandoval, "Genetic algorithms for optimal placement of phasor measurement units in electric networks," Electron. Lett., Vol. 39, No. 19, Sep. 2003, pp. 1403–1405.

[14] F. Aminifar, C. Lucas, A. Khodaei, and M. Fotuhi, "Optimal placement of phasor measurement units using immunity genetic algorithm," IEEE Trans. Power Del., Vol. 24, No. 3, Jul. 2009, pp. 1014–1020.

[15] J. Peng, Y. Sun, and H. F. Wang, "Optimal PMU placement for full network observability using Tabu search algorithm," Elect. Power Syst.Res., Vol. 28, 2006, pp. 223–231.

[16] N. C. Koutsoukis, N. M. Manousakis, P. S. Georgilakis, G. N. Korres, "Numerical observability method for optimal phasor measurement units placement using recursive Tabu search method", IET Gener. Transm. Distrib., Vol. 7, No. 4, 2013, pp. 347–356.

[17] S. Chakrabarti, and E. Kyriakides, "Optimal Placement of Phasor Measurement Units for Power System Observability" IEEE Trans. Power Syst., Vol. 23, No. 3, Aug. 2008, pp. 1433-1440.

[18] R. J. Albuquerque and V. L. Paucar, "Evaluation of the PMUs Measurement Channels Availability for Observability Analysis", IEEE Trans. On Power Syst., Vol. 28, No. 3, Aug. 2013, pp. 2536-2544.

[19] A. Ahmadi, Y. A. Beromi, M. Moradi, "Optimal PMU placement for power system observability using binary particle swarm optimization and considering measurement redundancy", Expert Systems with Applications, Vol. 38, 2011, pp. 7263–7269.

[20] M. Hajiana, A. M. Ranjbar, T. Amraee, B. Mozafari, "Optimal placement of PMUs to maintain network observability using a modified BPSO algorithm", Elec. Power & Energy Sys. Vol. 33, No. 1, 2011, pp. 28–34.

[21] M. Hurtgen and J.-C. Maun, “Optimal PMU placement using iterated local search”, Elec. Power & Energy Sys. Vol. 32, No. 8, 2010, pp. 857–860.
[22] C. Peng, H. Sun and J. Guo, “Multi-objective optimal PMU placement using a non-dominated sorting differential evolution algorithm”, Elec. Power & Energy Sys. Vol. 32, No. 8, 2010, pp. 886–892.
[23] N. M. Manousakis and G. N. Korres, “A Weighted Least Squares Algorithm for Optimal PMU Placement”, IEEE Trans. On Power Syst., Vol. 28, No. 3, Aug. 2013, pp. 3499-3500.
[24] E. Rashedi, H. Nezamabadi-pour, S. Saryazdi, "A gravitational search algorithm", Information Sciences, Vol. 179, pp. 2232-2248, 2009.
[25] H. R. Hassanzadeh, M. Rouhani, “A multi-objective gravitational search algorithm” Second ICCI, communication systems and networks conference, 2010, pp. 7–12.
[26] S. Duman, U. Guvenc, N. Yorukeren, “Gravitational search algorithm for economic dispatch with valve-point effects” IREE., Vol. 5, No. 6, Dec. 2010, pp. 2890-2995.
[27] Northern Regional Load Dispatch Center: available at: http//nrldc.org

Cationic Catalysis of Lime Treated Soft Soils

Greeshma Nizy Eujine*, S Chandrakaran** and N Sankar***
*Research Scholar, Department of Civil Engineering, National Institute of Technology Calicut, Kerala – 673601, India.
Email- nizy_123@yahoo.com
** ***Professor, Department of Civil Engineering, National Institute of Technology Calicut, Kerala– 673601, India

Abstract: Stabilization of soils is an effective method for improvement of soil properties and the pavement system performance. A study was conducted to examine the stabilizing effects of an enzyme and lime on improving the suitability of a natural soft clay as a subgrade material. The tests were carried out by introducing the additives into the soil in two different stages. The enzyme and lime were first mixed individually- Soil Lime mixtures and Soil Enzyme mixtures and later concurrently- Soil Enzymatic Lime mixtures. The initial part of the study determined the variation of index properties of the three soil combination. Laboratory tests were performed with different percentages of the additives on the soil and the modification of Atterberg Limits, Grain Sizes Distribution and Free Swell Index in the soil mixtures was studied week after week for a month. From the result of the plasticity indexes obtained, satisfactory values were noted and the respective dosages were chosen as the optimum. The latter part of the study examined the optimum dosages of enzyme, lime and enzymatic-lime that produced maximum strength improvement on the treated soil. The strength criteria were investigated by conducting unconfined compressive tests on 3.8mm diameter cylindrical specimens of samples cured with the dosages at and near the optimum combinations. The effect of density and moisture content was predominant in all cases and this enabled to identify the maximum dry density (MDD) and optimum moisture content (OMC) for the mixes. It was found that enzymatic-lime stabilized soils gave speedy stabilization as compared to enzyme stabilized and lime stabilized soils. The paper discusses the results obtained throughout the study. Under standard conditions, satisfactory reasons are found stating that Enzymatic Lime is a possible economical and innovative methodology for soil stabilization.

Keywords: Ground Improvement, Enzyme, Lime Stabilization, Soft Soil.

Introduction

In situ soil stabilization is commonly brought about by compaction, rearrangement of soil particles, addition of chemical reactants, or thermal process. Among these techniques, hydrated lime has proved to be a cost effective, easy to apply, favored method of soft soil improvement for most engineers. Rooted in the success of lime stabilized soils, various additives have been suggested to be incorporated with soil and with soil-lime and tried over the last few decades. Numerous research works are and have been done to understand the behavior of soil-lime mixtures in the presence of other salts/chemicals. But the effect of various inorganic additives such as calcium chloride and calcium sulphate with lime on the behavior of clay is yet to be studied in detail. Based on this idea, it was decided to integrate an enzyme into soils along with hydrated lime. The enzyme used in the work is an enzyme soil stabilizer produced in the United States by Nature Plus, Inc. A number of case studies have been reported [1] stating that enzyme by itself improves soil properties while no literature has been found on enzymatic lime soil stabilization. It is to be assumed in the work that like soils subjected to other modes of treatment, any improvements in the engineering behavior of enzymatic lime treated soil systems will mainly be attributed to the aggregation effects [2]The scope of the paper includes observation of the changes in Atterberg Limits of lime and enzyme treated soil systems and the unconfined compressive strength improvement of soil samples subjected to such treatments.

Mechanism of stabilization

As mentioned, the paper describes the use and effect of chemical agents lime and enzyme on the index properties of stabilized soils. The utilization of lime in soil modification is not a novel technology. It is a traditional means in a variety of construction applications since the time of Romans and has never entirely disappeared. When used in soil, lime modification describes an increase in strength brought by cation exchange capacity rather than cementing effect brought by pozzolanic reaction [3]. It alters the clay surface mineralogy, producing a reduction in plasticity and moisture holding capacity, and an improvement in soil stability. But the disadvantages of lime stabilization include lime carbonation and sulfate salt reactions which may lead to disintegration of bonds on aging. To account for the negative impacts, a number of salts and chemicals have been added with lime to soil and tested. Among these already tried and proved agents include cement, flyash, rice husk etc.

It was hence decided to mix an enzyme with lime for the purpose of soil stabilization and to study its effects on soil properties. Enzymes are organic molecules that catalyze very specific chemical reactions if conditions are conducive to the

reaction. They are typically used in low concentrations as they are not consumed by the reactions they make possible. Enzyme additives attach with large organic molecules that are attracted to the clay mineral's net negative surface charge [4] Enzyme used in the work like other enzymes is costly and a method to reduce the amount of enzyme, while obtaining the same degree of improvement, if achievable, would be greatly advantageous. And since lime and enzyme use the same mechanism of cation exchange to improve soil properties, the idea to add both lime and enzyme together in the soil and to investigate the alterations in soil properties seemed feasible. On addition of both lime and TerraZyme to soil, from laboratory tests it was observed that a higher degree of strength improvement was achieved by this method. The soils achieved maximum property enhancement in two weeks or less as compared to four/six weeks in other ground improvement techniques.

Materials used and methodology

The materials used in the work are, natural clay soil organized from Pantheerancavu at Calicut, lime purchased from local market in Kunnamagalam at Calicut and enzyme TerraZyme acquired from Avijeet agencies, Chennai. The soil and enzyme have been characterized and their properties are given in Tables 1 and 2.

The soil used in the work is montmorillionitic soft clay. High percentage of montomorillonite in soils renders high degree of expansiveness. When used as a subgrade material, these soils may cause the pavements to crack without any warning . Soil stabilization by the alteration of one or more soil properties, by mechanical or chemical means, to create an improved soil material possessing the desired engineering properties [5] is the sole remedy in such cases.

The key to success in soil stabilization is soil testing. All the specimens tested in this study were prepared and tested using standard procedures described in the Bureau of Indian Standards. Soil sample was air dried for a week, pulverized manually using weights, sieved through 425 micron sieve and preserved in large open containers in an enclosed room. Lime was sieved using 425 micron sieve and preserved in an air tight container to prevent carbonation. TerraZyme was preserved in an airtight bottle in its original liquid form. Soil and lime were mixed from 0%-10%.

The consequence of lime on the plasticity of clay soils is observed more or less instantaneous. In other words, the plasticity is reduced (this is brought by an increase in the plastic limit and reduction in the liquid limit of the soil), as is the potential for volume change. [6]Hence the variations in LL and OMC were observed for each fraction within 24 hours. The minimum amount of lime that did not further reduce the LL was chosen as optimum Lime content (OLC). The OMC increased by about 5% by OLC and later remained constant. The minimum lime required to initiate soil modification was 0.5 and OLC was 3%. However soil specimens were cast in UCS moulds (3.6mm diameter and 7.8mm height) with 1,2,3,5,7 % lime at their respective OMC and cured up to four weeks in airtight bags kept in an artificial desiccator. At least six samples of the same mix type were tested and the average of the values were recorded. The variations in Liquid Limit (LL), Plastic Limit (PL), Shrinkage Limit (SL), Unconfined Compressive Strength (UCS), Free Swell Index (FSI) and particle sizes were observed. From the results the OLC was once again inferred as 3%.

Table 1. Engineering Properties of Soil

Property	Value
Liquid Limit (%)	79
Plastic Limit (%)	48
Shrinkage Limit (%)	27
Bulk Density (kN/m^3)	13.25
Max. Dry Density (kN/m^3)	17.46
Optimum Moisture Content (%)	32

Property	Value
Soil Type: Montmorillitic	
Unconfined Compressive Strength (kPa)	64
Clay content (%)	22.17
Co-efficient of Curvature, Cc	1.053
Co-efficient of Uniformity, Cu	8.16

Table 2. Properties of TerraZyme

Property	Value
Boiling Point	212 F
pH	2.8 – 3.5
Vapor Pressure (mmHg)	As Water
Melting Point	Liquid
Vapor Density (Air = 1):	1
Solubility in Water	Infinite

Property	Value
Evaporation Rate	As Water
Specific Gravity (H_2O = 1)	1.00—1.10
Appearance and Odor	Lt. Gold Liquid, Characteristic Odor

Other parameters could also have been the objective of investigation. But this paper is only relates to observation of index properties and UCS tests. Similar tests were also done to determine the optimum TerraZyme content of soil. A dilution ratio chart was provided by Avijeet agencies that calculated the optimum TerraZyme content (OEC) for a particular soil based on particle size and plasticity index. Two soils were mixed with TerraZyme, the original soil used in the research work (Soil-A) and another montmorillitic soil obtained from Quilandy (Soil-B), which had similar properties as the former. Both soils showed significant improvement in strength.

The OEC was found to be same as that computed from the dilution ratio chart. The chart was thus validated. The OEC for soil used in research was obtained as 80ml/m^3. There was no variation in OMC on addition of TerraZyme to the soil. Soil was mixed at OMC itself for all curing tests. Various combinations of soil+lime+enzyme mixtures were cast, cured and tested. The amount of TerraZyme was kept close to OEC predominantly in all cases and percentage lime was varied from 1-6%. The changes in Atterberg limits and UCS were observed within 24 hours of mixing enzymatic solutions to the soil. It was observed that the strength gain was constant or remained same when the TerraZyme content was slightly increased above 70 ml/m^3. On TerraZyme dosage above 90 ml/m^3 the strength was found to decrease drastically. Thus another set of tests were done with the new OEC and varying percentages of lime. The best combination was obtained on adding 2% lime. Further refinement tests were performed by varying lime between 1.5 and 2.5%.

Results and discussion

Soil - Lime system

Bell found that the optimum addition of lime needed for the stabilization of the soils is between 1% and 3%, while the other researchers suggested the use of lime between 2% and 8% lime by weight [6] The addition of lime results in a decrease in the liquid limit and plastic limit, a commonly observed consequence of lime addition (Sweeney et al, 1965). In the current work, the optimum lime content was found out indirectly by relying on the variation of LL values of soil, when treated with lime. The lowest LL was obtained when 3% of lime by weight was added to the soil. The OMC of soil increased by 4% at the dosage. Later air dried and pulverised, 425 micron sieved soil was mixed with lime at 1,2,3,5,7% lime at their respective OMC and cast in to cylindrical specimens in UCS moulds (3.6mm diameter and 7.8mm height) and cured up to four weeks. The specimens were tested every week for a month. Curing period significantly influenced the engineering properties of lime stabilized soil. Table 3 and figure 1 through 4 represent the respective modifications in Index Property and Unconfined Compressive Strengths. The clay contents quoted for the natural and treated soils are the percentages of particles finer than 2 micron, determined by sedimentation after pretreatment and dispersion with sodium hexametaphosphate.

Table 3. Engineering property of soil treated with various percentages of lime at four weeks

Property		Liquid Limit	Plastic Limit	Shrinkage Limit	UCC (kPa)	Free Swell Index	Silt (%)	Clay (%)
Untreated		79	48	27	90	20	23.98	22.17
Lime (%)	1	55	40	24	188	16	19	5
	2	52	37	23	202	16	18	4.9
	3	48	35	23	272	16	18.5	1.78
	5	54	38	23	192	16.5	16.1	0.726
	7	54	37	23	190	16	18	0.632

Soil mixed with low lime content attains a maximum strength in less time than that to which a higher content of lime has been added. Strength does not increase linearly with lime content and in fact excessive addition of lime reduces strength. This decrease is because lime itself has neither appreciable friction nor cohesion. The optimum lime content tends to range from 4.5% to 8%, the higher values being required for soils with higher clay fractions. [6]. On successful addition, curing and testing of soil samples with lime, the optimum lime content (OLC) for soil was verified as 3%. At this dosage the LL, PL, SL and FSI reduced to 48, 35, 23 and 16 respectively. The Plasticity Index (PI) reduced from 31 in the untreated soil to 13. The UCS value was obtained as 272kPa at the end of four weeks.

Literature reveals that reaction products formed due to soil-lime reactions show the formation of the reaction product CSH (gel) in a reticulated network (well-knit framework) which binds the individual clay particles together to form aggregations. Locat [2] reported formation of platy CASH and reticular CSH cementitious compounds in the lime-treated soil system from SEM analysis. This may be one of the reasons for the increase in UCS values of soils treated with lime.

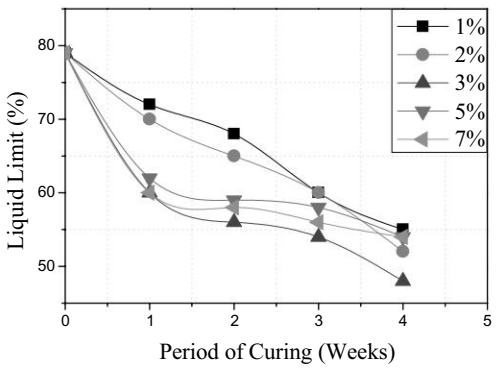


Figure 1. Variation of Liquid Limit Limits of soil treated with varying percentages of lime

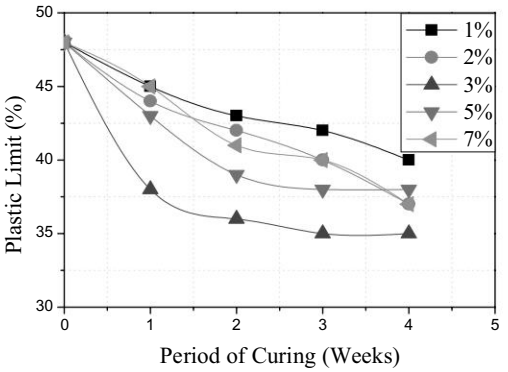


Figure 2. Variation of Plastic Limits of soil treated with with varying percentages of lime

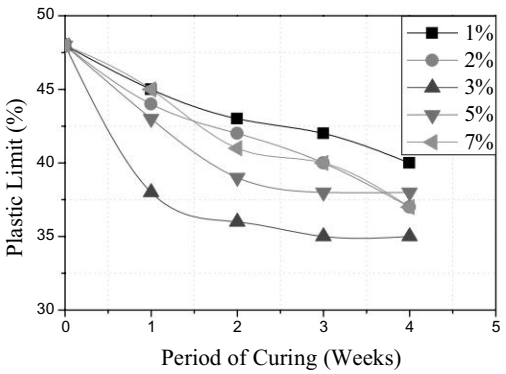


Figure 3: Variation of Effective Clay Size Particle Distribution of soil treated with with varying percentages of lime

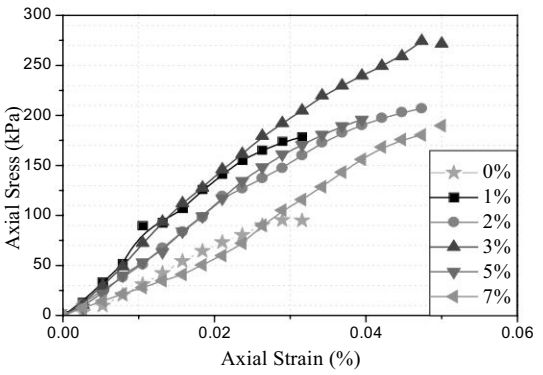


Figure 4: Stress Strain Curve of soil after 4 weeks of curing with with varying percentages of lime

Soil - Enzyme System

The dilution chart provided by NaturePlus was used to obtain an optimum enzyme dosage for soil. In order to verify the accuracy of the dilution chart, the original soil sample used in the research work (Soil-A) and another montmorillionitic soil of similar characteristics (Soil-B) were mixed with TerraZyme. The dosages chosen were 80% to 120% of the optimum dosage given by the chart. Air dried and pulverized, 425 micron sieved samples of the soil each were taken and moulded in UCS moulds (3.8mm diameter and 7.6mm height), and cured up to a month. The cured samples were subjected to unconfined compressive strength tests every week for a month. Based on the results, it was found that maximum unconfined compressive strengths was obtained for samples cast with optimum TerraZyme dosage given by chart. The samples cast with 80% and 120% optimum dosage has reduced strength with the latter percentage much lower strength values. For the next step refined TerraZyme dosages of 60 ,80, and 100 ml/m^3 were chosen, dosages near to optimum dosage. Similar to soil-lime system, air dried and pulverized, 425 micron sieved samples were mixed with the three refined TerraZyme dosages at pure soil OMC and cast in to cylindrical specimens in UCS moulds (3.6mm diameter and 7.8mm height) and cured up to four weeks. The specimens were tested every week for a month. The parameters tested were the Index Property and Unconfined Compressive Strengths.

It is observed that LL, PI and UCS alters appreciably for all stabilized soils. While during the stabilization of aggregate clay mixes with enzymes, the plasticity characteristics may fluctuate either way [7] on addition to an increase in strength characteristics. In the present study, the montmorillitic clay on reaction with TerraZyme shows decreased LL and PI. On successful addition, curing and testing of soil samples with TerraZyme, the optimum enzyme content (OEC) for soil was verified as 80 ml/m^3. At this dosage the the LL increased by 91, PL, SL and FSI reduced to 42, 17, and 16 respectively. PI increased to 49 and UCS value was obtained as 306kPa at the end of four weeks.

Dumbleton observed that cementing or aggregation of the particles of a soil will modify the effect which the clay fraction has on soil properties. On curing such a soil has reduced plasticity characteristics [8]The increase in strength and reduction in plasticity of soil on addition of TerraZyme may be due to the aggregation of soil particles. Some literature suggests formation of rough elephant skin like textural modification of soil and reduction in surface area of clay minerals, when treated with TerraZyme. Taha revealed that ESEM images for the treated and untreated samples confirmed that the treated samples appear

more aggregated than the corresponding untreated sample, and the clay features are less visible [9]This cohesiveness and interlayer friction may add to the cause behind the strength increase. More detail into the mechanism of TerraZyme soil stabilization may be obtained from SEM studies, but that is beyond the scope of this paper, and is not mentioned here.

Table 4. Engineering property of soil treated with various percentages of TerraZyme at four weeks

Property		Liquid Limit	Plastic Limit	Shrinkage Limit	UCC (kPa)	Free Swell Index	Silt (%)	Clay (%)
Untreated		79	48	27	90	20	23.98	22.17
TerraZyme(ml/m^3)	70	88	43	17	182.5	16	19.5	10.2
	80	91	42	17	306.2	16	18	6.8
	90	86	42	16	135.6	16	16	6

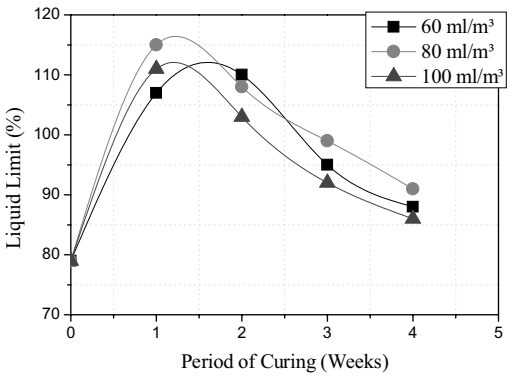


Figure 5. Variation of Liquid Limit of soil treated with with varying percentages of TerraZyme

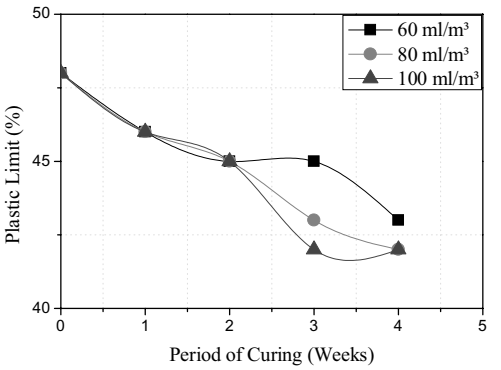


Figure 6. Variation of Plastic Limit of soil treated with with varying percentages of TerraZyme

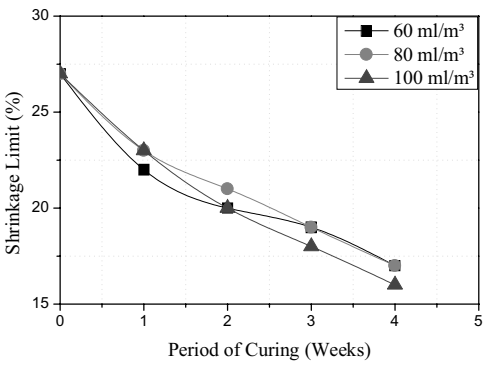


Figure 7. Variation of Effective Clay Size Particle Distribution of soil treated with with varying percentages of TerraZyme

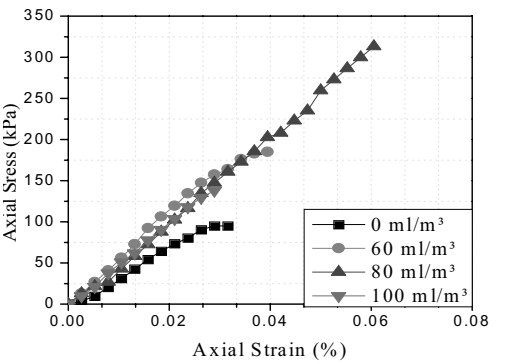


Figure 8. Stress Strain Curve of soil after 4 weeks of curing with with varying percentages of TerraZyme

Soil-Enzyme-Lime system

The results obtained in the soil-lime and soil-enzyme were positive. And it has been confirmed that both lime and TerraZyme use the mechanism of cation exchange for soil stabilization. Hence varying percentages of lime and enzyme were mixed together and tested. Variation in the values of Atterberg Limits and UCS in the first 24 hours were regarded to understand the consequence of enzymatic lime system in the soil. The effect of TerraZyme was limited when dosage was increases above 70 ml/m^3 in the presence of lime. Thus further tests were done with the new OEC as 70 ml/m^3 and varying percentages of lime from 1 and 5%. Air dried and pulverized, 425 micron sieved samples were mixed with the new OEC, and various lime percentages at pure soil OMC and cast in to cylindrical specimens in UCS moulds (3.6mm diameter and 7.8mm height) and cured up to four weeks. The specimens were tested every week for a month. Based on the values of UCS, lime percentages were further refined to be between 1.5 and 2.5 %.

Table 5. Engineering property of soil treated with various percentages of enzymatic Lime at four weeks

Property		Liquid Limit	Plastic Limit	Shrinkage Limit	UCC (kPa)	Free Swell Index	Silt (%)	Clay (%)
Untreated		79	48	27	90	20	23.98	22.17
Optimum TZ and Varying % Lime	1.5	88	43	17	182.5	16	19.5	10.2
	1.75	91	42	17	318.2	16	18	6.8
	2	86	42	16	135.6	16	16	6

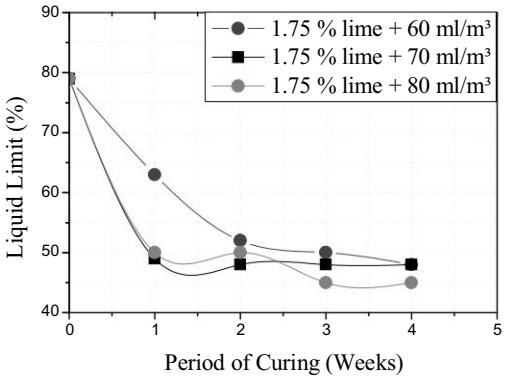


Figure 9. Variation of Liquid Limit of soil treated with with varying percentages of enzymatic lime

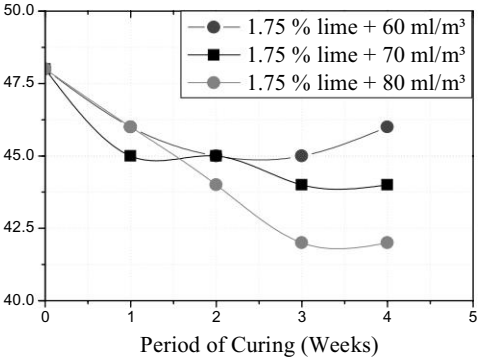


Figure 10. Variation of Plastic Limit of soil treated with varying percentages of enzymaticlime

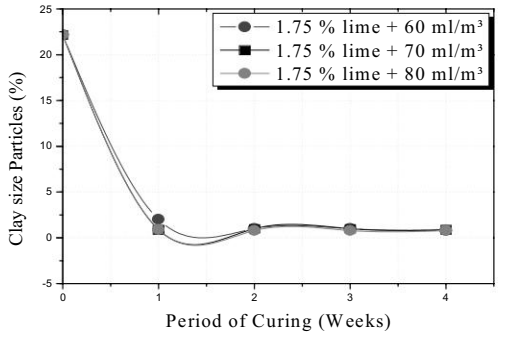


Figure 11. Variation of Effective Clay Size Particle Distribution of Soil treated with varying percentages of enzymatic lime

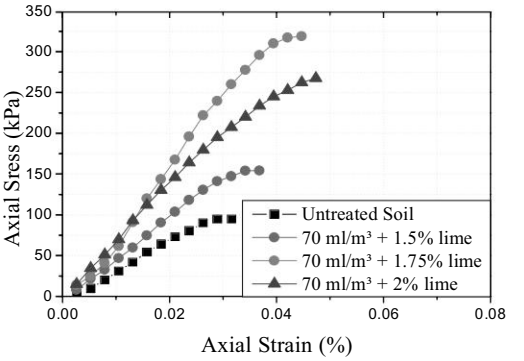


Figure 12. Stress Strain Curve of soil after 4 weeks of curing with varying percentages of enzymatic lime

The parameters tested were the Index Property and Unconfined Compressive Strengths in this step. The end result verified that 1.75% lime presented optimal improvement. Thus final optimal mix was obtained as 70ml/m^3 of TZ On successful addition, curing and testing of soil samples with enzymatic lime, the optimum additive content (OLT) for soil was verified as 70 ml/m^3 of TerraZyme and 1.75% of lime. At this dosage the the LL, PL, SL and FSI reduced to 44, 40 18 and 17 respectively. PI decreased to 4 and an UCS upto 300 kPa was achieved in the 2nd week itself. The rate of strength gain was however slow after the 2nd week and reached only 330 kPa at the end of 4th week.

It can be observed from the graphs that when soil is treated with enzymatic lime, the strength increase is predominant in the first two weeks and slower afterwards. The value of UCS obtained in two weeks on treatment with enzymatic lime is higher than the strength obtained at the end of four weeks on treatment with lime and enzyme. The plasticity characteristics are marginally affected much when lime and enzyme are added together to the soil as compared to individual agent stabilizations. In short the new system speeds up the stabilization of soils and reduces the curing and construction time radically.

Decrease in clay content – GSD

From figures 3, 7 and 11, it is observed that stabilization of soils bring about an apparent decrease in the percentage of clay particles in soil. This is mainly due to the agglomeration effect of stabilizers. When the clay contents are measured by sedimentation, visual and microscopic examinations reveal that several undispersed particles remain as a result of persistent aggregation. Hence when the results give low clay content, the actual mineral contents are in fact higher. In such cases low values of liquid and plastic limit are observed in relation to large values of actual clay mineral content [8]. Usually a relation between soil properties and clay content may be correlated. But this is impossible for treated soils, due to the drastic disparity in their nature and cannot be applied for the case mentioned in this paper. The other factors affecting the same may be understood only on a detailed examination of the treated soils and a fuller understanding of the origin of the soil materials, but that is again outside the scope of the paper.

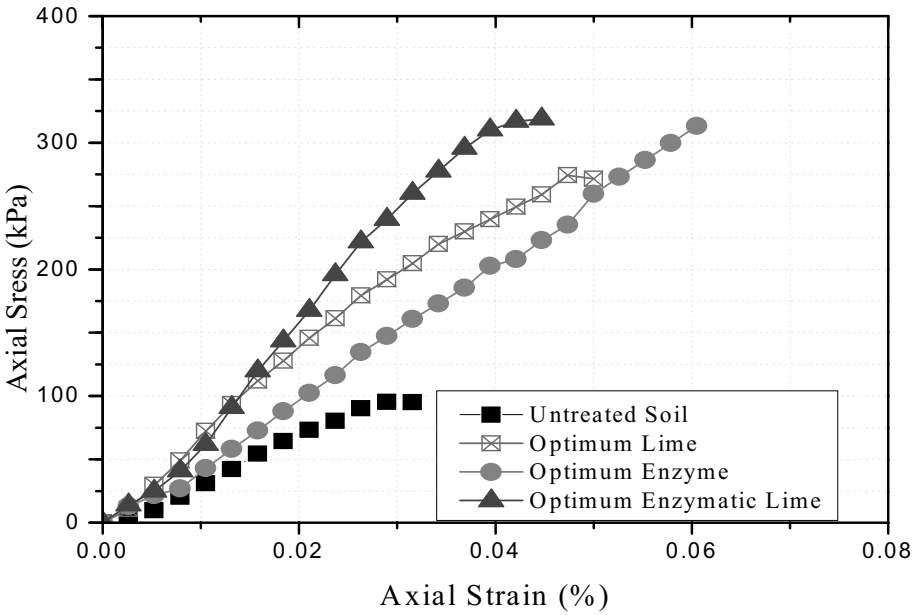


Figure 13. Stress Strain Curve of soil after 4 weeks of curing with OLC, OEC and OLT

Conclusion

Both lime and TerraZyme are independently used extensively for subgrade stabilization. The technique of introducing enzymatic lime into soil is a novice procedure which indicates to impart better and quicker stabilization from the laboratory studies conducted in the work. The result is the transformation of soil into a stronger permanent soil matrix in half the time as that required by lime or TerraZyme alone. Lower plasticity index indicates reduced swelling and shrinking, an adaptation that will alleviate common problems in roadways. By improving the quality of roads, the travel along unpaved roads maybe assured pleasant as well. The practice may be consequently integrated into stabilization of soils for subgrade.

The nature of enzymatic lime stabilized soils has not been studied as on date. Hence the behavior presented in this work may differ when applied to other soils under the same conditions. Enzymes catalyze very specific chemical reactions. As such, it is difficult to discern a general stabilization mechanism due to variations in the soil-specific reactions [10]. Montmorilliic clay on application of enzymatic lime showed decreased plasticity characteristics with improved strength. The catalystic action of TerraZyme appears to have increased in the presence of lime. A higher degree of stabilization has occurred with shorter curing period. Therefore a possibility of accomplishing soft soil stabilization has been recognized. The points to be noted include:

1. An increased aging effect on enzymatic lime stabilized soils has not been studied.
2. The improvement in strength may be attributed to the compaction effect and may vary in field conditions
3. Salts present in field conditions may affect the catalystic nature of enzymes, and cannot be predicted now
4. The proposed stabilization technique is not authenticated, but literature tends to support the same. In general clayey soils may be assumed to respond positively to this system

References

[1] M. Vedula, P. N. G, and B. P. Chandrashekar, "a Critical Review of Innovative Rural Road Construction Techniques and Their Impacts," 2002.
[2] J. Berube, M. A., Choquette, M., Locat, "Effects of lime on common soil and rock forming minerals," Appl. Clay Sci., vol. 5, pp. 145–163, 1990.
[3] P. Sherwood, "Soil stabilizationwith cement and lime. State of the art review.," Tech. Rep. Transp. Res. Lab. HMSO; London., 1993.
[4] D. E. Scholen, "Nonstandard Stabilizers," FHWA-FLP-92-011, no. July, 1992.
[5] J. B. Oza and P. J. Gundaliya, "Study of black cotton soil characteristics with cement waste dust and lime," Procedia Eng., vol. 51, no. NUiCONE 2012, pp. 110–118, 2013.
[6] F. . Bell, "Lime stabilization of clay minerals and soils," Eng. Geol., vol. 42, pp. 223–237, 1996.
[7] R. H. Mihai O. Marasteanu and R. V. Timothy R. Clyne, "Preliminary Laboratory Investigation of Enzyme," 2005.
[8] M. J. Dumbleton, G. West, and G. Dumbleton, M. J., & West, "Some factors affecting the relation between the clay minerals in soils and their plasticity," Clay Miner., vol. 6, no. 3, pp. 179–193, 1966.
[9] M. R. Taha and T. A. Khan, "Recent Experimental Studies in Soil Stabilization with Bio-Enzymes – A Review," Ejge, vol. 18, pp. 3881–3894, 2013.
[10] J. S. Tingle, S. L. Larson, C. A. Weiss, J. K. Newman, J. F. Peters, B. Tardy, and E. S. B. Iv, "4a 4 4 4," 2004.

Behaviour of Square Model Footing on Sand Reinforced with Woven Coir Geotextiles

Dharmesh Lal*, N Sankar**and S Chandrakaran***
*Research Scholar, Department of Civil Engineering, National Institute of Technology Calicut, Calicut-673601, India
E-mail: dharmeshlal34@gmail.com
**Professor, Department of Civil Engineering, National Institute of Technology Calicut
***Professor, Department of Civil Engineering, National Institute of Technology Calicut

Abstract: This paper presents the results of laboratory model tests carried out on square footings supported on woven coir geotextile reinforced sand beds. The influence of various parameters such as depth of reinforcement, length and number of layers of reinforcement were studied. On the whole, the results indicate that up to four fold increase in strength and about eighty percent reduction in settlement can be obtained by providing two or more layers of reinforcement. The optimum value for depth of reinforcement was found to be 0.25 times the width of foundation. It was also found out that bearing capacity does not proportionately increase with length and number of layers of reinforcement. There exist an optimum layout of placement of geotextile.

Keywords: Coir geotextile, Depth of reinforcement, Number of layers, Bearing capacity, Reinforcement length.

Introduction

Geosynthetic applications are widely used as soil reinforcement and is a proven alternative to conventional ground improvement technique under appropriate conditions. These synthetic materials and products generally have a long life, but are costly and may create environmental problems in the future. In countries where the availability and cost of the synthetic reinforcing materials are a major constraining factor and where natural materials are available in plenty, the potential of natural materials for use as soil-reinforcing elements is worth examining. Amongst the naturally occurring materials, coir, which is the pro-cessed husk of ripe coconuts, is reputed to be the strongest and most durable.

Research on the use of coir as a soil-reinforcing material started in the 1990s. Triaxial shear tests carried out on sand reinforced with discrete, randomly oriented coir fibers of different lengths and fiber contents indicated that coir has great potential as a soil- reinforcing material in field applications [8]. Reference [9] conducted a series of triaxial compression tests on sand reinforced with randomly oriented coir fibers and evaluated the strength and stiffness response of the reinforced soil system. This paper reports a series of laboratory test results of plate load test on sand bed reinforced by woven coir geotextiles. The objectives of the study were to investigate the effects of depth, length and number of layers of reinforcement on pressure versus settlement behaviour of sand bed.

Literature Review

The use of coir as a reinforcement material has been studied by researchers [6-12]. Review of various literatures shows that the use of woven coir geotextiles in reinforcing shallow foundations are rare.

Reference [9] Conducted studies on a strip footing on sand which was reinforced with randomly distributed coir fiber and mesh elements. Results indicated an increase in ultimate bearing capacity values. Larger mesh size elements produced better performance compared to other inclusions.

Reference [8] conducted a series of triaxial shear tests on sand reinforced with discrete, randomly oriented coir fibers and indicated the potential of coir as an efficient reinforcing material.

Reference [6] conducted a series of laboratory model tests were conducted on square footing resting on loose sand reinforced with braided coir rope. The test results indicated a substantial reduction in normalized settlement with the introduction of braided coir rope. Three layers of braided coir rope reinforcement yielded optimal performance.

Materials and Methodology

Materials used in the study

The test soil used in the study was dry sand whose particle size distribution characteristics are shown in Fig.1.The basic and index properties of sand, determined according to ASTM standards are summarized in Table 1. The tests were done at 60%

relative density. Woven coir geotextiles were used as reinforcement in the present study. The physical properties of coir geotextiles according Indian standards are given in Table 2.

Table 1. Properties of sand used in the study

Properties	Value
Specific Gravity	2.63
D_{10} (mm)	0.32
D_{30} (mm)	0.48
D_{60} (mm)	0.82
C_u	2.56
C_c	0.88
I.S Classification	SP
Maximum Dry density (g/cc)	1.69
Minimum Dry density (g/cc)	1.46
Angle of Internal Friction at 60 % relative density	38.5^0

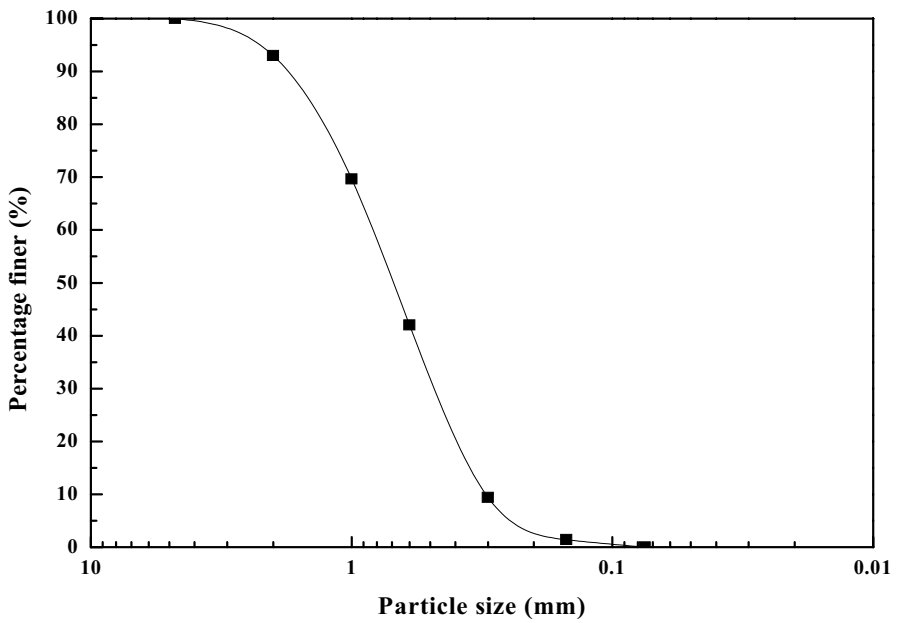


Fig.1. Particle size distribution characteristics of sand used in the study

Table 2. Properties of woven coir geotextile used in the study

Properties	Value
Mass per unit area	720 g/ m^2
Ultimate Load & Failure strain (Warp direction)	15 KN/m & 25%
Ultimate Load & Failure strain (Weft Direction)	9 KN/m & 8%
Aperture size (mmxmm)	5x12
Thickness	6mm

Methodology

For conducting model tests, sand was poured in to the test tank using a raining technique [1-3]. The height of fall to achieve the desired relative density was determined prior by conducting a series of trials with different heights of fall. The relative density achieved was confirmed by collecting samples in small cans of known volume placed at different locations in the test

tank and determining the density of sand at the time of filling. After preparing the bed, the surface was leveled, and the footing was placed exactly at the center of the loading jack to avoid eccentric loading. The footing was loaded by a hand-operated hydraulic jack supported against a reaction frame. A recess was made in the footing plate at its center to accommodate a ball bearing, through which vertical loads were applied to the footing. A precalibrated proving ring was used to measure the load transferred to the footing. The load was applied in small increments. Each load increment was maintained constant until the footing settlement was stabilized. Two displacement dial gauges were positioned on either side of the footing by means of magnetic bases attached to the steel section of the tank. Average of the two displacement dial gauges were recorded [1-3]. The arrangement of the test series is shown in Fig.2.The objectives of the test series were to study the influence of depth, length and number of layers of reinforcement on the overall performance of the foundation. Test series were conducted by varying the number of layers (N), depth of reinforcement layer from the top of footing (u), and length of reinforcement layer (L).

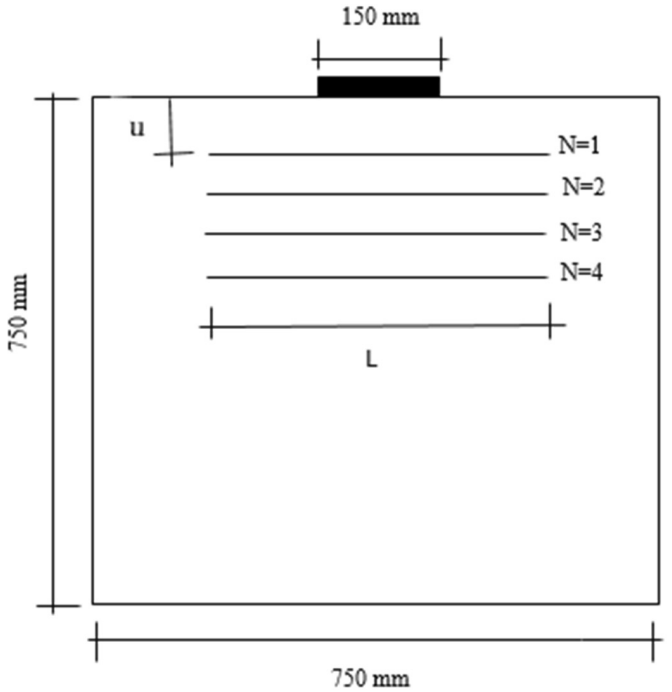


Fig. 2. Arrangement of geotextile reinforcement

Results and Discussion

Effect of depth of reinforcement

Fig.3 (a) shows the load settlement behavior of footing underlain by woven coir geotextile reinforced sand bed. For comparison result obtained from non-reinforced sand bed is also presented. The performance improvement for foundation at various depth of reinforcement is shown in figure 3(b). A non-dimensional term improvement factor is introduced to quantify the performance improvement [1-6]. Improvement factor is the ratio of footing pressure with reinforcement at a given settlement to the pressure on the unreinforced soil at the same settlement. It can be seen from Fig. 3(a) that pressure versus settlement behavior of woven coir geotextile reinforced foundation is better than that of unreinforced case. This is mainly due to the interface frictional resistance between sand and coir geotextile. Improvement factor versus u/B shows that bearing capacity of the footing increases up to a depth of 0.25 times the width of footing, there after it decreases because the reinforcement will be located out of the most effective zone. Here the improvement factor is taken for a settlement of 25 mm, which is considered threshold settlement for foundations. Maximum beneficial effect was obtained in this study by placing the woven coir geotextile reinforcement at a depth of 0.25 times the width of footing. It is also seen that placing the geotextile beyond 0.5 times the width of foundation, soil behavior is similar to unreinforced case.

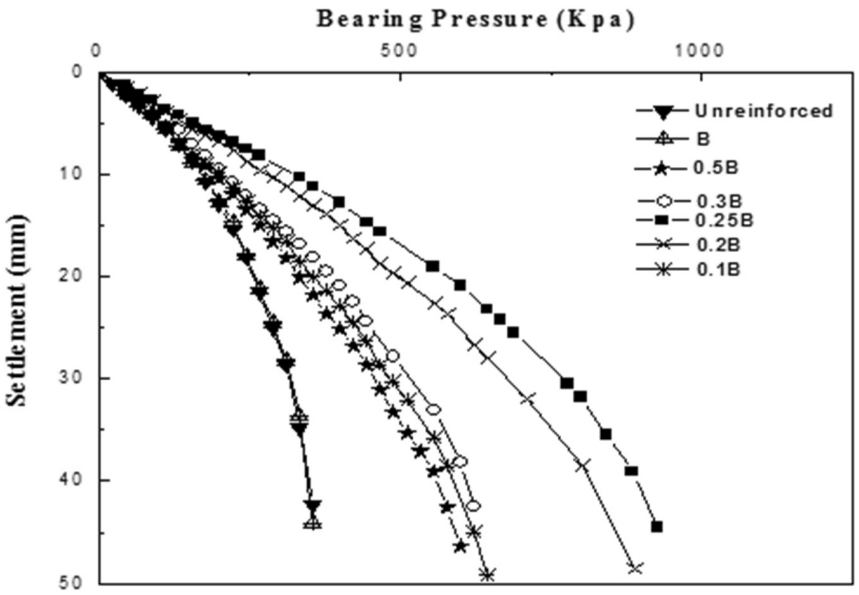


Fig.3 (a). Variation of applied bearing pressure with footing settlement for various depth of reinforcement

Effect of number of layers of reinforcement

Fig.4 (a) shows the applied pressure versus settlement behavior for various number of reinforcement layers (N). It can be seen that with increase in number of layers of coir reinforcement, the bearing capacity increases appreciably. However with an increase in number of layers above three, there is not much bearing capacity improvement. A settlement reduction factor is introduced by which improvement by reduction of settlement can be quantified. Settlement reduction factor = $(S_0\text{-}Sr)/S_0$, where S_0 is the settlement of non-reinforced sand bed at a given pressure and Sr is the settlement of sand reinforced with coir geotextile at the same pressure [6]. Fig. 4(c) shows the variation of Settlement reduction factor with improvement factor for different layers of reinforcement. Here in this case settlement reduction factor and improvement factor for different number of layers of reinforcement is taken for a normalized settlement (Settlement / width of foundation) of 30% of the non-reinforced sand bed [6]. It is seen that about 80% reduction in settlement and about four fold increase in bearing capacity can be obtained by providing two or more layers of reinforcement. It is also seen that beyond 3 layers there is not much improvement which may be due to the reinforcement layer positioned out of the effective zone.

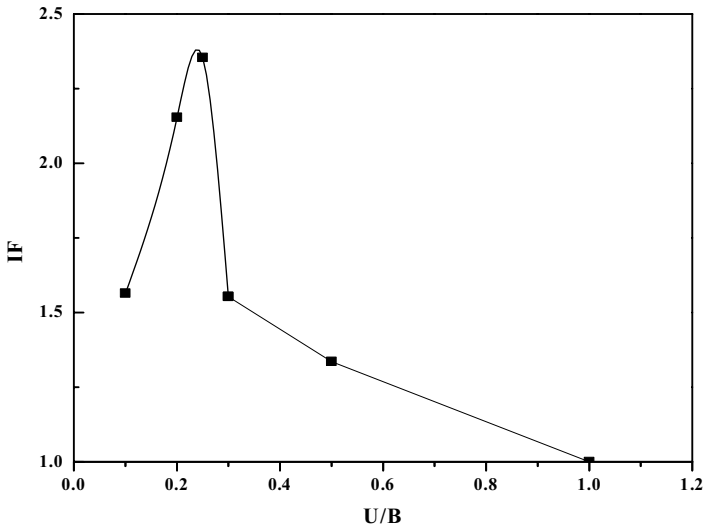


Fig.3 (b). Variation of Improvement factor with u/B

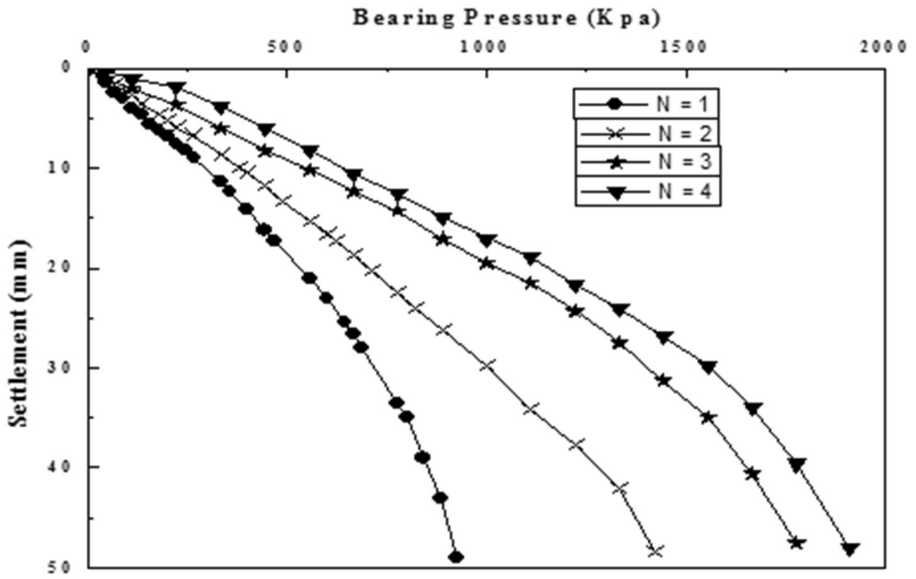


Fig.4 (a). Variation of applied bearing pressure with footing settlement for different layers of reinforcement

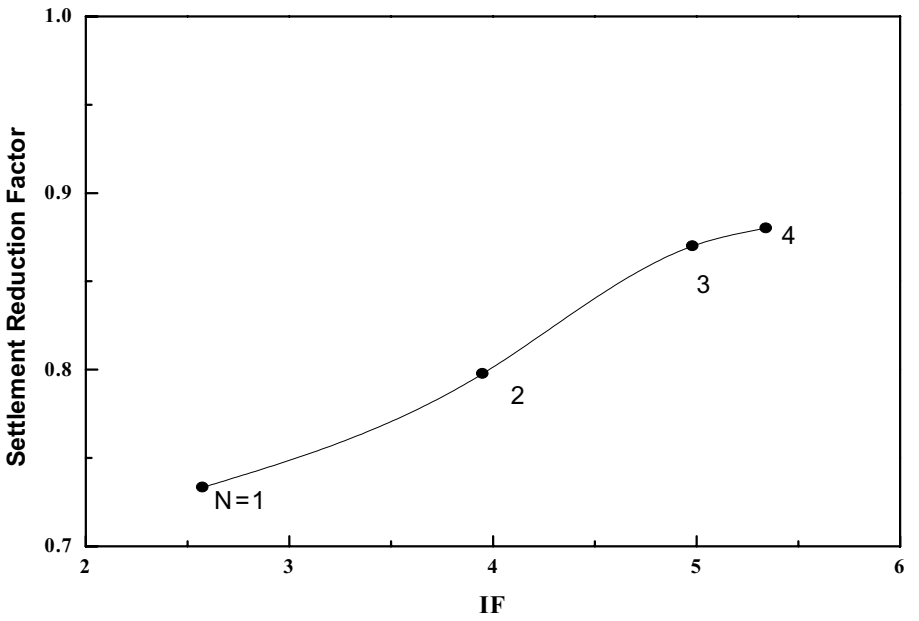


Fig.5 (a). Variation of settlement reduction factor with IF for different layers of reinforcement

Effect of length of reinforcement

Fig.5 (a) depicts the variation of applied pressure versus the settlement for different L/B ratios of reinforcement kept at optimum depth of 0.25B. Fig.5 (b) shows the variation of Improvement factor with normalized settlement values at different L/B ratios. From these figures it is clearly seen that improvement is not appreciable with increase in length. However providing a greater length will be beneficial regarding shear failure point of view.

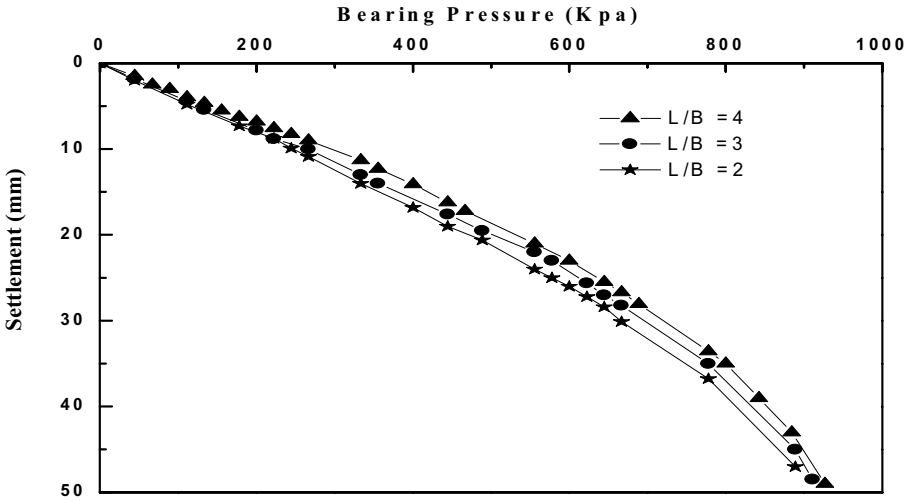


Fig.5 (a).Variation of applied pressure with footing settlement for different L/B ratio

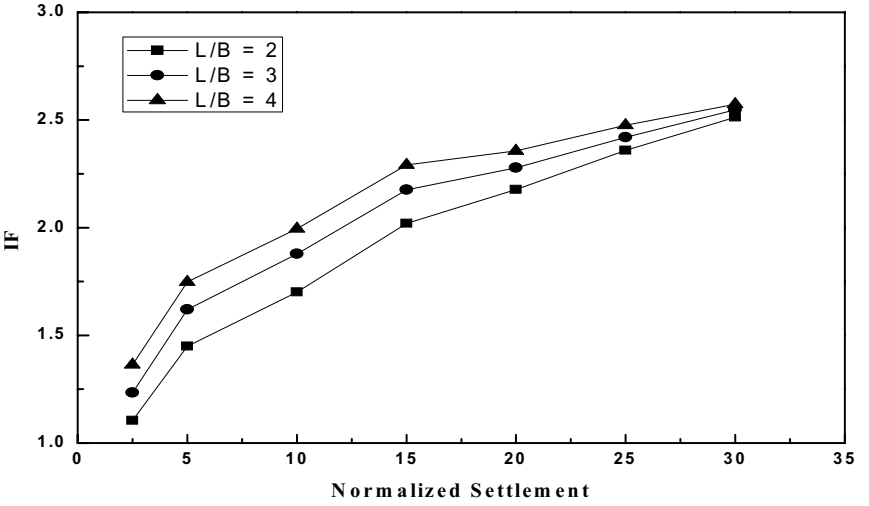


Fig.5 (b). Variation of Improvement factor with Normalized settlement for various L/B ratios

Conclusions

The use of natural fibres for erosion control and improvement of surficial stability of embankments is practiced in the state of Kerala in India, and this paper attempts to explain the degree of improvement obtained in terms of strength and stiffness by using this material. The Various parameters that were studied are length, depth and number of layers of reinforcement. Based on the results obtained the following conclusions can be obtained.

1) Provision of coir geotextile reinforcement significantly increases the load carrying capacity and reduces the footing settlement.
2) Optimum depth of the topmost layer of reinforcement is 0.25 times the width of foundation.
3) About 80% reduction in footing settlement and four fold increase in bearing capacity can be obtained by providing two or more layers of reinforcement.
4) The performance of foundation with increase in length is not appreciable, however a larger length will be beneficial regarding shear failure point of view.
5) However extensive research on large scale field tests is essential before actually applying them in the field.

Acknowledgement

The authors would like to thank National Institute of Technology Calicut for providing the necessary facilities for this experimental research. Thanks are also due to valuable comments from Greeshma Lamento, Femy Yamani and Mohanlal.

References

[1] S. N. M. Tafreshi and A. R. Dawson, "Comparison of bearing capacity of a strip footing on sand with geocell and with planar forms of geotextile reinforcement," Geotext. Geomembranes, vol. 28, no. 1, pp. 72–84, 2010.
[2] G. M. Latha and A. Somwanshi, "Bearing capacity of square footings on geosynthetic reinforced sand," Geotext. Geomembranes, vol. 27, no. 4, pp. 281–294, 2009.
[3] G. Madhavi Latha and A. Somwanshi, "Effect of reinforcement form on the bearing capacity of square footings on sand," Geotext. Geomembranes, vol. 27, no. 6, pp. 409–422, 2009.
[4] N. R. Krishnaswamy, K. Rajagopal, and S. K. Dash, "Performance of different geosynthetic reinforcement materials in sand foundations," Geosynth. Int., vol. 11, no. 1, pp. 35–42, 2004.
[5] M. Abu-Farsakh, Q. Chen, and R. Sharma, "An experimental evaluation of the behavior of footings on geosynthetic-reinforced sand," Soils Found., vol. 53, no. 2, pp. 335–348, 2013.
[6] P. Vinod, A. B. Bhaskar, and S. Sreehari, "Behaviour of a square model footing on loose sand reinforced with braided coir rope," Geotext. Geomembranes, vol. 27, no. 6, pp. 464–474, 2009.
[7] Y. Wasti and M. D. Bütün, "Behaviour of model footings on sand reinforced with discrete inclusions," Geotext. Geomembranes, vol. 14, no. 10, pp. 575–584, 1996.
[8] G. Rao and K. Balan (2000). "Coir geotextiles—Emerging trends", Kerala State Coir Corporation Limited, Alappuzha, Kerala
[9] G. L. Sivakumar Babu and a. K. Vasudevan, "Strength and Stiffness Response of Coir Fiber-Reinforced Tropical Soil," J. Mater. Civ. Eng., vol. 20, no. 9, pp. 571–577, 2008.
[10] K. R. Lekha, "Field instrumentation and monitoring of soil erosion in coir geotextile stabilised slopes - A case study," Geotext. Geomembranes, vol. 22, no. 5, pp. 399–413, 2004.
[11] K. R. Lekha and V. Kavitha, "Coir geotextile reinforced clay dykes for drainage of low-lying areas," Geotext. Geomembranes, vol. 24, no. 1, pp. 38–51, 2006.
[12] E. A. Subaida, S. Chandrakaran, and N. Sankar, "Laboratory performance of unpaved roads reinforced with woven coir geotextiles," Geotext. Geomembranes, vol. 27, no. 3, pp. 204–210, 2009.
[13] M. Harikumar, N. Sankar, and S. Chandrakaran, "Response of Sand Reinforced with Multi-Oriented Plastic Hexa-Pods," Soil Mech. Found. Eng., vol. 52, no. 4, pp. 211–217, 2015.
[14] Y.W. Yoon, S.H. Cheon, and D.S.Kang, "Bearing capacity and settlement of tire- reinforced sands. " Geotext. Geomembranes, vol 22, no. 5, pp. 439–453, 2004
[15] A. Hegde and T. G. Sitharam, "Experimental and Analytical Studies on Soft Clay Beds Reinforced with Bamboo Cells and Geocells," Int. J. Geosynth. Gr. Eng., vol. 1, no. 2, p. 13, 2015.
[16] S. M. Haeri, R. Noorzad, and A. M. Oskoorouchi, "Effect of geotextile reinforcement on the mechanical behavior of sand," Geotext. Geomembranes, vol. 18, no. 6, pp. 385–402, 2000.
[17] A. Hamidi and M. Hooresfand, "Effect of fiber reinforcement on triaxial shear behavior of cement treated sand," Geotext. Geomembranes, vol. 36, pp. 1–9, 2013.
[18] E. A. Subaida, S. Chandrakaran, and N. Sankar, "Experimental investigations on tensile and pullout behaviour of woven coir geotextiles," Geotext. Geomembranes, vol. 26, no. 5, pp. 384–392, 2008
[19] A. Marto, M. Oghabi, and A. Eisazadeh, "The Effect of Geogrid Reinforcement on Bearing Capacity Properties of Soil Under Static Load; A Review," Electron. J. Geotech. Eng., vol. 18 J, pp. 1881–1898, 2013.
[20] G. Rao and K. Balan (2000). "Coir geotextiles—Emerging trends", Kerala State Coir Corporation Limited, Alappuzha, Kerala

Upgradation of a Building to Higher Certification Levels as per LEEDv4 - Case Study

N. Amrutha Sudhakar* and S. Shrihari**
*PG Student, NIT Karnataka
sudhakaramrutha@gmail.com
**Professor of Civil Engineering, NIT Karnataka
s.srihari@gmail.com

Abstract: There is a general perception among the developers and clients that green buildings shall incur additional costs. For the same reason, they always prefer conventional buildings over green buildings. If at all they are going for green buildings, they always tend to go for lower rating levels. Most of the clients and contractors are either ignorant or unaware of the possibilities of a higher certification level without much increase in the total cost. Suppose they are educated about the possibilities of going for a higher certification level, cost of doing so and also a budgeting methodology for controlling the cost, it will definitely be an incentive for clients, owners, designers, and users to develop and promote highly sustainable construction practices. In this paper, a building that has attained a particular certification level in LEED India v1 is evaluated using LEED v4 to estimate the rating that it could achieve under LEED v4 and to determine the possibilities of upgrading the building to higher certification levels within v4. Credit wise evaluation has been done to determine the elements required for up gradation and also the cost of up gradation (in the Indian context) and on the basis of this evaluation as well as some literature review, a budgeting methodology is developed.

Keywords: Green Buildings, LEED India v1, LEED v4, Preliminary Analysis, Budgeting Methodology.

Introduction

The idea of green building can be considered a major reform in the construction industry. The attitude of the public as well as their views towards green buildings may affect its application. An important reason why people tend to go for the conventional buildings or green buildings of lower ratings is that they perceive a huge percentage increase in the total construction cost. Hence in this paper, all the prerequisites and credits of LEED v4 are evaluated in the context of a selected building in India to determine the possibilities of a higher certification level. The building has achieved silver rating under LEED India v1. Therefore this evaluation also helps in understanding the difference in credit requirements that have occurred in LEED over the years. Benefits of certification through the latest version (v4) is that it can be a valuable educational and marketing tool for owners and design and construction teams, through the process of creating a more sustainable building. We are not just focusing on "What extra is it going to cost?" but on "How can we do it?" as well. I.e. in addition to the cost, we are also considering the factors for upgradation of a building. After the evaluation, a general budgeting methodology is developed for cost control.

Literature Review

Social acceptance of green buildings definitely affects the clients or owners willingness to build environmentally responsible structures. Ref. [1] has demonstrated the level of acceptance of green buildings and social problems related to green buildings. From their survey, they have concluded that, when educated about green buildings, 90% of the sample would pay more for a green building over the standard building. Hence analysing the cost of green features is an important research theme.

Cost of green features varies depending on the rating system chosen. Ref. [2] did a comparative review of the five rating systems namely BREEAM, LEED, CASBEE, GREEN STAR and HK-BEAM. Categories considered for review were popularity and influence, availability, methodology, applicability, data collecting process, accuracy and verification, user friendliness, development, results presentation. As per the study, LEED and BREAM achieved the highest scores. Ref. [3] analysed BOMA Go Green, BREEAM (UK), Green Star, Passive House, The Living Building Challenge and LEED. This study also shows that LEED has a dominance above its counter parts. Ref [5] has evaluated the operational expenditures of certified buildings. The installation and disposal costs have not been considered. The results of the comparison of the selected certified buildings also shows that LEED certified buildings achieve energy and water savings because of the LEED mandatory requirements. Hence LEED may be considered as one of the best choices for pursuing green building certification.

Many developers have a common perception that green buildings have high development costs. I.e. they perceive that the construction would incur additional costs. Ref. [5] divides the cost of construction into land, hard and soft costs and has analysed the soft costs associated with development of a green building. Though only soft costs were studied, it was identified that the experts had a unanimous opinion that international green rating tools such as LEED are easier to achieve at less capital cost compared to the Malaysian GBI rating tool. This is an incentive to determine the cost of green features in green buildings.

Ref. [6] in a report submitted to California's Sustainable Building Task Force, The cost and financial benefits of green buildings, demonstrates conclusively that sustainable building is a cost-effective investment, and its findings should encourage communities across the country to "build green." This report represents one of the most definitive cost benefit analysis of green building ever conducted. The report also evaluates the various problems in determining cost.

Ref. [7] recognizes the need for building an appropriate cost model for the buildings. The report also says that as a part of staying on track, it is necessary to update and monitoring of the LEED check list.

Even though several studies have been conducted in determining the soft costs and hard costs associated with green buildings, the cost of upgradation of a building to higher certification levels by incorporating additional green or sustainable features at the beginning of the project itself is a less researched topic.

What is a green building?

Green building is a concept that agrees upon the fact that buildings have profound effects on the surrounding environment as well the occupants. Green buildings plays a major role in intensifying the positive effects and eliminating the negative effects.

LEED

Leadership in Energy and Environmental Design (LEED) developed by the U.S. Green Building Council (USGBC), is a set of rating systems for the design, construction, operation, and maintenance of green homes, buildings and neighbourhoods. LEED is organized into a set of environmental categories and these categories are subdivided into prerequisites and credits. Prerequisites are mandatory and must be achieved and therefore should be immediately addressed by the team. Credits are based on desired performance goals within each category.

LEED India v1

In LEED India v1, six major environmental categories are considered for sustainable development. Namely sustainable sites, water efficiency, Energy and atmosphere, Materials and resources, Indoor environmental quality & innovation. The environmental categories are subdivided into the established LEED - INDIA credits, which are based on desired performance goals within each category. The scores are certified: 26-32, silver rating: 33-38, gold rating: 39-51, platinum rating: 52+.

LEED v4

LEED v4 is the most recent version of the world's premier benchmark for high-performance green buildings. USGBC claims this version to be more specialized, bolder and designed for an improved user experience. Unlike LEED India v1, eight major categories are considered for sustainable development, namely Location and Transportation, Sustainable Sites, Water Efficiency, Energy and Atmosphere, Materials and Resources, Indoor Environmental Quality & Innovation. A comparison of LEED India v1 and v4 shows that the points are more difficult to achieve in the case of LEED v4 and also LEED v4 gives more importance to high performance cost effective outcomes, reducing environmental impacts and improving liveability and human health. The scores in LEED v4 are certified: 40-49, silver rating: 50-59, gold rating: 60-79, platinum: 80-110. It can be seen that the scores have been raised for each rating level, making it even more difficult to achieve.

Preliminary Analysis and Cost Based Assessment

An office building having 3B+G+8 floors with two towers, location: Chennai, total project value of 225Cr and having a silver rating under LEED India v1 was chosen for the case study. A feasibility checklist was prepared as per LEED v4 and the building was evaluated on the basis of this checklist. The preliminary analysis showed that, the building which achieved a silver rating under LEED India v1, is eligible for basic certification level only (47 points) when evaluated under v4. The summary of the preliminary analysis is shown in table 1.

From the table, it can be seen that the project is readily eligible for certification level only as per LEED v4. Also it can be seen that, with some additional effort, a maximum of 33 points can be earned extra. Thus with some minor/major additions to the project, it could have aimed for platinum rating.

The cost based assessment of the credits that are readily achievable as per preliminary analysis, is given below.

Location and Transportation

Sensitive Land Protection: There are typically no construction or soft costs associated with this credit. It is an outcome of site selection. However to get the CDMA permit (construction permit), which approves that the development footprint is not on a land that meets the criteria for sensitive land, CDMA scrutiny charge of INR 1.5 per square meter of the land area need to be considered.

Surrounding Density and Diverse Uses: This credit is usually a result of site selection rather than a driver. The building being in an urban area with high surrounding density, no construction or soft costs incurred.

Table 1. Summary of the points earned, pending and not eligible for

Major Credit Category	Points achieved under preliminary analysis	Points that could be achieved by minor or major additions to the project	Points that the project is not eligible for
Integrative process (This credit doesn't fall under any major credit category).	0	1	0
Location and Transportation	12	1	3
Sustainable Sites	8	2	0
Water Efficiency	8	1	2
Energy and Atmosphere	6	11	16
Materials and Resources	2	6	5
Indoor Environmental Quality	5	7	4
Innovation	3	3	0
Regional Priority	3	1	0
Total	47	33	30

Access to Quality Transit: This credit is also usually a result, rather than a driver, of site selection. Project is located on a site where public transportation is already available nearby. Hence no extra costs involved.

Green Vehicles: In order to achieve this credit, 5% of the parking space is designated as preferred parking space for green vehicles. Also electric vehicle supply equipment are installed to meet the requirement (26000 INR per piece).

Sustainable Sites

Preq: Construction activity pollution prevention: As a part of construction activity pollution prevention, an erosion and sedimentation control plan was formulated and implemented. Top 20 cm thick soil was stripped and stacked at site (120 INR per cum for excavation and stacking). Also while construction, temporary seeding was done to prevent erosion or sliding.

Site Assessment: As a typical approach to this credit, a site survey was performed. The soft costs associated with this credit are not very high.

Site Development - Protect or Restore Habitat: The site is on previously developed area and not on a green field area. Hence 30% of all portions of site identified as previously disturbed were restored using native vegetation costing Rs 600 per sqft.

Rainwater Management: There are a number of options for reducing the runoff at site and cost depends upon your choice. Here in this case, a rainwater storage sump of 100 cum capacity (18 lakh INR) has been provided. Additionally, concrete interlock pavement with permeable joints (have an underlying of open graded bedding course, open graded base course and open graded subbase on compacted soil subgrade, 80 INR per sqft) is provided Then we have vegetated open space and the cost for the same has already been considered. We also have roof garden (950 INR per sqft). Also pervious pavers have been provided for roof garden.

Heat Island Reduction: Both the options (non-roof and roof, parking under cover) have been satisfied. First option is roof and non-roof measures and on achieving this credit you get 2 points

Area of non-roof measures = Area of pervious pavers or open grid pavement system + Area of trees providing shade
= 43211.03 sqft+ 2081.55 sqft.

Cost of both have already been considered for achieving other credits like rain water management, protection and restoration of habitat etc.

Area of high reflectance roof = 23583 (tower A) + 23736 (tower B)
= 47319 sqft.

Cost of providing albedo coating is 45 per sqft. Area of vegetated roof = 20696.59 sqft and the cost has already been considered under the credit rain water management. Hence the total cost coming is only 21, 29, 350 INR.

Second option is parking under cover. If you are going for this option, you get only 1 point. Due to lack of space, instead of providing it at ground level and providing cover for the same, basement parking has been provided. This actually satisfies the credit and provides 100% shade compared to the ground level parking system. This would cost around 2500 INR per sqft. If

we go for the normal ground level parking system, it would cost around 1500 INR per sqft. Therefore the difference per sqft for greening comes to be Rs 1000 per sqft. Total parking area = 322099.26 sqft. Therefore additional cost = 322099.26 x 1000 = Rs 32, 20, 99, 260. Considering the huge cost difference and also the points achieved, first option has been attempted.

Water Efficiency

Prereq: Outdoor Water Use Reduction: With the intention of reducing the water usage, STP 400 cum per day capacity was constructed. This onsite treatment of water would add considerably to the total cost (1.2 Cr INR). Also for reducing the outdoor water usage, native species have been used in outdoor vegetation. The cost of the same has already been considered.
Prereq: Indoor Water Use Reduction: Typical approach to go for this credit is to go for low flow plumbing fixtures having water consumption less than 20% from the baseline. Through careful choice of low flow plumbing fixtures and elimination of cooling towers the project has achieved 100% indoor water use reduction.
Prereq: Building-Level Water Metering: As a typical approach to achieve this prerequisite, permanent water meters (1.2 lakh INR per meter inclusive of installation charge) have been installed for measuring input water to the STP, output from STP to HVAC system, flushing system and irrigation.
Outdoor Water Use Reduction: In the process of satisfying the prerequisite, 50% reduction was already achieved. Hence we needn't consider an extra premium for this credit.
*Indoor Water Use Reduction***:** In the process of satisfying the prerequisite, 100% reduction was achieved. Therefore here also we needn't consider an extra premium.

Energy and Atmosphere

Prereq: Fundamental Commissioning and Verification: Third party commissioning agent has been assigned to carry out the commissioning process.
Prereq: Minimum Energy Performance: Since energy efficient design, right sizing of the equipment and improvements in basic building systems etc. are an integral part of this project, similar to majority of the multi crore projects, this credit requirement of 5% improvement in energy efficiency, is easily achieved within the base cost. Hence there is no added cost. However we need to consider the soft costs for demonstrating this improvement in energy performance using whole building energy simulation system.
Prereq: Building-Level Energy Metering: As a typical approach to achieve this prerequisite building level energy meters and sub meters have been installed. Individual meters are quite inexpensive. As per the prerequisite, we require 1 HT energy meter, then we require sub meters for measuring the supply to the plug points, lighting & HVAC systems. Each tower requires these separately. Therefore we need a total of 6 sub meters
Prereq: Fundamental Refrigerant Management: The cooling load works out to be around 2216 TR in total. Therefore a system consisting of 6 numbers of 370 TR (working) and 1 number of 370TR (stand by) capacity air cooled screw chillers have been provided. Refrigerant used is R134A which is free of CFC, HCFC and halogens. A chiller with refrigerant such as R134A would cost almost Rs 20 lakh more than the less environmentally responsible ones.
Enhanced Commissioning: Cost already considered for meeting the prerequisite.
Enhanced Refrigerant Management: Cost already considered for meeting the prerequisite.

Materials and Resources

Prereq: Storage and Collection of Recyclables: The credit has no soft cost impact. For the purpose of collection of recyclables, a bin center 150sqm was constructed (Rs 750 per sqft). In addition to this, individual collection bins are provided to all occupants
Prereq: Construction and Demolition Waste Management Planning: As a part of construction and demolition waste management planning, some of the materials have been reused within the site. For e.g., concrete debris and concrete masonry blocks are reused within the site for making temporary roads, infrastructure fill etc. Steel scrap is reused within the site for precast item, grade slab, drain, lintels etc. Whereas materials such as card board, wood etc. are sold to the external agencies for recycling. Hence no additional cost is incurred.
Construction and Demolition Waste Management: As explained in the prerequisite, no added cost.

Indoor Environmental Quality

Prereq: Minimum Indoor Air Quality Performance: The technologies and standards required for this prerequisite are standard to this project also like most projects and in majority of the cases, this prerequisite has low cost impact. The prescribed standards for ventilation are usually met without any added cost. However in order to satisfy the prerequisite, for the purpose of monitoring, outdoor airflow measurement device for mechanically ventilated spaces and also CO2 sensors (15000 INR) are provided. For direct outdoor airflow measurement, we can use airflow measuring stations (50000 INR per piece) are used.

Prereq: Environmental Tobacco Smoke Control: Since the company has opted for a no smoking policy, the costs associated with providing designated areas, with adequate ventilation, for smoking are eliminated. However while adopting the non-smoking policy, the prerequisite requires that signage be posted within 3 meters of building entrances.
Enhanced Indoor Air Quality Strategies: As per the credit requirement, entryway systems have been provided at all regularly used exterior entrances. As entry way system, we are using air curtains, 3 m long and 2.5 meters high (25000 INR).
Construction Indoor Air Quality Management Plan: Air handling equipments used already meets the standards, hence no added cost. Also tobacco products are banned at site and hence no expense in this regard also. However some amount has been set aside for protecting the absorptive materials stored on site, Miscellaneous works (like covering up of ducts, harmful materials etc.), housekeeping works etc.
Indoor Air Quality Assessment: As building flush out was done, the credit requirement is satisfied. It costs in the range of Rs 15 to Rs 25 per sqft
Thermal Comfort: This project also like most of the projects are already designed to comply with the ASHRAE standards. Hence no added cost.
Acoustic performance: The maximum background levels from HVAC systems are as per 2011 ASHRAE Handbook, HVAC Applications. Hence this credit is achieved with no cost impact.

Innovation
Innovation: Since we have achieved exemplary performance in other credit categories like indoor water use reduction of 100%, heat island reduction, there is no added cost and we are getting two points also
LEED Accredited Professional: The cost component coming under this credit is the professional fee.

Regional Priority
Regional Priority: The three points achieved in this category are for the achievement of credits that address geographically specific environmental, social equity and public health priorities. There is no cost implication for this credit as such as you would have already considered it for those specific credits.

Summary of the additional cost incurred (green cost) per major environmental credit category
Based on the preliminary analysis, the cost increment (green cost) per satisfied prerequisite and earned credit was calculated and the average cost increment (from the conventional practice/ process/ construction) per major credit category is tabulated in table 2
It can be seen from the table that the green cost for achieving certification in LEED v4 for this particular building is 9.5 Cr, corresponding to just 4.22% of the total cost of the project.

Assessment of the pending credits to determine the opportunities and cost of upgradation to higher certification levels
*Integrative Process***:** This credit is basically for the identification and use of opportunities to achieve synergies across disciplines and building systems. It doesn't involve any construction cost. However soft costs associated with modelling are to be met. The credit requires that a preliminary simple box energy modelling be performed before the completion of schematic design. This credit also requires that a preliminary water budget analysis be performed and documented. Therefore by spending on simple box energy modelling and water budget analysis it is possible to achieve this credit.

Table 2. Green cost per major credit category

Major Credit Category	Cost of green in lakhs
Location and Transportation	2.65
Sustainable Sites	432.5
Water Efficiency	202.1
Energy and Atmosphere	148.3
Materials and Resources	158.9
Indoor Environmental Quality	15
Innovation	6
Regional Priority	0
Total	950

Location and Transportation
LEED for Neighbourhood Development Location: Projects attempting this credit are not eligible to earn points under other location and transportation credits. Since the project has already attempted other credits under this major credit category, we are not eligible to this.

High Priority Site: The project is not in a historic district or priority designation and it is not a brown field remediation also. Hence not eligible for this credit.
Bicycle Facilities: The credit has a very low price impact and hence we can readily go for this credit. This credit requires that bicycle storage is within 200 yard (180 m) from a bicycle network. Retrofitting a common area for bike storage can be pricey, costing anywhere from Rs 10000 to Rs 15000 per bike. Common areas with underused space, such as under stairwells or in parking garages, are good options for allocating two-wheeler parking. We can provide floor to ceiling bike rack which is less space consuming. Also shower rooms need to be provided. Providing so in the required numbers, this credit is easily achievable.
Reduced parking footprint: As per local zoning requirement, 1 car parking space must be provided per 100 sqm.
FSI Area = 42263.76 sqm. Therefore 423 parking spaces have been provided. The credit requires that 40% reduction below this baseline be achieved. That parking spaces must be limited to 254 nos. In two towered office with a total occupancy at around 5115, this is quite impossible to achieve.

Sustainable Sites
Open space: The credit requires that outdoor space greater than or equal to 30% of the total site area including the building footprint be provided. Also 25% of this outdoor space need to be vegetated.
Open space = 65821.42 sqft = 38% of the total site area.
Vegetated space = 22788.6 sqft = 34.6 % of the total open space.
Hence the above two conditions are already satisfied, Now we just need to incorporate some physical site elements that accommodate social activity as specified in the credit. Therefore this credit is achievable with less cost implications.
Light Pollution Reduction: This credit is basically to improve the night time visibility and reduce the consequences of development for wild life and people. The credit requires that the up light and trespass requirements be met as per the BUG rating method or the calculation method. Proper designing of the exterior lighting fixtures by taking into account the site security factor as well can help achieve this credit without much cost increment. Before lighting design, initially we must classify the project under one lighting zone. Based on the definition provided in MLO User guide, the project falls under the zone LZ3. Accordingly a set of fixtures may be chosen to meet the credit.

Water Efficiency
Outdoor Water Use Reduction: Additional reduction in outdoor water use, 100% reduction, is possible by an irrigation audit and irrigation scheduling on the basis of the findings from audit. Soil, plant or atmosphere based measurements are used by agencies in preparing irrigation schedule. Then for large lawn areas, we can opt for sprinkler irrigation system. The sprinkler output may be tested to adjust the time. Rain or soil moisture sensors may be used. Once it detects a designated amount of water, it shuts down the supply. Then some inexpensive methods may be used like locating sprinklers so that they are between 4 and 6 inches (10-15cm) from the edge of sidewalks, curbs, patios, etc. in lawn areas. In shrub areas they can often be 12 inches (30cm) from the edge, especially with a mature landscape. This will reduce the amount of spray onto the paved surface and will not create a dry area along the edge of the lawn. It will also reduce the amount of damage that trimmers cause to the sprinkler heads. Sprinkler system should not be installed on a sloping area as this reduces the efficiency and they spit and spew when the valve is turned on. Then for the shrub area, drip irrigation may be used as its 20% more efficient than sprinkler irrigation system and very little water is lost from evaporation or runoff. The total cost of achieving this credit is calculated by considering that all the above mentioned methods shall be adopted.
Cooling Tower Water Use: Air coolers are used in the campus for achieving exemplary performance in indoor water use reduction (100% reduction). Hence this credit cannot be attempted.

Energy and Atmosphere
Optimize Energy Performance: This credit involves substantial construction costs. There are two options in achieving points under this credit. Under the first option, you can demonstrate a percentage improvement in the energy performance or you can take the data from analysis of similar buildings in published data such as Advanced Energy Design guides. Another option is to go for prescriptive compliance for building envelope, interior lighting, exterior lighting, plug loads, it is possible to achieve a maximum of 5 points in this method. The total points that can be achieved in the event of compliance with the first option (i.e. based on published data) is less than that from prescriptive compliance. Hence the second option has been considered in the cost calculations.
Advanced Energy Metering: The cost of providing individual meters is not very significant, however, the cost of providing a good reporting system could be substantial. Still, the credit is feasible.
Demand response: The typical approach to achieve this credit is to enter into a contract with a demand response provider. There are no cost implications, instead the building owner gets paid for the energy curtailment he makes. There is compensation even for participation in demand response program. You are also eligible for a substantial one time incentive to

pay for control system upgrades. Hence this credit considered as achievable under the assumption that a demand response provider is available.
Renewable Energy Production: The project can achieve a maximum of 3 points under this credit by using renewable energy systems to offset building energy costs. Since we are attempting the credit Green Power and Carbon offsets, showing 50% of the buildings energy as from renewable source, we are eligible for all the 3 points.
Green Power and Carbon Offsets: The credit could be achieved by entering into a green power contract is usually in the range of $.02 or Rs 2 per kWh.

Materials and Resources
Building Life-Cycle Impact Reduction: This credit is difficult to achieve in this project as it does not come under historic building reuse or renovation of abandoned or blighted building, building. Also there is very less opportunity for reuse of reused, refurbished or salvaged materials in the project.
Building Product Disclosure and Optimization – Environmental Product Declarations, Sourcing of Raw Materials, Material ingredients
With more and more manufacturers supplying materials complying with this credit, experts in procurement say these credits could be achieved with an average cost increment of .1% over the total cost of materials.

Indoor Environmental Quality
Enhanced Indoor Air Quality Strategies: One extra point could be achieved for additional IAQ strategy such as sensors for additional source control and monitoring (for specific pollutants).
Low-Emitting Materials: If proper planning and study are done before finalizing the materials to be procured, low emitting materials could be bought with an average additional cost of 15% over the cost of conventional materials. However this credit becomes hard to achieve for some projects when suitable materials or products are less readily available.
Thermal Comfort: Since the project complies with the ASHRAE standards, there is no added construction cost. However we need to properly design the HVAC systems and the soft costs for the same may be considered.
Interior lighting: There are two options to satisfy this credit, lighting control and lighting quality. Lighting control requires that for 90% of individual occupant spaces, individual lighting controls be provided and for multi occupant spaces, multi zone control systems be provided. The cost can be moderate to substantial for this option. This option is difficult to pursue in this case as the occupancy number is very high, 5115 persons. Option 2, lighting quality requires that the lighting fixtures be as per the specifications in LEED. To satisfy this, you can expect an overall increase of 30 to 40% in cost over the total cost of lighting fixtures.
Daylight: As the annual sunlight exposure1000, 250 ASE1000, 250 exceeds 10% threshold mentioned in LEED, this credit cannot be attempted. Otherwise we should consider the soft costs for performing spatial daylight autonomy simulations.
Quality Views: We already have high performance vison glazing and in order to achieve this credit, we just have to go for thoughtful arrangement of the interior space.

Innovation
Innovation: Through easy methods like green education, green cleaning etc., this credit could be achieved with low or moderate cost impact.

Regional Priority
Regional Priority: By crossing the threshold for renewable energy production, this credit could be achieved.

Summary of the additional cost incurred (green cost) per major environmental credit category
Based on the above evaluation, the green cost required per pending credit to achieve a maximum possible score of 80 (corresponding to Platinum rating), for this project, was calculated and the cost per major environmental credit category is summarized in table 3.
It can be seen that by spending an extra premium of just Rs 8.7 Cr corresponding to 3.9% of the total cost of the project (225 Cr), in the initial stages of the project itself, it is possible to achieve platinum rating for this project.

Guidelines for cost control
A number of studies have proved beyond doubt that the green buildings shall payback during its lifetime itself. There are a number of factors that influence the total cost of green features. If left unchecked, the total cost could shoot up affecting the project completion or project delivery or could even result in huge losses. The following standard set of procedures could help us go for higher certification levels without much increase in the total cost.

Table 3. Green cost per major credit category for the pending credits

Major Credit Category	Cost of green in lakhs
Integrative process (Doesn't fall under any major credit category)	8.75
Location and Transportation	30.2
Sustainable Sites	7.92
Water Efficiency	31.59
Energy and Atmosphere	577.1
Materials and Resources	47.78
Indoor Environmental Quality	17.4
Innovation	0
Regional Priority	0
Total	877.65

Defining the scope and requirements

From the very beginning, the project requirements must be well conveyed and the scope made clear to all the members of your team.

Team work

The ease with which you can blend in the green features, satisfy the prerequisites which are mandatory and earn credits, depends to a great extent on the expertise of your team in the corresponding certification program. Whether or not the members of the team, both the design and construction team, are acquainted with the certification system, sustainable practices defined in the system, certification process etc. can have profound effects on the project value. Good team work is a secret key to effective cost control

Feasibility and cost based analysis – Finalize the rating to be targeted.

A feasibility checklist should be prepared at the project start itself and this must be updated regularly. Feasibility checklist provides a platform for credit wise cost and feasibility analysis. It helps to identify the possibilities of a higher rating. Accordingly the design and the construction teams may be informed to make cost effective changes to the project.

Budget allocation

Now you have in hand, a target rating level and also total available budget. Check if the funds are sufficient to meet the target. This is where the cost based analysis of credits gain importance. If an insufficiency is recognized, the goals must be either reduced or the funds increased. The project teams must not proceed forward with the same set of goals without finding a solution to insufficiency of fund as this could even affect the project completion to a large extend. Also going forward with insufficient funds will lead to more drastic scope cuttings in future than you would expect.

To align the budget with the program, a cost model should be prepared. Majority of the cost elements, in quantity as well as quality, will be available from the program and from that it is possible to build a cost model. This shall serve as a communication tool for the project team, letting a clear understanding of any limitations in budget.

Another important finding from the analysis is that costs are not necessarily cumulative. In many cases, a design/ green feature that allows a project to meet one sustainable design criteria will also allow that project some other criteria also without any additional cost impact or with minimal cost impact. This understanding also helps in budgeting to a great extent.

Being up to date

It is very important to be up to date to have a firm grip on the project. The steps for being on track are:

Paper works: Start any documentation work as early as possible, and stick on to it as you proceed with the project work.

*Energy model*s: Prepare energy models to inform the design team about changes to be made to cross the thresholds mentioned in the certification systems.

Monitoring and updating of checklist: Monitor and update the checklist regularly to see if the sustainability goals are being met.

Cost models: Prepare cost models, validate the same and use them to predict the green costs.

Conclusion

The best possible way to budget for green features in order to achieve the highest rating possible for the building is to identify the goals, perform energy modelling, prepare a feasibility checklist and to build an a cost model for them in the initial stage of the project itself.

The project chosen for case study could have achieved the highest possible platinum with just 3.9% cost increment over the total cost, had it done a feasibility and cost based analysis in the predesign phase itself.

References

[1] Dong-Xue Zhao, Bao-Jie He, Christine Johnson & Ben Mou, "Social problems of green buildings: From the humanistic needs to social acceptance", Renewable and sustainable energy reviews, Vol. 51, 2014, Pages 1594 - 1609
[2] Binh K. Nguyena, Hasim Altan, "Comparative review of five sustainable rating systems", Procedia Engineering, Vol. 11, 2011, Pages 376 - 386
[3] Michael Driedger, "Choosing the right green building rating system", Perkins + Will Research Journal, Vol. 01.01, 2009, Pages 22-41
[4] Jiri Dobias, Daniel Macek, "Leadership in Energy and Environmental Design and its impact on building operational expenditures", Creative Construction Conference, 2014, Pages 162-168
[5] Nurul Zahirah Mokhtar Azizi, Nazirah Zainul Abidin, Azlan Raofuddin, "Identification of Soft Cost Elements in Green Projects: Exploring experts' experience", Procedia Social and Behavioural Sciences, Vol. 170, 2014, Pages 18-26
[6] Greg Kats, "The cost and financial benefits of green buildings", 2003
[7] USGBC. Lisa Fay Matthiesen and Peter Morris, "Costing green: A comprehensive cost database and budgeting methodology", 2004
[8] Davis Langdon, "Cost of Green Revisited".

Advanced Oxidation Process (AOP) for Removing High Concentration of Iron in Drinking Water Sources

Lakshmy A Kumar* and V Meera**
* PG Student, Calicut University
lakshmyajith1991@gmail.com
**Associate Professor, Calicut University
meerav17@hotmail.com

Abstract: Iron can be removed from drinking water by advanced oxidation process (AOP) especially when the concentration of the metal is high and variable. The study focusses on the oxidation of iron using UV lamp of 8 watts and H_2O_2 in 1:1 proportion with the iron concentration present in water. Operating parameters investigated are iron concentration and reaction time. Experiments were conducted with tap water spiked with iron concentrations from 0.3 -10 mg/l. The results shows that iron can be brought down to the IS limits using AOP, provided proper contact time is given. For concentrations of range 0.3-1, 1-5 and 5-10 mg/l, the time required for attaining the permissible limits prescribed by IS 10500 2012 is 30 minutes, 90 minutes and 180 minutes respectively. This study reveals that AOP is an efficient method that can be adopted for treating water high in iron and to bring down its concentration to potable limits.

Keywords: Drinking water, Iron, Advanced Oxidation, H_2O_2, UV lamp.

Introduction

Iron can be a troublesome chemical in drinking water. Rainwater, as it infiltrates the soil and underlying geologic formations, dissolves iron, causing it to seep into aquifers that serve as sources of groundwater for wells. In drinking water, iron is seldom found at concentrations greater than 10 mg/l. However, as little as 0.3 mg/l can cause water to turn a reddish brown colour. Iron is mainly present in water in two forms: either the soluble ferrous iron or the insoluble ferric iron. Water containing ferrous iron is clear and colour less because, iron is completely dissolved. When exposed to air, the water turns cloudy and a reddish brown substance begins to form. This sediment is the oxidized or ferric form of iron that will not dissolve in water.

Iron has become an important water pollutant now a days. Oxidation, aeration, ion exchange etc are some of the common methods adopted for the removal of iron from water. Ion exchange involves the use of synthetic resins, dolomite coated manganese dioxide [1] or other similar materials, where an ion on the solid phase is exchanged for the unwanted ions in the water [2]. Ion exchange method is effective for iron removal but it releases sodium ions which increase its concentration in treated water. So it is not a practical water treatment option for those who are concerned with sodium intake. Also, this process is not effective on organic iron. Polyphosphate treatment is limited to low levels of contamination up to 1 mg/l. At this level the dissolved iron is stabilized and dispersed, and cannot react with the oxygen. Since the metals are not removed, odour and taste of the iron remain in the water. Phosphate compounds are not stable at high temperatures. If the water is boiled it becomes destabilized and iron will come out of solution. High levels of dissolved iron concentration, up to 5 mg/l, can be treated with aeration followed by filtration. Aeration mixes oxygen rich air with treated water and converts soluble iron into insoluble form [3]. Oxidation is the most effective method for the removal of iron from water. It oxidizes soluble ferrous in to ferric form and it gets precipitated out. It can be easily removed from water by giving filtration as a post treatment. Oxidants like potassium permanganate [4], sodium hypochlorate [5] hydrogen peroxide etc. can be used. Oxygen can be injected by air bubbling other than adding a chemical oxidant [6]. The methods like ultracentrifugation [7] adsorption in granular activated carbon [8] has been already investigated for the removal of iron from water. Using ultracentrifugation 76% of iron was removed with an influent concentration of 10 mg/l. Biological methods of metal recovery termed as biosorption have been suggested as a cheaper, more effective alternative to existing treatment techniques. The process of adsorption is a well-established and powerful technique for treating domestic and industrial effluent. In this method a low cost adsorbent, generally an agricultural waste is used [9]. Different forms of inexpensive, plant materials such as Bengal gram husk [9], rice husk [10], sugarcane baggase [10, 11], reed mace [12], coconut shell charcoal [13], brown sea weed [14], coconut coir [11], bamboo charcoals [15], water hyacinth [16] have been widely investigated as a potential biosorbents for iron. A removal efficiency of 70% was obtained with an influent iron concentration of 100 mg/l using Bengal gram husk. The above mentionesd conventional methods may not be able to meet the current IS standards when the amount of iron is in the

effluent is very high. So the need for developing an efficient method for the removal of iron has become necessary in the present scenario.

Advanced oxidation process (AOP) comprises a range of similar but different chemical processes aimed at tackling pollution in water, air and soil [17]. The term advanced oxidation refers specifically to processes in which oxidation of contaminants occurs primarily through reactions with hydroxyl radicals. In water treatment applications, AOPs usually refer to a specific subset of processes that involve O_3, Fe^{2+}, H_2O_2, and/or UV light. It can be used for the removal of both organic and inorganic pollutants. Iron based sulphite photo oxidation process can be used for the removal of arsenic [18]. O_3, O_3/UV, H_2O_2/UV, O_3/H_2O_2/UV, Fe_2/H_2O_2 can be used for chemical oxygen demand (COD), colour and pesticide removal [19, 20, 21, 22]. Advanced oxidation process (AOPs) involve the two stages of oxidation: 1) the formation of strong oxidants (e.g., hydroxyl radicals) and 2) the reaction of these oxidants with organic and inorganic contaminants in water. These hydroxide ions are an efficient oxidizing agent which can also oxidize iron at higher oxidation rate. Thus iron can be converted into ferric hydroxides which can be precipitated out.

Among the various methods Hydrogen peroxide/UV light process is having low cost and H_2O_2 is a common oxidizing agent. This process includes H_2O_2 injection and mixing followed by a reactor that is equipped with UV light. During this process, ultraviolet radiation is used to cleave the O-O bond in hydrogen peroxide and generate the hydroxyl radical. The reactions describing UV/H_2O_2 process are presented below

$$H_2O_2 + h\nu \rightarrow 2\ HO\bullet \quad (1)$$

$$H_2O_2 + HO\bullet \rightarrow HO_2\bullet + H_2O \quad (2)$$

$$H_2O_2 + HO_2\bullet \rightarrow HO\bullet + H_2O + O_2 \quad (3)$$

$$2\ HO\bullet \rightarrow H_2O_2 \quad (4)$$

$$2\ HO_2\bullet \rightarrow H_2O_2 + O_2 \quad (5)$$

$$HO\bullet + HO_2\bullet \rightarrow H_2O + O_2 \quad (6)$$

In the aforementioned equations, Eq. 1 is the rate limiting reaction. Theoretically in UV/H_2O_2 process, the higher initial hydrogen peroxide concentration produces higher hydroxyl radical concentration (Eq. 1), which decomposes more target compound. However, an optimal hydrogen peroxide concentration exists because over dosing of hydrogen peroxide would lead to reaction with hydroxyl radical and formation of $HO_2\bullet$ (Eq. 2). UV/H_2O_2 process is efficient in mineralizing organic pollutants. A disadvantage of this process is that it cannot utilize solar light as the source of UV light due to the fact that the required UV energy for the photolysis of the oxidizer is not available in the solar spectrum. Moreover, H_2O_2 has poor UV absorption characteristics and if the water matrix absorbs a lot of UV light energy, then most of the light input to the reactor will be wasted. The major factors affecting this process are the initial concentration of the target compound, the amount of H_2O_2 used, wastewater pH, presence of bicarbonate and reaction time. UV/H_2O_2 process has been widely used for dyes removal. The combined use of UV/H_2O_2 with ultrasonic waves (US) has also been reported in order to enhance colour removal during dyeing wastewater treatment [21].

The main objective of the present study was to examine the feasibility of UV/ H_20_2 advanced oxidation process in treating water contaminated by iron. The effects of reaction time and influent iron concentration were examined.

Materials and methods

Water sample preparation

The water samples used throughout the study were prepared from the stock iron solution with an iron content of 0.2 mg/ml using ferrous ammonium sulphate $(NH_4)_2Fe\ (SO_4)_2 \cdot 6H_2O$ and 1N potassium permanganate. The required amount of stock solution was mixed with tap water to get the iron concentrations of three ranges 0.3-1, 1-5 and 5-10 mg/l. The initial physico chemical characteristics of the tap water before spiking are shown in table 1. Experiments were done at 3 ranges of iron content. 1000 ml of sample prepared was adopted for the batch study.

Experimental setup

The experiments were carried out in a reactor of dimensions 30x25x15 cm. The AOP were designed to test the influence of operating parameters such as influent iron concentration (0.3- 10 mg/l) and treatment time. The pH was adjusted to neutral as the water was intended for drinking. It was adjusted using either 1M NaOH or 1M HNO_3. Hydrogen peroxide dosing was done in the ratio 1:1 with the iron concentration. 30% H_20_2 was adopted and it was diluted to get the required amount. Mixing

Table 1. Influent water characteristics

Parameter	Value
Hardness	264 mg/l
pH	6.4
Residual chlorine	0 mg/l
Chloride	7.9 mg/l
Dissolved oxygen	5 mg/l
Iron	BDL*
TDS	54.2 mg/l
Alkalinity	16 mg/l

*before spiking with stock solution of iron

was done by a magnetic stirrer at 450 rpm. UV lamp of 8 watt capacity was adopted and it was kept at the top of the reactor in such a way that the entire radiation fell within the reactor. The samples were collected in half hour intervals and tested for iron content. The process was continued till it reached the permissible limit. The supernatant alone is taken for analysis of iron.

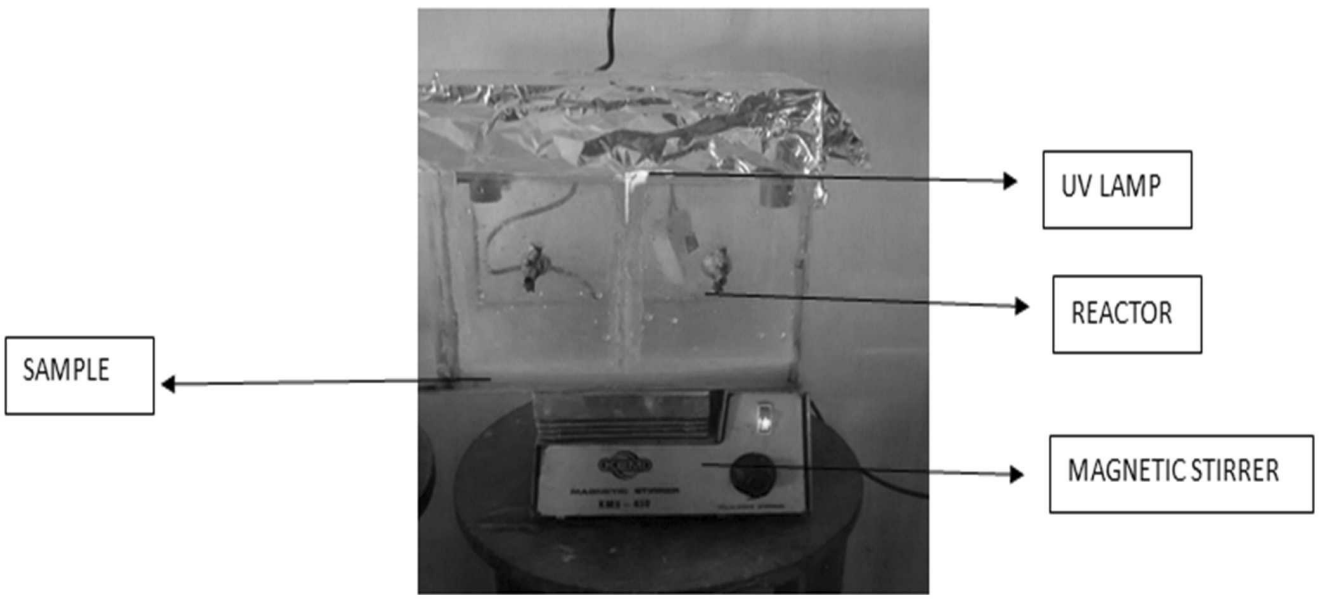


Figure 1. Experimental set up

Analytical techniques

Iron was measured by phenanthroline method as per APHA. Absorbance was measured using Systronics double beam spectrophotometer 2202 at 510 nanometer wavelength.

Results

The experiments were conducted until the residual concentration reached the IS 10500 2012 limits (0.3 mg/l). It indicates that higher the concentration of iron, the required time to achieve the IS limits is also higher. For 0.3 and 0.65 mg/l, the iron content reduced to 0.16 mg/l and 0.22 mg/l respectively after 30 minutes. But for 1 and 3 mg/l of iron content, the IS limits was reached only after 90 minutes and it has been reached up to 0.28 mg/l and 0.30 mg/l respectively. The effluent concentration reported for 5 mg/l was 0.12 mg/l after 120 minutes of reaction. Higher concentrations like 10 mg/l took 180 minutes to reduce the iron concentration to 0.25mg/l. A line of best fit was drawn with the values obtained after treatment and the co-efficient of determination was obtained as 0.88.

Table 2. Result of AOP

Influent iron concentration (mg/l)	Iron content at different time intervals (mg/l)					
	30 minutes	60 minutes	90 minutes	120 minutes	150 minutes	180 minutes
0.3	0.13	0.06				
0.65	0.22	0.09				
1.00	0.51	0.31	0.28			
3.00	1.2	0.61	0.30			
5.00	3.7	0.99	0.36	0.12		
10.00	6.89	3.37	1.59	0.69	0.49	0.25

Table 3. Summary of result

Influent iron concentration (mg/l)	Concentration of H_2O_2 (mg/l)	Time for achieving desirable/permissible limit as per IS 10500 2012 (Minutes)
0.3	0.3	30
0.65	0.65	30
1	1	90
3	3	90
5	5	120
10	10	180

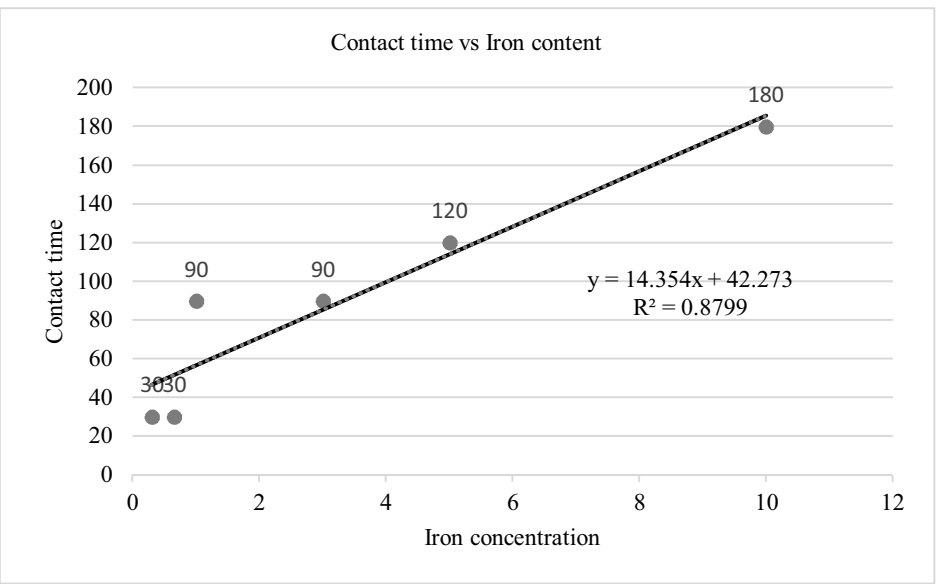


Figure 2. Line of best fit

Conclusion

This study has shown the potential of advanced oxidation process for the removal of iron from contaminated water (influent concentration ranging from 0.3-10 mg/l) for potable purpose. Advanced oxidation process adopted was using UV and hydrogen peroxide. The pH was kept at neutral as treatment was intended for drinking purposes. It was seen that the contact time to reach the permissible limit increased with concentration of iron showing its significant effect on treatment. For a range of 0.3-10 mg/l of influent iron, the maximum time required for advanced oxidation to reduce the concentration to permissible limits is180 minute and for 0.3- 1 mg/l the maximum time obtained was 90 minutes. Coefficient of determination was obtained as 0.88. The study thus showed AOP as an effective method in treating influent water containing iron for a wide range of concentration often reported in drinking water sources.

References

[1] M Milka, I Trajkovic, S Rogan, V Petrovic and S Jovanic, "Removal of Manganese and Iron from Groundwater in the Presence of Hydrogen Sulfide and Ammonia", Journal of Water Resource and Protection, Vol.6, Issue.1, 2014, Page.1781-1792

[2] A T Lemley, J J Schwartz, L P Wagenet, "Iron and Manganese in Household Drinking Water", Journal on Water Treatment, Vol.1, Issue.2, 2005, Page.1 -5.

[3] K Teunissen, A Abrahamse, H Leijssen, L Rietveld and A DijK, "Removal of both Dissolved and Particulate Iron from Groundwater", Journal of Drinking Water Engineering and Science, Vol.1, Issue.1, 2008, Page. 87–115

[4] Zogo L M Bawa, H Soclo and D Atchekpe,"Influence Of Pre-Oxidation With Potassium Permanganate On The Efficiency Of Iron And Manganese Removal From Surface Water By Coagulation-Flocculation Using Aluminium Sulphate: Case Of The Okpara Dam In The Republic Of Benin",Journal of Environmental Chemistry and Ecotoxicology Vol. 3, Issue.1, 2008, Page. 1-8.

[5] C Kan, W Chen, M Wan, P Phatai, J Wittayakun and K Li "The Preliminary Study of Iron and Manganese Removal from Groundwater by Naocl Oxidation And Microfiltration", Journal of Sustainable Environmental Resourses ,Vol. 22, Issue.1, 2005, Page. 25-30

[6] D Ellis, C Bouchard and G Lantagne, "Removal of Iron And Manganese from Groundwater by Oxidation and Microfiltration", Journal Of Desalination Vol. 130, Issue.1, 2000, Page.255-264

[7] E A Boyle, J M Edmond and E R Sholkovitz, "The mechanism of iron removal in estuaries ", International Journal of Environmental Management, Vol. 41, Issue.4, 1977, Page. 1313- 1324

[8] M E Goher, A M Hassan, Ibrahim A. Abdel-Moniem, A H. Fahmy, M H. Abdo and S M El-sayed, "Removal of Aluminum, Iron and Manganese Ions from Industrial Wastes Using Granular Activated Carbon and Amberlite IR-120H", Egyptian Journal of Aquatic Research, Vol. 41, 2015, Page. 155-164

[9] N Ahalya, R Kannamadi and T V Ramachandra, "Biosorption of Iron (III) from Aqueous Solutions Using The Husk Of Cicer Arientinum", Indian Journal of Chemical Technology, Vol.13, 2006, Page. 122-12

[10] C Dalaia, R Jhab and V R Desaic, "Rice Husk and Sugarcane Baggase Based Activated Carbon for Iron and Manganese Removal", Journal of Aquatic Procedia, Vol.4, 2015, Page. 1126 – 1133

[11] R Balaji, S Sasikala and G Muthuraman, "Removal of Iron from Drinking / Ground Water by Using Agricultural Waste as Natural Adsorbents", International Journal of Engineering and Innovative Technology, Vol. 3, Issue.12, 2014, Page.567-580

[12] N T A Ghani, A K Hegazy, G A El-Chaghaby and E C. Lima, "Factorial Experimental Design for Bio-Sorption Of Iron and Zinc Using Typha Domingensis Phytomass", Journal of Desalination, Vol. 249, 2009, Page.343-347

[13] K S Beenakumari, "Removal of Iron from Water Using Modified Coconut Shell Charcoal as Adsorbent", Journal Current World Environment, Vol. 4 Issue.2, 2009, Page. 321-326

[14] M M Figueira, B Volesky, and H J Mathieu, "Instrumental Analysis Study of Iron Species Biosorption by Sargassum Biomass", Journal of Environmental science and technology, Vol. 33, 1999, Page.184-1846

[15] B K Baruah, B Das,A Haque, K Misra and A K Misra, "Iron Removal Efficiency of Different Bamboo Charcoals: A Study on Modified Indigenous Water Filtration Technique in Rural Areas of Assam ", Journal of Chemical and Pharmaceutical Research, Vol.3, Issue.2, 2011, Page.454-459

[16] D T Win, M M Than and S Tun, "Iron Removal from Industrial Waters by Water Hyacinth", Journal of Science and Technology, Vol. 6, Issue. 2, 2002, Page. 55-60

[17] C Comninellis, A Kapalka, S Malato, S A Parsons, I Poulios and D Mantzavinos, "Advanced Oxidation Processes for Water Treatment: Advances and Trends for R&D", Journal of Chemical Technology and Biotechnology, Vol. 83, 2008, Page.769–776

[18] Z Myint and T Maree, " Arsenic Removal from Water using Advanced Oxidation Processes", Journal of toxicology, Vol.133, Issue.5, 2008, Page.113–118

[19] N Azbar, T YonarVand K Kestioglu, " Comparison of Various Advanced Oxidation Processes and Chemical Treatment Methods for COD and Color Removal from a Polyester and Acetate Fiber Dyeing Effluent ", International Journal of Chemosphere, Vol. 55, 2008, Issue.1, Page.35–43

[20] A Alaton , A Isil and W B Detlef, " Advanced Oxidation of a Reactive Dyebath Effluent: Comparison of O3, H2O2/UV-C and Tio2/UV-A Processes", , Journal of Water Research Vol.36,Issue.1, 2002, Page.1143–1154

[21] A S Stansinaki, "Use of Selected Advanced Oxidation Processes (Aops) for Wastewater Treatment – A Mini Review", Global Nest Journal, Vol. 10, Issue. 3, 2008, Page 376-385

[22] M I Badawy, M Y Ghaly and T A Gad-Allaha, "Advanced Oxidation Processes for the Removal of Organophosphorus Pesticides from Wastewater", Journal of Desalination, Vol. 194, 2006, Page.166–175

[23] G A Sodamade and L F Pearse, "Removal of Iron and Manganese Using Rapid Sand Filtration", Journal Of Technical Transaction, Vol. 47, Issue. 3, 2013, Page. 46-51
[24] American Public Health Association "Standard methods for examination of water and wastewater", 2nd edition, 1971, Pages 3.67-3.69.
[25] D Gosh, C R Medhi and M K Purkait, "Treatment of Fluoride, Iron and Arsenic Containing Drinking Water by Electro Coagulation Followed by Microfiltration", International Journal on Environmental Science and Technology, Vol.1, Issue.1, 2011, Page 338-345
[26] M G. Salem, M H El-Awady and E Amin, "Enhanced Removal of Dissolved Iron and Manganese from Nonconventional Water Resources in Delta District, Egypt", International Journal of Energy Procedia, Vol. 18, 2012, Page.983 – 993.

Waste Footprint of Kochi City, Kerala – An Analysis

Athira Ravi* and Subha Vishnudas**
*Research Scholar, Cochin University of Science and Technology, Kochi, India
ravi.athira@gmail.com
**Associate Professor, Division of Civil Engineering, Cochin University of Science and Technology, Kochi, India
subhakamal@gmail.com, v.subha@cusat.ac.in

Abstract: Common waste produced during pre modern era was mainly ashes and human waste, and these were released back into the ground locally, with least environmental impact. Following the onset of industrialization and the sustained urban growth of large population centers, the buildup of waste in the cities caused a rapid deterioration in levels of sanitation and the general quality of urban life. A lot of solutions arose like land filling, composting, incineration, pyrolisis etc. for handling the problem. But all of these either had an environmental impact or a public protest. What we have to do is to have a proper account of waste generated and its impact. Waste foot printing is one such technique which quantifies the impact of waste generated by an individual. This paper gives an overview of the waste foot print of residential areas of Kochi city, its analysis, findings and discusses the various sustainable options to reduce the waste footprint of the residential areas of Kochi city.

Keywords: Urbanization, sustainable waste management, ecological footprint, waste footprint, Kochi city.

Introduction

Urbanization is the movement of people from rural to urban areas [1]. The urbanization trend nowadays and the modern life style have increased the waste load on the earth and thereby polluting the urban environment to uncontrollable and dreadful limits [2]. The existing land fill sites and waste dumping sites are full beyond capacity and under unhygienic conditions leading to pollution of water sources, proliferation of vectors of communicable diseases, foul smell and odors, release of toxic chemicals, unaesthetic feel and ambience etc [3]. In earlier days, municipal wastes, comprised mainly of biodegradable matter, did not create much problem to the community as the quantity of wastes generated was either recycled/reused directly as manure or was within the assimilative capacity of the local environment [4] . The biodegradable wastes of the urban centers were accepted by the suburban rural areas for bio composting in the agricultural areas. With increasing content of plastics and non-biodegradable packaging materials, municipal wastes became increasingly offensive to the farmers and cultivators. As a result, the excessive accumulation of solid wastes in the urban environment poses serious threat not only to the urban areas but also to the rural areas. Now, dealing with waste, is a major challenge in many of the local bodies or government. There are two aspects to the challenge, the social mind set and technology application [5]. The social mind set is a very important aspect to be considered in this challenge. People are having the notion that the government is the authority to dispose whatever waste they are generating. This is very pathetic situation. Only the generators can manage waste. Though there are campaigns and awareness programmes to reduce the waste generation and source reduction, it is very hard to maintain the enthusiasm after the campaigns. In these circumstances we have to think of an alternative which is to be enforced by laws or rewards to reduce the amount of waste generation. A system, which gives the waste impact on earth quantified, just as we take the current bill, water bill etc and an amount to be paid based on the quantity, should be imagined. Or on the other hand the waste generators which are causing low impact should be rewarded or appreciated. There should be clear cut limit for this quantified value based on the locality we live in and its biocapacity to assimilate the waste. Waste foot printing is one such tool which can reach these goals to some extent [5].

Kochi, the commercial capital of Kerala and the second most important city next to Mumbai on the Western coast of India, is a land having a wide variety of residential environments. Central city extends to an area of 275.85 sq.km and the area jurisdiction of the city corporation is 94.88sq.km [6]. The population of the corporation area as per 2001 Census is 5,95,575 and the gross density is about 6277 persons/sq.km. As per 2011 Census the population is 6, 01,574 [6]. From the ecological footprint studies in Kochi city, it is revealed that the consumption rate (EF=2.19gha) of the population in the city is very high and it is far exceeding the national average (0.8gha) and the nations biocapacity (0.4gha) and the available bio capacity per person in the world (1.8gha). The study also revealed that shelter footprint, which mainly depends on the house area usage and number of occupants, is very high in the city. The improper waste disposal at the source (residential units) is increasing the waste footprint of the population which results in the high goods and services footprint [6].

This paper gives a detailed analysis (general and statistical) of the waste footprint of Kochi city, findings and discusses the various sustainable options for reducing the waste footprint of the residential areas of Kochi city.

Concept of Waste Footprint

Before detailing the waste footprint concept, the concept of ecological footprint analysis [7] is briefed since waste footprint is a subset of ecological footprint. Ecological footprint analysis is a quantitative tool that represents the ecological load imposed on the earth by humans in spatial terms. Ecological footprint analysis was invented in 1992 by Dr. William Rees and Mathis Wackernagel at the University of British Columbia [8]. The ecological foot print of a defined population is the total area of land and water ecosystems required to produce the resources that the population consumes, and to assimilate the wastes that the population generates, wherever on earth the relevant land / water are located. The footprint is expressed in global hectares. A global hectare is one hectare of biologically productive space with world average productivity [9].

By the waste footprint or the ecological footprint of waste generation, the measurement of biologically productive land like fossil, energy land, forest land, pasture land, built up area etc, to assimilate the generated waste is meant [10]. Waste footprint can provide the per capita land requirements for waste generation. By calculating the waste footprint, the local authority can determine the land required assimilating the waste generated in present and future, selection of disposal site and disposal site characteristics, the land fill site design and the importance of recycling of different waste categories in order to reduce the footprint [10].

In calculating the ecological footprint for household waste generation, methodology to assess the household ecological footprint, developed by Mathis Wackernagel, Ritik Dholakia, Diana Deumling and Dick Richardson, Redefining Progress v 2.0, March 2000, was used. The methodology utilized the resource consumption and waste generation categories and the land use categories for those consumption and waste generation. [10]. The land use categories are summarized as

- Energy Land: The area of forest that would be required to absorb the CO_2 emissions resulting from that individual's energy consumption.
- Crop Land: The area of cropland required to produce the crops that the individual consumes.
- Pasture Land: The area of grazing land required to produce the necessary animal products.
- Forest Land: The area of forest required to produce the wood and paper.
- Sea Space: The area of sea required to produce the marine fish and seafood.
- Built Area: The area of land required to accommodate housing and infrastructure.

To calculate the ecological footprint of waste generation, the generated waste is categorized as paper, plastic, glass, metal, and organic waste (food waste). The sum of the total land required for different waste categories the biologically productive land required for waste assimilation can be obtained, which means the per capita ecological footprint of waste generation. The methodology presents all results in per capita figures. Multiplying the per capita data by the selected area's population gives the total waste footprint of that area.

Waste Footprint of Kochi City

Methodology

The city accommodates a population of 6,01,0574 as per 2011 census. For the detailed study of waste footprint of the city, representative samples were selected from the residential areas of the Kochi Corporation and outskirts. The samples were selected based on the following criteria

- density of population(high and low)
- location(away and near of CBD and major transportation nodes)
- mode of waste disposal(household level or community level)
- type of housing unit(individual plots, low rise building, row housing units high rise building)
- ownership of the building(individuals, government, builders)

The survey was conducted using a structured questionnaire containing questions concerning the socio economic profile of the households, quantity of waste generation of each category of waste, type of waste disposal etc. The objective of the questionnaire was to analyze the variation in waste footprint values depending on the socio economic profile of the people, quantity of waste generation, daily variations and the type of waste. Questionnaire survey was conducted for 500 samples during three different seasons namely dry season (April 2010 and December 2010-January 2011), wet season (July 2010) and festival season (August 2010), inside the Corporation boundary and random samples in the outskirts.

The year 2010 was taken as the base year. For tracking the waste generation and the recycling methods in the residences after the primary survey, survey was repeated in 7% of the primary survey samples in 2011, 2012 and 2013. The criteria for house selection in the second stage survey were the response from the inhabitants in the base year survey and level of cooperation. 15 enumerators participated in the base year survey. Households were requested to segregate the wastes generated per day and store for one day. The wastes generated from samples were categorized into paper, glass, plastic, metal and organic waste (mainly food waste). The amount of paper waste was indirectly taken from the data of periodicals in the houses. The amount of glass and metal waste generated in a week was taken in account.

Analysis of the Waste Foot Print of Kochi City

The analysis was done to find the yearly variation of waste with respect to the criteria selected during the survey. The analysis was done using the waste footprint analyzer which is a program developed based on the equations of ecological footprint of waste generation developed by William Rees and Mathis Wackernagel (1996), the authors of the concept for inputting the survey data and estimating the footprint values in a visual basic platform. The analyzer generated the footprint value in hectares per capita. Fig. 1 gives a display of the analyzer which communicates mainly through 3 windows.

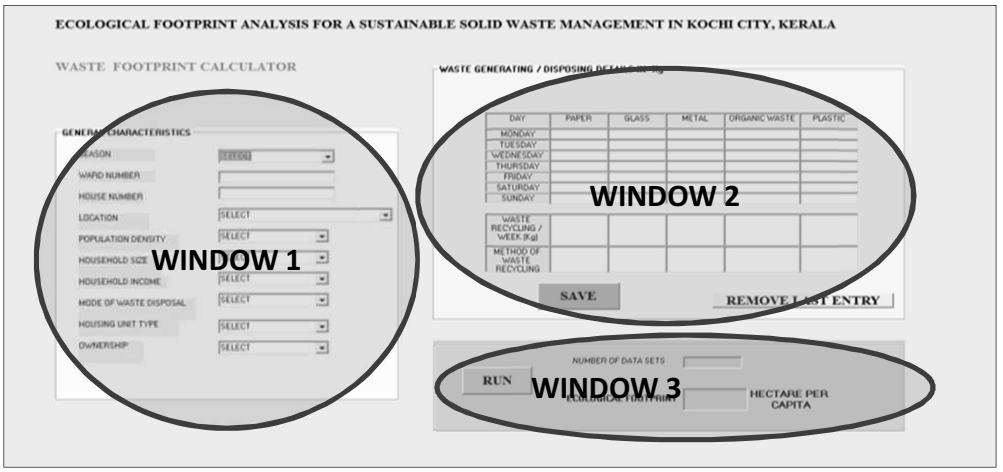


Figure 1. Windows/components of waste footprint analyzer

Window 1 and Window 2 are data input windows and Window 3 is execution cum output window. Window 1 has 10 sub windows which feeds the socio economic characteristics of the household under survey. The entries regarding season, ward number, house number, location, population density, household size, household income/month, mode of waste disposal, housing unit type and ownership details can be entered through these sub windows. Window 2 deals with the waste generation characteristics of the household. Window 3 is an execution cum output window. It consists of RUN button and a sub window. The button 'RUN' is an execution button which triggers the program execution. The sub window displays the number of datasets entered and gives the footprint value in hectares per capita.

The 500 samples' questionnaires in three different seasons were entered and the programme is executed to get the waste foot print of the residents of the city. 1500 datasets were created on this account for waste footprint calculations. The analyzer displayed the waste footprint in hectare per capita.

Statistical analysis and method

For combined analysis of the data over years in order to analyze the variations in quantity and footprint values in different conditions, homogeneity of error variance across all years were tested for significance by doing Bartlett's chi-square test (Gomez et al., 1984) for each variable. The test results showed that except for a very few cases the error variances were homogenous. Therefore the pooled analysis (Gomez et al., 1984) of variance could be conducted across the years to test if the variable was significant over the years and whether the interaction between year and the variable was significant. Since the sample size of each case was different, this was done by curtailing the sample size to the minimum size of the cases. The data was selected at random. The pooled analysis has been done in split plot manner.

Results and Findings

- The waste generation in the residential areas of Kochi City as on 2013 is 0.51kg/capita/day with an average household size of 3.72.
- On an average the organic waste constitutes 80.1%, 10.5% metal waste, 5.1% glass waste, 2.6 % paper waste and 1.9% plastic waste.
- In order to assimilate these wastes an area of 0.013 hectare per capita is required in the dry seasons, 0.016 hectare per capita for the festival seasons and 0.015 hectare per capita for the wet seasons.
- An average of 132.04 m^2 per capita of energy land, 0.08 m^2 per capita of forest land and 16.47 m^2 per capita of built up land is required to assimilate the waste generated by the residents of Kochi city.

- Even though the percentage of plastic in the solid waste is low compared to the other components, its percentage share of total waste footprint is relatively higher than other components except for metals. Metals also contribute to higher footprint. This is evident from the Fig. 2.

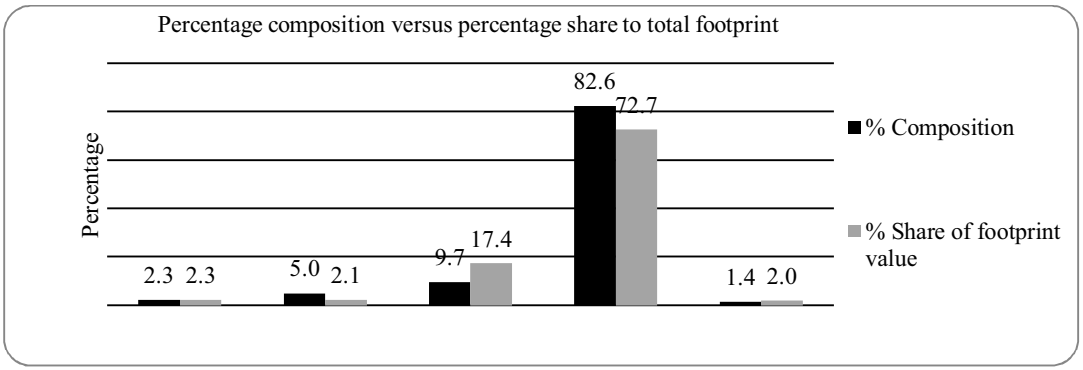


Figure 2. Percentage composition versus percentage share to total foot print

- The temporal variations of the waste footprint of the residential areas of Kochi city shows that the waste footprint has been increasing from 0.129 hectares per capita in 2010 to 0.0163 hectares per capita in 2013. This accounts for 26.35% increase within 4 years (Fig. 3).

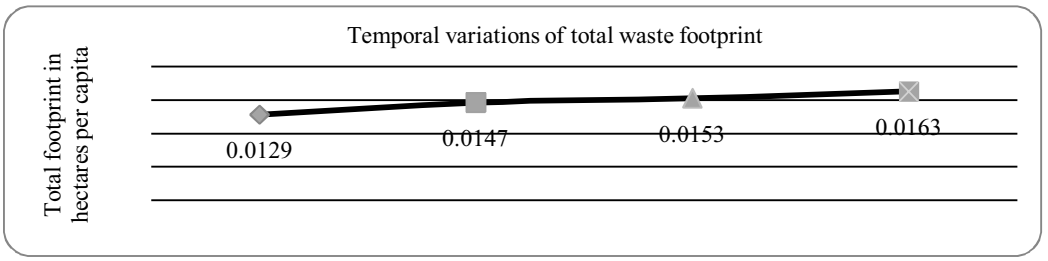


Figure 3. Temporal variations of waste footprint

- The analysis of ecological footprint of waste generation in the residential areas of Kochi city showed that with the present trend of waste generation and an assumed population growth rate of 4.5% as per the census studies, by 2051 the population will need about the full area of the city to assimilate the generated waste. This is shown in Table 3.

Table 3. Land requirement for waste management of the city

Year	Population	Waste footprint per person	Area (hectares) required for the total population
2001	595575	0.0129	7674.6
2011	601574	0.0129	7751.9
2021	628645	0.0129	8100.7
2031	656934	0.0129	8465.2
2041	686496	0.0129	8846.2
2051	**717388**	**0.0129**	**9244.3**
2061	749671	0.0129	9660.3
2071	783406	0.0129	10095.0
2081	818659	0.0129	10549.2

Paper waste

- On an average, the paper waste constitutes 2.6 % of total waste generated by the residents of Kochi city which includes magazines, newspapers, paper for packing, notices, information bulletins, paper items related to school, offices from the house.

- The average paper footprint of the residential areas of Kochi city is 4.06 m^2 per capita which implies that about 4.06 m^2 land area per person is required to assimilate the paper waste generated. This requires 3.53 m^2 of energy land, 0.08 m^2 of forest land and 0.44 m^2 of built up land.
- The paper waste constitutes maximum during the wet and festival season compared to dry season. The reasons can be attributed to the high level purchase of commodities in the festival season and the publication of periodicals, notices and other information bulletins in the festival season.
- The amount of paper waste generation is more in locations near to CBD and high density areas.
- The paper waste is more in houses with household size 5, followed by household size 3, 4 more than 5 and 2. But over the years the paper waste has not been varying with the size of the household. However year wise variation alone is significant in the case of paper waste.
- The quantity of waste generation in the houses which depends on community level disposal methods is more when compared to houses which opt for household level waste disposal.
- The paper waste show their maximum for samples with household income 10000-15000, followed by above 20000, 15000-20000, 5000-10000 and less than 5000. There are significant mean variations within year and between HH income classes. The variations based on household income shows that the quantity of generation of paper waste is directly proportional to the household income up to the class 15000-20000 and then decreases for the income class above 20000.
- For paper waste the row housing units generated more waste followed by individual plots, low rise buildings and high rise buildings. There are significant mean variations within year and between housing units. The quantity of paper waste generation has been on the increase from year to year.
- The paper waste is more in individual owned buildings followed by builder owned and government owned. There are significant mean variations within year and between ownership types. The quantity of paper waste generation has been on the increase from year to year. The amount of paper waste generation is more in individual owned houses followed by builder owned and government owned.

Glass waste

- On an average the glass waste constitutes about 5.1% of total waste which mainly constitutes bottles, storage jars, crockery etc.
- The average glass footprint of the residential areas of Kochi city is 3.28 m^2 per capita which implies that about 3.28 m^2 land area per person is required to assimilate the glass waste generated. This requires 2.92 m^2 of energy land and 0.36 m^2 of built up land.
- The glass waste constitutes maximum during the dry season followed by wet and festival season.
- The amount of glass waste generation is more in locations near to CBD and low density areas.
- The glass waste is more in houses with household size 5, followed by household size more than 5, 4, 3 and 2. The glass waste has been varying significantly with the size of the household. Also the year wise variation is significant.
- The quantity of waste generation in the houses which depends on community level disposal methods is more when compared to houses which opt for household level waste disposal.
- The glass waste show their maximum for samples with household income above 20000 followed by 5000-10000, 10000-15000, 15000-20000 and less than 5000. There are significant mean variations within year and between HH income classes. There are significant mean variations within year and between HH income classes. The temporal variations in the amount of glass waste generation over the years shows that the quantity of glass waste generation has been on the increase from year to year up to 2012 and then shows a decline.
- For glass waste the high rise buildings generated more waste followed by low rise buildings, row housing units and individual plots.
- The glass waste is more in government owned buildings followed by builder owned and individual owned. There are significant mean variations within year and between ownership types.

Metal waste

- On an average the metal waste constitutes about 10.5% of total waste and it includes utensils, equipment parts etc.
- The average metal footprint of the residential areas of Kochi city is 29.12 m^2 per capita which requires 25.89 m^2 of energy land and 3.23 m^2 of built up land for assimilating the wastes generated.
- The metal waste constitutes maximum during the festival season followed by wet and dry season.
- The amount of metal waste generation is more in locations near to CBD and high density areas.
- The metal waste is more in houses with household size 3, followed by household size 5, 4 more than 5 and 2. The solid waste in the form of metal (in kg) has not been varying with the size of the household and there is no variations within the year and both within year and between HH size classes.

- The quantity of waste generation in the houses which depends on community level disposal methods is more when compared to houses which opt for household level waste disposal.
- The metal waste show their maximum for samples with household income above 20000 followed by 10000-15000, 5000-10000, 15000-20000 and less than 5000. There are significant mean variations within year and between HH income classes. The temporal variations in the amount of metal waste generation over the years shows that the quantity of metal waste generation has been on the increase from year to year up to 2012 and then shows a decline. It shows that the metal waste generation is increasing up to income level 10000-15000 and then decreases.
- For metal waste, the low rise buildings generated more waste followed by row housing units, individual plots and high rise buildings. There are no significant mean variations between housing units or within year.
- The metal waste is more in government owned followed by individual owned buildings and builder owned. There are significant mean variations between ownership classes. The amount of metal waste generated is more for individual owned buildings followed by government owned and builder owned.
- There are significant mean variations within year and between ownership types. The quantity of metal waste generation has been on the increase from year to year. The amount of metal waste generation is more in individual owned houses followed by builder owned and government owned.

Organic waste

- On an average the organic waste constitutes about 80.1 % of total waste which mainly include the food waste.
- The average organic footprint of the residential areas of Kochi city is 107.88 m^2 per capita which implies that about 107.88 m^2 land area per person is required to assimilate the organic waste generated. This requires 95.92 m^2 of energy land and 11.96 m^2 of built up land.
- The organic waste constitutes maximum during the festival season followed by wet and dry season. In all the seasons the organic waste constitutes the maximum.
- The amount of organic waste generation is more in locations near to CBD and low density areas.
- The organic waste is more in houses with household size more than 5 followed by household size 5, 3, 4 and 2. There are significant mean variations in the generation of organic waste between the HH size classes. But there are no temporal variations.
- The quantity of waste generation in the houses which depends on household level disposal methods is more when compared to houses which opt for community level waste disposal.
- The organic waste show their maximum for samples with household income 10000-15000 followed by above 20000, 5000-10000, 15000-20000 and less than 5000. There are significant mean variations within year and between HH income classes. The temporal variations in the amount of organic waste generation over the years show that the quantity of organic waste generation has been on the increase from year to year. The variations based on household income shows that the organic waste generation is increasing with the income level up to the income group 10000-15000 and then shows a decline.
- For organic waste the high rise buildings generated more waste followed by row housing units, low rise buildings and individual plots. There are no significant mean variations between housing units or within year.
- The organic waste is more in government owned followed by individual owned buildings and builder owned. there are no significant mean variations between ownership classes or within year

Plastic waste

- On an average the plastic waste constitutes about 1.9 % of total waste and mainly includes the carry bags, utensils, storage bins etc.
- The average plastic footprint of the residential areas of Kochi city is 4.25 m^2 per capita which implies that about 4.25 m^2 land area per person is required to assimilate the plastic waste generated which requires 3.78 m^2 of energy land and 0.47 m^2 of built up land.
- The plastic waste constitutes maximum during the festival season followed by dry and wet season.
- The amount of plastic waste generation is more in locations near to CBD and low density areas.
- The plastic waste is more in houses with household size more than 5 followed by household size 5, 3, 4 and 2. There are significant mean variations within year and between HH size classes. The temporal variations in the amount of plastic waste generation show that the plastic waste that a household is emanating has been on the increase from year to year. The variations based on household size shows that the plastic waste generation is inversely proportional to the household size except for the lowest class and highest class.
- The quantity of waste generation in the houses which depends on household level disposal methods is more when compared to houses which opt for community level waste disposal.

- The plastic waste show their maximum for samples with household income 10000-15000 followed by above 20000, 5000-10000, less than 5000 and 15000-20000. There are significant mean variations within year and between HH income classes. The temporal variations in the amount of plastic waste generation over the years show that the quantity of plastic waste generation has been on the increase from year to year. The variations based on household income shows that the plastic waste generation is highly flexible with income levels.
- For plastic waste the high rise buildings generated more waste followed by individual plots, low rise buildings and row housing units. There are no significant mean variations between housing units or within year.
- The plastic waste is more in builder owned followed by government owned and individual owned buildings. There are significant mean variations between ownership classes. The variations based on ownership shows that the amount of plastic waste generated is more for builder owned buildings followed by government owned and individual owned.

Paper footprint

- On an average the paper footprint comes to 4.06 m^2 per capita which constitutes 2.74% of the total footprint.
- The percentage composition of paper waste in the total waste is 2.6% and the % share to the total footprint value is 2.74%.
- The paper footprint is more in the wet season followed by dry season and festival season.
- The footprint is more in locations near to CBD and in high density areas.
- The paper footprint is inversely proportional to the household size. Also there are significant mean variations within year and between HH size classes. The temporal variations in the paper footprint values over the years show that the paper footprint values have been on the increase from year to year.
- Paper footprint is highest for the income group 5000-10000 followed by the groups above 20000, 15000-20000, 10000-15000 and less than 5000. There are significant mean variations within year and between HH income classes. The temporal variations in the paper footprint values over the years show that the paper footprint values have been on the increase from year to year. The variations based on household income shows that the paper footprint is directly proportional to the household income up to the class 15000-20000 and then decreases for the higher class
- The paper footprint in the houses which depends on community level disposal methods is more when compared to houses which opt for household level waste disposal.
- The paper footprint is more for samples in individual plots, followed by low rise buildings, row housing units and high rise buildings. There are significant mean variations within year and between housing units. The paper footprint values have been on the increase from the year 2010 to the year 2012 and then decreases in the year 2013. The variations based on housing units shows that the paper footprint is high for individual plots followed by row housing units, low rise buildings and high rise buildings.
- There are significant mean variations within year and between ownership classes. The paper footprint values have been on the increasing over the years. The variations based on ownership classes shows that the paper footprint is high for individual owned buildings followed by builder owned and government owned buildings.

Glass footprint

- On an average the glass footprint comes to 3.38 m^2 per capita and constitutes 2.23% of the total footprint.
- The percentage composition of glass waste in the total waste is 5.1% and the % share to the total footprint value is only 2.23%.
- The glass footprint is more in the dry season followed by wet season and festival season.
- Also the footprint is more in locations near to CBD than locations away from CBD.
- The glass footprint is more in high density areas compared to low density areas.
- For glass footprint the footprint values are the highest for samples with household size 2, followed by household size 5, 3, 4 & more than 5. There are significant mean variations between HH size classes. The variations show that the glass footprint is inversely proportional to the household size except for the household size 5.
- The glass footprint show their maximum for samples with household income above 20000 followed by 5000-10000 10000-15000, 15000-20000 and less than 5000. There are significant mean variations within year and between HH income classes. The glass footprint values have been on the increase from the year 2010 to 2012 and then decreases in the year 2013. The variations based on household income shows that the glass footprint is directly proportional to the household income up to the class 10000-15000 and then decreases.
- The glass footprint in the houses which depends on household level disposal methods is more when compared to houses which opt for community level waste disposal.
- The glass footprint is more for low rise buildings followed by high rise buildings, individual plots and row housing unit. There are significant mean variations between housing units. The glass footprint values are high for low rise buildings followed by high rise buildings, individual plots and row housing units.

- The glass footprint shows maximum for the government owned buildings followed by individual owned and builder owned. There are significant mean variations within year. Glass footprint values have been on the increase from year to year up to 2012 and then show a decline.

Metal footprint

- On an average the metal footprint comes to 29.12 m^2 per capita which constitutes 29.12% of the total footprint.
- The percentage composition of metal waste in the total waste is 10.5% and the % share to the total footprint value is only 29.12%.
- The metal footprint is more in the festival season followed by wet season and dry season.
- The footprint is more in locations near to CBD and in high density areas.
- The metal footprint the footprint is inversely proportional to household size. There are significant mean variations within year and between HH size classes. The temporal variations in the metal footprint values over the years show that the metal footprint values have been on the increase from year to year. The variations based on household size shows that the metal footprint is inversely proportional to the household size.
- Metal footprint is highest for the 10000-15000 group followed by the groups 5000-10000, above 20000, 15000-20000 and less than 5000. There are significant mean variations between HH income classes. The variations based on household income shows that the metal footprint is directly proportional to the household income up to the class 10000-15000 and then decreases.
- The metal footprint in the houses which depends on community level disposal methods is more when compared to houses which opt for household level waste disposal.
- The metal footprint values shows maximum in low rise buildings followed by row house buildings, individual plots and high rise buildings. There are significant mean variations between housing units. The metal footprint values are high for low rise buildings followed by individual plots, row housing units and high rise buildings
- The metal footprint shows significant mean variations between ownership classes. The footprint is maximum for the individual owned buildings followed by government owned and builder owned.

Organic footprint

- On an average the organic footprint comes to 107.88 m^2 per capita which constitutes 70.25 % of the total footprint.
- The percentage composition of organic or food waste in the total waste is 80.1% and the % share to the total footprint value is only 70.25 %.
- The organic footprint is more in the festival season followed by wet season and dry season.
- Also the footprint is more in locations near to CBD and high density areas.
- There are significant mean variations between HH size classes. The variations based on household size shows that the organic footprint is inversely proportional to the household size except for the class >5.
- Organic footprint shows maximum for the group 10000-15000 followed by the groups 5000-10000, above 20000, less than 5000 and 15000-20000. There are significant mean variations between HH income classes. The variations based on household income shows that the organic footprint is directly proportional to the household income up to the class 10000-15000 and then decreases.
- The organic footprint in the houses which depends on community level disposal methods is more when compared to houses which opt for household level waste disposal.
- The organic footprint values shows maximum in low rise buildings followed by row house buildings, individual plots and high rise buildings. There are no significant mean variations between housing units or within year.
- The organic footprint shows significant mean variations between ownership classes. The footprint is maximum for the individual owned buildings followed by government owned and builder owned.

Plastic footprint

- On an average the plastic footprint comes to 4.25 m^2 per capita. This constitutes 2.76 % of the total footprint.
- The percentage composition of plastic waste in the total waste is 1.9% and the % share to the total footprint value is only 2.7 %.
- The plastic footprint is more in the festival season followed by wet season and dry season.
- The footprint is more in locations near to CBD and in low density areas.
- The plastic footprint shows maximum value for household size 3, followed by 2, more than 5, 4 and 5. There are significant mean variations within year. Temporal variations in the plastic footprint values over the years shows that the plastic footprint value has been on the increase from year to year.

- There are significant mean variations within year and between HH income classes. The temporal variation in the plastic footprint values shows that the footprint has been on the increase from year to year. The variations based on household income shows that the plastic footprint is directly proportional to the household income up to the class 10000-15000 and then decreases.
- The plastic footprint in the houses which depends on community level disposal methods is more when compared to houses which opt for household level waste disposal.
- The plastic footprint values shows maximum in row house buildings followed by individual plots, low rise buildings and high rise buildings. There are significant mean variations within year and between housing units. The plastic footprint value has been on the increase over the years. The variations based on housing units shows that the plastic footprint is high for row housing units followed by individual plots, low rise buildings and high rise buildings.
- The footprint is more for the builder owned followed by government owned and individual owned buildings. There are no significant mean variations between ownership classes or within year.

Season

In all the seasons the organic waste (food waste) constitutes maximum followed by metal waste, glass waste, paper waste and plastic waste. And so are the footprint values.

Location

The amount of wastes generated in different locations in different seasons is given in Table 4. Table shows that the amount of almost all the wastes in all seasons is more in locations near to CBD/MTN. Over consumption or unnecessary purchases may be reason for increasing the waste generation. Food wastes constitute the highest, followed by the metal wastes and glass wastes.

Table 4. Amount of wastes in kg/day/household based on location in different seasons

Season	Location	Paper	Glass	Metal	Organic waste	Plastic
Dry	Away from CBD/MTN	0.035	0.086	0.166	1.122	0.013
	Near to CBD/MTN	0.036	0.097	0.185	1.192	0.020
Festival	Away from CBD/MTN	0.037	0.076	0.136	1.421	0.026
	Near to CBD/MTN	0.042	0.077	0.176	1.565	0.031
Wet	Away from CBD/MTN	0.036	0.080	0.149	1.319	0.025
	Near to CBD/MTN	0.041	0.086	0.167	1.577	0.027

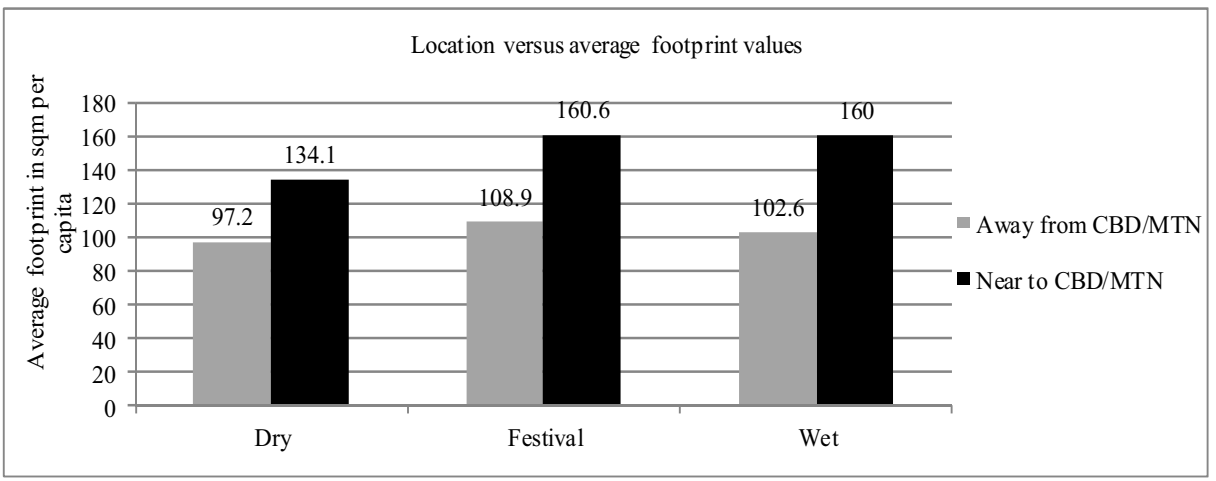


Figure 4. Location versus average foot print

Fig. 4 shows that the average foot print values are high in locations near to CBD/MTN. Residences which are near to CBD/MTN show high footprint values in the wet season and festival season. The footprint value is about 20% more when compared to that of the dry season. A similar trend is also noticed in the residences away from CBD/MTN.

Population density

The analysis showed that as the density of population is increasing the footprint is increasing as the amount of waste generated is more in high density areas as shown in Fig. 5.

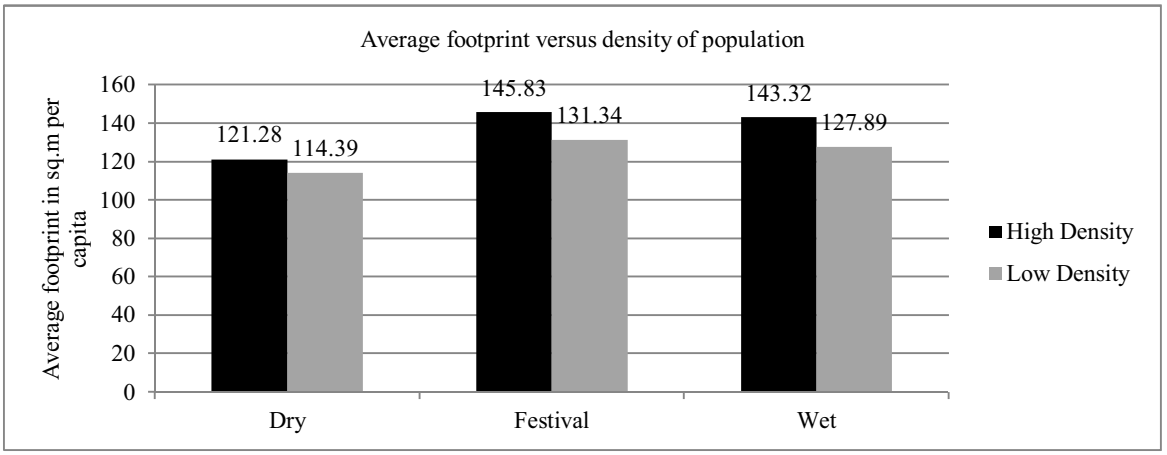


Figure 5. Average footprint versus density of population

The average footprint is more in the festival season. The per capita average footprint in sqm is about 145.83 in the festival season. Table 5 shows the composition of wastes in the high and low density areas in different seasons. The paper content in the dry seasons is high in the high density areas. But during the other seasons the composition of paper waste is more in the low density areas. The glass and metal content shows high composition during the dry and wet season in the high density areas. Organic waste and plastic content is more in the low density areas in all seasons.

Table 5. Density and % composition of waste in different seasons

	High Density			Low Density		
	% Composition			% Composition		
Season	Dry	Festival	Wet	Dry	Festival	Wet
Paper	2.60	2.16	2.15	2.25	2.28	2.24
Glass	6.29	4.00	4.96	6.04	4.30	4.78
Metal	11.32	9.79	10.56	12.37	7.18	7.58
Organic waste	78.89	82.57	81.32	78.07	84.62	83.38
Plastic	0.90	1.48	1.00	1.26	1.61	2.01

Household size

Analysis based on household size and average footprint value showed that the household size is inversely proportional to the average footprint values in all season (Fig. 6). In most cases the footprint value is high in the festival season. The reason can be explained from the Table 6.

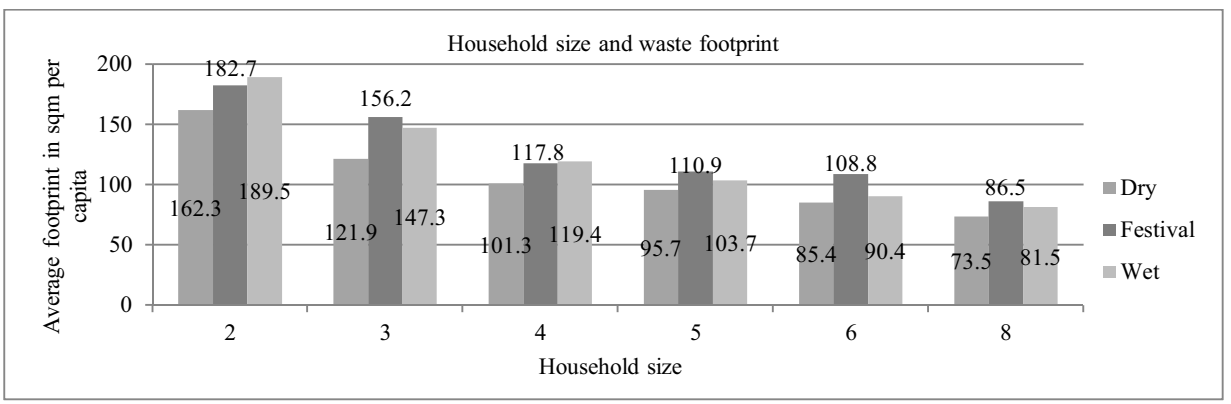


Figure 6. Household size and waste footprint

Table 6. Household size and per capita footprint

Household size	Total quantity of waste generated	Waste generated per person	Average footprint per person
2	9.30	4.65	178.13
3	11.09	3.70	141.59
4	11.78	2.94	112.84
5	13.73	2.75	103.54
6	15.09	2.51	94.82
8	17.30	2.16	80.52

Table 6 shows that the waste generated per person in low household size families are more when compared to families with large household size. This is contributing the high footprint values in families with small household size.

Household income

Analysis based on household income shows that the average footprint is comparatively high for the income group 10000 to 15000 followed by the above 20000 group. The footprint is high in the festival season for all income groups. The lowest contributor to waste footprint is the less than 5000 income group. The comparisons are given in Fig. 7.

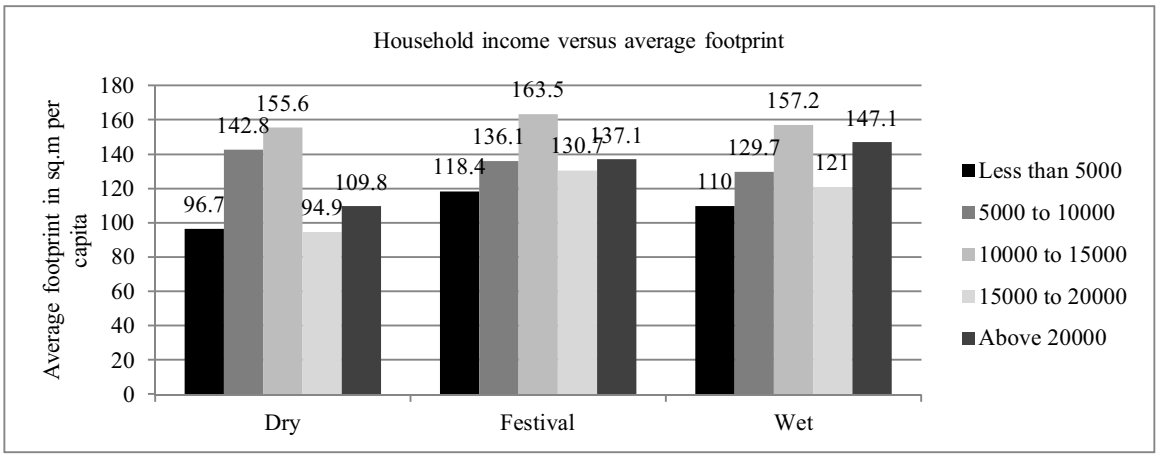


Figure 7. Household income versus average footprint

Waste disposal

Community level disposals show high waste footprint values when compared to household level disposal methods (Fig. 8). The waste footprint values are high in festival season. The low waste footprint values for household level disposals shows that the waste disposal at source itself is a sustainable option for proper solid waste management.

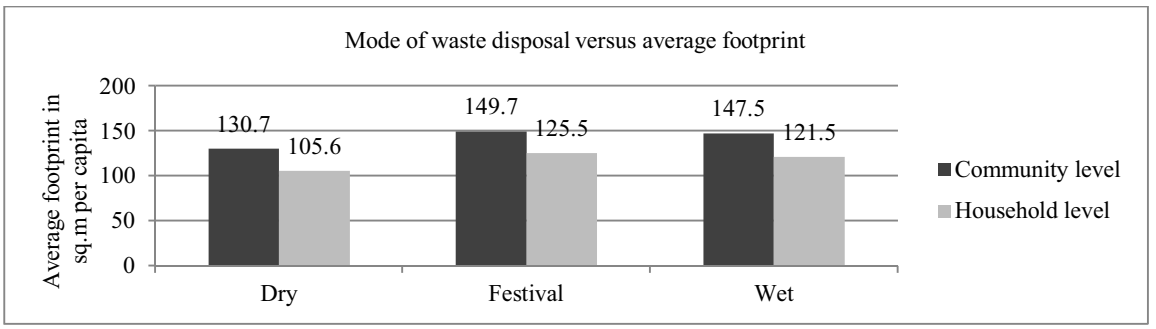


Figure 8. Mode of waste disposal versus average footprint

Housing unit

Analysis based on the housing type is given in Fig. 9. Almost all the housing types show high footprint values in the festival season. The average waste foot print value is comparatively low for individual plots except in festival season.

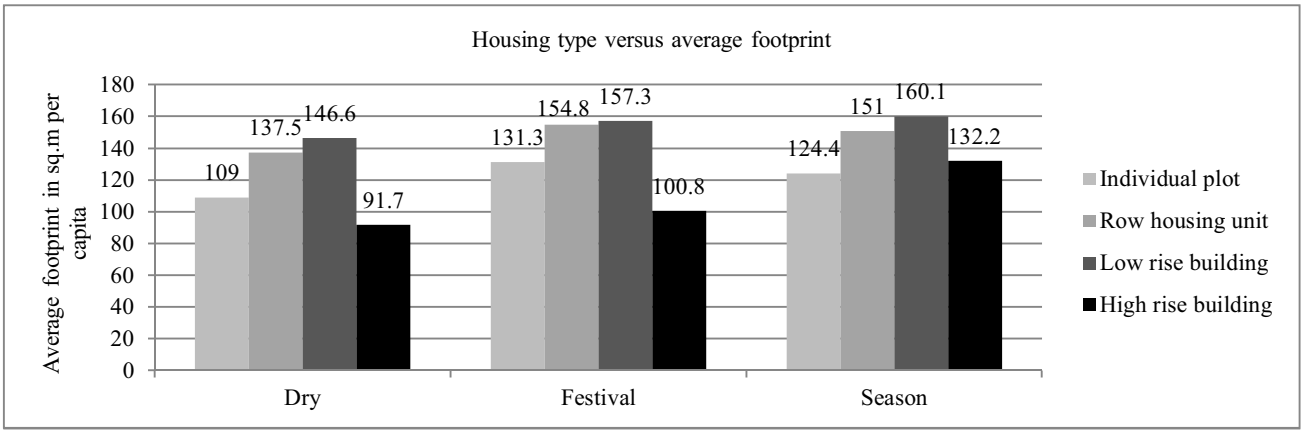


Figure 9. Housing type versus average footprint

Ownership

Residences owned by builders showed low waste footprint values when compared to others. The footprint values in the wet and festival season are also high (Fig. 10).

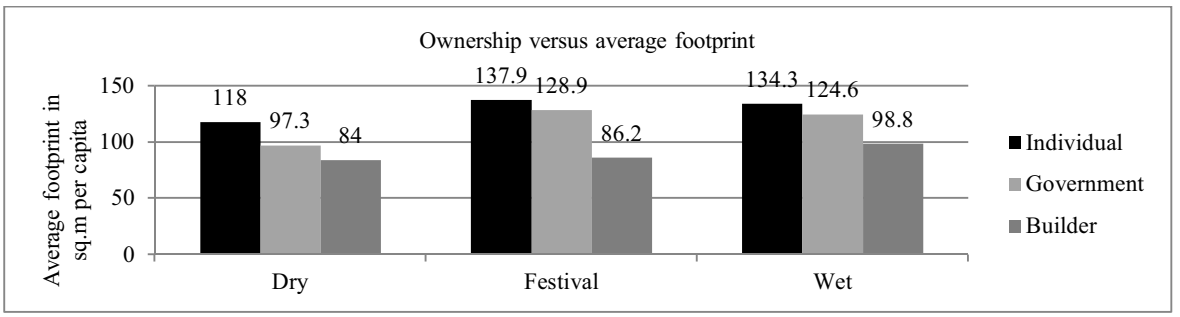


Figure 10. Ownership versus average footprint

Sustainable waste management options for reducing the waste footprint of Kochi city

Waste categories and different recycling levels affecting footprint

Table 7 shows the waste categories and their recycling levels affecting footprint. The 'present recycling' in the table meant the recycling rate during the time of primary survey. Since only 53 samples reported paper recycling it is assumed that the 0% paper is recycled. During the primary survey in 2010 and in the surveys conducted in the consecutive years and based on other secondary surveys, it was observed that many recycling initiatives are in the pipeline and at the anvil going to launch in the residential areas of the city. Some of them were the biogas production by Kerala Suchitwa mission, vermi-composting in residential flats by CREDAI Kochi and other programmes by NGOs. The 'targeted recycling' is meant in this regard. The 'projected recycling' rate is assumed considering the maximum recycling levels practiced in other urban areas that can reduce the waste footprint to considerable levels.

Table 7. Waste categories and different recycling levels affecting waste footprint

Waste Category	Recycling (%)			Waste Footprint (in sqm per capita)		
	Present	Targeted	Projected	Present	Targeted	Projected
Paper	0	60	90	3.26	2.36	1.92
Glass	0	30	50	2.85	2.58	2.42
Metal	0	30	60	23.35	16.69	10.04
Organic waste	0	75	90	96.76	54.67	46.25
Plastic	0	25	50	2.64	2.18	1.72
Total waste footprint (m^2/capita)				128.86	78.48	62.35

The research anticipates a 39% reduction in footprint value through the programmes going to get launched in the city and suburbs. A maximum of 51% reduction in footprint value can be attained through the high optimistic value of recycling.

Different waste generation levels and footprint values

Table 8 which shows the waste categories and different waste generation levels affecting waste footprint showed a proportional decrease in the footprint value with decrease in waste generation. As per the waste management hierarchy theories the source reduction proved the first order hierarchy in waste management in terms of waste footprint values.

Table 8. Waste categories and different waste generation levels affecting waste footprint

Waste Category	Waste generation			Waste Footprint		
	Present kg/capita/day	Targeted reduction (%)	Projected reduction(%)	Present footprint	Targeted footprint	Projected footprint
Paper	0.01	50	80	3.26	1.63	0.651
Glass	0.03	30	50	2.85	1.99	1.42
Metal	0.05	30	50	23.35	16.35	11.68
Organic waste	0.42	50	90	96.76	48.38	9.68
Plastic	0.01	50	75	2.64	1.32	0.66
Total				128.86	69.67	24.09

Combined analysis of waste reduction and recycling

According to the analysis with the recycling techniques proposed to launch in the city and a 50% reduction in paper, organic and plastic and 30% reduction in glass and metal waste generation can cause a 66.5% reduction in the waste footprint value. And in the maximum optimistic level 80% reduction in paper waste generation and 90% recycling of paper, 50% reduction in glass waste generation and with 50% recycling, 50% reduction in metal waste generation and with 60% recycling, 90% organic waste reduction and 90% recycling, 75% reduction in plastic waste and 50% recycling can reduce 91% of the present waste footprint of the city (Table 9).

Table 9. Combined analysis of waste reduction and recycling

Waste Category	Present			Targeted (%)			Projected (%)		
	Generation (kg)	Recycling	Footprint	Reduction in Generation	Recycling	Footprint	Reduction in Generation	Recycling	Footprint
Paper	0.01	0	3.26	50	60	1.18	80	90	0.38
Glass	0.03	0	2.85	30	30	1.81	50	50	1.21
Metal	0.05	0	23.35	30	30	11.68	50	60	5.02
Organic waste	0.42	0	96.76	50	75	27.33	90	90	4.62
Plastic	0.01	0	2.64	50	25	1.09	75	50	0.43
Total	128.86			43.09			11.66		

Conclusion

The paper illustrated that waste foot printing can be used as a tool to assess the impact of waste generation in an area, thereby focusing on the appropriate waste management technique suitable for the area. The waste footprint figures of Kochi city and its analysis showed that by 2051 the whole area of the city corporation will be required to assimilate the waste generated by the residents if the present trend of waste generation exists. This pointed out the highly unsustainability dilemma existing in the residential areas of Kochi city in the case of waste management. The analysis based on the different criteria showed that several social and economic factors are also affecting the waste footprint in addition to the technical or engineering factors. In addition to technical or engineering innovations in waste management a shift in the mindset of the people or an awareness creation is absolutely essential for the sustainable waste management of the area.

The analysis of various options which can reduce the waste footprint of the city highlights that waste reduction practices especially for organic waste and paper and the options for recycling can reduce the waste footprint of the residential areas of Kochi City to a considerable extent. Therefore organic waste reduction and recycling techniques should be encouraged in the city. Further studies in this regard can be framed to assess the sustainability of various organic waste management and recycling techniques.

The waste footprint can calculate the impact of waste generation of a single individual. This aspect of the concept is to be utilized and a model for calculating the waste footprint of a single individual in the city can be developed and the model is to

be made available through the social media. Thus the individual households in the city can calculate their waste footprint from their home by simply entering their amount of waste generation and their social and economic aspects. This will make them able to compare the footprint values with their neighboring houses, friends and colleagues and act in a positive manner.
On the other hand the policy makers can set strategies for sustainable solid waste management based on the waste footprint values or monitor the waste footprint values of the individuals so as to set limits for waste generation. This will become more effective if the equations for waste footprint calculations are generated specifically for the area under study. For this research and development studies should be initiated at the country level.
With this the paper concludes that waste foot printing technique is an apt tool for quantifying the waste generated which will help in technical and engineering innovations in waste management. At the same time it can used as an awareness or mind set change tool for sustainable solid waste management. The waste footprint study of Kochi city can be used as a pilot project in the country and more research and development projects regarding waste foot printing could be initiated at the country level, state level and local body level.

References

[1] Davis, K., "The Urbanization of the Human Population", J. Scientific American, 213, 3, 1965, 41-53.
[2] Bader A. Hakami , El-Sayed Sedek Abu Seif, “Household Solid Waste Composition and Management in Jeddah City, Saudi Arabia: A planning model”, International Research Journal of Environment Sciences, Vol. 4, 1, January (2015), 1-10.
[3] Athira Ravi , Subha V., “Waste Foot Printing For Waste Management – The Need Of The Hour”, Proceedings National Seminar on Green Environment Theme : Waste Management, December 2013, 19-27.
[4] Beena Puthillath, Dr.R.Sasikumar, “Integrated Solid Waste Management Score Board-A tool to measure performance in Municipal Solid Waste Management”, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 4, Issue 5(2), September - October 2015, 109-114.
[5] R. Ajayakumar Varma, “Technology Options for Treatment of Municipal Solid Waste with Special Reference to Kerala” Available online www.sanitation.kerala.gov.in/pdf /workshop/techno_2. pdf.
[6] Athira Ravi, Dr. Subha V, “Sustainable Development through the Environment Management Tool Ecological Footprint Analysis – A Study in Kochi City, India”, International Journal of Scientific & Engineering Research, Volume 4, Issue 4, April-2013, 776-782.
[7] W.E.Rees, “The concept of ecological footprint”, Academic press, 2001.
[8] Athira Ravi, Dr. Subha V, Ecological Foot Print Analysis-A Sustainable Environmental Management Tool for Kochi City”, Proc. of Int. Conf. on Advances in Civil Engineering, 2010, 6-8.
[9] Akkucuk, Ulas, “Handbook of Research on Developing Sustainable Value in Economics, Finance, and Marketing”, IGI Global, 2014.
[10] Dr. Md. Salequzzaman, Umme Tania Sultana, Asif Iqbal, Md. Ahasanul Hoque, “Ecological Footprint Of Waste Generation: A Sustainable Tool For Solid Waste Management Of Khulna City Corporation Of Bangladesh”, Available Online Http://Www.Necsi.Edu/Events/Iccs6/Papers/Df915143c9a8eec7ba94bbe57d41.Pdf.

Effect of Infill on Performance of Steel Frame Buildings

Ashitha K. Ikbal * and Satyabrata Choudhury **
* PG scholar, Department of Civil Engineering, NIT Silchar;
Email: ashiikbal@gmail.com
** Professor, Department of Civil Engineering, NIT Silchar;
Email: scnitsilchar@gmail.com (corresponding author)

Abstract: In India steel buildings are less common, though steel buildings may be economical in many instances. The study of the performance of steel buildings needs to be carried out to understand the seismic behaviour of steel buildings. One such study involves the effect of infill on the steel building performance. The Infill has significant influence on the response and performance of the buildings. In general it increases the stiffness of the system. However, it may shift the action from the column to beam. In the present study, the infill effect has been modeled by equivalent strut as per FEMA 356. SAP2000 version 14 has been used to evaluate the performance of the buildings. The study highlights the effect of infill on steel buildings including that on time period and nonlinear performance.

Keywords: Steel buildings; building performance, time period, design drift; performance level.

Introduction

Introduction of infill in steel structures will change the properties of steel structures. It is supposed to perform positively under seismic loading. Thought it has so many advantages, very less investigation were done under this topic. Infill in steel structures are supposed to increase the lateral stiffness, strength, fire resistance, wind resistance, weather resistance etc. Seismic force acting on the structures shall be carried by infill walls and frames, so that stiffness of the structure with infill will be much higher than stiffness of structure without infill.

Seismic performances of reinforced concrete structures with infill were analyzed before. It is used to give better performance than the structure without infill under seismic loading, so many researches have been done under that topic. Steel structures with and without infill has been analyzed by this work. In this research, a detail evaluation for seismic performance of infill steel frames has conducted by software SAP2000. Infill effect is modeled by equivalent struts. Performances of both the types of buildings are done by linear and non linear dynamic analysis (time history analysis) and non linear static analysis (push over analysis).

Past Works

Analytical and experimental researches on RC buildings and Steel buildings are started before. A very little research has been done on steel buildings comparing to RC buildings. Various research studies have shown that buildings with infill perform better than buildings without infill and stiffness increased after the infill introduction.

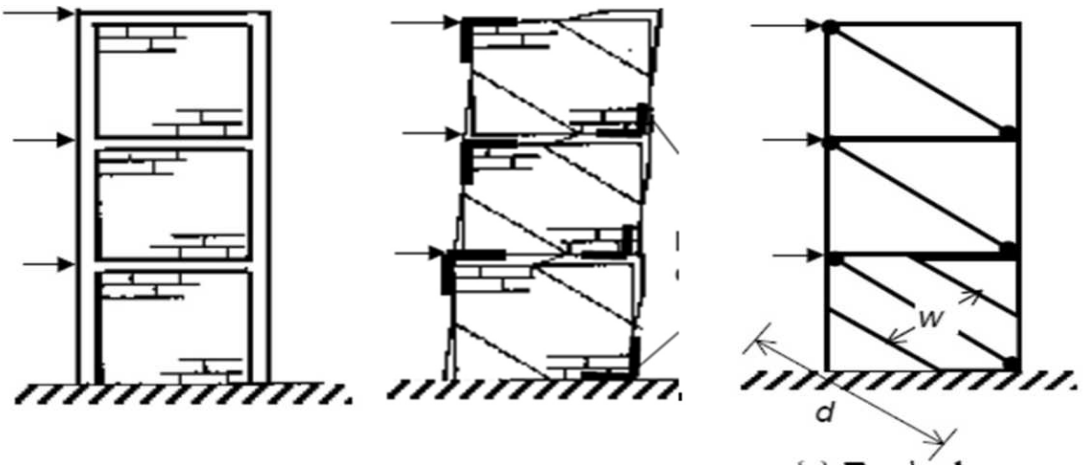
Davis R (2004) conducted seismic performance of masonry infill RC frames. For those he choosed two typical buildings located in moderate seismic zones of India. For knowing performance response spectrum analysis and non linear push over analysis done over these buildings. After these analyses he concluded that the presence of infill increases the stiffness of building it modifies the structural force distribution, also bending moments in the ground floor column increases and the mode failure by soft storey mechanism.

The typical behavior of an infilled frame subjected to lateral load as shown in figure 1.

This figure illustrates how equivalent struts are providing. The results of the study have shown that after adding infill the stiffness of the building increasing with respect to increase in storey shear. Time period of the building got decreased by the addition of infill.

Tasnimi A A (2011) described an experimental program to investigate the in-plane seismic behavior of steel frames with brick masonry infills having openings. Test specimens included masonry infills having a central opening, strong pier-weak spandrel, weak pier-strong spandrel and a door opening. All infills were unreinforced and all lateral deformations were imposed in the plane of the frames. He concluded that the ductility of infilled frames with openings depends on the failure mode of infill piers, and by experimentally found that diagonal tension failure in their piers and do not exhibit more ductility than solid infilled frames.

Raza S (2014) conducted a detailed evaluation of masonry infilled steel structures under seismic performance, which is concerned with the evaluation of local and global parameters of a structure when it is subjected to seismic loads. By that he concluded that For infilled steel frame structures the fundamental time period, roof displacement and interstorey drift ratios



(a) Infilled frame (b) Deformed shape (c) Equivalent strut

Figure 1. Behavior of infilled frames. (Robin Davis, 2004)

decrease with the increase in number of bays due to increase in the lateral stiffness, and the parameters increase with the increase in number of stories due to increase in the mass of the structure and larger height to base width ratio.

Methodology

Building Discription

The building Nomenclature for 4 steel buildings used in the present study is as shown below. Roman number indicates plan number and the last numerical figure indicates the number of stories.

Table 1. Building Nomenclature

Sl No	Building Name	No.of Stories	Description
1	**PI-4**	4	Plan I having 4 stories
2	**PII-5**	5	Plan II having 5 stories
3	**PI-4 INF**	4	Plan I having 4 stories and with infill
4	**PII-5 INF**	5	Plan II having 5 stories and with infill

Methodology

In this study, steel frame buildings with and without infill has been designed as per IS 800:1998 and IS 1893:2002. Also it has been conducted detail evaluation for seismic performance of infill steel frames by software SAP2000 version 14. Infill effect modeled by equivalent struts. Perform the analysis over time history and non linear analysis and find out the results like base shear distribution, response spectra, interstorey drift, performance level(push over analysis), roof displacement function, floor spectra, hinge pattern, push over curve. It gives an idea about overall structural behavior during an earthquake.

In nonlinear seismic analysis the ground motions has to be represented through time histories. Five Spectrum Compatible Ground Motions (SCGMs) has been generated with the following name with refers to the design spectrum at IS 1893: 2002 (Part 1) for zone 5. The SCGMs has been developed by software named Kumar (2004).

Table 2. SCGMs

Sl. No	**Name of SCGM**	**Background earthquake**	**Duration(sec)**
1	**GM1**	Petrolia-Cape Mendocino(1992)	60
2	**GM2**	Lome Prieta-Santa cruz mtns(1989)	40
3	**GM3**	Kerncounty-California(1952)	55
4	**GM4**	Elcentro-Imperial valley(1940)	54
5	**GM5**	Coalinga(1983)	65

Flow Chart of present study

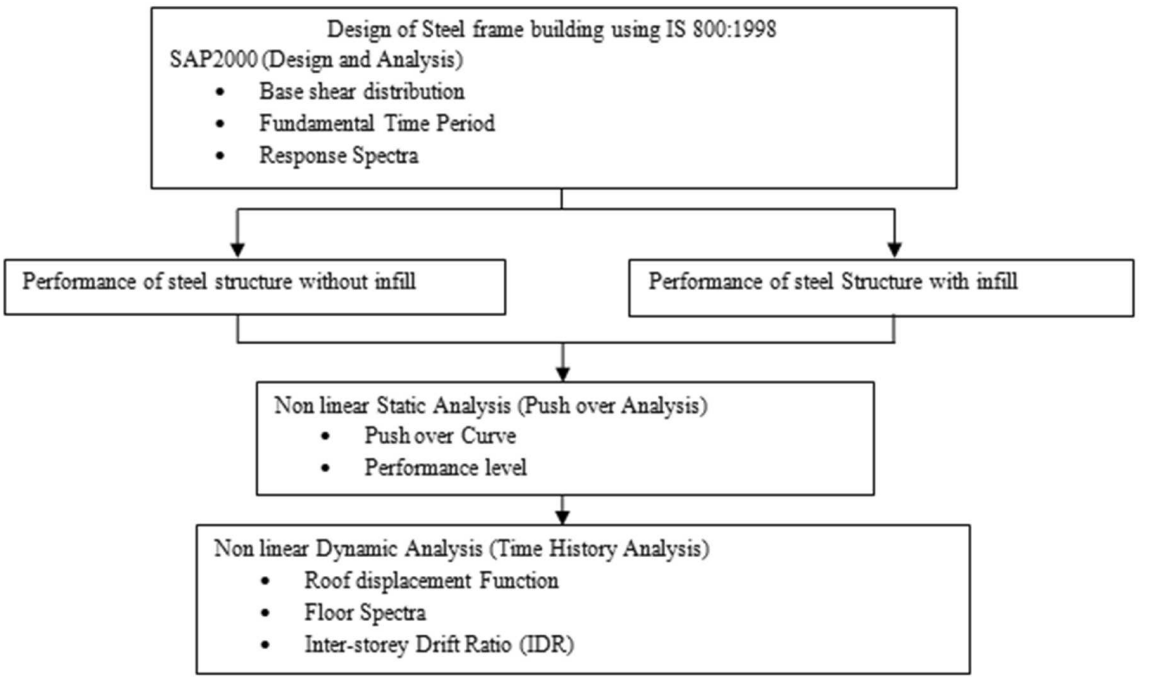


Figure 2. Flow chart of present study

Introduction of infill in Steel buildings

FEMA 356 is used for the introduction of infill in steel buildings. The elastic in-plane stiffness of a solid unreinforced masonry infill panel prior to cracking shall be illustrated with an equivalent diagonal compression strut of width, *a,* given by Equation. The equivalent strut shall have the same thickness and modulus of elasticity as the infill panel it represents,

$$a = 0.175(\lambda_1 h_{col})^{-0.4} r_{inf}$$

Where, $\lambda_1=[(E_{me}\, t_{inf}\, sin2\Theta)/(4E_{fe}\, I_{col}\, h_{inf})]^{1/4}$, and h_{col} = Column height between centerlines of beams (in), h_{inf} = Height of infill panel (in), E_{fe} = Expected modulus of elasticity of frame material (ksi), E_{me} = Expected modulus of elasticity of infill material (ksi), I_{col} = Moment of inertia of column (in^4), L_{inf} = Length of infill panel (in), r_{inf} = Diagonal length of infill panel (in), t_{inf} = Thickness of infill panel and equivalent strut (in),Θ = Angle whose tangent is the infill height-to length aspect ratio (radians), λ_1= Coefficient used to determine equivalent width of infill strut.

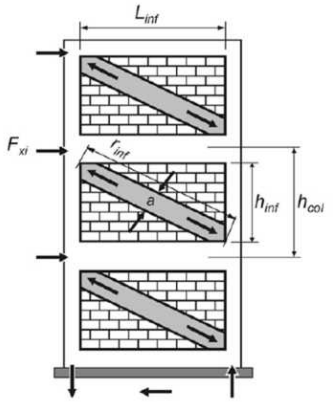
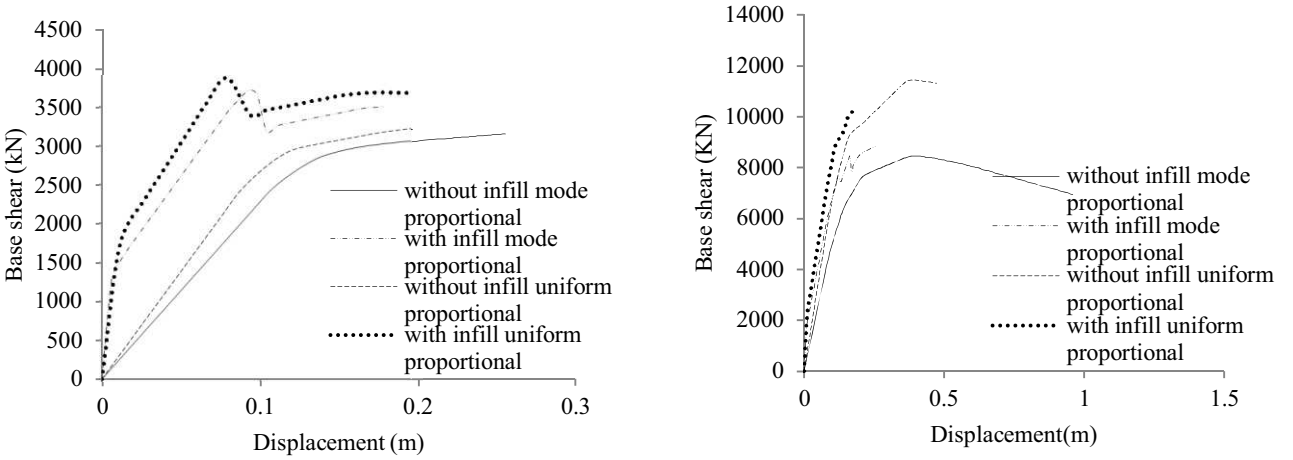


Figure 3. Compression Strut Analogy (FEMA 356)

Non linear static and non linear dynamic analyses are done as per FEMA 356 using SAP 2000. Results have been given in following chapter.

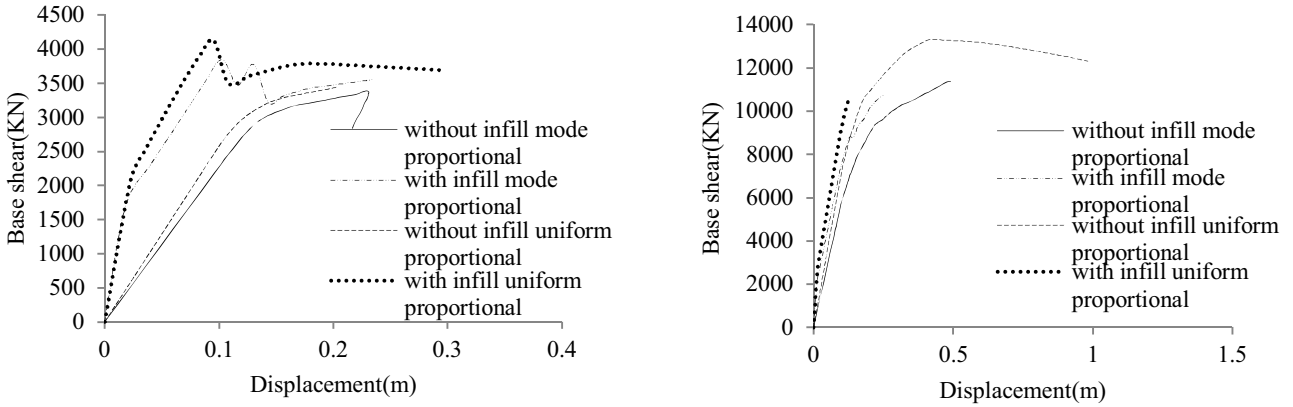
Results and Discussions

Both the buildings were analyzed using non linear static, non linear dynamic analysis and response spectrum method using IS 1893:2002 (part 1). The analyses were done by software SAP2000. The fundamental time period of the buildings were obtained from software. Steel buildings are rigid than the concrete buildings. According to this it will perform better than concrete buildings. When infill is considered, the fundamental time period of the building reduces and the building attracts more base shear. It is performing far better than the reinforced concrete buildings and bare steel frame buildings. Figure 1 and 2 shows the building changes after and before adding infill in terms of base shear vs. displacement.



a) Long direction b) Short direction

Figure 4: Push over curve for PI-4 and PI-4 INF

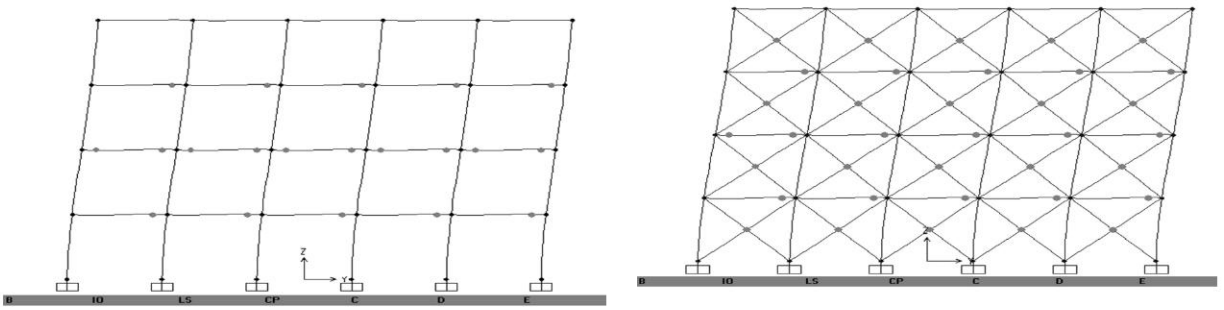


a) Long direction b) Short direction

Figure 5: Push over curve for PII-5 and PII-5 INF

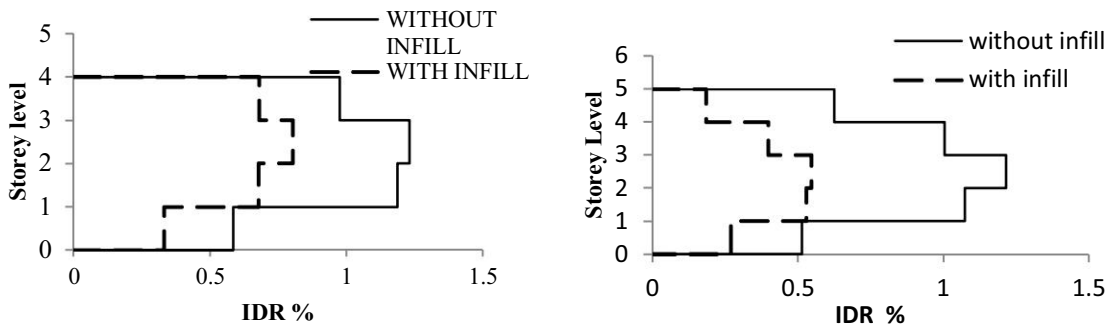
From these two figures it is clear that the infill increases the stiffness of building with respect to increase in base shear. Here nonlinear static results for both mode proportional and uniform proportional are given. Mode proportional push over curve means here it is taking vertical loads (gravity loads) acting on the building. Uniform proportional means it is taking only horizontal loads coming to buildings.

Figure 5 shows the hinge pattern for the plan I 4 storey building. It is clear that after adding infill hinge is changing from LS (life safety) to IO (immediate occupancy). Immediate Occupancy, is defined as the post-earthquake damage state that remains safe to occupy, it retains the pre-earthquake design strength and stiffness of the structure. Life Safety, means the post-earthquake damage state in which significant damage to the structure has occurred, but some margin against either partial or total structural collapse remains.



(a) Hinge pattern for PI-4 without infill (b) Hinge pattern for PI-4 INF with infill

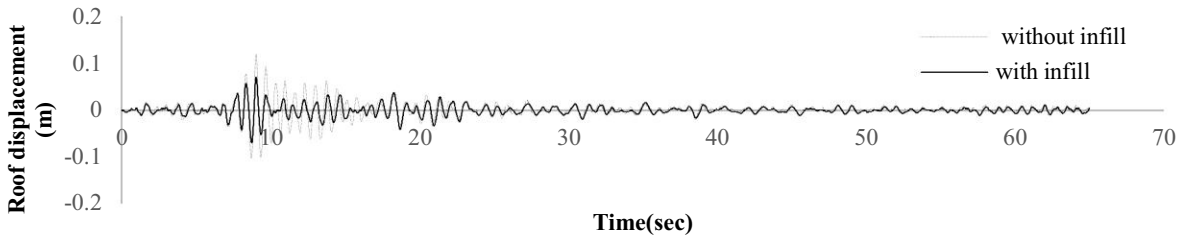
Figure 6. Hinge pattern for PI-4 with and without infill



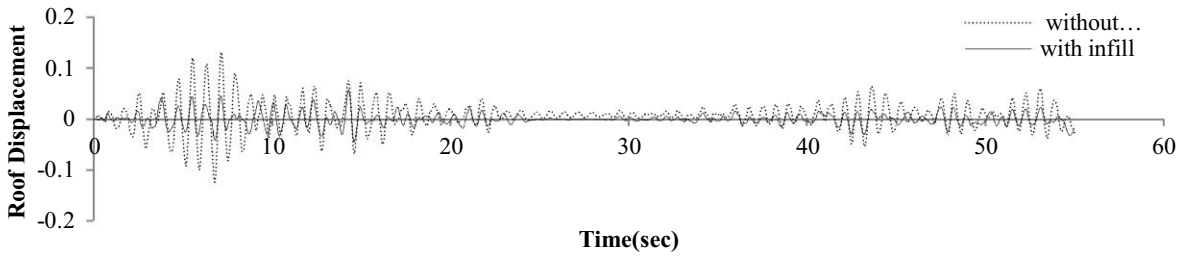
a) IDR for PI-4 and PI-4 INF b) IDR for PII-5 and PII-5 INF

Figure 7: Storey level vs. IDR%

Figure 6 shows the maximum IDR obtained for the buildings after nonlinear dynamic analysis (time history analysis). Figure 6 shows the roof displacement with respect to time period. From figure 6 and 7 it is clear that introduction of infill reduces the drift of the building. Infill gives strength and stiffness to the building.

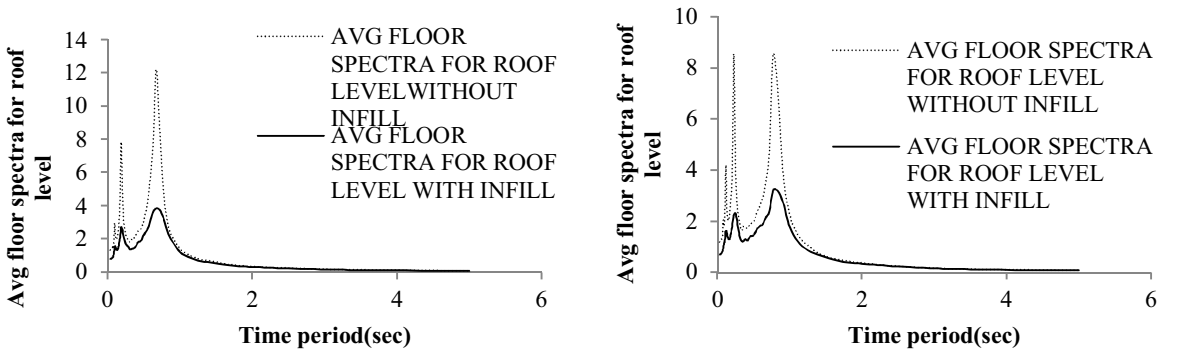


a) Roof Displacement for PI-4 and PI-4 INF



a) Roof displacement for PII-5 and PII-5 INF

Figure 8: Roof displacement vs. Time period for both buildings



a) Average floor spectra for PI-4 b) Average floor spectra for PII-5

Figure 9. Average floor spectra for PI-4 and PII-5

Floor spectra are the measure of maximum response characteristic of floor motion. Figure 8 indicates the average floor spectra for roof level of both the buildings. From figure it is shown that after infill addition the floor response also getting decreased. That figures out building is gaining more stiffness with infill.

Conclusion

The performances of steel building with and without infill are carried out by static non-linear pushover analysis and non-linear dynamic time history analysis. It is concluded that steel buildings with infill performed much better than steel buildings without infill. The stiffness of the building is increasing with decreasing time period. While time period is decreasing spectral acceleration of the building is increasing. Base shear and spectral acceleration have direct proportional relation and the base shear behavior with displacement obtained from push over analysis. From that it concluded that building stiffness increased with infill. After non linear dynamic analysis, ie., time history analysis it found that interstorey drift ratio of infilled frame building got decreased. Drift is indirectly proportional to stiffness. Roof displacement and floor spectra for roof level also gave the clear idea about building behavior.

Acknowledgment

I would like to take this opportunity to express my sincere gratitude and regards to my supervisor Prof. Satyabrata Choudhury of Civil Engineering Departmant of National Institute of Technology, Silchar for their guidance and constant inspiration throughout the course of my project. I also like to thank Dr. Upendra Kumar (Head of the Department of Civil Engineering, NIT Silchar) who allowed me to use the departmental computer laboratory to carry out my project work. I also express my deep sense of gratitude to all the faculty members of Civil Engineering Department, My family, seniors and my friends for their encouragement, help and support.

References

[1] Raza S, Khan M K I et al, “Seismic Performance Assessment of Masonry Infilled Steel Frame Structures”, International Journal of Innovation and Applied Studies, v 8, 2014, pp. 1470-1484.

[2] Doudoumis I N, “Improving Lateral Stiffness in The Diagonal Strut Model Of Infilled Frames”, The 14th world conference of earthquake engineering, 2008

[3] Banihashemi M R et al, “Reliability and fragility curve assessment of steel concentrically braced frames”, European Journal of Environmental and Civil Engineering,2015

[4] Davis R, Praseetha K et al, “Effect Of Infill Stiffness On Seismic Performance Of Multi-Storey Rc Framed In India”, 13TH World Conference ON Earthquake Engineering Vancouver, 2014, pp: 1-6.

[5] Elghazouli A. Y, “Assessment Of European Seismic Design Procedures For Steel Framed Structures”, Bull Earthquake Eng, v: 8, 2009, pp:65–89.

[6] Hariri-Ardebili M.A, Samani H A et al, “Free and Forced Vibration Analysis of an Infilled Steel Frame: Experimental, Numerical, and Analytical Methods”, Hindawi Publishing Corporation Shock and Vibration Volume 2014.

[7] Syed Farooquddin, Renuka Devi M.V et al, “Lateral Stiffness Of Infilled Frame With Door & Window Openings For Varying Modulus Of Masonry”, International Journal of Advanced Engineering Technology, v 3, 2012, pp 123-125.

[8] Tasnimi A A, Mohebkhah A, “Investigation on the behavior of brick-infilled steel frames with openings, experimental and analytical approaches”, Engineering Structures, 2011, pp 968-980.

[9] Barghi M, Azadbakht M, “Evaluating The Effect Of Masonry Infills On Natural Period Of Buildings With Moment-Resisting Frame”, The Structural Design Of Tall And Special Buildings, 2011, Pp 649-660.

[10] Abdelkareem K H, Abdel Sayed F K, "Equivalent Strut Width For Modeling R.C. Infilled Frames", Equivalent Strut Width For Modeling R.C. Infilled Frames, 2013, Pp. 851 – 866.
[11] Adukadukam A, Sengupta A K, "Equivalent Strut Method for the Modelling of Masonry Infill Walls in the Nonlinear Static Analysis of Buildings", Springer, 2013, pp 99-108.
[12] Razzaghi M S, Javidnia M, "Evaluation of the effect of infill walls on seismic performance of RC dual frames", Springer, 2015, pp 49-54.
[13] Chatterjee A K, Bhowmick A K, "An equivalent strut and tie model for steel plate shearwall Systems", WCEE, 2012.
[14] FEMA 356

A Study on Dynamic Shear Amplification in RC Frame Structures

Uday Sinha*and Satyabrata Choudhury**
*PG scholar, Department of Civil Engineering, NIT Silchar;
Email: udaysinha58@gmail.com
**Professor, Department of Civil Engineering, NIT Silchar;
Email: scnitsilchar@gmail.com (corresponding author)

Abstract: In this paper a study has been carried to know the dynamic shear amplification in the structures. Unified Performance-Based Seismic Design (UPBD) has been used to design the buildings where both target drift of buildings is considered along with the performance level. Two different buiding with different plan is used. The performance objectives of the buildings are selected in terms of interstorey drift ratio (IDR) and member performance level. The target performance objectives for the two buildings considered are 2% IDR with Life Safety (LS) performance level, and 3% IDR with Collapse Prevention (CP) performance level. The buildings have been modeled using SAP2000 V 14.0.0 software. The performance parameters have been evaluated by performing nonlinear analysis.

Keywords: Unified Performance-Based Design (UPBD), Interstorey Drift Ratio (IDR), Reinforced concrete frames, Building performance level, Dynamic shear amplification.

Introduction

With rise in population, the importance of tall structures come into effect to accommodate many number of people with lesser land use. Therefore, it is important to design the tall structural adequately keeping in mind the natural hazard to which it is surrounded and one such hazard are the earthquakes. Seismic design of buildings has been traditionally force based. In the force-based codal method of design, the base shear is computed based on perceived seismic hazard level, importance of the building and probable reduction in demand due to nonlinear hysteresis effects. The computed base shear is distributed at floor levels with some prescribed or estimated distribution pattern. Through force based method of design, an engineer cannot deliberately design structure for an intended performance level. The alternative approaches are displacement-based design (DBD) and performance-based design (PBD) which are gradually becoming popular in recent times. In these methods the design is done for intended displacements or, some intended performance objectives under a perceived hazard level. The DDBD (Direct Displacement Based Design) method for frame buildings by Pettinga and Priestley is evolved out of DBD method. In this method, the building is converted to an equivalent single degree of freedom (ESDOF) system and the design is aimed at satisfying some specified interstorey drift limit. If the required interstorey drift is not satisfied, the trial interstorey drift is reduced and redesign is done so that finally the target interstorey drift may be achieved. In this method, however, the member performance levels are not considered. Choudhury and Singh in 2013 reported a method which is an improvement towards the DDBD method for frame buildings, which tried to satisfy the target drift and performance level, and christened it as unified PBD (UPBD) approach. In this paper buildings are designed with UPBD method for desired target performance objectives and the dynamic behaviour of the buildings is observed from which dynamic amplification of shear force is studied.

Literature studies

Priestley and Kowalsky (2000) presented direct displacement based seismic design of concrete buildings in which design was done to achieve a specified damage level under specified earthquake. This method computed the required base shear to achieve specified damage level state.

Pettinga and Priestley (2005) applied direct displacement-based design method to six RC frame structures and evaluated the interstorey drift ratio which exceeded assumed drift limit through nonlinear time history analysis. They proposed a revised form of the modified modal superposition to account for higher-mode amplification of column shear forces, while a simple intensity dependent scaling factor to be applied in the capacity design process was developed for column bending moments.

Nilupa Herath, Priyan Mendis, Tuan Ngo, Nicholas Haritos (2010) presented a study on the seismic performance of super tall buildings. In the study the effect of higher modes on the performance of super tall buildings is discussed and current methods of analysis of super tall buildings are reviewed in the study.

Priestley (2000) presented performance based seismic design to develop an alternative approach to achieve an expected strain or drift performance level under an expected seismic hazard level.
Choudhury and Singh (2013) has improved Pettinga and Priestley (2005) method to incorporate both member performance level and drift limit simultaneously in displacement-based design and reported a new UPBD method.

Methodology

Design Philosophy
The Unified Performance Based Design (UPBD) method as stated earlier is an improvement over the DDBD method in which both the target drift and performance level of the building is taken care of. In this method the design of the structure can be done for a target performance objectives in terms of interstorey drift ratio (IDR) and member performance level. The member performance level is expressed in terms of plastic rotation allowed in beams corresponding to the desired damage state. Weak-beam strong-column concept is followed thereby allowing plastic hinges to develop only in beams and not in columns and hence columns remain elastic up to the performance point
Target objectives in the form of design drift and member performance level is achieved by estimating a design beam depth by the following considerations:
The frame yield rotation θ_{yF} is given by Eq. (1) in which ε_y is yield strain of rebar, l_b is length of beam and h_b is depth of beam.

$$\theta_{yF} = 0.5\varepsilon_y l_b / h_b \quad (1)$$

Eq. (1) gives the maximum elastic angular drift of a frame building. This includes the contribution of beam rotation, joint rotation, column end displacement and shear deformation. The design angular drift θ_d of the system is the sum of frame yield rotation and plastic rotation in the system (θ_p), as expressed by Eq. (2) and explained in Fig. 1. In Fig. 1, H_e effective height of ESDOF system. With capacity design, only beam is allowed to yield and plastic rotation comes from beams (that is, $\theta_p = \theta_{pb}$), where θ_{pb} is the plastic rotation in beams.

$$\theta_d = \theta_{yF} + \theta_{pb} \quad (2)$$

Substituting Eq. (1) in Eq. (2) and rearranging, Eq. (3) is obtained.

$$h_b = \frac{0.5\varepsilon_y l_b}{\theta_d - \theta_{pb}} \quad (3)$$

Eq. (3) gives a beam depth that shall satisfy the interstorey drift and target performance level of the building corresponding to plastic rotation allowed in beams. Here, θ_{pb} is average plastic rotation in beams corresponding to the performance level desired, and can be obtained from FEMA-356

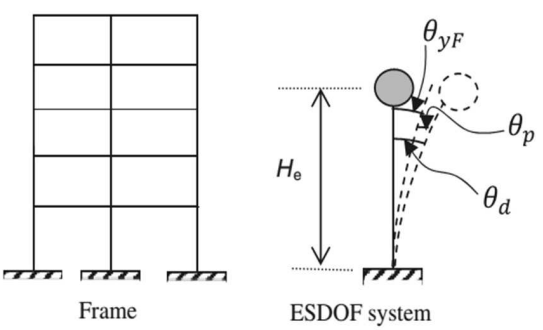


Figure 1. Equivalent SDOF system (Source: Choudhury and Singh)

UPBD method suggests that IO performance level can be combined with design drift from 1 to 1.5 %. It has been further assumed that the LS performance level can be combined with drift from 1.5 to 2.5 % and that the CP performance level can be combined with drift beyond 2.5 %. Eq. (3) can be used to express ratio of beam depth to beam length (h_b/l_b) for various drifts and performance levels (i.e. beam plastic rotations)
The design objectives include interstorey drift limit and member performance level. After choosing the of design objectives, suitable beam depth is obtained. The column sizes are chosen by trial such that the column steel from demand imposed lies approximately from 3 to 4 % of gross section area of column. The other steps of design are vastly after DDBD method of Pettinga and Priestley. The following steps are taken in designing as furnished below:

1. The target design drift and performance level of the building are decided. The beam depths (from h_b/l_b ratio) are found out from Eq. (3). The beam width has been kept from one-third to half of beam depth as per general design practice

2. The column sizes are preliminarily adopted from experience. In final design stage the column sizes are so adjusted that the column steel from demand imposed is restricted from 3 to 4 % of column sectional area.
3. The ESDOF system properties are determined,

$$\Delta_d = \frac{\sum m_i \Delta_i^2}{\sum m_i \Delta_i}, m_e = \frac{\sum m_i \Delta_i}{\Delta_d}, H_e = \frac{\sum m_i \Delta_i h_i}{\sum m_i \Delta_i} \quad (4)$$

Here m_i, h_i and Δ_i are respectively the mass, height from base and displacement for *i*-th storey, Δ_d is target (spectral) displacement, $m_e$is equivalent mass, He is the effective height of the ESDOF system

4. The displacement spectra corresponding to design acceleration spectra are generated for various damping. This is the specified hazard level for design. Here, EC-8 design spectra of 0.45g level is used for type B soil.
5. The damping in the system is computed from ductility as given below. The yield displacement (Δ_y) of ESDOF system, frame ductility (μ) and equivalent effective damping (ξ) in the system is obtained from the following equations,

$$\Delta_y = \theta_{yF} H_e, \mu = \frac{\Delta_d}{\Delta_y}, \xi = 5 + 120\left(\frac{1-\mu^{-0.5}}{\pi}\right)\% \quad (5)$$

6. The design base shear is computed as detailed below. The effective time period (Te) is obtained from displacement spectra corresponding to the curve for equivalent damping ξ and the value of target displacement Δ_d. Effective stiffness for ESDOF (K_e) system and base shear (V_b) is given as follows

$$K_e = 4\pi^2 \frac{m_e}{T_e^2}, V_b = K_e \Delta_d \quad (6)$$

7. The base shear is now distributed at floor levels as per the equation given below. If the building is more than 10-storey high, then to take into account the effect of higher modes, F_t is typically 10 % of base shear put at roof level.

$$F_i = V_b \frac{\Delta_i m_i}{\sum \Delta_i m_i}, \; F_i = F_t + V_b \frac{\Delta_i m_i}{\sum \Delta_i m_i} \quad (7)$$

8. The design is done with expected (mean) strengths of materials. As per FEMA-356 provisions the expected strength of concrete is 1.5 times of the 28-days characteristic strength and, that for steel is 1.25 times the yield strength of rebar. The load combinations are as below:

$$D+L,\ D+L\pm F_x,\ D+L\pm F_y$$

here, *D* is the dead load, *L* is the live load, *Fx* and *Fy* are storey forces in mutually perpendicular directions.
9. After designing, the performance of the building is checked through Non-linear analysis.

Modelling Aspects

Two reinforced concrete frame buildings have been modeled using SAP2000 v.14.0.0. The building frames have been assumed to stand on fixed supports. The floor slabs have been modelled as rigid diaphragms. As the stiffness is proportional to strength, the effective stiffness values of the members is evaluated after design stage and incorporated in the model for nonlinear analyses. The yield moments for beam sections are obtained from SAP2000 corresponding to design steel. The effective beam flexural rigidity is given below, in which E is modulus of elasticity of concrete, I_{eff},beam is effective moment of inertia of beam, M_{yb} is beam yield moment and ϕ_{by} is yield curvature of beam. Effective flexural rigidity of column is given below, in which I_{eff},column is effective moment of inertia of column, M_{cy} is column moment capacity. M_{cy} is read from interaction diagram of column constructed at expected strength level, corresponding to gravity axial load in the column. Yield curvature of column ϕ_{yc} is given below, in which h_c is column depth in the direction of earthquake under consideration.

$EI_{\text{eff,beam}} = \frac{M_{by}}{\phi_{by}}$, $\phi_{\text{by}} = 1.7\frac{\varepsilon_y}{h_b}$, $EI_{\text{eff,column}} = \frac{M_{cy}}{\phi_{cy}}$, $\phi_{\text{cy}} = 2.1\frac{\varepsilon_y}{h_c}$

The default hinge properties available in SAP2000 v.14.0.0 has been used for both column and beam members. Consequently the post-elastic force deformation behaviour for the members has been adopted in the modelling as per FEMA-356 (2000).

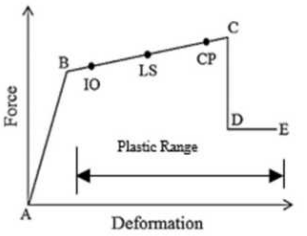


Figure 2. Typical force-deformation behaviour as per FEMA-356 (2000)

Building description

The RC frame buildings have been designed as per UPBD method. Two different type of plans are used.P1 denotes plan 1 and P2 denotes plan 2. The grade of concrete is taken as M25 and grade of steel as Fe500. The target objectives, member size considered and design parameters are given in the following tables.

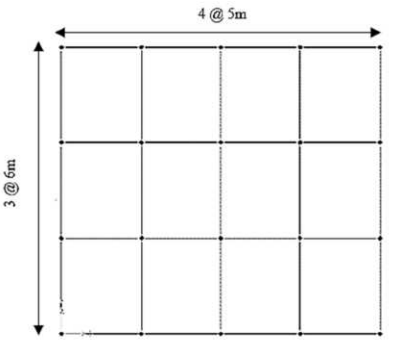


Figure 3. Plan 1

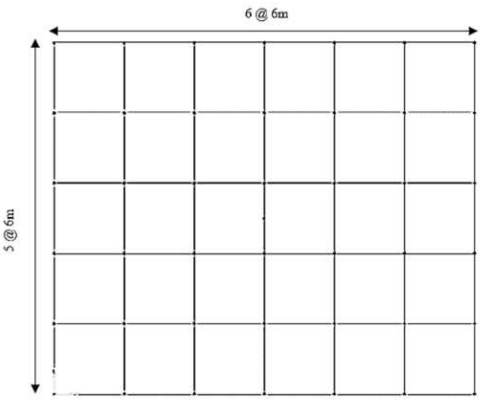


Figure 4. Plan 2

Table 1. Building name and design considerations

Building name	No. of storeys	Target performance objectives	
		θ_d	Performance level
P1_10_CP	10	3%	CP
P2_12_LS	12	2%	LS

Table 2. Member sizes of the building considered

Building name	Column size (mm x mm)	h_b/l_b		Beam size (mm x mm)	
		Upper bound	Lower bound	Long Direction	Short Direction
P1_10_CP	575 X 575 to 625 x625	0.13	0.04	450 x 250	540 x 300
P2_12_LS	675 x 675 to 925 x 925	0.14	0.06	750 x 400	750 x 400

Table 3. Design parameters for the building

Design parameters	Building considerd	
	P1_10_CP	**P2_12_LS**
Θd (%)	0.03	0.02
Δd (m)	0.5475	0.4316
Me (kg)	4259.69	16012.2
He (m)	21.94	25.99
μ	1.79	1.66
ξ (%)	14.7	13.55
Te (sec)	4.35	3.48
Ke	8887.08	52197.9
Vb (KN)	4865.85	22527.7

The base shear (Vb) obtained is then distributed through the storeys and the storey shear forces Fx and Fy are obtained and the building is designed. After designing performance check is done through non-linear analysis.

Analysis and results

Non linear static analysis (Pushover case)

Nonlinear static analysis is performed to check the performance of the designed buildings. Pushover analysis with lateral load patterns as mode proportional and uniform load have been used where the performance of the structure has been decided by the maximum value of response out of ATC 40 (1996), Capacity spectrum method, FEMA 356 (2000) DCM, FEMA 440 (2004) Equivalent linearization method and FEMA 440 (2004) Displacement modification method.

The perpormance point (PP) denotes the performance level of the building. The curve shown is for CP level buiding and after performing Non Linear static analysis, it is seen that the building is adequaltely designed to perform in CP level.Similarly, the other building has been designed to perform in the LS level.

Non linear dynamic analysis (Time history analysis)

Nonlinear time history analysis (NLTHA) has been carried out to determine the performances of the buildings. Time history method shall calculate building response at discrete time steps using discretized recorded or synthetic time histories as base motion.The NLTHA has been carried out with five spectrum compatible ground motions (SCGM) after EC-8 design spectrum at 0.45g level of seismicity. The arificial earthquales were generated using Kumar (2004) software.

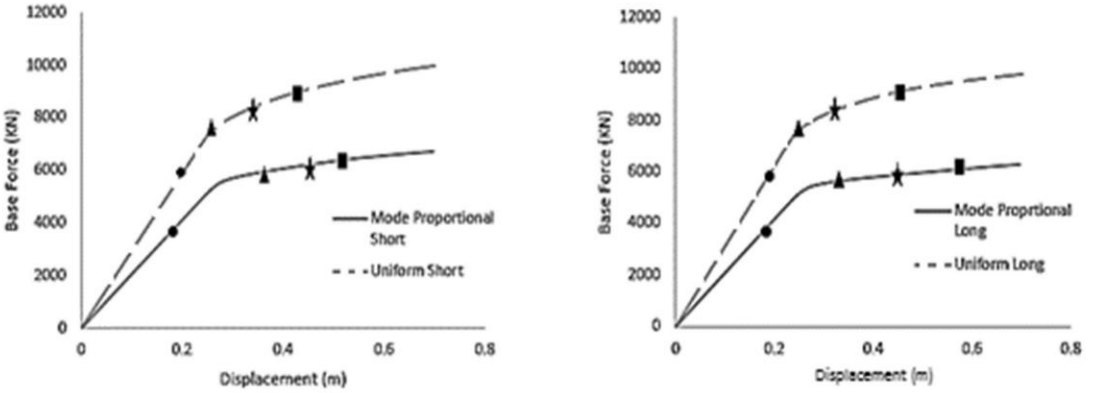


Figure 5. Typical pushover curves for the building **P1_10_CP** is shown for (a) short direction and (b) long direction

Table 4. Ground motions considered

Sl. No	Name	Background Earthquake	Direction	PGA	Duration (sec)
1	SCGM1	Duzce 1999	Duzce, 270 (ERD)	1.37	20.2
2	SCGM2	EL Centro 1940	N-S Component	1.18	31.8
3	SCGM3	Gazli 1976	Karakyr, 90	1.05	16.3
4	SCGM4	Kocaeli 1999	Sakarya, 90(ERD)	1.14	45
5	SCGM5	N.Palm Spring 1986	0920, USGS station 5070	1.64	20

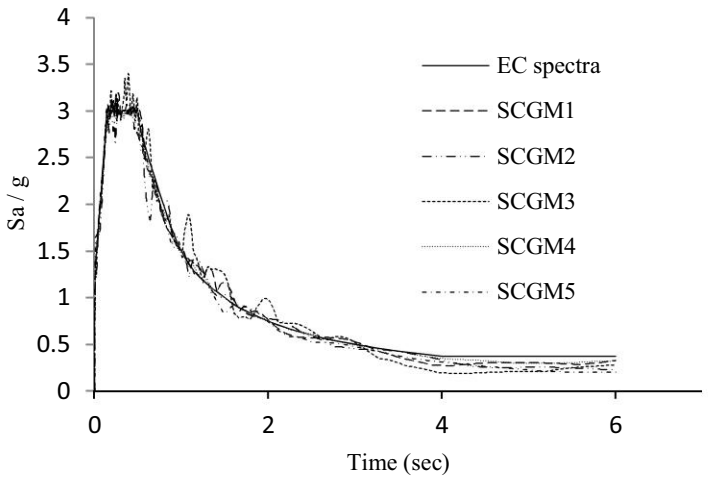


Figure 6. Match of EC 8 design spectrum with response spectra of SCGMs used for Non-Linear time history analysis

SAP 2000 v 14.0.0 software, the formation of plastic hinge in structural elements is shown as such that the pink color indicates Immediate Occupancy (IO) hinge, blue color represents LS hinge and cyan color represents CP hinge. At the end of time history analysis, structure P1_10_CP revealed formation cyan hinges which suggests it is in CP performance level. Similarly, for P2_12_LS blue hinges were developed making it a LS performing building.

The inter-story drifts (IDR) have been obtained under five SCGMs and the maximum drift has been reported. The dotted line represents the design drift for the building.

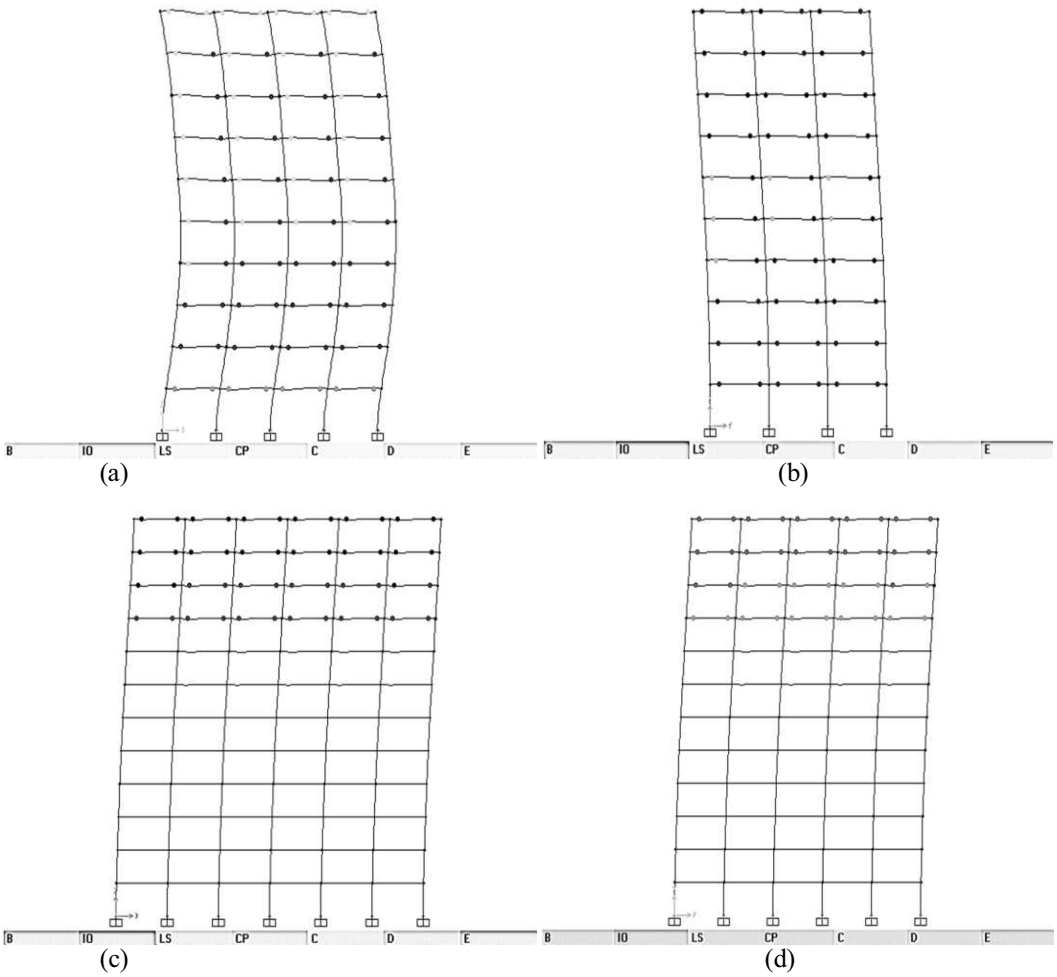
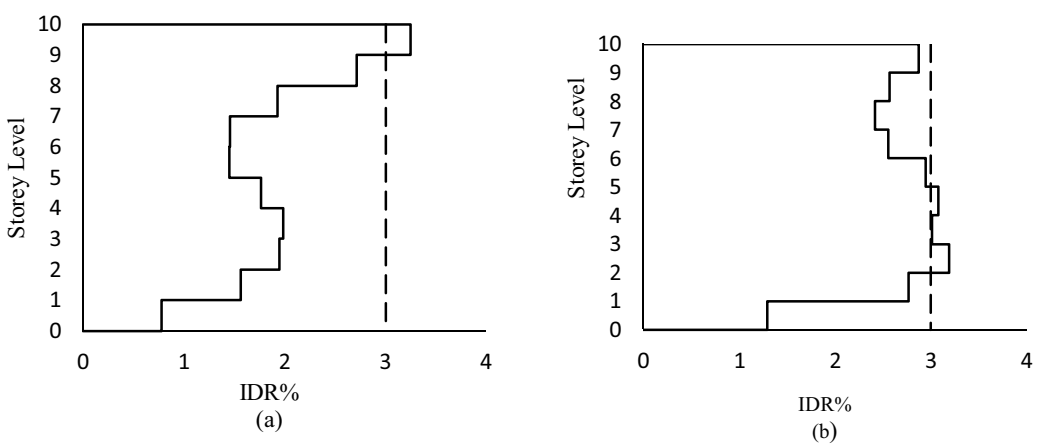


Figure 7. Typical hinge results for the buildings after time history analysis
(a) P1_10_CP_long, (b) P1_10_CP_short, (c) P2_12_LS_long, (d) P2_12_LS_short



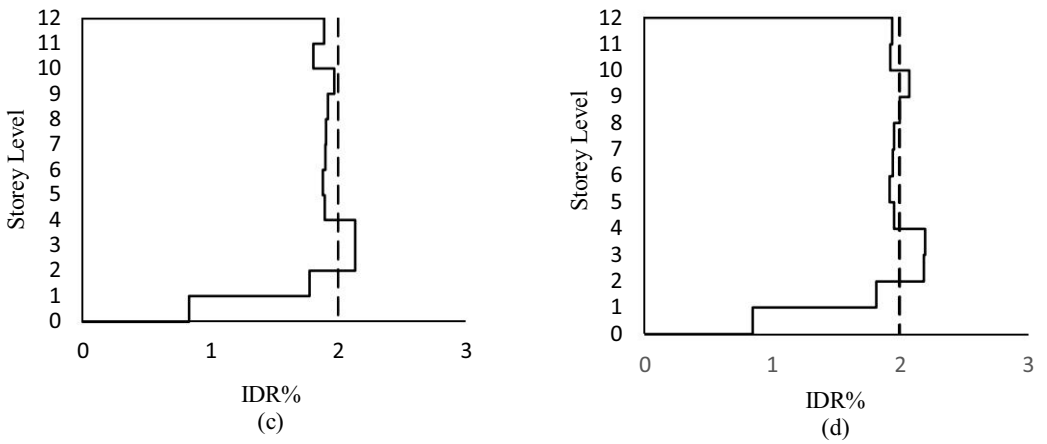


Figure 8. Typical inter storey drift ratio of the buildings after time history analysis (a) **P1_10_CP_long**, (b) P1_10_CP_short, (c) P2_12_LS_long, (d) P2_12_LS_short

The interstorey drift results have been shown above. The figures indicates that the building have achieved the target drift for which it is designed.

The base shear variation of the buildings under the SCGMs is reported below. Out of the five SCGMs considered, base shear variation with maximum peak amplitude is shown below.

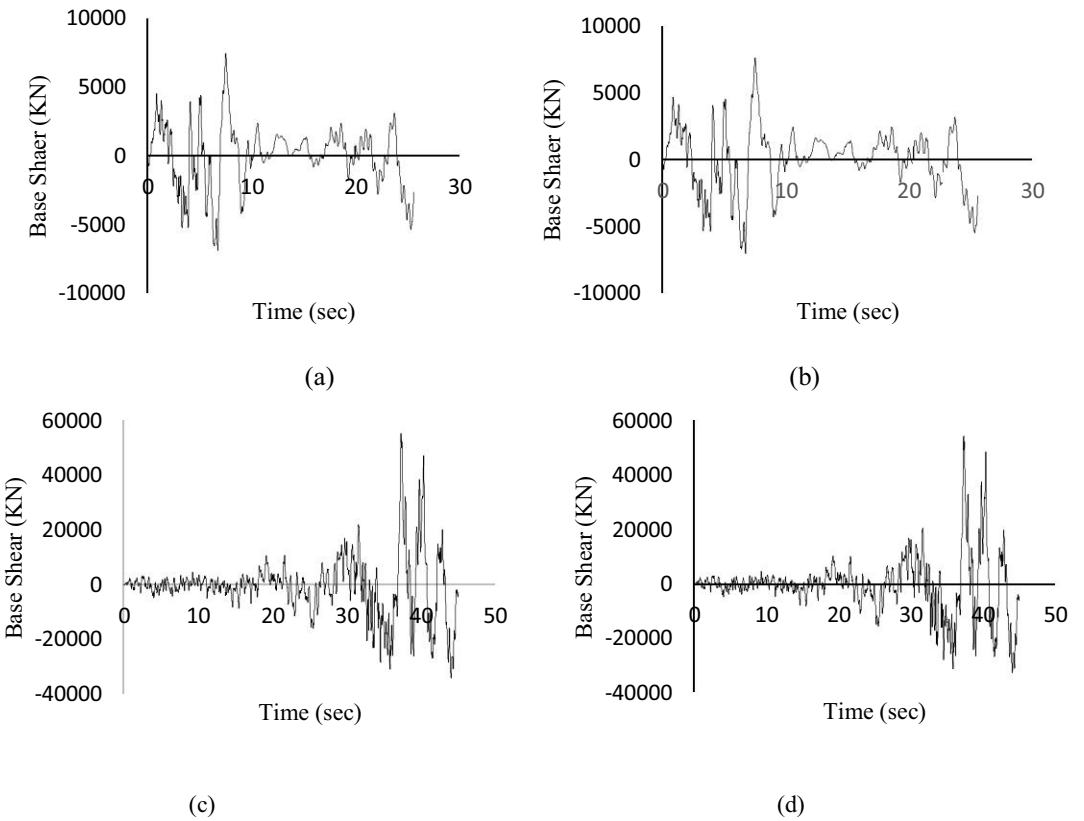


Figure 9. Base shear variation of the buildings after time history analysis (a) P1_10_CP_long, (b) P1_10_CP_short, (c) P2_12_LS_long, (d) P2_12_LS_short

From the base shear variation maximum amplitude of base shear is noted and the amplification of the storey shear forces from the design shear forces is found out. The difference between the shear forces between the design and time history results is shown below.

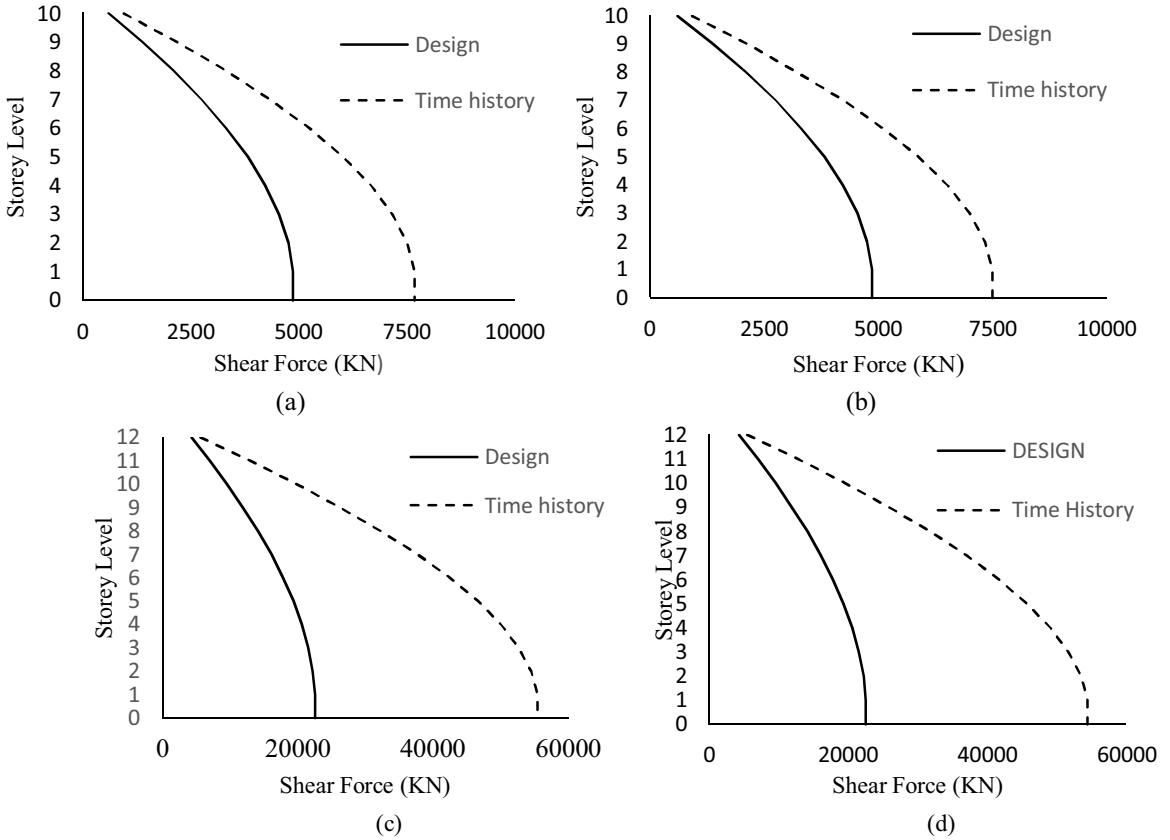


Figure 10. Difference in storey shear forces between design and time history analysis results (a) P1_10_CP_long, (b) P1_10_CP_short, (c) P2_12_LS_long, (d) P2_12_LS_short

Conclusion

The buildings as stated earlier has been designed with UPBD method to get the desired target performance objectives which is not possible with forced based approach i.e., Codal method. The main purpose of designing through UPBD method is to know the actual behaviour of the designed building which is checked by performing Non- linear analysis. The two buildings as mentioned earlier has been designed to their respective performance level i.e., the buildings have met the desired target performance objectives which is clear from Fig. 5, 7 and 8. The main study carried out here is to know the dynamic shear force amplification that occurs in the structure under the action of earthquakes. The maximum base shear variation of the structures under the five SCGM mentioned is noted. Storey shear forces are distributed with base shear results obtain after time history analysis. A comparison is made between the design and time history results. Fig.10 clearly depicts that there is significant amplification in shear force from the designed shear force. Proper amplification factors needs to be developed to accurately predict the amount of amplification by carrying out the study to many number of buildings.

References

[1] M. J. N. Priestley and M. J. Kowalasky, “Direct Displacement-Based Design of Concrete Buildings”, Bulletin of New Zealand National Society of Earthquake Engineering, Vol.33, No.4, 2000, pp. 421-444.

[2] J. D. Pettinga and M. J. N. Priestley, “Dynamic Behavior of Reinforced Concrete Frames Designed with Direct Displacement-Based Design”, Journal of Earthquake Engineering,2005,pp. 309-330

[3] M. J. N. Priestley, “Performance-Based Seismic Design”, Bulletin of the New Zealand Society for Earthquake Engineering, Vol.33, No.3,2000, 2000, pp.325-346.

[4] M. J. N. Priestley, “Myths and Fallacies of Earthquake Engineering”, The Mallet Milne Lecture, IUSS Press, Pavia, Italy, 2003.

[5] S. Choudhury, S. M. Singh, "A Unified Approach to Performance-Based Design of RC Frame Buildings", Journal of the Institution of Engineers (India)": Series ~ DOI I0.1007 /s40030- 013-0037-8, 2013.
[6] B.R. Rad and P. Adebar, "Dynamic Shear Amplification in High-Rise Concrete Walls: Effect of Multiple Flexural Hinges and Shear Cracking", The 14th World Conference on Earthquake Engineering, October 12-17, 2008, Beijing, China.
[7] "ATC-40: Seismic Evaluation and Retrofit of Concrete Buildings", Applied Technology Council, Redwood City, California, 1996.
[8] "FEMA-356: Prestandard and commentary on the seismic rehabilitation of Buildings", Federal Emergency Management Agency, Washington, D.C., 2000.
[9] "FEMA-440: Improvement in Nonlinear Static Seismic Analysis Procedures", Federal Emergency Management Agency, Washington, D.C., 2000.

Prediction of Compressive Strength for Different Curing Stages using Steepest Descent ANNs

K. Rajasekhar* and Gottapu Santosh Kumar**
* Assistant Professor, Department of Civil Engineering, Andhra University College of Engineering, Andhra University, Visakhapatnam-530003, India
** Research Scholar, Department of Civil Engineering, Andhra University College of Engineering, Andhra University, Visakhapatnam-530003, India
E-mail: kumar.santou@gmail.com

Abstract: A new ANNs training model is proposed to predict the compressive strength of Slurry Infiltrated Fibrous Concrete (SIFCON) with manufactured sand for various curing periods in this paper. All over the world, availability of natural sand is dwindling day by day due to its limited availability in the river beds. The alternative to this natural resource is the manufactured sand. The ANNs model is constructed, trained and tested for predicting the compressive strengths of various ages using Steepest Descent based weights updating method. The experimental data corresponding to compressive strength of the SIFCON specimens are obtained using the Universal Testing Machine (UTM) of capacity 300T. A total of 80 different SIFCON cube specimens of M20 grade mix using the manufactured sand is cast for different fibre fraction (8%, 10% and 12%) and tested for a different curing period such as 7, 28 and 56 days. In this paper, the compressive strength of different curing ages (14, 21, 35, 42, 49, 63, 70, 77, 84 and 91 days) is predicted for 8%, 10% and 12% fibre volume fraction. For the ANN training phase, different fibre configurations, curing of SIFCON cubes corresponding to various time periods, number of neurons, learning rate, momentum and activation functions were considered. The results show that the relative percentage error in the training set was 6.02% and the testing set was 15%. Research results demonstrate that the proposed Steepest Descent based ANN model is practical, gives high prediction accuracy and beneficial.

Keywords: Manufactured sand, Steepest Descent, ANN, compressive strength, SIFCON.

Introduction

Traditionally, concrete has been fabricated from a few well defined components: cement, water, fine aggregate, coarse aggregate, etc. In this paper, a new approach on concrete has been made, named as SIFCON, which is considered as a relatively new high performance concrete. SIFCON is a special type of steel fibre reinforced concrete. It differs from steel fibre reinforced concrete in two ways by mix and percent volume of fibres [1]. The matrix in SIFCON has no coarse aggregates, but a high cementitious content [2]. The percentage of fibres by volume can be anywhere from 4 to 20% even though the current practical range is from 4 to 12%. The proportions of cement and sand generally used for making SIFCON are 1:1, 1:1.5 or 1:2. The cement slurry made up of the stated proportions alone has some applications. The water cement ratio varies in-between 0.3 to 0.4. Percentage of super plasticizers vary from 2 to 5% by weight of cement [3].

The compressive strength of concrete is a major and important mechanical property, which is generally obtained by measuring concrete specimens after a standard curing of 28 days [4]. Conventional methods of predicting 28-day compressive strength of concrete are basically based upon statistical analysis by which many linear and nonlinear regression equations have been constructed to model such a prediction problem.

Conventional prediction models have been developed with a fixed equation form based on a limited number of data and parameters. If the new data is quite different from the original data, then the model should be updated to include its coefficients and also its equation form. ANNs do not need such a specific equation form. Instead of that, it needs sufficient input-output data. Also, it can continuously re-train the new data, so that it can conveniently adapt to the new data [5, 6]. ANN has been investigated to deal with problems involving incomplete or imprecise information (Noorzaei et al., 2007) [7]. Several authors have used ANNs in structural engineering. For example, Yeh (1998), Kasperkiewicz et al. (1995), Lai and Sera (1997) and Lee (2003) applied the NN for predicting properties of conventional concrete and high performance concretes [8-11]. Bai et al. (2003) developed neural network models that provide effective predictive capability with respect to the workability of concrete incorporating metakaolin (MK) and fly ash (FA) [12]. Guang and Zong (2000) proposed a method to predict 28-days compressive strength of concrete by using multilayer feed forward neural networks [13]. Dias and Pooliyadda (2001) used back propagation neural networks to predict the strength and slump of ready mixed concrete and high strength concrete, in which chemical admixtures and mineral additives were used [14]. In this paper, a feed forward and back

propagation ANN using the Steepest Descent method is used for the prediction of the SIFCON cube compressive strengths for different curing periods.

This paper is presented in the following three sections. It begins with the basic concepts of ANNs and implementation of Steepest Descent Algorithm based ANN in Section 2. Section 3 shows the test results and discussion on the predicted compressive strengths for different curing ages due to *Steepest Descent algorithms based* ANNs and the paper's conclusions in section 4.

Implementation of Steepest Descent algorithm for neural network training process

An Artificial Neural Network (ANN) is an interconnection of processing elements called neurons. Neurons are the basic computing elements that perform data processing in a network. The neurons incrementally learn from their experimental data to capture the linear and non linear trends in the complex data, so that it provides reliable predictions for new situations containing even partial information. The processing ability of the network is stored in the interneuron connection strengths called parameters or weights. The weights are acquired by adaptation process.

The Steepest Descent algorithm which is also known as an error back propagation algorithm is the first neural network training algorithm and is still widely used today. The implementation of neural network training using the Steepest Descent algorithm involves two steps: i) Calculation of observation or Jacobian matrix and ii) training process design.

The training process is designed using the following steps:

i) Estimate the error with the initial weights that are randomly generated

ii) Do an update using the following objective function, f (x), to adjust the weights.

$$f(x) = \frac{1}{2}\sum_{i=1}^{m} r_j^2(x) \qquad (1)$$

Where, r_j are called residuals and r_j = target value - ANN output

Each residual (r_j) represents the difference between computed quantities and their measured counterpart, and therefore represent the function of sought parameters

iii) With the new weights, estimate the mean squared error (MSE)

iv) Do an update using the following update equation of Steepest Descent algorithm to adjust the weights

$$[X_{K+1}] = [X_K] - \alpha_k [J^T R] \qquad (2)$$

Where, α_k is the learning constant (step size) and X is the vector consisting of x_1, x_2.....x_n.

R is the residual vector ($\hat{R}(x)) = [r_1(x), r_2(x), ..., r_m(x)]^T$

'g' is the gradient of the objective function.

K is the instant of time or epoch time

$\hat{J}$ is the Jacobian matrix which, in the case when $\hat{x}$ is n-dimensional vector will be m × n matrix,

v) With the new weights evaluate the mean square error

vi) If the current MSE is increased as a result of the update, then retract the step such as reset the weight vector to the previous value and decrease combination coefficient, α_k by a factor of 10 or by the same factor as iv.

vii) Go to step ii with new weights until the current mean square error is smaller than the required value.

Test Results and Discussions

The compressive strengths of the three specimens of different fibre configuration (8%, 10% and 12%) are tested using the UTM of capacity 300T in the laboratory (Fig.1). The compressive strength of the one each of average of the 3 specimens of SIFCON cube specimens referred to as S8 (SIFCON-MS 8%), S10 (SIFCON-MS 10%), and S12 (SIFCON-MS 12%) for 7 days, 28 days and 56 days are presented in the Table 1. Table 1 also illustrates the percent of increase in compressive strengths (SIFCON-MS 8%, SIFCON-MS 10% and SIFCON-MS 12%) of 28 days and 56 days when compared with 7 days.

From Fig.2, it can be seen that 12% fibre fraction has given the optimum strength volume. For 7 days, the compressive strength of the specimens of 12% fibre fraction is 11% more than that of the 8% fibre fraction and more than that of 1% more strength than that of the 10% fibre fraction. For 28 days, the compressive strength of the specimens of 12% fibre fraction is 17% more than 8% fibre fraction and 6% more strength than 10% fibre fraction. For 56 days, the compressive strength of the specimens of 12% fibre fraction is 21% more than 8% fibre fraction and 14% more strength than 10% fibre fraction.

Table 2 illustrates the predicted compressive strengths of the specimens obtained due to the proposed Steepest Descent based ANN (1-2-15-1 design). The proposed ANN consists of one input, 2 hidden layers with 15 neurons each and one output. The compressive strength is predicted for 14, 21, 35, 42, 49, 63, 70, 77, 84 and 91 days from the available compressive strength of 7, 28 and 56 days, which are obtained from testing the cube specimens in UTM of 300T in laboratory. From the Table 2 it is seen that, the predicted values are 88% accurate. The compressive strength obtained has gradually increased as the curing

period increases, an increase in strength is seen from 7 to 63 days and the compressive strength has seen to be stabilized after 63 days that is from 70 to 91 days.

Fig 1. Test procedure of the SIFCON specimens

Table 1. Values of compressive strength of SIFCON specimens obtained in the laboratory

S.No.	Mix	Compressive strength (MPa)			% increase or decrease in Compressive strength compared with 7 days		
		7 days	28 days	56 days	7 days	28 days	56 days
1	SIFCON-MS 8%	29.30	35.71	43.70	-	21.87	49.18
2	SIFCON-MS 10%	35.19	42.54	47.11	-	20.88	33.87
3	SIFCON-MS 12%	43.85	46.15	53.63	-	5.24	22.30

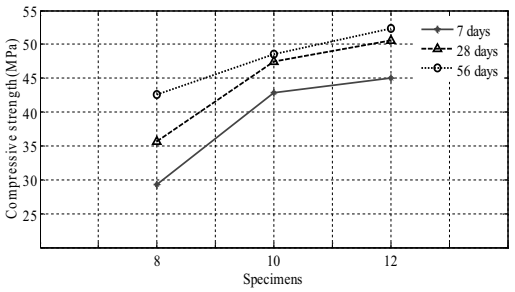


Fig 2. Variation of Compressive strength of SIFCON with fibre fraction

Table 2. Values of Predicted compressive strengths for different curing ages

Period of Curing (Days)	Compressive Strength of SIFCON Specimens for Percent Steel Fibers(MPa)		
	8%	10%	12%
7	26.6	35.19	43.85
14	29.43	39.71	44.76
21	32.47	41.53	45.83
28	35.56	42.54	46.14
35	40.92	42.68	46.14
42	41.69	43.84	46.37
49	42.93	44.11	48.32
56	45.78	47.11	53.63
63	46.89	51.82	56.62
70	46.89	51.85	56.62
77	46.89	52.94	57.16
84	46.89	52.94	57.18
91	49.69	53.19	57.18

Conclusions

The experimental investigation of the compressive strength of SIFCON with manufactured sand for various curing ages was carried out and the compressive strengths of concrete specimens are predicted using the Steepest Descent ***based ANN*** and the following conclusions are made.

- From the study, it can be stated that addition of steel fibres in SIFCON significantly increased the Compressive strength.
- In comparison with 8%, 10% and 12% steel fibres addition in concrete, 12% fibre addition showed the optimum value in compressive strength for all the curing periods (7, 28 and 56 days).
- The compressive strength is predicted for 14, 21, 35, 42, 49, 63, 70, 77, 84 and 91 days from the available compressive strengths of 7, 28 and 56 days, which are obtained from testing the cube specimens in UTM of 300T in laboratory, for 8%, 10% and 12% fibre volume fraction.
- The compressive strength obtained has gradually increased with the curing period i.e an increase in strength is seen from 7 to 63 days and compressive strength has got stabilized after 63 days curing period.
- The predicted values are 88% more accurate when compared to experimental results. Hence, it is concluded that this ANN can be used in the complex concrete structures design such as high raised buildings, dam, bridges etc.,

References

[1] P Thamilselvi and P Devadas Manoharan, "Study of Behaviour of Reinforced Concrete Frames with Slab Using SIFCON in Beams Column Joints" Journal of Engineering and Applied Sciences, Vol. 3 issue 1, pp. 104-108, 2008.

[2] Vinayaka N M, "An Experimental Investigation on SIFCON using Mono Fibres", International Journal of Civil and Structural Engineering Research, vol. 3 issue 1, pp. 267-272,April- September 2015.

[3] Jong-In Kim; Doo Kie Kim; Maria Q. Feng and Frank Yazdani, "Application of Neural Networks for Estimation of Concrete Strength" Journal of Materials in Civil Engineering, Vol. 16, No. 3, June 1, 2004.

[4] Arun Aniyan Thomas and Jeena Mathews, "Strength and Behaviour of SIFCON with Different Types of Fibres", International Journal of Civil Engineering and Technology (IJCIET) vol. 5, issue 12, pp. 25-30, December 2014.

[5] Seyed Jamalaldin Seyed Hakim, "Application of Artificial Neural Network to Perdict Compressive Strength of High Strength Concrete", International Journal of the Physical Sciences Vol.6 Issue 5, pp. 975-981, 4th March 2011.

[6] P S Lande and A S Gadewar, "Application of Artificial Neural Network in Prediction of Compressive Strength of Concrete by Using Ultrasonic Pulse Velocities", Journal of Mechanical and Civil Engineering (IOSR-JMCE) Vol. 3 Issue 1, pp.34-42, September-October 2012.

[7] Noorzaei J, Hakim SJS, Jaafar MS, Abang AAA, Thanon WAM, "An optimal architecture of artificial neural network for predicting compressive strength of concrete", Jornal of Indian Concrete, 81 (8): 17-24, 2007.

[8] Yeh IC, "Modelling of strength of HPC using ANN". Cem. Concrete Res. J., 28(12): 1797-1808, 1998.

[9] Kasperkiewics J, Racz J, Dubrawski A, "HPC strength prediction using ANN", ASCE. J. Comp. Civil Eng., 4: 279-284, 1995.

[10] Lai S, Sera M, "Concrete strength prediction by means of neura network", Journal of Construction Building Material, 11(2): 93-98, 1997.

[11] Lee SC, "Prediction of concrete strength using artificia neural networks", J. Eng. Struct., 25: 849-857, 2003.

[12] Bai J, Wild S, Ware JA, Sabir BB, "Using neural networks to predict workability of concrete incorporating metakolin and flyash", J. Adv. Eng. Software, 34: 663-669, 2003.

[13] Guang NH, Zong WJ, "Prediction of compreiive strength of concrete by neural networks". J. Cem. Concrete Comp., 16: 287-298, 2000.

[14] Dias Wps, Pooliyadda SP, "Neural networks for predicting properties of concrete with admixtures". J. Constr. Build. Mater., 15: 371-379, 2001.

Study of Factors Affecting Labour Efficiency at Construction Sites

Purnima Bajpai*, Megha Kalra** and Garima Srivastava***
*Assistant Professor, The Northcap University, Gurgaon
purnimabajpai@ncuindia.edu
**Assistant Professor, The Northcap University, Gurgaon
meghakalra@ncuindia.edu
***Assistant Professor, The Northcap University, Gurgaon
garimasrivastava@ncuindia.edu

Abstract: Construction is a labour intensive industry and the labour productivity is considered as a direct measure of their efficiency. A higher productivity can help to ensure better competitiveness in the industry which in turn may help to meet the project goals and values. The study aims at identifying and analyzing the key factors which are likely to affect the labour efficiency at construction sites and also analyzing their impact. The study also aims at providing suggestions in order to reduce the variation in labour productivity. A questionnaire survey approach has been used for data collection and multi variate regression has been used to analyze the major factors affecting labour productivity and the results indicate six main factors which significantly impact on the labour efficiency at construction sites.

Keywords: Construction sites, Labour Productivity, Multiple variate regression, efficiency.

Introduction

Productivity is an average measure of the efficiency of production. It can be expressed as the ratio of output to inputs used in the production process

The Construction industry is labour intensive and labour is one of the major and the most complex of all resources need at the Construction sites for execution of the work. Labour costs account for almost 30 % to 55 % of overall project cost. In general labour productivity is usually expressed as output per hour.

Among the various factors affecting productivity at construction sites. Hence it becomes extremely important to study the factors that are likely to affect the labour productivity at site and further there is also a need to quantify the affect of these factors so that inefficiencies can be reduced and productivity can be maximized. The study aims at quantifying these factors and to generate a model for predicting labour productivity and efficiency.

Objectives of The Study

The study is focused on identifying the key attributes affecting labour productivity at construction site and their affect on the labour efficiency and further model the impact of these factors through multi variate regression so that appropriate recommendations can be made in to reduce the variations of labour productivity in construction projects.

Literature Review

Labour is one of the most important resources in the construction industry and accounts for majority of the cost. The quantity and quality of labour available for performing a task in construction is much more susceptible to the influence of management than materials or capital. However, it is extremely important to note that the labour productivity is a measure of the overall effectiveness of the system in effectively utilizing labour, equipment and capital so as to convert the labour efforts into useful output.

25 critical factors were identified by Homyun Jang et al (2009) which were grouped into 4 main heads, mainly management of work at sitet, work technique, characteristics of work and worker component.A similar study carried out by Durdyev and Mbachu (2011) identified 56 variables which affected the labour productivity, and further categorized them into 8 factor heads.

Enshassi (2007) classified the attributes affecting labour productivity and efficiency in the construction into 10 main groups, namely: characteristics of internal workforce, leadership, work motivation time, materials and equipment, supervision, project characteristic, site security, quality and external factors.113 variables were identified by Soekiman et al (2011) which effect the labour productivity sites these were grouped under 15 heads according to their characteristics, namely:, execution ,

design, equipment & material , health and safety, supervision, working times, project factors, quality of work and environment, financial factors, leadership and coordination, organization structure, and many external factors.
Nabil Ailabouni et al (2007) highlighted 32 significant factors affecting the productivity in the construction industry and were then grouped into 4 groups based on their common characteristic, namely: environmental factors, organization related factors,dynamics of the group and personal factors.
On the basis of Literature review, 34 factors have been selected and further based on their characteristics these have been grouped into 8 heads namely: Manpower, Managerial, Environmental, safety, Material/Equipment, Schedule, motivation, and Quality
The **first** Factor considered is MANPOWER characteristics and the variables studied under this are lack of skills, increase in the age, labour absenteeism, lack of training and personal problems of the labour. The **Second** Factor under consideration is MANAGERIAL factors this includes poor site management, poor communication at site, misunderstanding and misrepresentation of information between labour and supervisors, and lack of periodic meetings with labour. Further the **third** factor is ENVIRONMENTRAL FACTORS , these include variables like weather conditions, location of project, working in confined places, and size of the project site. The **fourth** factor is MOTIVATION which includes the study of variables like delays in payments, facilities of transportation to and from site, financial motivation system, availability of appropriate spaces for eating at relaxing at site. Etc.
Further the fifth and a very important factor was that of MATERIAL AND EQUIPMENT which mainly and discussion on variables like material shortages at site, lack of proper storage facilities, old and inefficient machines and shortages of tools and equipments.
The last three factors are very important namely SCHEDULE, QUALITY AND SAFETY which include parameters like Working days and hours, holidays, schedule ,safety precautions, first aid,accidents at site, availability of safety engineer at site, sufficient lighting, quality of raw materials, quality of food available, quality of work required, and quality inspections etc

Hypothetical Model

The hypothetical model consists of eight hypotheses from H1 to H8, the factors are the independent variables which affect labour productivity directly, a dependent variable here is the labour productivity wheras the independent variables under this study are Manpower related factors, managerial factors, environmental factors, motivational factors, materials and equipments used, safety parameters and quality issues.

Research Methodology

The study proposes to use a questionnaire survey approach to identify the key factors affecting the labour productivity at construction sites. The questionnaire consisted of 56 questions which were circulated to different construction sites and the panel consisted of respondents from various levels of the organization from project managers, site supervisors, labour, quality inspectors, etc.A reliability analysis was then carried out in order to test the reliability of the survey data collected
Cronbach's alpha is used as the most common method to measure reliability. It is most commonly used when we have multiple likert scale type of questions in a survey/questionnaire. As the number of variables is large we need to narrow down our data for the multivariate regression, hence a factor analysis has been carried out on the 56 variables and which were ultimately grouped into 8 main factors. Then a multivariate regression is carried out on the 8 factors which are believed to have a direct impact on the labour productivity
A questionnaire shall be developed to identify the major factors affecting the productivity of labour at construction site. The questions formed would be to produce responses for the factors identified from literature survey and expert interviews. Some questions would also be included on the Initiation, Monitoring & Controlling and Closing process areas of these projects. Obtaining “high-quality” responses with adequate disclosures would be a key task and the researcher would work in collaboration with project stakeholders to achieve this. The target is to obtain at least 30% responses would be targeted with reasonable number of responses from three identified Organizations.
Factor analysis is carried out in order to reduce the number of variables under common factors and make the data handling a easier process .The initial classification was done purely on the basis of literature review (and expert opinion from industry experts.
A reliability analysis was then conducted on the factors for each classified component so as to verify the reliability of the components. A multi variate regression analysis was then carried out on the 8 factors This was the last stage of the statistical analysis in which the factors that were found to be significant in affecting the labour productivity at construction sites were identified and modelled.

Results of Data Analysis and Discussion

60 questionnaires were distributed to different construction firms in the population. Out of these, 53 questionnaires were successfully filled and returned. The response rate obtained was 88.33%.

Reliability Analysis

Cronbach's alpha approach has been used for carrying out reliability analysis on the responses obtained through the questionnaire survey. This method is used to assess the basic consistency on the basis of the average correlation between the data that was measured in an identical manner.

The value of cronbach's alpha and total correlation must lie between 0.6 and 0.3 respectively. Three factors were deleted due to low correlation after considering the result of the first reliability analysis,

Table 1. Results for Reliability Studies

Sr No	GROUP	CRONBACH'S ALPHA VALUE
1	Manpower issues	0.6935
2	Managerial issues	0.663
3	Environmental concerns	0.626
4	Motivation factors	0.665
5	Material & equipment	0.747
6	Schedule aspects	0.778
7	Safety issues	0.672
8	Quality concerns	0.645

Factor Analysis

The aim behind carrying out a factor analysis is to group larger number of related variables into factor heads and also for data reduction Factor analysis is one of the most commonly used statistical technique which explains the common dimension that compose variables (called components) by analyzing the existing correlation between them.

In this study 34 independent variables were identified which affect the labour productivity on construction sites. After the reliability analysis three factors were eliminated due to low total correlation.

The Kaiser Meyer Olkin value obtained was 0.59 which is greater than 0.5, which indicates that the data will factor well and in the Bartlett's test the significant value obtained is lower than 0.05, which clearly indicated that the correlation matrix is different from an identity matrix.

Further the eigen value which is an index that represents the explanatory power of the components, and is extracted from the number of the components that have a value of 1 or more. From the analysis there are seven components whose initial eigen values are found to be greater than one and these components were extracted during analysis. These seven components were able to explain 65.822% of the information contained in the original factors.

Multi Variate Regression Analysis

The coefficient of determination denoted by R^2 came out to be to 0.822. This indicates that factors considered under the study are able to explain 82.2 percent of the total variations in labour productivity at sites. The P value obtained of 0.000 implies that the labour productivity is significant at the 5 percent level of significance.

Table V indicates that about 82.2% of the total variation in the labour productivity can be explained by the seven independent variables considered under this study. This indicates that the model is adequate in the prediction.

A F value of 35.38 in the table VII below indicates that the overall regression model is significant and has some explanatory value (P- value $p=0.00<0.05$). This also indicates that there exists a significant relationship between the predictor variables.

LPF = -3.104 + 0.233F1 + 0.199F2 + 0.344F4 + 0.225F5 + 0.124F6 + 0.218F7+0.168F8

(F1: manpower related group, F2: managerial factors group, F4: motivational factors group, F5: material/equipment group, F6: schedule group, F7: safety group, F8: quality group)

Table 2. Total Variance

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.615	21.597	21.597	5.615	21.597	21.597	3.058	11.762	11.762
2	2.751	10.583	32.180	2.751	10.583	32.180	2.856	10.984	22.746
3	2.125	8.173	40.352	2.125	8.173	40.352	2.752	10.583	33.329
4	1.977	7.604	47.956	1.977	7.604	47.956	2.223	8.551	41.881
5	1.803	6.935	54.892	1.803	6.935	54.892	2.152	8.278	50.159
6	1.550	5.961	60.853	1.550	5.961	60.853	2.151	8.275	58.434
7	1.292	4.970	65.822	1.292	4.970	65.822	1.921	7.389	65.822
8	.970	3.730	69.553						
9	.902	3.469	73.021						
10	.811	3.121	76.143						
11	.757	2.911	79.054						
12	.710	2.732	81.786						
13	.625	2.406	84.192						
14	.585	2.248	86.440						
15	.528	2.032	88.472						
16	.473	1.818	90.290						
17	.449	1.726	92.016						
18	.377	1.451	93.467						
19	.354	1.362	94.830						
20	.339	1.305	96.135						
21	.292	1.125	97.260						
22	.203	.780	98.040						
23	.186	.715	98.755						
24	.149	.571	99.326						
25	.105	.405	99.731						
26	.070	.269	100.000						

Extraction Method: Principal Component Analysis.

Table 3. Summary for Multivariate Linear Regression Analysis

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.920[a]	.846	.822	.415	.846	35.380	7	45	.000

a. Predictors: (Constant), Quality, Material, Managerial, Safety, Manpower, Motivation, Schedule

Table 4. Anova Table for Multivariate Linear Regression Analysis

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	42.567	7	6.081	35.380	.000[b]
	Residual	7.734	45	.172		
	Total	50.302	52			

a. Dependent Variable: LPV

b. Predictors: (Constant), Quality, Material, Managerial, Safety, Manpower, Motivation, Schedule

Table 5. Coefficient of Multiple Linear Regression Analysis

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.105	.455		-6.826	.000
	Manpower	.328	.102	.233	3.205	.002
	Managerial	.291	.094	.199	3.111	.003
	Motivation	.489	.101	.344	4.825	.000
	Material	.305	.094	.225	3.263	.002
	Schedule	.169	.104	.124	1.631	.110
	Safety	.275	.081	.218	3.398	.001
	Quality	.312	.116	.168	2.693	.010

a. Dependent Variable: LPV

Table 6. Model Summary

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.920[a]	.846	.822	.415	.846	35.380	7	45	.000

a. Predictors: (Constant), Quality, Material, Managerial, Safety, Manpower, Motivation, Schedule

Table 7. Hypothesis Summary

HYPOTHESIS	GROUP	T	Sig	Decision
H1	Manpower	3.205	.002	Reject hypothesis H0
H2	Managerial	3.111	.003	Reject hypothesis H0
H4	Motivation	4.825	.000	Reject hypothesis H0
H5	Material/Equipment	3.263	.002	Reject hypothesis H0
H6	Schedule	1.631	.110	Accept Hypothesis H0
H7	Safety	3.398	.001	Reject hypothesis H0
H8	Quality	2.693	.010	Reject hypothesis H0

Conclusion and Recommendations

The results obtained for the multivariate regression analysis resulted in six independent groups, which have a positive impact on labour productivity and efficiency, and one group, which had no significant impact. The Multiple linear regression analysis indicated that the model generated explains 82.2% of the variance in labour productivity and efficiency.

Among the six independent groups, the motivational factors group has obtained the highest score indicating maximum impact on labour productivity.

Hence through this study it be concluded that the motivational factors have the highest impact on labour productivity which has the highest standardized beta value of 0.344. It can also be seen from the study that the low labour satisfaction possibly has a negative impact on labour productivity and morale. Hence it is important that the construction companies increase labour satisfaction by incorporating different factors in their policy like paying a reasonable salary, and developing financial reward programs and also by improving the living condition at construction site.

Also it is highlighted through the study that the issues related to manpower play an important role during the project execution. A bad workforce is a major problem and affects the schedule of the project. Hence it is important to improve the labour commitment and the relationship among workers by increasing labour benefit and also by team building programs. Lastly in order to reduce the variations in labour productivity, the construction companies must take steps to improve supervision conduct periodic meetings at site and ensure proper selection procedure for supervisors.

References

[1] Allmon, E., Haas, C.T, Borcherding, J.D., Allmon, E. and Goodrum, P.M. (2000), ―US construction labor productivity trends, 1970-1998‖, Journal of Construction Engineering and Management, Vol. 126 No. 2, pp. 97-104.

[2] AbouRizk, S., Knowles, P. and Hermann, U. (2001), ―Estimating labor production rates for industrial construction activities‖, Journal of Construction Engineering and Management, Vol. 127 No. 6, pp. 502-11.

[3] Bernstein, H.M. (2003), ―Measuring productivity: an industry challenge‖, Civil Engineering, Vol. 73 No. 12, pp. 46-53.

[4] Building Futures Council (2006), Measuring Productivity and Evaluating Innovation in theUS Construction Industry, BFC, Alexandria, VA.

[5] Dai, J., Goodrum, P.M. and Maloney, W.F. (2009), ―Construction craft workers‘ perceptions of the factors affecting their productivity‖, Journal of Construction Engineering and Management, Vol. 135 No. 3, pp. 217-26.

[6] Dissanayake, M., Fayek, R.A., Russell, A.D. and Pedrycz, W. (2005), ―A hybrid neural network for predicting construction labour productivity‖, Proceeding of ASCE International Conference on Computing in Civil Engineering, 12-15 July, Cancun, Mexico.

[7] Durdyev, S., & Mbachu, J. (2011). On-site Labour Productivity of New Zealand Construction Industry: Key Constraints and Improvement Measures. Australasian Journal of Construction Economic and Building, 1(3), 18-33.

[8] Enshassi, A., Mohamed, A., Mustafa, Z. A., & Mayer, P. E. (2007). Factors affecting labour productivity in building project in the Gaza strip. Journal of Civil Engineering and Management, 8(4), 245-254.

[9] Henry, M. J., Jackson, A. M., Bengt, H. (2007). Factors affecting the productivity of building craftmen – studies of Uganda. Journal of Civil Engineering and Management, 8(3), 169-176.

[10] Homyun Jang, Kyonghoom Kim, Juhyung Kim, and Jaejun Kim. (2011). Labour productivity model for reinforced concrete construction projects. Construction Innovationm Process, Management, 11(1), 92-113.

Comparative Analysis of Motif Discovery Methods in Time Series

Sukriti Parnami* and Veenu Mangat**
* M.E. IT Research Scholar, UIET, Panjab University, Chandigarh
sukritiparnami@yahoo.com
** M.E. Supervisor, Assistant Professor, IT, UIET, Panjab University, Chandigarh
veenumangat@yahoo.com

Abstract: Due to wide use of Information Technology, substantial amount of information is being gathered for exploratory analyses, business operations and online networking in the big data era. Due to the gathering of information for several events at distinctive time periods, huge datasets are formed. A time series can be defined as a series of numeric values obtained at various points, occurring after regular intervals. An interesting research problem in time series is motif discovery. Motif discovery subroutine can be utilized as a part of algorithms for classification and summarization. In this paper, comparative analysis of various motif discovery methods in time series datasets has been done and a method for improving the same has been suggested.

Keywords: Data mining, motifs, match, subsequences, time series.

Introduction

Due to wide use of Information Technology, substantial amount of information is being gathered for exploratory analyses, business operations and online networking in the big data era. Due to the gathering of information for events at distinctive time periods, huge datasets are formed. A time series can be defined as a series of numeric values obtained at various points, occurring after regular intervals. E.g.: time series sequence may represent network flow, exchange rates or weather conditions over time. The data is substantial in size, has high dimensionality and needs to be updated after continuous period of time.

A major reason for representation of time series data is to minimize the dimension of original data. The most common method is to sample the data. However, if the rate is low, this method has the disadvantage of disfiguring the shape of time series sampled. Another technique is to utilize the mean estimation of every fragment for representation of the set of data points. For a given time series $X = (x_1, x2_{,....},x_m)$ and n, reduced dimension, the compressed time series $Q = (q_1,q_2,….q_n)$ after dimensionality reduction can be obtained by

$$q_k = \frac{1}{e_k - s_{k+1}} \sum_{i=s_k}^{e_k} x_i \qquad \text{(i)}$$

where the starting and ending data points are denoted by s_k and e_k of the kth point in the time series X, respectively (Figure1.). The technique is known Piecewise Aggregate Approximation (PAA).

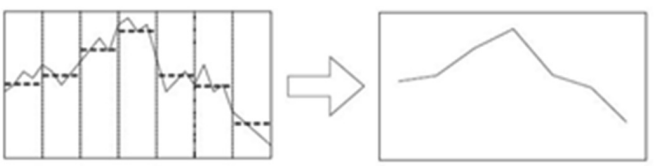


Figure 1: Dimension reduction by PAA

Another approach for representation of time series is to convert the numeric data in time series to symbols. First discretization of the time series into fragments is done, and then each fragment is converted into a symbol. A common technique for this is symbolic aggregate approximation (SAX) the reduced time series from PAA is converted to string of symbols. The y-axis is divided into equiprobable regions. Symbols are used to represent regions and every segment is mapped to a symbol based on the region in which it lies. The compressed series, q utilizing PAA is changed over to a symbol string S $(s_1,...,s_w)$.

Based on the representation of time series, various mining tasks can be found and can be grouped into fields such as clustering and pattern discovery, classification, rule discovery and summarization [1]. Among these, detection of previously

unknown, frequently occurring patterns is more interesting problem. Such patterns are called motifs. In Figure2. An example of motif discovered in an electroencephalogram (EEG) time series dataset is shown. An efficient motif discovery algorithm can be used as a tool for summarizing and envisioning enormous time series databases. Likewise, it can be utilizes as a subroutine as a part of different mining tasks, for instance:

- Motif discovery is required for the discovery association rules. Motifs are mentioned as primitive shapes in [2] and frequently occurring patterns in [3].
- A few algorithms for classification work by developing prototypes of every class [4, 5]. These prototypes may be viewed as motifs.
- Many time series inconsistency/interestingness discovery algorithms basically comprise of demonstrating normal behavior with a set of motifs, and identify future patterns that are not at all like these motifs [6].
- In robotics, a method has been introduced by Oates et al. [7], to permit an autonomous agent to sum up from a set of qualitatively diverse experiences gathered from sensors. These "experiences" are seen as motifs.
- Much work on discovering estimated occasional patterns in time series can be considered as an effort to find motifs that happen at compelled intervals [8].

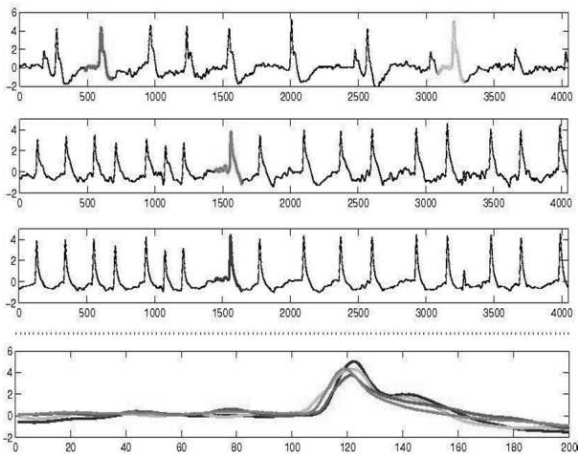


Figure 2: Example of motifs in their original context (up), and in the same referential for similarity observation (down) for EEG time series data

Definitions and Notations

Time Series
A time series can be defined as an ordered list T = $\{t_1,t_2,\ldots t_m\}$ of real-valued variables, where m represents the length of time series.

Subsequences
For a given time series T of length m, a subsequence $T_{i,k}$ of T is a time series of length k < m, which starts from position I, i.e., $T_{i,k} = \{t_i,t_{i+1},\ldots,t_{i+k-1}\}$, $1 \leq i \leq m-k+1$.
The distance between two subsequences is often compared by Euclidean Distance (ED).

Euclidean Distance
For two time series (or subsequences) $X=\{x_1,x_2,\ldots x_n\}$ and $Y=\{y_1,y_2,\ldots y_n\}$ both of length n,
Euclidean Distance (ED) between them is given by:

$$D(X,Y)=\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad \text{(ii)}$$

Non-Trivial Match
For a given time series T containing two subsequences $S_{i,k}$ and $S_{j,k}$ with the length k, If their distance $D(S_{i,k}, S_{j,k})$ is smaller than the specified threshold r, then $S_{j,k}$ is the matching subsequence of $S_{i,k}$,
If the two subsequences begin at essentially different positions, they are said to be a nontrivial match. We say that the starting positions are essentially different if there exists x, such that. i < x <j and the distance D ($S_{i,k}$, $S_{x,k}$) is larger than r.
Counting of trivial matches should be excluded by the algorithms.

K motifs
For a given time series T, subsequence length n,range r, the most significant motif in T (known as 1-Motif) is a subsequence

S_1 that has the highest count of non-trivial matches. The most significant motif in T (called K-Motif) is the subsequence S_k having the highest count of non-trivial matches and satisfies D (S_k,S_i)> 2r, for all $1 \leq i \leq K$.

Motif Type: Motif subsequence is a subsequence that matches the motif type within a time series *T*. A motif type often has multiple motif subsequences.
Taking the CBF dataset as an example, three similar subsequences of a particular motif type are detected, as seen in Figure3.

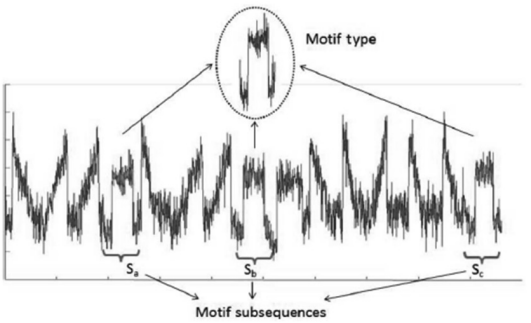


Figure 3: Motif type and three subsequences S_a, S_b and S_c

Various Motif Discovery Methods

Primarily motif discovery can be divided into exact and approximate methods. Exact methods are able to find motif exactly but are often inefficient while dealing with larger dataset. Approximate ones map segmentations into low dimensional words to reduce their computational complexity and execution time.

Exact Motif Discovery Method

J. Lin et al. [9], the paper introduce a Brute-Force (BF) algorithm for exact discovery of motifs, which require comparisons quadratic in the length of time series. For a subset t, of size n in time series BF compares it with each subset s in subsequence S to find the closest match. Its time complexity is O (m^2), where m represents the length of the time series.

Table 1: Brute Force Algorithm

Algorithm	Brute Force Motif Discovery
Procedure	$[L_1,L_2]$=*BruteForce_Motif*(D)
in:	D: Database of Time Series
Out:	L_1,L_2: Locations for a Motif
1	*best-so-far* = INF
2	**for** i = 1 to m
3	**for** j = i+1 to m
4	**if** $d(D_i,D_j)$ < *best-so-far*
5	*best-so-far*=$d(D_i,D_j)$
6	L_1=i, L_2=j

Mueen-Keogh (MK) [10], a tractable exact algorithm, faster than BF by using linear ordering of data to provide useful information to guide the search for motifs. Worst case of the algorithm is quadratic but can be reduced by three orders of magnitude.

Approximate Discovery Methods

The first approximate algorithm for motif discovery was proposed by Chiu et al. [11]. It uses probabilistic and iterative approach. Here time series is converted to SAX representation. For every iteration, random positions of words are selected as wildcards and whole list is traversed. For each match collision matrix entry is incremented. At the end largest entries in collision matrix are selected as motif candidates and ach candidate is checked for validity in original data.
H. Tang and S. S. Liao [12]. A novel algorithm is introduced in this paper that doesn't require exact w value, which is the length of the pattern. The idea is to keep w value relatively small to distinguish short pattern first, and then utilize a

concatenation routine to concatenate the short patterns found to generate the entire pattern .This approach improves the widely used K-motif algorithm, it is capable of discovering patterns with different lengths

Liu et al. [13], Here the issue of locating motif has been formalized as persistent top-K motif ball issue in a m-dimensional space and heuristic methodology has been suggested that can enhance quality of motifs. Comparative time series subsequences are assembled as m-dimensional data points in a ball of range r. A motif in the data set is then a thick ball after removal of the trivial matches in an m-dimensional space. The maximum radius of k ball being

$$\sqrt{\frac{1}{1/m+1}}\sqrt{2r}.$$

The above talked about methodologies are restricted to finding pattern occurrences of same length, neglecting to capture the likenesses of events being consistently scaled along the time axis. D. Yankov et al, [14]. This paper proposes identification of time series motif for uniform scaling. Here time series motif projection algorithm of [11] is extended to capture motif under uniform scaling distance d_u

Combining Exact and Approximate Methods

B. Liu et. al. [15], the paper introduced an efficient Motif Discovery method for Large - scale time series (MDLats) by combining the advantages of both exact and approximate methods. By computing standard motifs, MDLats eliminates a majority of redundant computation in related area and reuses existing information to the maximum. Figure4. Shows the main steps of MDLats

First step is the input i.e. the initial time series which is processed to obtain subsequence symbols by using SAX [16]. and PAA [17] methods. Segmentation Compression Module compresses subsequence symbols by abstracting repetitive symbols. Then standard motifs are calculated which are used by Final Motifs Discovery Module for generation of final motifs that have similar patterns. Since all motifs are included in candidates, whole motifs with different lengths from short patterns.

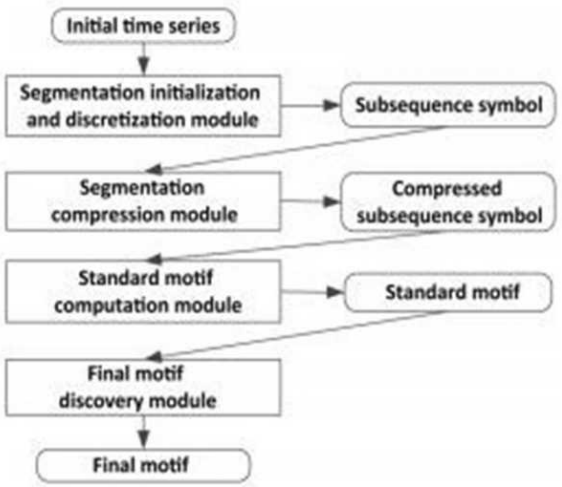


Figure 4: Main Steps of MDLats

Advantages of MDLats algorithm over the above discussed:

- It improves efficiency for motif discovery while retailing accuracy.
- By computing standard motif, it reduces majority of redundant computations and reuse existing information to the maximum by exploiting the relation between existing and newly arrived data.
- It is adaptive to different lengths of motif.
- It is scalable and is deployed in Hadoop for parallel computing.

Experimental Analysis

In this section we have compared two exact discovery methods i.e. BF [9], MK [10] and one approximate method RP [11] in terms of execution time. The results are shown in fig. 4. The dataset used in this section is constituted by random walk time series available in the MK algorithm [10] website. There are 10 sets of random walk series data, containing 10000 to 100000

time series, each of length 1024.
Note that, we have compared results with only one iteration of RP algorithm. However, as an iterative algorithm, several iterations are necessary in order to converge. Results with more than one iteration will have less execution time than the BF algorithm as BF algorithm is quadratic in nature whereas RP algorithm is linear with high constant factor. MK algorithm is up to three orders faster than BF.

Table 2: Comparison of various methods for motif discovery based on execution time

Size of Data set	BF	MK	RP
10000	83.703	3.45	53.54
20000	190.962	16.11	193.88
30000	399.765	32.58	404.41
40000	648.18	31.29	705.02
50000	965.047	49.92	1221.13
60000	1429.116	66.54	1613.53
70000	1861.126	90.03	2139.20
80000	2550	103.31	2708.53
90000	3277	108.70	3468.50
100000	4200	130.68	4357.39

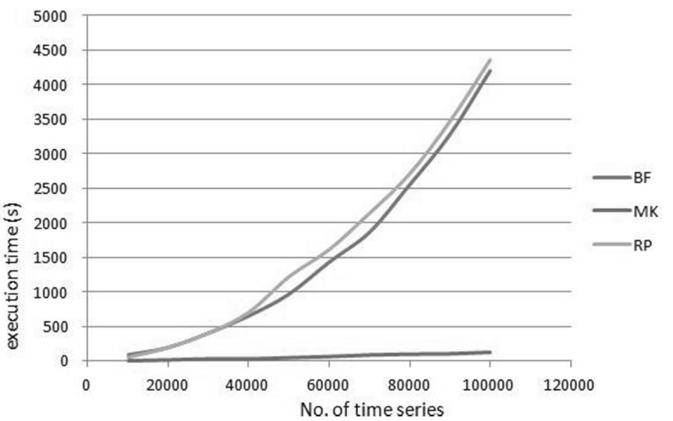


Figure 5: Execution time of the algorithms vs. No. of time series

Conclusion and Future Work

- We can improve the MDLats algorithm mentioned in [15] by optimizing the SAX representation by using algorithms like Genetic algorithm (GA), Particle Swarm Optimization (PSO).
- Motifs should be evaluated in such a way that they could be ranked in terms of their significance.
- Motifs can be used to describe the time series data in other data mining tasks like:
    - Classification
    - Abnormality Detection
    - Forecasting

Acknowledgement

The authors would like to thank UIET and Panjab University authorities for providing resources to carry out this research work.

References

[1] T. C. Fu, "A review on time series data mining" ,Engineering Applications of Artificial Intelligence, 24 (1) (2011), pp. 164–181
[2] G. Das, K. Lin, H. Mannila, G. Renganathan and Smyth, "Rule discovery from time series" in proceedings of 4th International Conference on Knowledge Discovery and Data Mining. New York, NY, Aug 27-31. pp 16-22, 1998.
[3] F. Hoppner, "Discovery of temporal patterns - learning rules about the qualitative behavior of time series", in the proceedings of fifth

European Conference on Principles and Practice of Knowledge Discovery in Databases. Freiburg, Germany, pp 192-203. 2001.
[4] E. Keogh and Pazzani, “An enhanced representation of time series which allows fast and accurate classification clustering and relevance feedback”, International Conference on Knowledge Discovery and Data Mining. New York, NY, Aug 27-31. pp 239-243
[5] M. Hegland, W. Clarke and Kahn, “Mining the MACHO dataset, Computer Physics Communications”, vol.142, pp. 22-28, 2002.
[6] D. Dasgupta and S. Forrest, “Novelty detection in time series data using ideas from immunology”, In Proceedings of the fifth International Conference on Intelligent Systems, 1999.
[7] T. Oates, M. Schmill and P. Cohen, "A Method for Clustering the Experiences of a Mobile Robot that Accords with Human Judgments", In Proceedings of the 17th National Conference on Artificial Intelligence, 2000, pp 846-851.
[8] J. Han, G. Dong and Y. Yin, “Efficient mining partial periodic patterns in time series database”, in Proc. of the fifteenth International Conference on Data Engineering, Sydney, Australia, pp 106-115,1999.
[9] J. L. E. L. S. Lonardi and P.Patel, ”Finding motifs in time series,” in Proceedings of 8th ACM SIGKDD International Conference on Knowledge Discovery Data Mining, Edmonton, AB, Canada, 2002, pp.53-68.
[10] A. Mueen et al., “Exact discovery of time series motifs,” in Proceedings of SIAM International Conference on Data Mining ,2009,pp. 473-484.
[11] B. Chiu, E. Keogh, and S. Lonardi, “Probabilistic discovery of time series motifs,” in Proc. 9th International Conference Knowledge Discovery Data Mining, Washington, DC, USA , 2003, pp. 493-498.
[12] H. Tang and S.S. Liao, “Discovering original motifs with different lengths from time series,” Knowledge Based Systems., vol.21,pp.666-671.2008
[13] Z.Liu .et.al.,"Locating motifs in time series data ,"in Proceedings of SIAM International Conference on Data Mining,2009,pp.473-484.
[14] D. Yankov et al.,"Detecing time series motifs under uniform scaling” in Proc. 9th ACM SIGKDD International Conference Knowledge Discovery Data Mining, 2003, pp. 493-498.
[15] B. Liu, J. Li, C. Chen, W. Tan, "Efficient motif discovery for large-scale time series in healthcare "IEEE Transactions on Industrial Informatics., vol. 11, no. 3, pp. 583-590, June 2015
[16] J. Lin et al.,"A symbolic representation of time series, with implications of streaming algorithms," in Proc 8th ACM SIGMOD Workshop Res. Issues Data Mining Knowledge Discovery,2003,pp. 2-11
[17] E. Keogh et al., "Dimensionality reduction for fast similarities search in large time series databases," J. Knowledge Inf. Syst., vol. 3 pp.263-286,2001

Biometric Identification using Lip Imprint with Hybrid Feature Extraction Techniques

Semonti Chowdhury*, Joydev Hazra** and Satarupa Bagchi Biswas***
* Rajasthan College of Engineering for Women, India
E-mail:semonti.chowdhury@gmail.com
** ***Heritage Institute of Technology, Kolkata, India
E-mail:joydev.hazra@heritageit.edu
E-mail:satarupa.biswas@heritageit.edu

Abstract: This paper is representing a hybrid feature extraction technique for biometric identification using lip imprint. This research work combined the shape and texture features of human lip imprint. Canny Edge Detection Algorithm is used here to extract the shape features and to extract texture features GLCM (Gray Level Co-occurrence Matrix) is calculated. The whole set of complied features are then inputted to the Support Vector Machine (SVM) classifier to identify and/or verify the corresponding person.

Keywords: Shape Features, Texture Features, Biometric Identification, Support Vector Machine (SVM), Canny Edge Detection, GLCM.

Introduction

Lip imprints are proven to be used for biometric identification and well accepted globally for this purpose. The science which includes the study of lip imprint and the use of lip imprint as biometric identification of human being is called "Chelioscopy" which is derived from the Greek word "Chelian" for English word "Lip". There are many different biometric identification techniques adapted worldwide based on fingerprint, palm-imprint, DNA sequence, iris and retina pattern, signature etc. Among them Lip imprint based identification technique is having some specific advantages. In has been proven that, lip imprint of a person doesn't change over age of a person. So this technique of biometric identification will get better result over fingerprint based or palm imprint based identification. Moreover, gender and blood group of a person also can be identified from lip imprint. So, the lip imprint based biometric detection has been adapted globally and also accepted as unique identifier of people legally as well as may be accepted as court room evidence. This technique is presently used in Criminal Forensics to identify criminals from their lip imprints collected from crime scene.

The fact of lip imprint to be used as biometric identification was first recommended by French Criminologist Edmond Locard in 1932. Then in 1950 in his book titled "Homicide Investigation", LeMoyne Snyder mentioned the lip imprint as a potential biometric identifier. Afterwards many researchers, criminologists and dentists researched to prove acceptability of lip imprint as biometric identifier. Some of the researches even tried to extract other information, like sex, age, family, blood group etc, about the owner from lip imprints.

The research on lip imprint based identification system actually got a mileage when in 1970 Japanese doctor Suzuki [1] and in 1974 Tsuchihasi proved the claim of lip imprint to be used as biometric identification in their corresponding researches on different age grouped Japanese people. In 1975 Suzuki and Tsuchihasi identified five different kinds of groove patterns using which further lip imprints could be easily categorized and identified [2]. Grooves are the fine lines present in a lip imprint. These five groove patterns are known as Suzuki's Classification. The five types of groove patterns are:

Type I: A clear-cut line or groove running vertically across the lip.

Type I': Straight grooves that disappear half way into the lip, instead of covering entire breadth of the lip, or partial-length groove of Type I

Type II: Grooves that fork in their course or a branched groove (like Y)

Type III: An intersected groove (crisscross pattern)

Type IV: A reticular groove

Type V: Grooves that doesn't fall in any of the above categories and cannot be differentiated morphologically

Figure 1 is depicting all the above-mentioned types of groove patterns mentioned above.

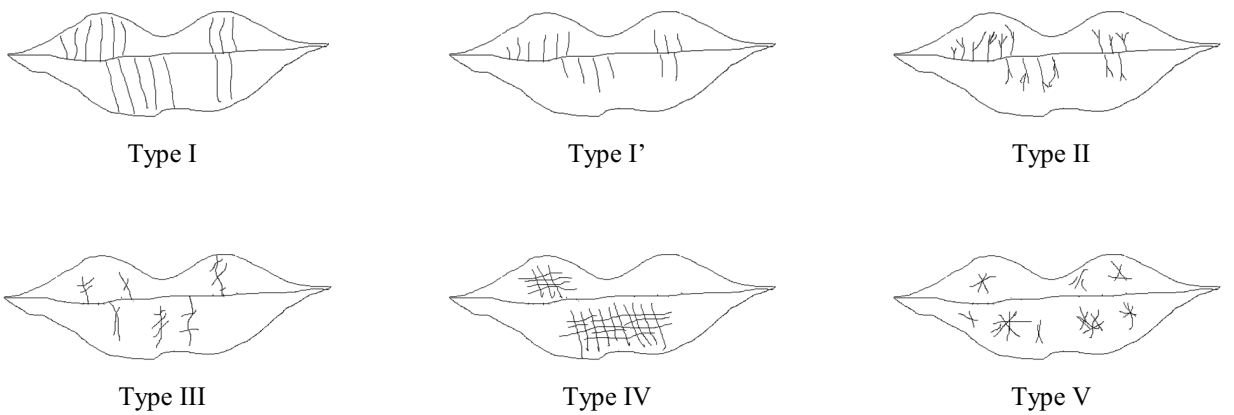


Figure 1: Suzuki Classification: different groove patterns

Lots of researches are carried out then after on lip imprint detection. Some of the researches are carried out manually, by registering the imprint on transparent sheet or glass using some sticky colouring agent like petroleum jelly or lipstick. The measurement and classification of grooves are also done manually using rulers and magnifying glasses. On the other hand some researchers took the computer algorithms for image processing and different classifiers to identify lip imprints in automated way. Automated techniques are advanced than the manual technique as the precision and accuracy of similarity measurement are more in case of automated techniques. There is no room for human error. Only the imprint taken on transparent sheets are to be converted as digital image and fed to the algorithm as input; the algorithm further analyze the imprint automatically.

The manual lip imprint identification techniques are generally based on Suzuki's Classification. The study of grooves and presence of groove patterns are matched between different lip-imprints are similarity is measured between them. The study of lip imprint is done manually using high power magnifying glass. Besides finding matches between lip-imprints, in different research works, people tried to gather more information about the person using lip-imprint. In 2009 a group of researchers proved that in female lip imprints Type I and Type II grooves are majorly identified, whereas in male lip imprints Type III, Type IV and Type V type of grooves are majorly found [3]. In the same year, another group of scientists proved the power of lip imprint as biometric identifier in their research on 300 North Indian people aged between 18 – 65 years [4]. In 2010 another research tried to prove that, by studying groove patterns of lip imprint gender, blood group and family of human being can be identifies [5]. Later in the same year, another research proved that, by studying lip imprint, blood group and gender can be identified, but family of a person cannot be detected [6]. They decided from the result of their research that, there is no correlation between the lip imprints of siblings, parent and children and thus no correlation among the lip imprints of the people belonging to the same family.

The automated machine learning based approaches for lip imprint detection generally takes two different approaches: structural information oriented approaches where different structural measurements of the lip imprint are taken in account and statistical analysis based approaches where the presence of grooves are represented in terms of features and different statistical analysis are carried out on those set of features.

Some structural information oriented approaches

In 2009, Michał Choraś extracted different geometric patterns from lip imprint and used them to identify the imprint [7]. Thus he proved that besides lip imprint, lip structures also can be used to identify a human being. In 2010 another research used Variance Based Haar-like Features and Kalman Filter and further used SVM on those features to identify lip imprints [8]. In the same year another research adapted "Generalized Hough Transform" to identify lip imprint [9]. Here they have used GHT on the pre-processed image to form R table, Accumulator Array and Hyper-surface. Using these extracted features they computed the matching. Another research in 2013 adapted GHT to find out distances between groove-patters [10]. In this research, the GTH is used for extracting the grooves from lip imprint and further the imprints are matched based on linear distance and angle formation between neighbouring grooves. In 2011 another group of scientists extracted the local structural features like Lip's Width to Perimeter ratio, Upper to Lower Lip height ratio, Upper Lip Height to Width ratio, Lower Lip Height to Width ratio, Inner to Outer Circle ratio, Width to Middle Height ratio, Left side Upper to Lower Lip Convexity ratio, Right side Upper to Lower Lip Convexity ratio, and Indent ratio to identify lip imprints [11]. Whereas in the same year another group applied "Brute Force Algorithm" to reduce False Acceptance Rate and False Rejection Rate in the

lip imprint identification technique [12]. FAR and FRR are the measurement of inefficiency of any biometric identification system.

Some statistical analysis based approaches

The Mean Differences Similarity Measurement was adapted by a research [13] where the features were extracted by forming segments of lip patterns and afterwards by applying Hough Transformation on each segment. GLCM [14], Scale Invariant Feature Transform (SIFT) [15], Speeded Up Robust Features (SURF) [16] and combination of SIFT and SURF [11] was used by the researchers to extract the features of an image. On the extracted features the Mean Differences Similarity Measurement was applied to find out the match.

In 2011 a research implemented the lip imprint identification system using Dynamic Time Warping algorithm (DTW) [17]. In this research three steps were adapted to extract features: Image normalization, Lip pattern extraction, and Feature extraction. The feature extraction part is done here by DTW algorithm for both the upper and lower lip. This process detects the vertical, horizontal and diagonal projections of the lip imprint images. In 2012 another research divided the lip imprint into two segments: upper lip and lower lip imprint [18]. Further each of these segments is subdivided into four quadrants as shown in Figure2. They applied Gaussian Filter, Canny Edge Detection and four different Sobel filters to extract the groove patterns horizontally, vertically and both diagonally. The matching is performed as average match of groove patterns in all the eight segments. Another research applied Dynamic Time Wrapping Algorithm and Coperland Vote Counting Method to identify similarity between lip-imprints [19]. The first algorithm is used to compare between imprint features and the second one is used for detecting maximum similarity within a group. In 2015 another statistical analyses based research used Bi-function Analysis Method to detect lip imprint [20]. All these different lip imprint detection and identification techniques discussed above are carefully reviewed in a research paper [21] published in the year of 2015, where all the different approaches are discussed on a single platform.

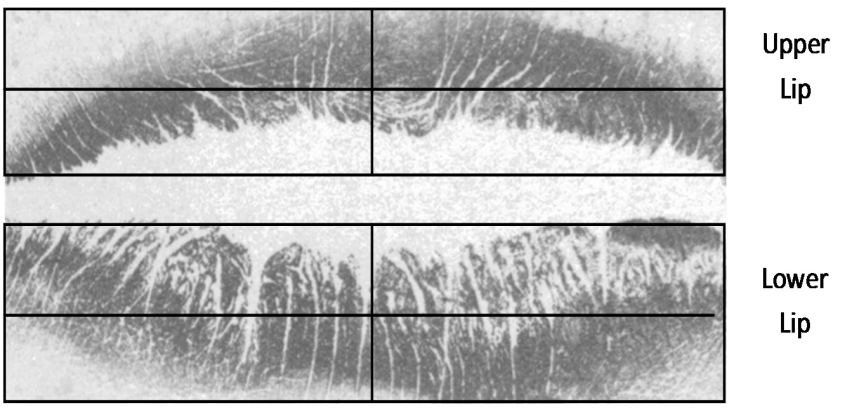


Figure 2: Lip-imprint segments

In this paper, different features are extracted in three levels namely local, global and shape. As described in the next sections, GLCM is used for both local and global level features and Canny edge detector is used for shape features.

Prerequisite

Canny Edge Detection

Canny Edge Detection Algorithm [22] results into providing optimal edge output based on three criteria namely: Flawless Edge Pixel Detection, Localization and Avoidance of False Responses. The basic aim of this algorithm was to increase the detection of actual edge pixels and reduction of false acceptance. Another aim of this algorithm is to provide seamless contour of the edge. Figure3 shows an input image and its corresponding output of Canny Edge Detection Algorithm as an example.

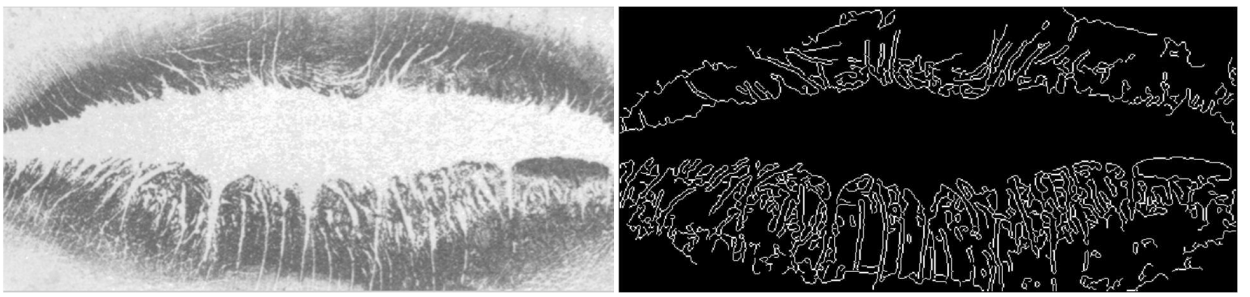


Figure 3: Source and Output of Canny Edge Detection Algorithm

The Canny Edge Detection Algorithm works in the following five steps:

Step 1: In the preliminary step any kind of noise gets filtered out from the source image by applying Gaussian Filter. In this step, a tactfully chosen Convolution Mask is applied to the image which imposes a blurring effect and thus helps to reduce noise and to detect edge more flawlessly.
Step 2: In the current step, the gradient of the image is computed as at the edge strength can be identified where gradient of the image is maximum. The gradient of the image is computed using the following formula:

$|G| = |Gx| + |Gy|$

[where Gx and Gy are gradient of the image in horizontal and vertical direction respectively. The Gx and Gy are shown in Figure4.]

-1	**0**	**+1**
-2	**0**	**+2**
-1	**0**	**+1**

Gx

-1	**-2**	**-1**
0	**0**	**0**
+1	**+2**	**+1**

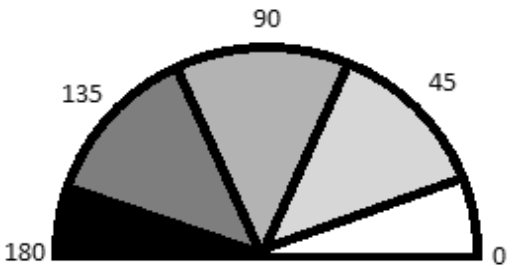
Gy

Figure 4: Gx and Gy gradients

Step 3: In this step, the edge direction gets computed using the following equation:

$\theta = \tan^{-1} (Gy/Gx)$

[where θ indicates direction of the edge, only 0, 45, 90, 135 and 180 degrees are considered. All other values are approximated to this set. The approximation is done as shown in Figure5.
In Figure5, different shaded regions are approximated as the values given in the index.]
If Gx becomes 0, decision gets based on value of Gy. If Gy is 0, θ is considered as 0 degree, otherwise θ is considered as 90 degree.
Step 4: To find out perfect this edge, after computing direction, candidature of all non-maximum pixels (a pixel for which gradient is not at local maxima) got involved into the detected edge, are suppressed.
Step 5: Finally at this last step, a unique feature of Canny Edge Detection, called Hysteresis, is applied. Generally by involving single threshold value the detected edge may have dashed line contour representation. To get rid of this, Canny Edge Detection uses double threshold values T1 and T2, one high and the other is low. The starting pixel is accepted based on T1, but the continuing pixel selection is dependent on both T1 and T2.



Region	Approximated Angle
	0 degree
	45 degree
	90 degree
	135 degree
	180 degree

Figure 5: Approximation of detected edge direction

Gray Level Co-occurrence Matrix (GLCM)

Statistical based features are describing the texture of an image. Gray Level Co-occurrence Matrix (GLCM) [23] is well known method to describe the texture of an image. The co-occurrence matrix provides spatial distribution of intensity or gray value of neighbor pixels in an image. GLCM could be calculated to consider one pixel, two neighbor pixels and three or more neighbor pixels, i.e. first order, second order and third order or higher order statistic of gray value of an image. GLCM find out the frequency of reference intensity of a pixel i, with the intensity of neighbor pixel j either in horizontal, vertical or diagonal direction with a defined radius. It has been proved by the researchers that large radius value is not suitable to detect fine texture information of an image.

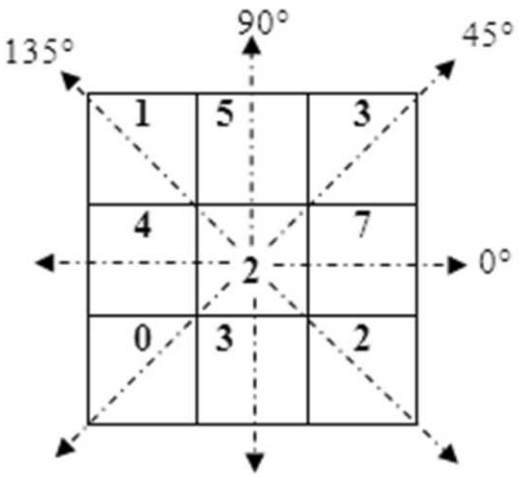


Figure 6: Angles for GLCM calculation

For extracting the fine texture, the value of radius is considered 1 or 2 for eight neighbors of reference pixel in angles of , 45°, 90°, 135°, 180°, 225°, 270° or 315°. The calculation of GLCM is similar in 0° and 180°, 45° and 225° and so for others. So, we consider the angles in 0°, 45°, 90°and 135° for calculation of GLCM as shown in Figure6.

GLCM Features

G(i, j) is the (i, j)th entry and G_n denotes the dimension of the normalized co-occurrence matrix. There are fourteen features defined by Haralick. Six of them are used in our propose work for classification of lip imprint images listed below:

i. $\text{Energy} = \sum_{i,j} G(i,j)^2$

ii. $\text{Homogeneity} = \sum_i \sum_j \frac{1}{1+(i-j)^2} G(i,j)$

iii. $\text{Entropy} = -\sum_{i,j} G(i,j)\log(G(i,j))$

iv. $\text{Inertia} = \sum_{i,j} (i-j)^2 G(i,j)$

v. $\text{Dissimilarity} = \sum_{i=0}\sum_{j=0} G(i,j)(i-j)$

vi. $\text{Correlation} = \sum_{i=0}\sum_{j=0} G(i,j)\frac{(i-\mu_x)(j-\mu_y)}{\sigma_x\sigma_y}$

Support Vector Machine (SVM)

Support vector machine (SVM) [24] [25] is a supervised learning classifier which creates a separating hyperplane that maximizes the margin between two data sets in a space. SVM minimizes the empirical classification error as well as maximize the geometric margin between two classes. In multiclass classification, the problems are divided into several binary classifications. There are two common approaches to design such binary classifier as (i) One-Versus-All and (ii) One-Versus-One. In One-Versus-All, the classification is done by "winner takes all strategy"; the highest output function assigns the class. The classification of one-versus-one approach is done by a max-wins voting strategy. We have used SVM tool "SVM*light*" for classification [26].

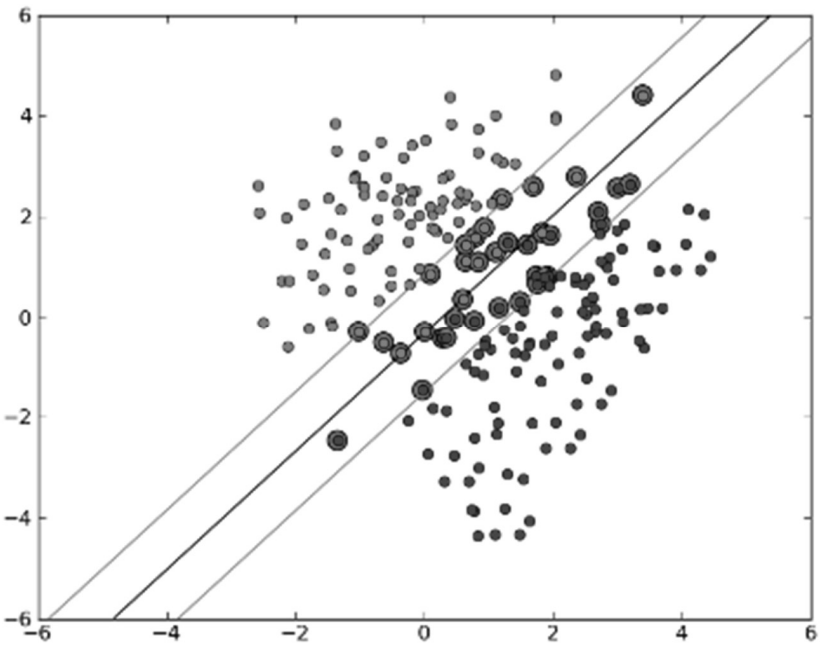


Figure 7: Support Vector Machine

Proposed Method

The proposed method can be divided into two parts: feature extraction and classification. The accuracy of lip imprint image classification depends on extracting discriminate features from the preprocessed image and the nature of classifier. In feature extraction part image features are extracted from three levels namely global, local and shape as depicted in Figure9. The texture features are extracted using GLCM in global and local level. The preprocessed lip imprint images are divided into four equal parts in local level as shown in Figure8. From each part we extract the features using GLCM to create feature vector at local level. Generally in global level whole lip imprint images are consider to extract the feature using GLCM. But the size of the images is scaled down to reduce the complexity of GLCM calculation. In the proposed method preprocessed lip imprint images are resized into 100 × 200. The Canny edge detection algorithm is used to extract the shape features.

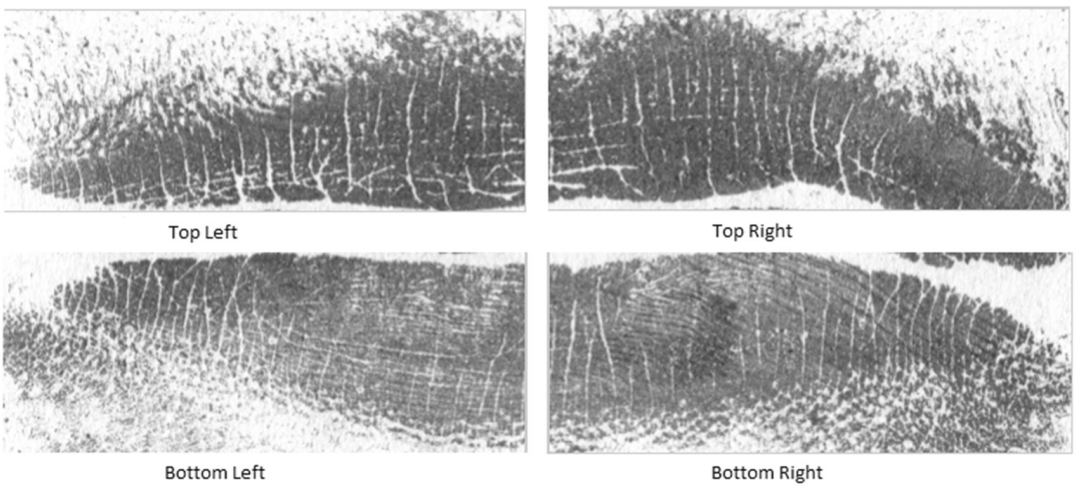


Figure 8: Four parts of lip imprint as used at local level feature extraction

For the lip imprint classification SVM is used. After extracting features, a feature vector is created which is passed to SVM for training. With help of the trained SVM we can identify or verify an unknown person. To do this we extract the features from the test image at three levels using GLCM for texture features and canny edge detector algorithm for shape features.

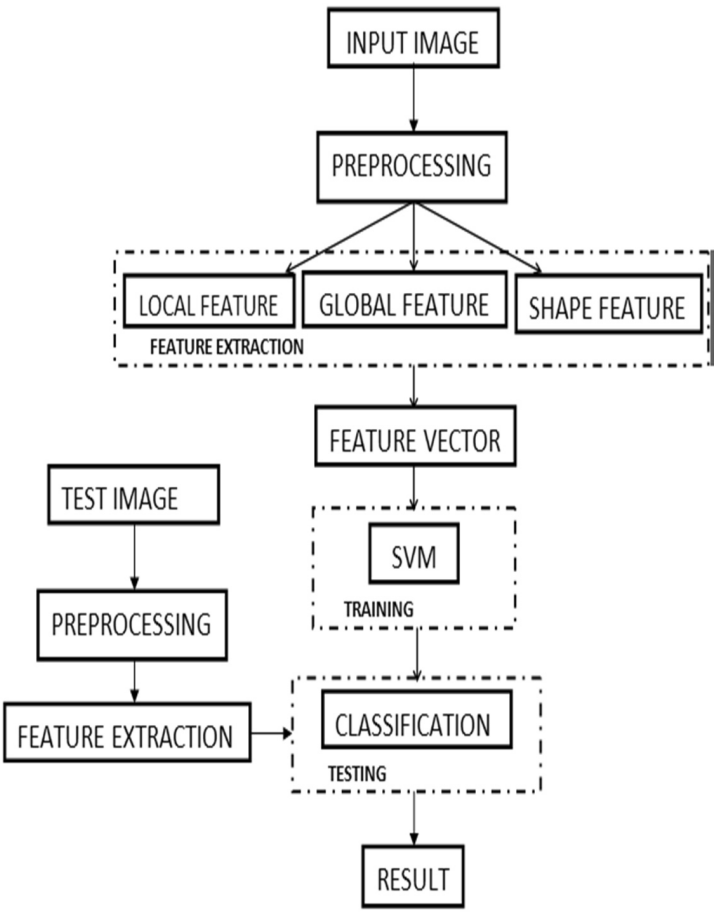


Figure 9: Flowchart of the proposed method

Result Analysis

To evaluate the performance of the proposed method, several experiments were performed on different lip images. We have performed different experiments on three lip-imprint image datasets – Lip-Print, Roller set 1 and Roller set 2 [http://biometrics.us.edu.pl (Lip imprint database from Biometric Research Centre, University of Silesia at Katowice, Poland)].

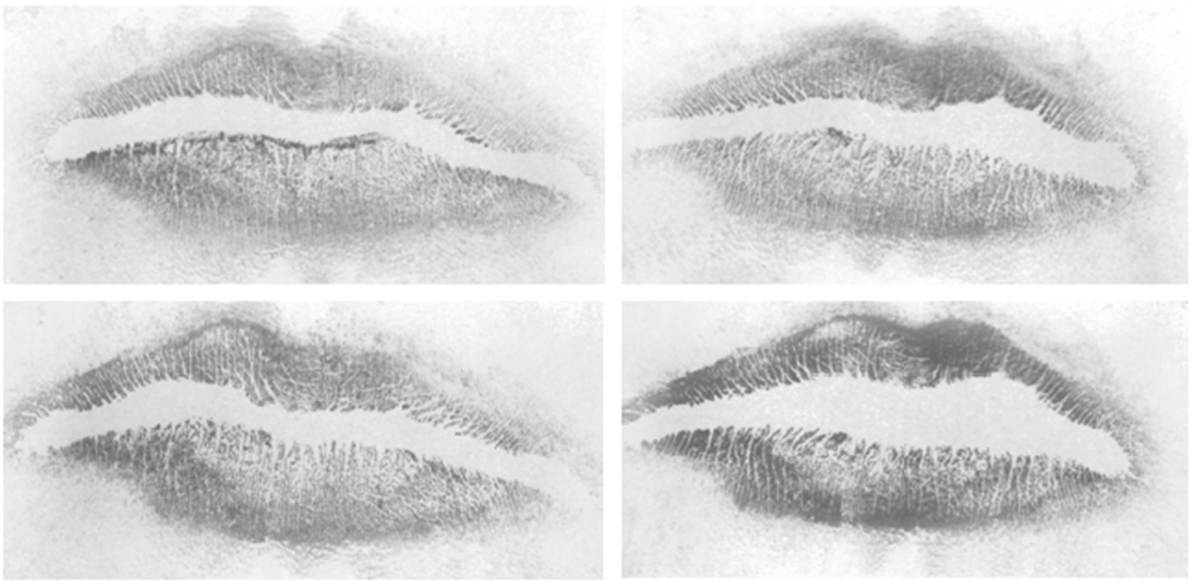


Figure 10: Data Set of Lip Imprint (Lip_print)

We have presented these three datasets as indicated in Figure10, Figure11 and Figure12 respectively. The first data set, Lip_print contains 4 lip imprint images of 15 people; hence all together 60 images. Further the second data set, Roller_set_1 represents the same 4 imprint images for 30 persons. So we applied this dataset to check the efficiency of the proposed method with more different types of images. Lastly Roller_Set_2 dataset provided us with more number of varied images. It contains the lip imprints of 40 people and for each of them 10 imprints are recorded. So totally 400 images are present in Roller_Set_2.

Figure 11: Data Set of Lip Imprint (Roller_Set_1)

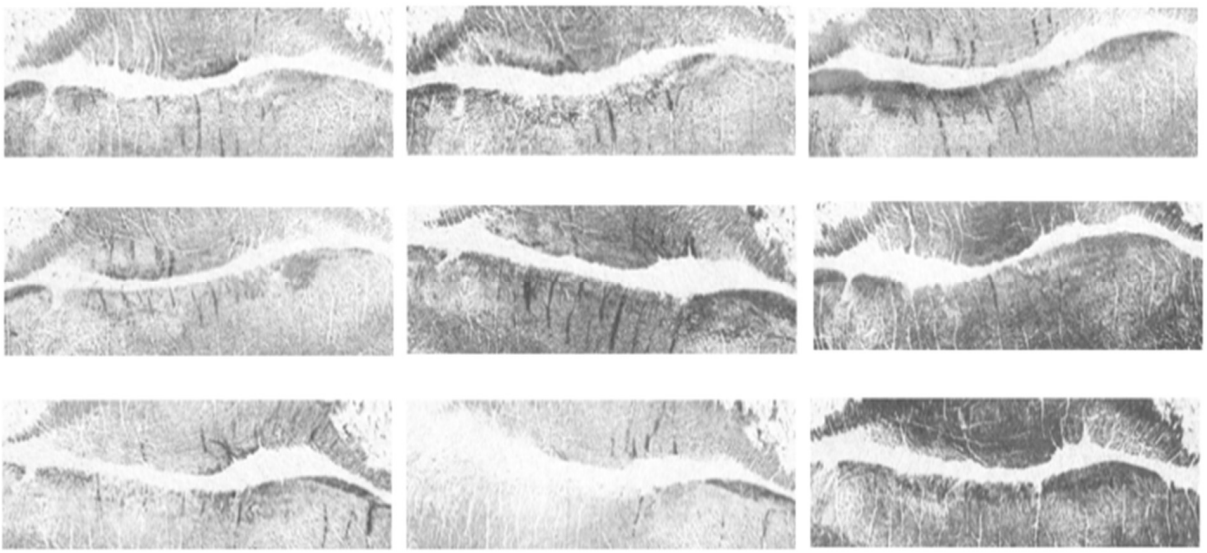


Figure 12: Dataset of Lip imprint (Roller_Set_2)

Experimental results of the proposed method on dataset given in Figure12 are given in Figure14. In this proposed method, we extracted features from three levels. In local level, lip-imprint images are equally partitioned into four parts namely Top Left, Top Right, Bottom Left and Bottom Right. From each part we calculated GLCM at angles of 0°, 45°, 90°and 135°. Six features i.e. Energy, Homogeneity, Entropy, Inertia, Dissimilarity and Correlation, are extracted from each GLCM. So, in total 24 features are extracted in every part. Images are resized into 100 x 200 in global level and resized images are considered to calculate GLCM. In global level 24 texture features are extracted. The proposed method 120 numbers of texture features are extracted from the lip-imprint images.

In this article, we also considered the shape features. Canny edge detection algorithm is used to find out the structural or shape information. We extracted shape information every 5° from lip-imprint images. Feature vector contained 120 texture features and 73 shape features as shown in Figure13. Total 193 features are used for classification the lip-imprint images.

In this paper we have applied two classifier – k-NN [27] and SVM. The result of our proposed algorithm is shown in Table1. The proposed method shows convincing results compared to both SURF [16] and SIFT [15] in respect to identify the owner of the lip imprint as tabulated in Table2.

Table 1: Result of the proposed method

Datasets	k-NN	SVM
Lip-Print (Figure9)	66.66%	86.67%
Roller set 1 (Figure10)	70%	79.92%
Roller set 2 (Figure11)	78.34%	91.67%

Table 2: Comparative Result Analysis

	Misclassification rate by SVM classifier
Proposed method	8.33%
SURF	16.66%
SIFT	11.11%

The proposed method shows convincing results compared to both SURF [16] and SIFT [15] in respect to identify the owner of the lip imprint.

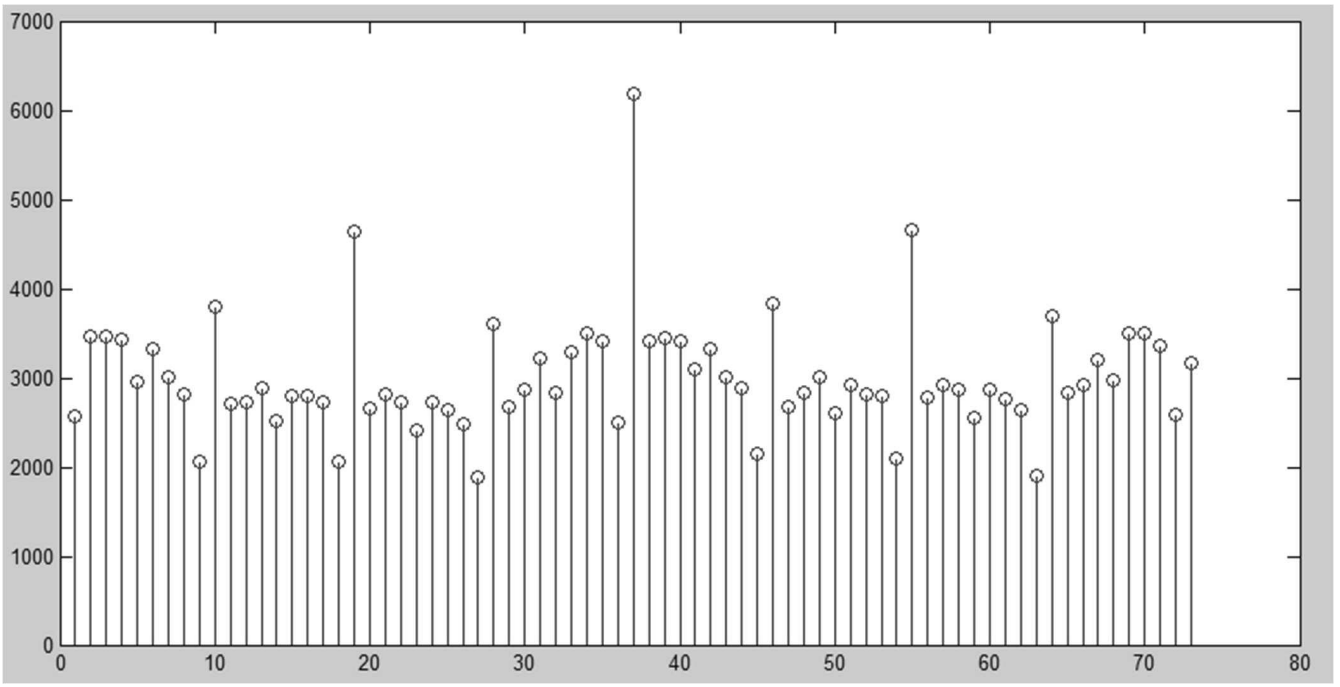


Figure 13: Features extracted using Canny Edge Detection Algorithm

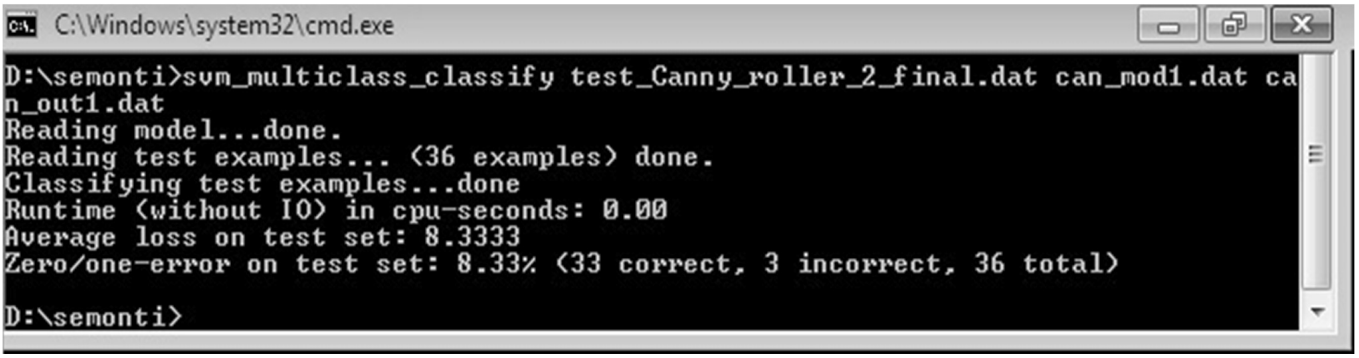


Figure 14: Result of proposed method

Conclusion

In this article, an effective technique is proposed to identify owner of the lip imprint using feature vector and SVM. GLCM and Canny edge detector method is used as feature extractors. Results of the proposed method show significant improvement over the existing algorithms with respect to identify owner of the lip imprint. The result generated by the proposed algorithm is compared with many other methods used for the same purpose and it has been identified that, relevant betterment is achieved with the newly proposed procedure.

The result has shown significant improvement in accuracy. As lip imprint is globally accepted as biometric identifier, accuracy of detecting the owner of the lip imprint has immense importance in criminology.

The proposed method has generated significantly less number of misclassifications compared with other globally acclaimed methods of classification as applied on lip-imprint data.

The proposed method is efficient to be used for biometric identification of human being by means of lip imprint using automated systems. In future we will further enhance the performance of the proposed algorithm in terms of accuracy.

Acknowledgement

We would like to thank University of Silesia at Katowice, Faculty of Computer Science and Material Science, Institute of Computer Science, Computer Systems Department for allowing us to use their datasets for the purpose of this research. We also thank T. Joachims for SVM^{light} which we have used as classifier in this research work.

References

[1] K Suzuki, Y Tsuchihashi, "A new attempt of personal identification by means of lip print", J Indian Dent Assoc, vol.–42, no.–1, January 1970, pp. 8-9
[2] K Suzuki, Y Tsuchihashi, "Two criminal cases of lip print", ACTA Criminol Japan vol.–41, 1975, pp. 61-64
[3] Preeti Sharma, Susmita Saxena, Vanita Rathod, "Cheiloscopy: The study of lip prints in sex identification", Journal of Forensic Dental Science, vol. – 1, issue – 1, 2009, pp. 24-27
[4] U. Bindal, S. L. Jethsani, N. Malhotra, R. K. Rohatgi, M. Arora, P. Sinha, "Lip prints as a method of identification in human beings", J Anat (India), vol. – 58, no - 2, 2009, pp. 152-155
[5] Shilpa Patel, IshPaul, Madhusudan.A.S., Gayathri Ramesh, Sowmya G.V, "A study of lip prints in relation to gender, family and blood group", International Journal of Oral & Maxillofacial Pathology, vol. – 1, no – 1, November 2010, pp 4-7
[6] T.N.Uma Maheswari, "Role of Lip prints in Personal Identification and criminalization", Forensic Medicine and Toxicology [serial online], vol – 12, no – 1, December 2010
[7] Michał Choraś, "Lips Recognition for Biometrics", Advances in Biometrics, vol. – 5558, September 2009, pp 1260-1269
[8] Lirong Wang, Xiaoli Wang, Jing Xu, "Lip Detection and Tracking Using Variance Based Haar-Like Features and Kalman filter", Proc. 5th Int.Conf. Frontier Comput. Sci. Technol. (FCST) , August 2010, pp 608 – 612
[9] Łukasz Smacki, Piotr Porwik, Krzysztof Tomasztcki, Sylwia Kwarcinska, "The lip print recognition using Hough Transform", Journal of Medical Informatics & Technologies, vol. 14, 2010, pp 31-38
[10] Krzysztof Wrobel, Rafal Doroz, "Method for Identification of Fragments of Lip Prints Images on The Basis of The Generalized Hough Transform", Journal of Medical Informatics & Technologies, Vol. 22, 2013, pp. 189-193
[11] S. Bakshi, R. Raman, P.K. Sa, "Lip pattern recognition based on local feature extraction", Proc. Annu. IEEE India Conference, December 2011, pp.1 -4
[12] Pawan Sharma, Shubhra Deo, S. Venkateshan, Anurika Vaish, "Lip Print Recognition for Security Systems: An Up-Coming Biometric Solution", Intelligent Interactive Multimedia Systems and Services, vol. – 11, 2011, pp 347-359
[13] Lukasz Smacki, Krzysztof Wrobel, "Lip Print Recognition Based on Mean Differences Similarity Measure", Computer Recognition Systems 4 of the series Advances in Intelligent and Soft Computing, vol. 95, pp. 41-49
[14] J Hazra, A Roy Chowdhury, P Dutta, "Immune-based feature selection in rigid medical image registration using supervised neural network", Hybrid Soft Computing Approches, Springer, Vol-611, 2015, pp. 429-454.
[15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vision, Vol. 60, Nov 2004, pp. 91-110.
[16] H. Bay, T. Tuytelaars and L. Van Gool, "SURF: Speeded up robust features", Vol. 3951, pp. 404-417.
[17] Lukasz Smacki, Krzysztof Wrobel, Piotr Porwik, "Lip Print Recognition Based on DTW Algorithm", Nature and Biologically Inspired Computing (NaBIC), 2011 Third World Congress, October 2011, pp. 594 - 599
[18] S. Bhattacharjee, S. Arunkumar, S. K. Bandyopadhyay, "Personal Identification from Lip-Print Features using a Statistical Model", International Journal of Computer Applications, Vol - 55, no. - 13, October 2012, pp. 30-34
[19] Piotr Porwik, Tomasz Orczyk, "DTW and Voting-Based Lip Print Recognition System", Computer Information Systems and Industrial Management, vol – 7564, 2012, pp. 191-202
[20] Krzysztof Wrobel , Rafał Doroz, Malgorzata Palys, "Lip Print Recognition Method Using Bifurcations Analysis", Intelligent Information and Database Systems, vol – 9012, 2015, pp. 72-81
[21] Semonti Chowdhury, Satarupa Bagchi Biswas, Joydev Hazra, "Lip Imprint Based Biometric Identification: A Survey", International Journal of Scientific & Engineering Research, vol - 6, Issue 11 ,November-2015, pp. 1098-1102
[22] J. Canny, "A Computational Approach to Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, November 1986, no. 6.
[23] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural features for image classification", IEEE transactions on System, Man and Cybernetics, SMC-3(6), 1973, pp. 610–621.
[24] Burges C.J.C, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, 1998, 2:121-167.
[25] Cristianini N.;Shawe-Taylor J., "An Introduction to Support Vector Machines", Cambridge University Press, UK, 2000.
[26] T. Joachims, "Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning", B. Schölkopf and C. Burges and A. Smola (ed.), 1999, MIT-Press.
[27] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification", IEEE Trans. Inform. Theory, Vol. IT-13, Jan 1967, pp 21-27.

Routing Protocols in Wireless Local Area Networks (WLANs)

Sachin D. Latkar * and S. A. Patil **
* Electronics & Telecommunication Engineering Department, Gharda Institute of Technology, Khed, India.
Email: latkarsachin@gmail.com
** Professor, Electronics & Telecommunication Engineering Department, Gharda Institute of Technology, Khed, India.
Email: sapatil.mah@gmail.com

Abstract: This work surveys routing protocols employed in Wireless local area Network. There are many applications of wireless LANS. The applications embrace in-home networks, campus-sized networks, complete mobile networks on airplanes and trains. Wireless local area Network (WLAN) hot-spots are created in restaurants and hotels and users can access the web from these hot-spots. Most of the time these public networks do not need a password to join the network. Different WLAN hot spots need registration and/or registration fee to access the network. The state of the system in such cases is not static and can change which ends into an amendment in physical and logical topologies used. The convergence by a typical routing protocol is required to handle the problem. This work reviews a number of the quality routing protocols in terms of their characteristic, practicality, economical route discovery mechanisms and comparison of their several merits and demerits in Wireless local area Network. The protocols are classified as a table driven (proactive), reactive (on-demand), power aware, hierarchical and geographical multicast routing protocols.

Keywords: Wireless local area Network (WLAN), Routing Protocols, Proactive, Reactive, Hybrid.

Introduction

A routing protocol is a standard which is used for communication between computer units, exchange packets and discover routes to destination hosts. Routing protocol distributes information in the network. [2] It uses wireless transceiver at the interface. The functions performed by routing protocol are route discovery; route keeping; convergence of network topology; harmonized exchange of packets.

Wireless LANs became popular within the home [1] thanks to simple installation and use, and in business complexes providing wireless access to their customers; usually for free. New York town, for example, has begun a pilot program to supply town employees in all 5 boroughs of the town with wireless web access. The wireless network is classified into two types viz. infrastructure network and ad hoc network.

Infrastructure network is a network with a stable and wired gateway. There is a centralized controlling agency in the infrastructure network. Typical applications [2] of this type include office wireless local area networks (WLANs). In ad hoc network there is no centralized controlling agency, all the nodes are capable of movement and can be connected dynamically. Ad hoc networks are useful in the emergency type of applications which include search-and-rescue operations, meeting and conventions which share information and data acquisition operation in hospital terrain.

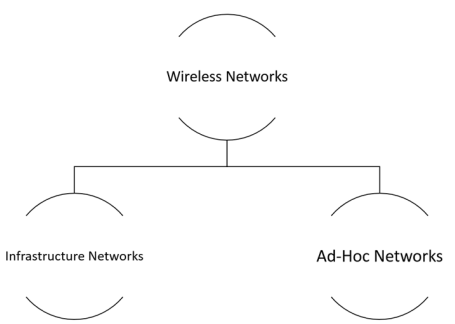


Figure 1: Types of Wireless Networks

Problem Statement

Limited and inadequate resources available in Mobile Ad hoc Network (MANET) build the planning of efficient and reliable routing protocol tough task. It is necessary to use these restricted resources. Intelligent routing algorithmic rule [3] is required that uses these restricted resources whereas at the same time being adaptable to propellant network conditions like network size, traffic density, a mobility of nodes and broken routes.

Quantitative properties embrace throughput, delay in routing, route discovery time, packet delivery ratio, jitter etc... Most of the routing protocols are quantitatively and qualitatively enabled.

Hence, it is to visualize fully totally different routing protocols so as that it will enable and aid the researcher or engineer to select best routing protocol in step along with his or her work. This work evaluates major routing protocols used in WLAN. The performance analysis of routing protocols is meted out.

Review of Related Research Efforts

In ad hoc Network cluster communication is a lot of vital, during which routing protocols play an important role for data transmission. Using or not using central server or access point, the Wireless network form a short-lived network with a collection of wireless nodes during which, every node changes indiscriminately at completely different times. So as to ascertain data transmission between nodes, multiple hops are required due to restricted range i.e. transmission rate.

Mobile Ad hoc network (MANET) routing protocols are classified according to several criteria, reflecting fundamental design and selection choice for implementation [2]. Numerous routing protocols have been proposed and developed for wireless local area networks.

Reference [21] analyses and simulates a planned Wireless local area Network (WLAN) using different routing protocols. The performances of various protocols are compared and analyzed using Optimum Network Performance (OPNET) simulator tool during which metrics like delay, throughput, packet delivery, load, Ethernet delay, are measured.

Reference [2] carries out the overview of routing protocol for wireless local area network. This work takes an overview of some important routing protocols used in WLAN by classifying them in terms of their characteristics and functionality, efficient route discovery mechanisms, and comparison of their particular advantages and disadvantages in a wireless local area network (WLAN).

Reference [3] presents review and a comparison of the typical representatives of MANETS (Mobile Ad Hoc Networks) routing protocols. An Ad hoc network is an assembly of wireless mobile nodes forming a temporary network without the aid of any centralized control or infrastructure. Such networks have no permanent topology due to the high degree of node movement. Hence, efficient and reliable routing is one of the key challenges in mobile ad hoc networks. Many routing algorithms have been proposed and evolved for accomplishing this task. Therefore, it is difficult to decide which protocol performs best under a number of dissimilar scenarios. Hence, the work undertakes the selection of routing protocol for MANETs.

The IEEE 802.11e is developed to supply (Quality of Service) QoS capabilities to WLAN, giving revelatory enhancements to multimedia system traffic. Since the widest deployed and used wireless interfaces are IEEE 802.11 based. Reference [22] expose results relative to the correlation of reactive routing protocols for MANETs and also the IEEE 802.11e technology. It is found that substantial enhancements in terms of throughput and normalized routing overhead are achieved because of enlarged routing responsiveness. The relation between the behavior experienced in every case and also the internal mechanisms of the routing protocol getting used are elaborated giving a holistic read of the phenomena. It is an insight into the interaction of routing protocols and also the MAC implementations of IEEE 802.11 and IEEE 802.11e.

Wireless Mesh Networks (WMNs) will offer wide range Wireless local area Networks (WLANs) area by connecting Access Points (APs) of WLANs with one another using radio communications. A routing protocol is extremely vital to stay communication quality over radio multi-hop communications as a result of radio waves are impacted a lot of by close surroundings. Once multi-user distributed applications like a video conference and an IP phone are used, it is foreseen that a large quantity of traffic flows on the network. Therefore, it is necessary to take into account network masses to use these applications. Reference [23] proposes a multicast routing protocol for WMNs that considers network loads and hop count. Moreover, the performance of the algorithm is assessed via simulation. It is shown via simulation results that the proposed algorithm has higher performance than a traditional multicast ad-hoc on demand protocol (MAODV) at the high loaded state of affairs.

Reference [24] surveys regarding IEEE 802.11 Wireless local area Networks thoroughly. Wireless Local Area network physical layer deals with transmission and reception of signals. The reference incorporates IEEE 802.11 standards, Wireless Local Area network protocol design, its benefits and its limitations.

Reference [25] is curious about developing a rapidly workable model of 802.11's result on network behavior. The interest is derived from investigations into routing algorithms for large scale ad-hoc networks, executing on parallel architectures. As their curiosity is not in the MAC layer but in the routing, they anticipate that a rapidly executed model of 802.11 can accelerate simulations targeted on routing problems whereas giving "good enough" estimates of packet latency, throughput, and loss here are vital performance advantages to simulating a wireless network's mac layer with a model that is easier than

true 802.11. For serial simulation, the key performance the benefit is due to reduction of events required. For parallel simulation, the overwhelming performance benefit is from better look ahead. The key contributions of this reference are to report on the implementation of 802.11 within the DaSSF (high-performance simulation kernel) framework, the recognition of 802.11 model simplification as a worthy goal, and preliminary results that ensure their intuition that such a simplification will yield the necessary performance gain.

A growing need to have ubiquitous connectivity has intended reference [26] to research to supply a nonstop association between varied wireless platforms like cellular networks, WLANs, and MANETs. Reference think about integration at the routing layer and suggest two adaptable routing protocols viz. Integrated Routing Protocol with reactive gateway discovery (IRP-RD) and with proactive gateway discovery(IRP-PD). The protocols exploit topology information stored at the fixed network elements (cellular Base Stations and WLAN Access Points) for the route discovery and maintenance processes. The proposed protocols can give connectivity to the cellular network and/or WLAN hotspots through multi-hop routing whereas disagree within the gateway discovery approach used. In IRP-RD, multi-hop routes to gateways to the cellular network or a WLAN hot spots are discovered on demand, whereas in IRP-PD out of coverage users proactively maintain routes to the gateways. Moreover, planned protocols will be utilized in any heterogeneous state of affairs, combining a cellular network and WLANs operative in infrastructure or ad-hoc (MANET) mode. They give simulation results that demonstrate the effectiveness of the planned integrated routing protocols and show the benefits and disadvantages of every gateway discovery approach in several heterogeneous situations.

Classification of Routing Protocols in WLANs

The routing protocols are essential in a wireless network. Many routing protocols have been designed to nullify the effect of topological changes that may result because of dynamic and random nature of MANETs. Every routing protocol is intended with specific functions and characteristics with improvement to existing protocol. Many studies have been carried out in wireless LAN routing protocols in order to define a set of protocols that will enhance bandwidth utilization, minimize energy consumption, higher throughput, less overhead, cost etc...

The routing protocol classification [2] is given below with their nomenclatures described in actual theory.

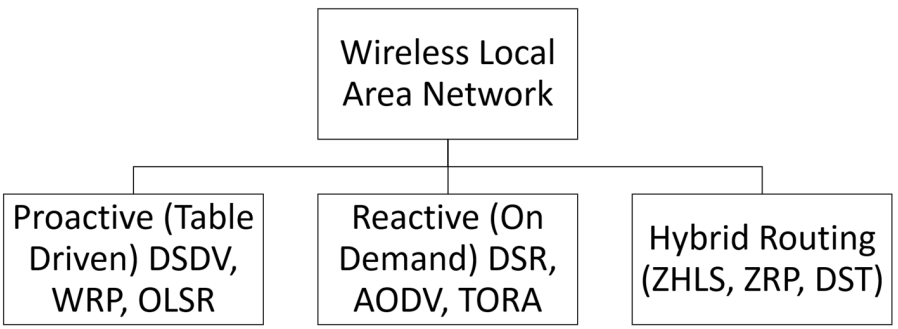


Figure 2: Classification of WLAN Routing Protocols

Proactive Routing Protocol (Table-Driven)

In table driven Routing Protocols [4] each node maintains routing tables. The tables contain routing information to other nodes in the network. All the nodes update these tables so as to maintain up-to-date and harmonies view of the network. The nodes propagate broadcast update messages when the network topology changes. The protocols differ across a number of tables and how the routing information is propagated. But it incurs additional overhead in propagating routing information.

Destination Sequenced Distance Vector Routing Protocol (DSDV)

The protocol [5] is inherited from Routing Information Protocol (RIP). Each node maintains a table which keeps distance or cost information. The routing table gives optimized distance to each destination and track to get there. The routing information is periodically updated to keep routing tables up to date. This protocol addresses poor looping properties of RIP when the links are broken. The modification adopted makes it suitable routing protocol for ad hoc networks.

DSDV protocol [27] guarantees loop-free methods. The protocol reduces the count to infinity problem. The extra traffic will be avoided with progressive updates rather than full dump updates. DSDV maintains solely the most effective path rather than maintaining multiple ways to each destination. With this, space in routing table is reduced.

The inessential advertising of routing data [27] results in wastage of bandwidth albeit there is no amendment within the topology. DSDV does not support Multipath Routing. It is troublesome to work out a time delay for the advertisement of routes. It is very tough to take care of the routing table's advertisement for a bigger network. Every and each host within the

network ought to maintain a routing table for advertising. Except for the larger network, this might result in overhead, that consumes additional bandwidth.

Wireless Routing Protocol (WRP)

It is table driven distance vector routing protocol [7]. Each node in the network maintains following tables

1. Distance Table (DT)
2. Routing Table (RT)
3. Link Cost Table (LCT)
4. Message Retransmission List (MRL)

Distance table (DT) contains matrix where each element contains distance and neighbor for a particular destination. The Routing Table (RT) contains up to date view of the network for all known destinations. The Least Cost table (LCT) contains the cost of percolating information through each link. The cost is calculated in terms number of hops to reach the destination. The MRL contains a record for every update message that is to be retransmitted and has a counter for each entry. The counter is decremented by one for each retransmission. Each update message contains updates that need to be carried out. Each node marks node in the RT table which needs to acknowledge the update for the message transmitted to that node. Once the counter reaches zero, the updates for which acknowledgments have not been received are to be retransmitted again. After successful transmission of the update message, the message needs to be deleted. After receiving update distance from update message, it also checks distance for its neighbors and hence achieves faster convergence.

WRP [28] has a similar advantage as that of DSDV. Additionally, it has quicker convergence and involves fewer table updates. however, the complexness of maintenance of multiple tables demands a bigger memory and larger process power from nodes within the ad hoc wireless network. At high mobility, the control overhead concerned in updating table entries is sort of a similar as that of DSDV and thus is not appropriate for extremely dynamic and also for a really large ad hoc wireless network. WRP demands large memory storage and resources in maintaining its tables. The protocol is not appropriate for large mobile ad hoc networks because it suffers from restricted scalability.

Optimized Link State Routing Protocol (OLSR)

It is a routing protocol [8] for mobile ad-hoc networks and is proactive in nature. It derives the stability from the original link state algorithm. Due to proactive nature it has the advantage of roots becoming available when needed. In the all the link state protocols, all the links with the neighbor nodes are declared and flooded in the entire network. OLSR is an optimization of pure link state protocol. First, it reduces the size of control packets. It only declares a subset of links with its neighbor nodes who are its multipoint relays. This reduces the number of retransmissions in flooding or broadcast procedure.

The protocol does not create extra control traffic in response to link failures and additions. The protocols keep routes for all the destinations in the network and hence are suitable for networks where large subsets of nodes are communicating with each other. It is also suitable for networks where source and destinations are changing with time. The protocol is suitable for large and dense networks where a selection of multipoint relays achieves more optimization.

The advantage of OLSR [30] is that it lowers control information and expeditiously minimizes broadcast traffic bandwidth usage. Though OLSR provides a path from supply to the destination, it is not essentially the shortest path, as a result of each route involves forwarding through an MPR (Multipoint Relay) node. An extra disadvantage is that OLSR additionally has routing waiting period and bandwidth overhead at the MPR nodes as they act as localized forwarding routers.

The protocol does not require reliable transmission of its control messages as nodes transmit their control messages periodically and hence can sustain a loss of packets. The protocol does not need in order delivery of messages as each control message contains the sequence number of recent information.

Source tree Adaptive Routing Protocol (STAR)

The protocol [6] uses least overhead approach compared to optimum routing approach. If a node 1 wants to communicate with node N and does not have a path in its source tree it sends an update message to all neighbors that there is no path to N. Neighbors that have path responds with update messages. Node 1 updates its source tree and may start transmission. In STAR, topology of network is modeled as directed graph

G (V, E); where V=set of nodes

E=set of links connecting edges.

A neighbor who has the path to destination sends its own tree in response. If the path is not available, then the node forwards the message until the alternate path is found. This is called link break maintenance mechanism in STAR.

Global State Routing Protocol (GSR)

It is similar to (Destination Sequence Distance Vector) DSDV protocol [7]. It uses link state routing but keeps away flooding of routing messages. Each node maintains Neighbor List, Topology table, Next Hop Table and Distance Table. The neighbor list contains the list of neighbors.

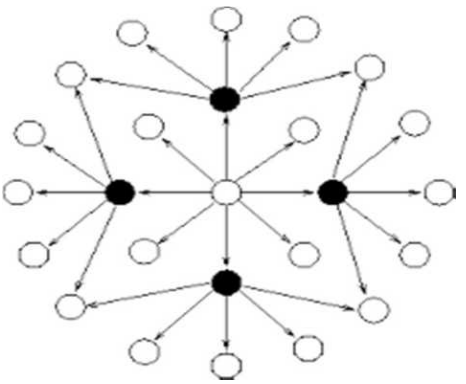


Figure 3: Multipoint Relay

For each Destination node, topology table contains link state information as announced by the destination and time stamp of the information. For each destination, the next hop table contains the address of next hop to which packets must be forwarded to reach the destination. The distance table contains the minimum distance to every node.
The drawbacks of GSR [31] are an out sized update message that consumes a substantial quantity of bandwidth and therefore the latency of the link phase change propagation, that depends on the update period.

Summary of Table-Driven Routing Protocols
Table-driven routing protocols keep routing table at each node [8]. The table is updated periodically. Routing Procedure is based on this data. Basically, Table-Driven Routing protocol has the main advantage [2] is that routing information is available to each node to initiate a session. As a result, nodes encounter higher overhead cost in order to maintain routing tables and convergence takes much time when there is link failure.

a. Link State Algorithms: Each node keeps a view of network topology with a cost for each link [8]. Link costs are broadcasted. Each node updates its topology and applies shortest path algorithm to find next hop to reach the destination.
b. Distance Vector Algorithms: Each node keeps for each destination a set of distances to get that destination [8]. The neighbor with minimum distance is selected for that destination. The node periodically broadcasts its routing table containing the best next node for each destination to each of its neighbors.

The comparison of proactive routing protocols is given below [11]

Table 1: Proactive Routing Protocols Comparison

Parameters	**DSDV**	**WRP**	**OLSR**
Routing Updates	Periodic	Periodic	Periodic
Loop- Free	Yes	Yes	Yes
Routing Overhead	High	High	Low
Caching Overhead	Medium	High	High
Throughput	Low	Low	Medium
Routing Tables	2	4	4

The proactive protocols [27] are not appropriate for bigger networks, as they have to keep up node entries for every and each node within the routing table of each node. This causes a lot of overhead within the routing table resulting in the consumption of a lot of bandwidths.

Reactive Routing Protocols (On Demand)
In Reactive Routing Protocols [9] route information for routes is maintained for active nodes only. Route discovery occurs by flooding route request packet within the entire network. Reactive routing protocols are classified into source routing and hop by hop routing. In source routing, each data packet carries complete source to a destination address. In Hop by Hop routing, nodes need not maintain neighborhood connectivity through periodic beacon messages.

Dynamic Source Routing (DSR)
It is a source based loop-free routing protocol [10]. Each node maintains route cache where it keeps the source routes learned by the node. The route discovery process is initiated when the source node does not have a valid route to the destination in

the route cache. The DSR protocol works well in small to medium size network with moderate mobility. It is beacon less protocol [2]. No HELLO messages are exchanged between nodes to notify neighbors in the network.
This protocol [32] employs a reactive approach that eliminates the necessity to periodically flood the network with table update messages that are needed in an exceedingly table-driven approach. The intermediate nodes conjointly utilize the route cache information with efficiency to scale back the control overhead. The disadvantage of this protocol is that the route maintenance mechanism does not regionally repair a broken link. Stale route cache information might additionally end in inconsistencies during the route reconstruction phase. The connection setup delay is more than in table-driven protocols. Even if the protocol performs well in static and low mobility environments. The performance degrades quickly with increasing mobility. Also, considerable routing overhead is involved owing to the source routing mechanism utilized in DSR. This routing overhead has directly proportionality with the path length.

Ad Hoc on-Demand Distance Vector Routing Rule (AODV)
It is the mix of each (Dynamic supply Routing) DSR and Destination Sequence Distance Vector (DSDV) routing protocol [11]. It ensures destination sequence number to ensure loop freedom at all times. AODV does not offer any kind of security rather than source routing [2], it depends on dynamically creating route table entries at Intermediate nodes.
AODV [27] can handle extremely dynamic behavior owing to its reactive nature. The protocol is used for each unicasts and multicasts communication.
The algorithm [27] expects that the nodes within the broadcast medium will observe every others' broadcasts. Overhead on bandwidth are going to be occurred compared to DSR, when a (Route Request) RREQ travels from node to node. While discovering the route information on demand, with the addresses of all the nodes through which it is passing, it sets up the reverse path in itself and it carries all this information all its way. AODV lacks an economical route maintenance technique. The routing information is usually obtained on demand, together with for common cause traffic. The messages are exploited for insider attacks together with route upset, route takeover, node isolation, and resource consumption. AODV is intended to support the shortest hop count metric. This metric is kind to long, low bandwidth links over short, high-bandwidth links. The AODV does not find a route until a flow is initiated. This route discovery latency result is high in large-scale mesh networks.

Temporally- Ordered Routing Algorithm (TORA)
It is an adaptive distributed algorithmic rule for multihop ad hoc networks. [12] It is purposely engineered for quick dynamical network topologies. It has 3 parts that are construction phase, maintenance phase, and destruction phase. It uses link reversal procedure.
Link reversal routing (LRR) protocols [13] are developed for fast changing topology networks where conventional routing protocols are not working anymore. But at the same time it is used for the networks where change is not so fast that it will flooding will be the only possibility. LRR protocols do not necessarily give optimal route from source to destination, but it does not matter in this kind of situations.
One of the merits of TORA [29]is that the more than one routes between any source-destination pair are supported by this protocol. Therefore, removal or failure of any of the nodes is quickly resolved while not source intervention by a change to an alternate route. TORA is additionally not free from disadvantages. One in every of them is that it depends on synchronic clocks among nodes within the ad hoc network. The dependence of this protocol on intermediate lower layers for certain practicality presumes that the link status sensing, neighbor discovery, in order packet delivery and address resolution are all readily offered. The solution is to run the web MANET Encapsulation Protocol at the layer now below TORA.This can create the overhead for this protocol tough to break away that obligatory by the lower layer.

Associatively-Based Routing (ABR)
It is bandwidth efficient distributed routing protocols [14]. It is source initiated on-demand routing protocol. ABR uses point to point and broadcast routing. The destination node in the ABR chooses the route based on associability. The ABR has three phases route discovery, route reconstruction, and route deletion.
ABR routing protocol defines a new kind of routing metric known as, "Degree of Association stability" [15]. In this protocol, the route is chosen based on "degree of association stability" of mobile nodes. every node periodically creates a beacon to announce its existence. Upon receiving the beacon, a neighbor node updates its own associativity table. For every received beacon, the associativity of the beaconing node with the neighboring node is increased. A high value of associativity tick for any beaconing node means the node is comparatively static. Associativity tick is reset once the neighboring node moves out of the neighborhood of the other node.
The ABR [33] avoids packet duplications. The route reconstructions are absent in the protocol. The protocol is complex in terms of operation and communication.

Signal Stability–Based Adaptive Routing Protocol (SSA)
The protocol performs on-demand route discovery by choosing longer lived routes [16]. The route discovery is predicated on signal strength and site stability. The signal strength criterion permits the protocol to form a distinction between robust and weak channels. Every channel is characterized by average signal strength at that packets are interchanged between hosts at either end of the channel. The location stability criteria select a channel that has existed for the longer amount of time. Along these 2 ideas select robust channels and exist for the longer period greater than some threshold. The protocol [34] reduces path failure by signal stability but the overhead is more.

Summary of Reactive Routing Protocols
The source node sends the request packet in order to find a route to destination node [2]. The source node floods the packet to all the nodes in the network. The route path followed by the request packet is saved and sends back to source node by the destination node. As the request packet is flooded it generates multiple reply packets and hence multiple routes. The shortest route is used. It is dynamic strategy since each node can update its routing table when topology information is received.
Fresh routes require fewer calculations for data transmission. The disadvantage in the method is that each route each intermediate node must store and maintain routing information for each route. Each node may be aware of its surrounding neighbors through the use of beaconing messages. As it is reactive (on-demand) routing protocol routing information is provided as per need and periodic updates are not required.
The reactive routing protocols comparison [11] is given below

Table 2: Reactive Routing Protocols Comparison

Parameters	**AODV**	**DSR**	**TORA**
Route Generation	By Source	By Source	Locally
Periodic Updating	No	No	No
Performance Metrics	Speed	Shortness	Speed
Routing Overhead	High	High	High
Caching Overhead	Low	High	Medium
Throughput	High	Low	Low
Multipath	No	Yes	Yes
Route Updating	Non -Periodic in nature	Non -Periodic in nature	High Routing Overhead

Hybrid Routing Protocols
Hybrid routing protocol is a combination of proactive and reactive routing protocols. Proactive and reactive techniques are used to route packets. The route is established with proactive techniques and uses reactive flooding for mobile nodes. Sometimes these protocols are also referred to as hierarchical routing protocols. The protocol divides the networks into clusters. Each cluster has a cluster head. Cluster head keeps information about other clusters. Other nodes maintain information about their own clusters. A collection of clusters is called as a supercluster. Such types of protocols have the advantage of both tables driven and on demand approaches. [17]

Zonal Routing Protocol (ZRP)
ZRP uses table driven approach inside the zone and on demand approach outside zone [17]. A zone is created based on radius. If the radius is equal to one, then nodes need to maintain a table of routes to reach one hop neighbor alone. If the radius is equal to 2 then concerned nodes need to maintain the table of routes to reach its two-hop neighbors. So, in this case, table driven routing with the distance of two is used inside the zone and on-demand routing is used outside the zone.

Zone-Based Hierarchical Link State Routing Protocol (ZHLS)
It is hierarchical protocol in which network is divided into non-overlapping zones [18]. In addition, mobile nodes are assumed to know their locations through systems like GPS (Global Positioning System). Each node knows the node connectivity within its node and zone connectivity of the entire network. All the nodes in the ZHLS maintain two routing tables viz. namely intra-zone routing table and inter-zone routing table. The protocol uses hierarchical address scheme which contains node ID and zone ID. There are two kinds of link state revisions- the node level LSP (Link State Packet) and Zone Level LSP.
A node periodically broadcasts its node level LSP to all other nodes in the zone in which it resides. Gateway nodes broadcast zone level LSP when a virtual link is broken or created. That is why every node knows the zone level topology of the entire network.

Before sending packets, a source checks its intra-zone routing table. If the destination node is in the same zone as the source the routing information is already present. If this is not the case, then source sends a location request to all other nodes through gateway nodes. The zone in which destination node is situated replies with location response containing zone ID of the destination. During the journey of the packets through zones, inter-zone routing table will be used and when the packet arrives the destination zone then intra-zone routing table will be used.

Sharp Hybrid Adaptive Routing based protocol (SHARP)
This protocol [18] maintains the balance between proactive and reactive routing. It adapts efficiently between proactive and reactive strategies. The protocol defines proactive zones around some nodes. A node specific zone radius determines the number of nodes within given proactive zone. Nodes within a proactive zone maintain pro-actively to a central node. SHARP maintains proactive routing zones around popular destinations and for the nodes that have little or no data traffic it will maintain will rely on purely reactive traffic. By increasing the radius of the proactive zone, the protocol can decrease the loss rate and variance in delay but will increase the packet overhead to maintain the routes in larger zone.

Distributed spanning trees based routing protocol (DST)
The nodes in the network are grouped into a number of trees [11]. Each tree has two types of nodes: route node and an internal node. The root node controls the structure of the tree and decides whether the tree can be merged with other trees. Other nodes serve as regular nodes. Router, merge and configure are the three states of the node and the node can be in one of the three states depending on a type of task it is trying to perform.
DST proposes two strategies to propose route between source and destination pair. The first approach is Hybrid Tree Flooding (HTF) in which source sends control packets to all neighbors and adjoining bridges in the spanning tree. Each packet remains static at these places for a specific holding time. The second approach is Distributed Spanning Tree (DST) shuttling in which source sends control packets to edges till each of them reaches a leaf node. When a packet reaches the leaf node it is forwarded to shuttling level.
The drawback of such architecture is an existence of a single point of failure for the entire tree. If the route node fails, then entire routing structure falls apart. Further holding time used to buffer the packets may introduce an extra delay in the network.

Summary of Hybrid Routing Protocols
Hybrid protocols are a new generation of protocols which is a combination of proactive and reactive routing protocols [11]. The protocols provide scalability and eliminate the single point of failure. Bottleneck node creation in the network is avoided by allowing any number of nodes to perform data forwarding if the path becomes available. The difficulty of hybrid protocols is to organize the network according to required parameters. The protocol attempts to keep down the number of rebroadcasting nodes by defining a zone which allows the node to work together. The best or suitable nodes can perform route discovery. The disadvantage of hybrid routing protocol is that nodes that have high topological information keep more routing information which requires more memory and power consumption.
The comparison of some of the hybrid routing protocols [11] is given in Table III.

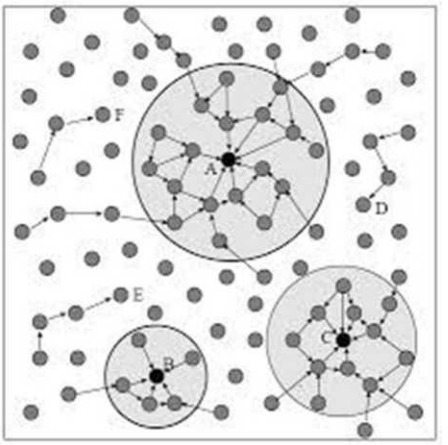


Figure 4: The SHARP Proactive and Reactive Zones

Preview of Other Routing Protocols

Other Routing Protocols

An efficient geographic multicast routing protocol (EGMP) is proposed in [19] which can scale to large group size and network size. The protocol efficiently implements multicasting delivery and management of membership for a group. The network piece of ground is split into geographical non overlapping square zones and a zone elects the leader to take charge of local group membership management. A zone based bidirectional multicast tree is formed to connect those zones having group members and such tree can utilize network resource effectively. The protocol [35] is an efficient multicast protocol which is suitable for dealing with large group size networks.

Geographical routing [20] is another type of routing which uses location information to formulate efficient route search towards a destination. Geographical routing requires propagation of single hop information to find a route towards the destination. The localized approach of the protocol reduces a need for maintaining the routing tables and reduces control overhead. It does not require flooding. The protocol is scalable, supports mobility and introduces minimal overhead.

Power-aware routing protocols [2] refer to set of protocols which are power aware. The nodes in the MANET are typically powered by batteries with limited energy supply. One in every of the foremost difficult issue in MANETs is to how to conserve energy and increase a lifetime of the nodes. It means that to increase the life of the network itself. Some routing protocols have been designed to take these challenges. Power-aware routing protocols are needed to be considered when energy saving is important criteria.

Minimum Net Transmission Power Routing (MTTPR) is a basic power aware routing protocol that forever tries to minimize the total power of the entire network. It does so by selecting the minimum hop count route. The metric “minimize the energy consumed per packet” is executed.

Every individual node battery life is not taken care of by MTTPR algorithm, so Minimum Battery Cost Routing (MBCR) algorithm is suggested by introducing extra battery cost function, that is, the inverse of the remaining battery capacity. This implies that if the remaining battery power decreases, the value operate can rise. This algorithm first finds the battery for each node of the network and then finds the battery cost function.

Table 3: Comparison of Hybrid Routing Protocols

Parameters	**ZRP**	**ZHLS**	**DST**
Routing Formation	Flat	Hierarchical	Hierarchical
Multiple Routes	No	Yes	Yes
Beacons	Yes	No	No
Route Information Stored in	Intra-zone and Inter-Zone Routing Tables	Intra-zone and Inter-Zone Routing Tables	Route Tables
Route Metric	Shortest Path	Shortest Path	Forwarding using Tree Neighbors
Advantages	Reduced Transmissions	Low Control Overhead	Reduced Transmission
Disadvantage	Overlapping Zones	Static Zone Map Required	Root Node

The comparison between three different categories of routing protocols [11] is given below

Table 4: Comparison of Routing Protocols On Classification

Parameters	**Table Driven**	**On Demand**	**Hybrid**
Storage Requirements	Higher	Dependent on number of routes maintained or required	Depends on size of each zone or cluster
Route Availability	Always available	Computed as per need	Depends on location of destination node
Periodic Route Updates	Required Always	Not required	Used inside each zone
Delay	Low	High	for local destinations it is low and high for inter-zone
Scalability	100 nodes	>100	>1000
Control traffic	High	Low	Lower than other two types
Routing Information	Keep stored in table	Does not store	Depends upon requirement.
Routing Philosophy	Mostly flat	Flat	Hierarchical

Conclusion

This work takes an overview of a number of routing protocols used in a wireless network. The protocols are chiefly classified as proactive, reactive and hybrid protocols. Other types like geographical routing and power aware routing are also discussed. Different types of comparisons have been carried out. The principal factor in all these protocols is finding an optimum route between source and destination with given conditions. The overview as presented in this work will go a long way in providing a platform for anyone to choose the best protocol for his/her work and to do necessary innovations if any.

References

[1] "https://en.wikipedia.org/wiki/Wireless_LAN"
[2] Nwoko T .C, Akanu. J .C, Alumona. T. L, "Overview of Routing Protocols in Wireless Local Area Networks", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 11, Year 2014, Pages 50-56.
[3] Wahi, Charu, and Sanjay Kumar Sonbhadra, "Mobile Ad Hoc Network Routing Protocols: A Comparative Study." International Journal of Ad Hoc, Sensor and Ubiquitous Computing, Volume3, Issue 2, Year 2012, Pages 21-31.
[4] "ttp://www.cse.wustl.edu/~jain/cis788-99/ftp/adhoc_routing/"
[5] He, Guoyou, "Destination-sequenced distance vector (DSDV) protocol",Networking Laboratory, Helsinki University of Technology ,Year 2002, Pages1-9.
[6] Hemagowri, J., C. Baranikumari, and B. Brindha, "A Study on Proactive Routing Protocol in Ad-hoc network", International Journal of Modern Engineering Research (IJMER), Year 2013, Pages 1- 4.
[7] Jacquet, Philippe, Paul Mühlethaler, Thomas Clausen, Anis Laouiti, Amir Qayyum, and Laurent Viennot. "Optimized link state routing protocol for ad hoc networks", In Multi Topic Conference, 2001, IEEE INMIC 2001, Technology for the 21st Century. Proceedings. IEEE International, IEEE, Year 2001, Pages 62-68.
[8] "http://users.cis.fiu.edu/~skala001/11-12-2004-DSDV-report.htm"
[9] Abolhasan, Mehran, Tadeusz Wysocki, and Eryk Dutkiewicz, "A review of routing protocols for mobile ad hoc networks", Ad hoc networks, Volume 2, Issue 1, Year2004, Pages 1-22.
[10] Tuli, Ruchi, "Ad hoc Mobile Wireless Network Routing Protocols-A Review", TECHNIA – International Journal of Computing Science and Communication Technologies, Volume 2, Issue 1, Year 2009, Pages 294-299.
[11] Gupta, Anuj K., Harsh Sadawarti, and Anil K. Verma, "Review of various routing protocols for MANETs", International Journal of Information and Electronics Engineering, Volume 1, Issue 3, Year 2011, 251-259.
[12] Kaur, Emanpreet, and Abhinash Singla, "To Compare the Performance of Tora and Dsr Based On the Size of the Network", Volume 3, Issue 7, Year 2015, Pages 44-54.
[13] Vainio, Miia, "Link reversal routing", Nokia Group, Year 2002.
[14] Geetha, K., and P. Thangaraj., "An Enhanced Associativity Based Routing with Fuzzy Based Trust to Mitigate Network Attacks."World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering, Volume 9, Issue 8, Year2015, Pages 1627-1635.
[15] Dhenakaran S, Parvathavarthini A, "An Overview of Routing Protocols in Mobile Ad Hoc Network", International Journal of Advance Research in Computer Science and Software Engineering", Volume 3, Issue 2, Year 2013, Pages 251-259.
[16] Dube, Rohit, Cynthia D. Rais, Kuang-Yeh Wang, and Satish K. Tripathi, "Signal stability-based adaptive routing (SSA) for ad hoc mobile networks", Personal communications, IEEE, Volume 4, Issue 1, Year 1997, Pages 36-45.
[17] Sangheethaa, S., and P. V. Shyamily, "Zone-based Proactive Source Routing Protocol for Ad-hoc Networks." Compusoft, Volume 4, Issue 1, Year 2015, Pages 1477-1482.
[18] Raheja, Kanishka, and Sunil Kr Maakar, "A Survey on Different Hybrid Routing Protocols of MANET", (IJCSIT) International Journal of Computer Science and Information Technologies, Volume 5, Issue 4, Year 2014, Pages 5512-5516.
[19] Xiang, Xiaojing, Xin Wang, and Zehua Zhou, "An efficient geographic multicast protocol for mobile ad hoc networks", In Proceedings of the 2006 International Symposium on World of Wireless, Mobile and Multimedia Networks, IEEE Computer Society, Year 2006, Pages 73-82.
[20] Battula, Rama Sundari, and O.S. Khanna, "Geographic Routing Protocols for Wireless Sensor Networks: A Review", International Journal of Engineering and Innovative Technology (IJEIT), Volume2, Issue 12, Year 2013, Pages 39-42.
[21] Kalamani, P., Kumar, M. V., Chithambarathanu, M., & Thomas, R."Comparison of RIP, EIGRP, OSPF, IGRP Routing Protocols in Wireless Local Area Network (WLAN) by using OPNET Simulator tool-A Practical Approach", Volume 16, Issue 4, Year 2014, Pages 57-64.
[22] Calafate, C. T., Manzoni, P., & Malumbres, M. P. (2005, February), "On the interaction between IEEE 802.11 e and routing protocols in mobile ad-hoc network" ,13th Euromicro Conference on Parallel, Distributed and Network-Based Processing, IEEE, Year 2005, Pages 110-117.
[23] Oe, K., Koyama, A., & Barolli, L., "Proposal and Performance Evaluation of a Multicast Routing Protocol for Wireless Mesh Networks Based on Network Load", Mobile Information Systems, Year 2015.
[24] Singh, Richa Singh Dr Uday Pratap, "IEEE 802.11 Wireless Local Area Networks: A Review", International Journal of Engineering, Management & Medical Research (IJEMMR), Volme 1, Issue 3, Year 2015, Pages 7-16.
[25] Liu, J., Nicol, D. M., Perrone, L. F., & Liljenstam, M., "Towards high performance modeling of the 802.11 wireless protocol"' In Simulation Conference, Proceedings of the Winter, Volume 2, IEEE, Year 2001, Pages 1315-1320.
[26] Cavalcanti, D., Cordeiro, C., Kumar, A., & Agrawal, D. P., "Self-adaptive routing protocols for integrating cellular networks, WLANs, and MANETs", Wireless Communications and Mobile Computing, Volume7, Issue 3, Year 2007, Pages 375-397.
[27] Gorantala, K., "Routing protocols in mobile ad-hoc networks", A Master 'thesis in computer science, Year 2006, Pages 1-36.

[28] “https://en.wikipedia.org/wiki/Wireless_Routing_Protocol”
[29] “https://essays.pw/essay/comparison-of-tora-and-aodv-computer-science-essay-55713”
[30] Annamalai, P., “Comparative Performance Study of Standardized Ad-Hoc Routing Protocols and OSPF-MCDS”, Doctoral dissertation, Virginia Polytechnic Institute and State University, Year 2005.
[31] Machwe, Parikshit, Neha Jha, Sunil Arora, and Raghav Mittal. "Routing Protocols for Ad-hoc Networks."
[32] “https://en.wikipedia.org/wiki/Dynamic_Source_Routing.”
[33] Raja, L., & BABOO, S., “Comparative study of reactive routing protocol (AODV, DSR, ABR and TORA) in MANET”, International Journal of Engineering and Computer Science, Volume 2, Issue 3, Year 2013, Pages 707-718.
[34] Manimegalai T, Dr. Jayakumar C, “A Conceptual Survey of MANET Routing Protocols”, International Journal of Innovative Research in Science,Engineering and Technology,Volume 2, Issue 11, Year 2013, Pages 6125-6130.
[35] Srikalyani Paashikanti, Rajashekar Thallapalli, “Advanced EGMP Based Multicast for Ad-Hoc MIMO”, International Journal of Computer Trends and Technology, Volume 5, Issue 2, Year 2013, Pages 55-62.

True Time based Simulation for AODV in MANET

L Usha* and D K Kumuda**
* Assistant Professor, Dept. of Telecommunication, SIT
lusharamesh@gmail.com
** Assistant Professor, Dept. of Telecommunication, SIT
kumudamurthy@gmail.com

Abstract: AODV routing protocol has been extensively used for data transfer in MANET's due to high mobility of the nodes and wireless technology. There are several problems associated in the network. It is essential procedure for MANET network to design and optimize. We proposed True Time simulator for MANET's in MATLAB/simulink. We use computational intelligent technique, paths to capture networks dynamic behavior and in-node behavior. The simulation result show True Time could accomplish MANET'S simulation and displayed node energy under several conditions.

Keywords: AODV, MANET, True Time, Node energy, MATLAB/simulink.

Introduction

Need for ad hoc wireless communication has increased with more use of portable devices knows as Mobile ad hoc network (MANET). MANET consists of independent nodes that can communicate to each other through radio waves. MANET is an infrastructure less network in which nodes move arbitrarily [1][2]. Due to this mobility nature of the nodes they themselves act as router. The primary challenge in building MANET is to establish path between the communicating nodes. This type of network has power constraints since they depend on limited battery resources.

MANETS are self organizing and self configuring multi hop wireless networks where the structure of the network changes dynamically because of mobility of the nodes, broadcast nature of the network frequent network reconfiguration. Continuous change in position and connection consumes more energy. Therefore it is necessary to save energy to maintain the life time of the network in order to enhance the life time of nodes [3][4][5]. There are several simulator used for MANET network research such as NS-2, OPNET, OMNET++, TOSSIM etc. NS-2 is widely used simulation tool. Table 1 gives some elements of the dependability for each simulator [1].

The rest of this paper is organized as follows. Section I describes Ad-Hoc on-Demand Distance Vector routing protocol used for simulation. Section II introduces True time Kernel toolbox of MATLAB/Simulink. Section III gives implementation and Simulation Model details. Section IV describes the result analysis and conclusion is drawn in section V.

Ad-Hoc On-Demand Distance Vector

The Ad-Hoc On-Demand Distance Vector (AODV) is a reactive routing protocol. The protocol is used to discover the route when source and destination nodes cannot communicate directly i.e on-demand. The source node generates route request packet (RREQ) and broadcast to its neighbor nodes. The neighbor nodes checks its routing table for the destination node, if found sends route reply packet (RREP) for path establishment between source and destination node else it further broad cast to RREQ to its neighbor. This process is continued until RREQ reaches destination node. The destination node generates RREP and unicast in the reverse path. Thus path is established [6].

True Time blocks

Lund university, Sweden developed a co-simulation tool TrueTime (TT) in MATLAB/simulink. This provides a simulation environment for network control. TrueTime could be used to simulate computation within the nodes and dynamic behavior of the network. [7][8][9]

A group of simulink blocks are provided in TrueTime library (figure1). In simulation we have used TrueTime kernel (figure2), TrueTime battery and TrueTime Wireless Network (figure3).

Table 1: Elements of dependability: ganularity and mobility[1]

Name	Granularity	Metropolitan mobility
ns-2	Finest	Support
DIANEmu	Application- level	No
Glomosim	Fine	Support
GTNets	Fine	No
J-Sim	Fine	Support
Jane	Application-level	Native
NAB	Medium	Native
OMNet++	Medium	No
OPNet	Fine	Support
QualNet	Finer	Support
SWANS	Medium	-

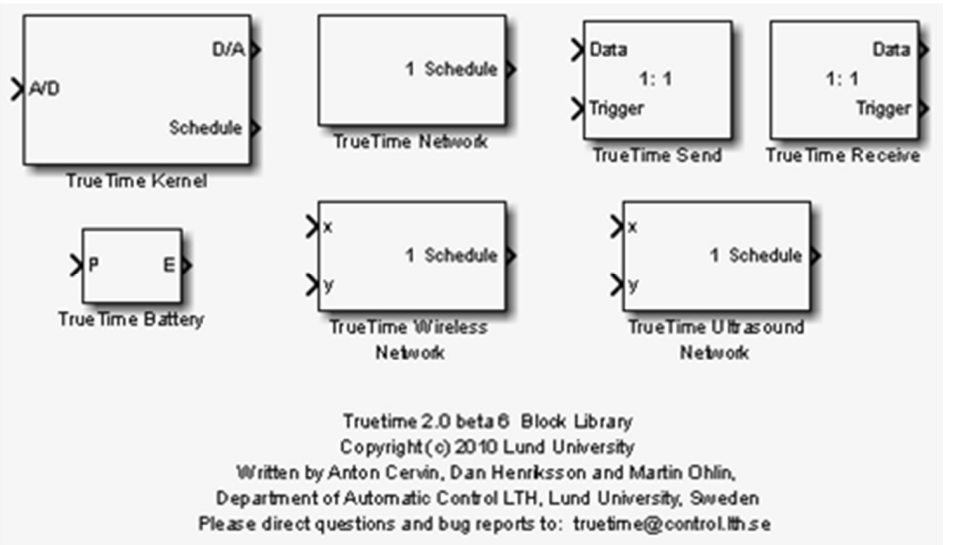


Figure1: TrueTime Block Library

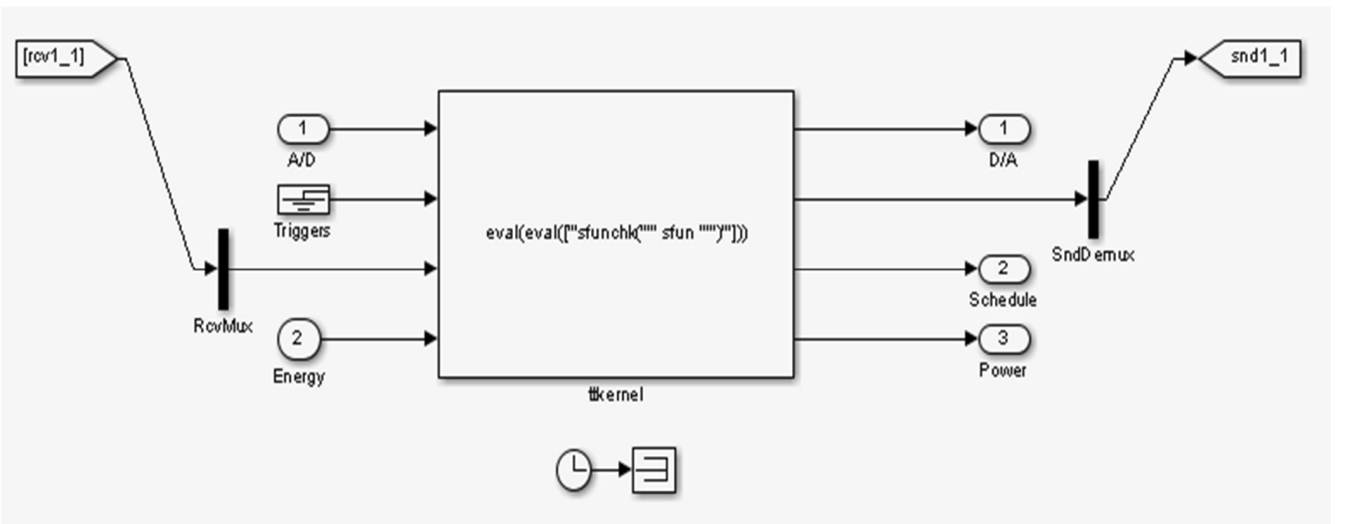


Figure 2: Internal structure of a TrueTime Kernel

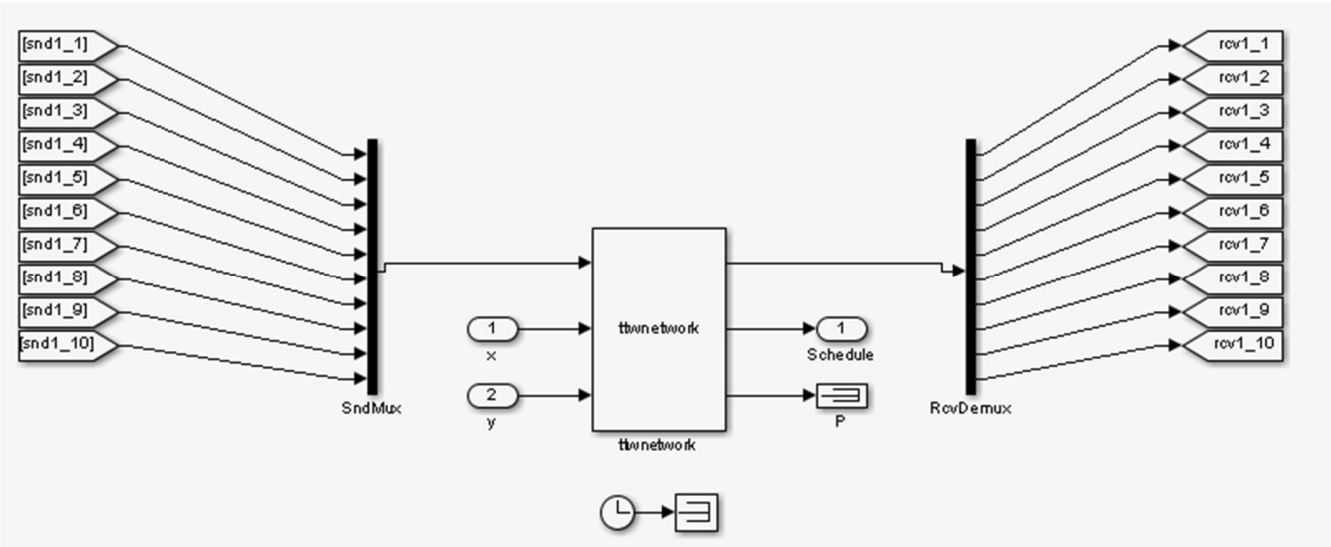


Figure 3: Internal structure of a TrueTime Wireless Network

Implementation and Simulation Model

The following parameters were considered for simulation (Table 2). Each node position was randomly given in the area of 400m*400m and each node was assigned with energy of 20J. Figure4 shows internal structure of one such node. Figure5 shows overall AODV implementation using TrueTime blocks.

Table 2: Simulation parameters

Transmission Power	-8dBm
Receiver Threshold	-48dBm
Data rate	1Mbps
Antenna	Omni directional
Path loss Exponent	3.5
Pause Time	0.5s
Hello message interval	1s

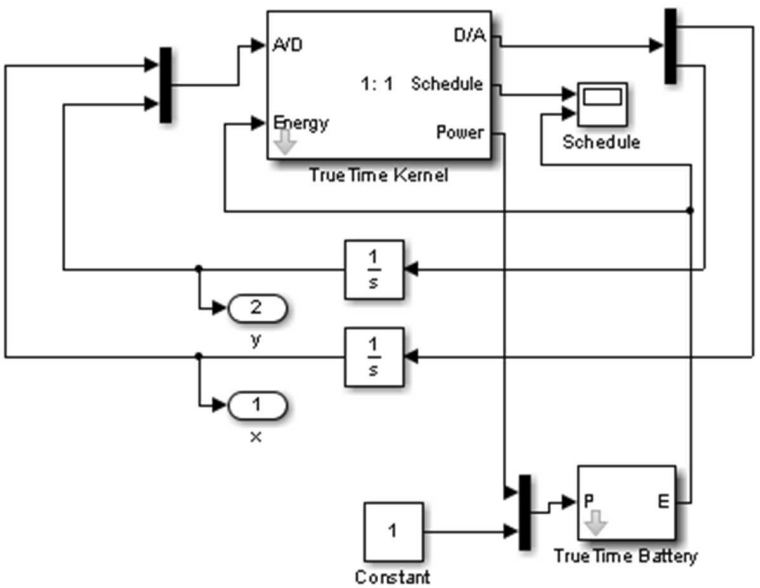


Figure 4: Internal structure of a Node

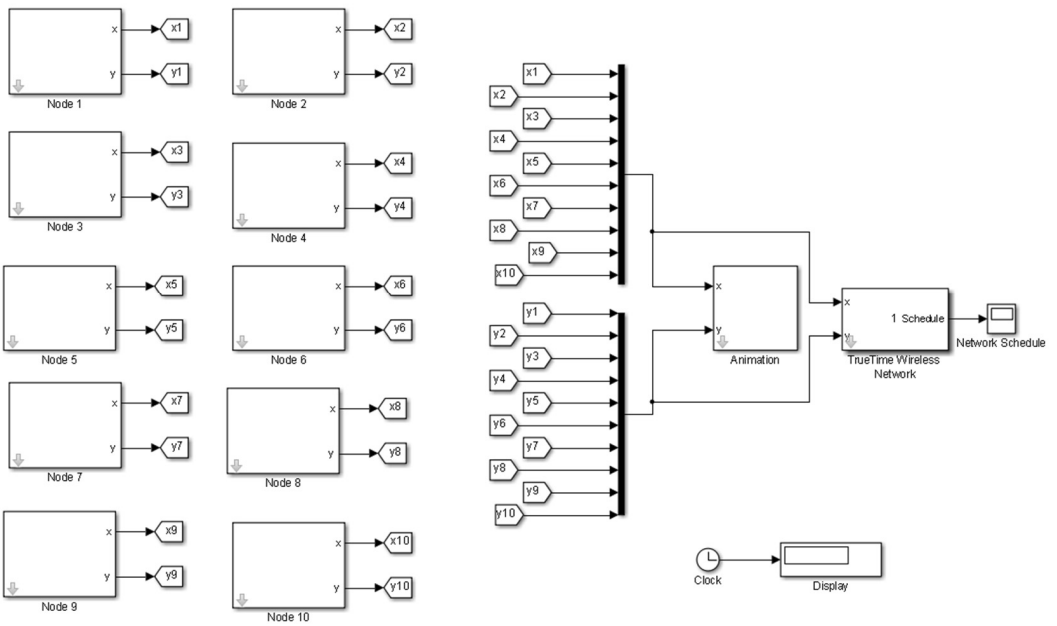


Figure 5: AODV TrueTime simulink model

Result Analysis

We simulated a MANET of size 10 nodes giving mobility for 2 nodes at different instants. Source and destination were randomly chosen. Figure 6 shows a scenario of node deployment. The instant at which RREQ is broadcasted, receiving RREP from destination; route establishment, link break etc are displayed in MATLAB command window as shown in Figure7.

The number of intermediate nodes between source and destination verses time taken for path establishment for 0%, 25% and 50% message loss in network is shown in figure8. Due to heavy load in the network message loss occurs; thereby time taken for path establishment in 25% and 50% cases is more compared to 0% loss. The packet loss during 0%, 25% and 50% were 6, 10 and 22 respectively.

The primary challenge in building MANET is to establish path between the communicating nodes. This type of network has power constraints since they depend on limited battery resources. Continuous change in position and connection consumes more energy. It is therefore required to known the energy status of the node.

The energy consumption of node for path establishment varies under different circumstances. For examples retransmission of RREQ or data or multipath establishment. Assuming 10%, 25% and 50% if its initial energy is consumed for path establishment. Energy consumption for one node is shown in figure 9 for a simulation time of 20s.

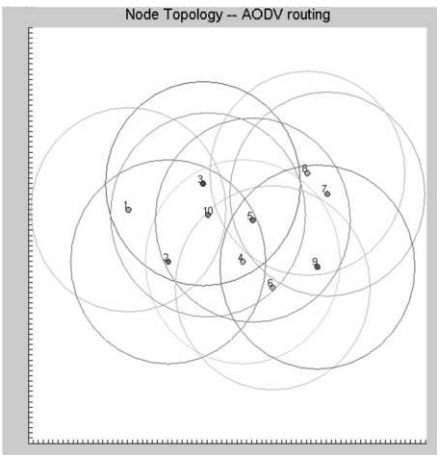


Figure 6: Node deployment scenario

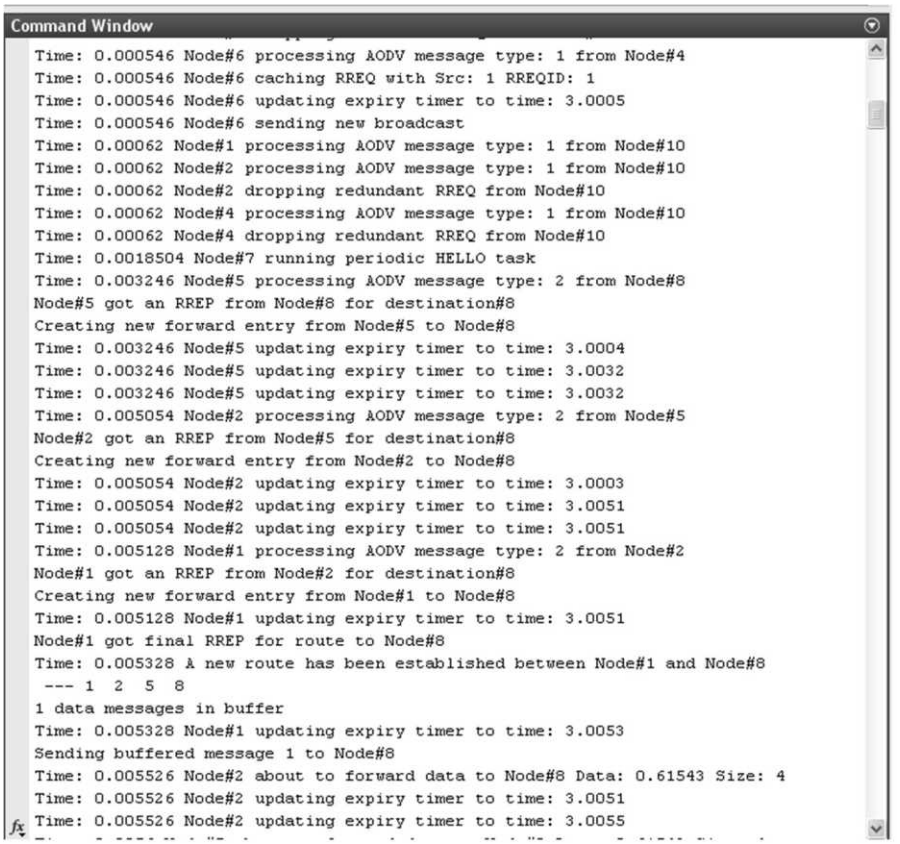


Figure 7: AODV routing display in MATLAB command window

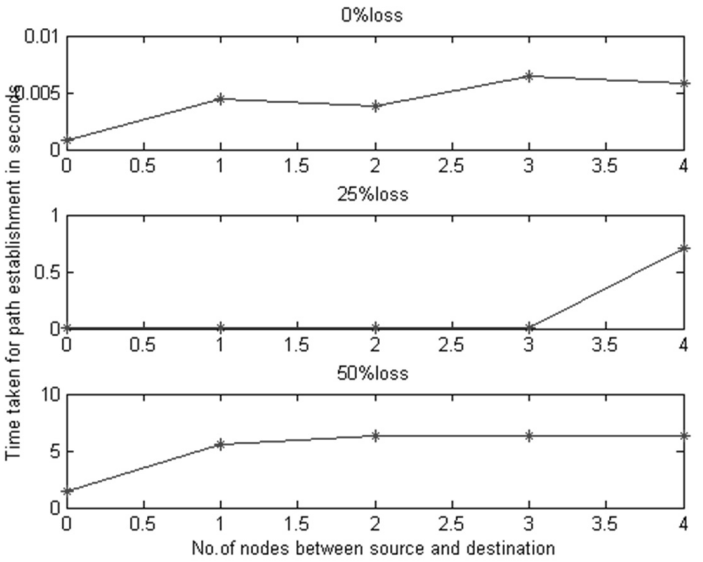


Figure 8: Time taken for path establishment

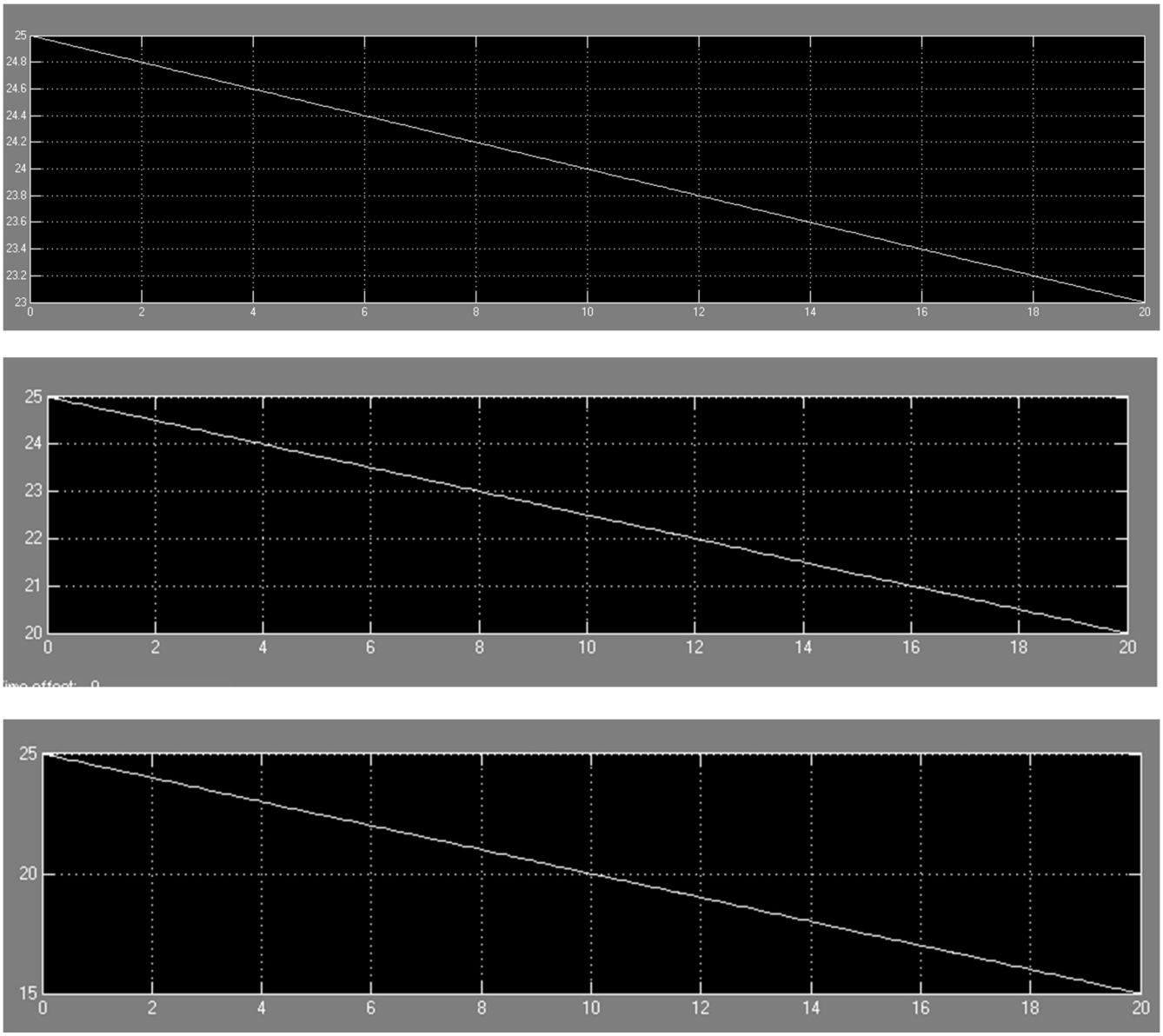


Figure 9: Energy consumption

Conclusion

In this paper, we have used TrueTime as a tool to implement AODV. We give an example of simulation which use 10 nodes deployed 400m*400m. We have simulated the network and in-node behavior. Energy is a constraint parameter in MANET's, it is necessary to save energy to maintain the life time of the network in order to enhance the life time of the MANET, energy efficient routing protocols are required. AODV does not consider any security schemes; it is vulnerable to various types of attacks. Security schemes can be considered in design of AODV in MANET's. If these parameters are included in AODV, source node can decide the data transmission in the path looking into the energy status and security from the route table.

References

[1] Luc Hogie, Pascal Bouvry, Fr´ed´eric Guinand, "An Overview of MANETs Simulation" Electronic notes in Theoretical Computer Science 150 (2006) 81-101

[2] Aditya Bakshi, A.K.Sharma, Atul Mishra, "Significance of Mobile AD-HOC Networks (MANETS)", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-2, Issue-4, March 2013

[3] Aarti, Dr. S. S. Tyagi, "Study of MANET: Characteristics, Challenges, Application and Security Attacks", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, May 2013
[4] K.Parkavi, Dr.P.Vivekanandan, "Energy aware secure routing protocol for mobile adhoc networks", International Journal of Advanced Research in Computer and communication Engineering, Volume 2, Issue 5, May 2013
[5] S.Nithya, S K Deepika, G Sindhu, "A review of energy aware routing protocol in MANET", International Journal of engineering and advanced technology, Volume 2, Issue 2, December 2012
[6] Charles E Perkins , Elizabeth M Royer, " Adhoc OnDemand Distance Vector Routing", In IEEE International Conference on Mobile computing syatems and applications, February 1999.
[7] http:// www.control.lth.se/truetime
[8] Shodhganga.inflibnet.ac.in:8080/jspui/bitstream/.../13_chapter%205.pdf
[9] Chang Chen, Zhenping Li, Ping Song and Kejie Li, "True Time based simulation for wireless Senor network", In IEEE International Conference on Computational Intelligence Software(CiSE), December 2010.

A Model to Enhance the Performance of Distributed File System for Cloud Computing

Pradheep Manisekaran* and M R Ashwin Dhivakar**
*Assistant professor, Department of Computer Science and Engineering, NIMS University, Jaipur, India
Email: pradheep45@hotmail.com
**Research Scholar Jaipur National University Jaipur
Email: ashdhiv@gmail.com

Abstract: Cloud computing is a new era of computer technology. Clouds have no borders and the data can be physically located anywhere in any data center across the network geographically distributed. Large scale distributed systems such as cloud computing applications are getting very general. These applications come with increasing challenges on how to transfer and where to store and compute data. The most current distributed file systems to deal with these challenges are the Hadoop file system (HDFS) and Google file system (GFS). But HDFS has some issues. The most factors are that it depends on one name node to handle the majority operations of every data block in the file system. As a result, it may be a bottleneck resource and one purpose of failure. The second potential problem with HDFS is that it depends on TCP to transfer data. Usually, TCP takes several rounds before it will send at the complete capability of the links in the cloud. This results in low link utilization and longer downloads times. In such file systems, nodes simultaneously serve computing and storage functions; a file is divided into a number of chunks allocated to distinct nodes so MapReduce tasks may be performed in parallel over the nodes. However, in a cloud computing, the crash is the commonplace, and nodes could also be upgraded, replaced, and added to the system. Files can even be dynamically created, deleted, and appended. This results in load imbalance in a distributed file system; that's, the file chunks aren't distributed as uniformly as potential among the nodes. Growing distributed file systems in production systems powerfully depend upon a central node for chunk reallocation. This confidence is clearly inadequate in a large-scale, failure-prone setting as a result of the central load balancer is put out vital workload that's linearly scaled with the system size therefore, it become the performance bottleneck a single purpose of failure. Suppose we tend to save the files in cloud information and a few third party accesses those files and adds some extraneous information which will damage our system. thus to boost the performance and security of cloud computing in this thesis we use a new approach called load balancing with round robin algorithm.

Keywords: Cloud computing, File system, Distributed System, Storage System, Load balancing.

Introduction

Cloud computing is a compelling technology .in cloud users can dynamically store and access their resources without sophisticated deployment and management of resources by means of internet. Cloud computing is emerging as a new paradigm of large-scale distributed computing.it has moved computing and users data away from desktop, portable devices into large data centers.it has the capability to utilize the power of internet and wide area network to access the resources that are available remotely (e.g. software, storage, data, network).cloud computing has two broad categories such as cloud and cloud technologies. The term “cloud” refers to a collection of infrastructure services such as software as a service, infrastructure as a service, and platform as a service. The term “cloud Technologies” refers to various cloud runtimes such as MapReduce framework [1], Hadoop Distributed File System (HDFS), Google File System (GFS), etc.

Cloud computing involves distributed technologies to satisfy a number of users and applications by providing functionalities like resource sharing, software, hardware, information through internet.in order to reduce the capital and operational cost, and to increase the performance in terms of response time and data processing time, maintain the system stability. Day by day the number of users, amount of data, structure of the network is increasing rapidly so that there are lot of technical challenges involves in this process such as virtual machine migration, data transfer, bottleneck performance, unpredictability, server consolidation, fault tolerance, scalable storage, high availability and major issue is the load balancing. Dealing with these challenges of large scale distributed data computer and storage intensive applications such as search engines, cloud storage applications, and social networks require robust scalable efficient algorithms and protocols.

The google File System (GFS) which is used by google and Hadoop Distributed File System (HDFS) is a most common algorithm deployed in Facebook and yahoo today. Distributed file system are key building blocks for cloud computing application. Based on the MapReduce framework in such file systems nodes simultaneously serve computing and storage functions; a file is partitioned into a number of chunks allocated to distinct nodes so that MapReduce task can be performed

in parallel over the nodes. And these file chunks are assigned to different cloud storage node known as chunk server.in such a distributed file system the load of a node is typically proportional to the number of file chunks the node possesses [4].because the files in a cloud computing can be haphazardly created, deleted, and appended in the file system[6], And nodes can be upgraded, replace and added in the file system .the file chunks cannot be distributed uniformly as possible among the nodes.in this case load, balance among storage nodes is a critical function in clouds.
However, GFS and HDFS has some potential problem. The first one is HDFS depends on a single name node to manage almost all operations of every data block in the file system. As a result, in can be a bottleneck resource and a single point of failures and once it fails to perform the action it takes a long time to recover. And the second one is it totally depends on TCP to transfer data. Usually, TCP takes several rounds to transfer data in cloud this results in low link utilization and longer download time
In this paper, we studies and address these problems with current system such as GFS and HDFS .in order to increase the system scalability we using a light weight end server to connect and share all requests with many name nodes. This makes a single name node to many name nodes. And it is stateless. If it goes down no data will be lost and we can bring it up instantly and another main feature of our system is that it uses an efficient load balancing algorithm to balance and split the load between the name node servers. Our proposed model can achieve full link utilization and also decreases download time. As a result, of this, there won't be any bottleneck failure and we can achieve lower chunk transfer times
The contribution of this paper involves:

- We propose a vertically distributed framework that defines bindings between client system and the name node servers
- We propose an approach to schedule jobs with CPU and resource requirements in shared heterogeneous cloud computing

The proposed policy is demand driven and it improves overall resource utilization. The proposed scheduling policy is studied under various system and workload parameters

Background

Cloud technologies

The cloud technologies such as MapReduce and Dryad, Google File System, Hadoop Distributed File System, Microsoft Dryad and CGL-MapReduce have created new trends [26]. Distributed file systems such as GFS and HDFS are used to access data through distributed storage system built on heterogeneous compute nodes and the Dryad and CGL-MapReduce used to read data from local disks

Cloud computing services

Cloud computing offers three major services such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [8].

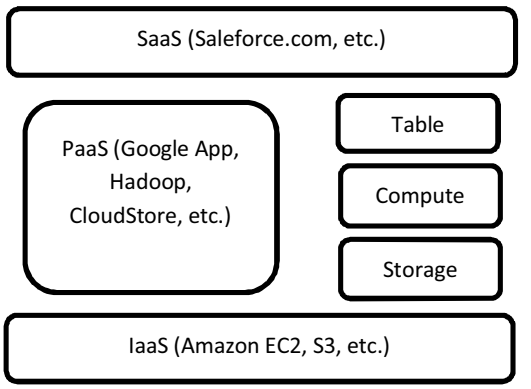


Fig 1: Cloud computing services

IaaS provides computing resources such as storage, servers, other low-level networks, and hardware resource virtually over the internet based on the demand, for example, GoGrid, Amazon's EC2 serving infrastructure to the IT industries [26]

PaaS provides application framework and a set of application programming interface that can be used for developing purposed as well as for developing applications for the cloud. The Google and Microsoft are the major companies which providing PaaS
SaaS provides fully executable software delivered over the internet. Through the internet, the users can operate and make utilize the functionality of the product

Existing Load Balancing Algorithms

Equally Spread Current Execution Algorithm(ESCE)
Here the load will be given equally to all the virtual machine connected to the data center. If any request coming from client/node .the load balancer scans the index table which contains information about the virtual machines as well as a number of jobs currently assigned to the virtual machine. If the request comes from the data center to allocate the VM [13]. It scans the index for least loader VM. The first identified VM is selected for handling the request from the client.After completing the assigned task, the load balancer will update the index table by decreasing the allocation count for identified VM. Here scanning process of index table again and again is a major issue

Round Robin Algorithm
This is very simple and lightweight algorithm that works on the concept of time quantum here time is divided into multiple slices and each node is given a particular time quantum.in this time quantum, the node has to perform all the operations allocated to it.in Round Robin scheduling algorithm time quantum play a very major role for scheduling the resources to the client. The time quantum should not be extremely smaller or extremely bigger than the RR scheduling algorithm.it is very light weight and simple algorithm but there is an additional load on the scheduler to decide the size of quantum

Throttled Load Balancing Algorithm(TLB)
In this algorithm, the load balancer maintains an index table same like ESC algorithm which contains a list of virtual machines as well as their states. The client will first pass a request to the data center to find a suitable virtual machine and then data center will query the load balancer for allocation of VM. The load balancer scans the index table from the top until finding the suitable VM and job will be allocated for that VM. If the VM is not found the load balancer will return -1 to the data center. After completion of the job the details will be updated in index table to free the VM and make it ready for next job allocation.Here also the scanning process for VM will makes lot of delays.

Proposed Work

Architecture

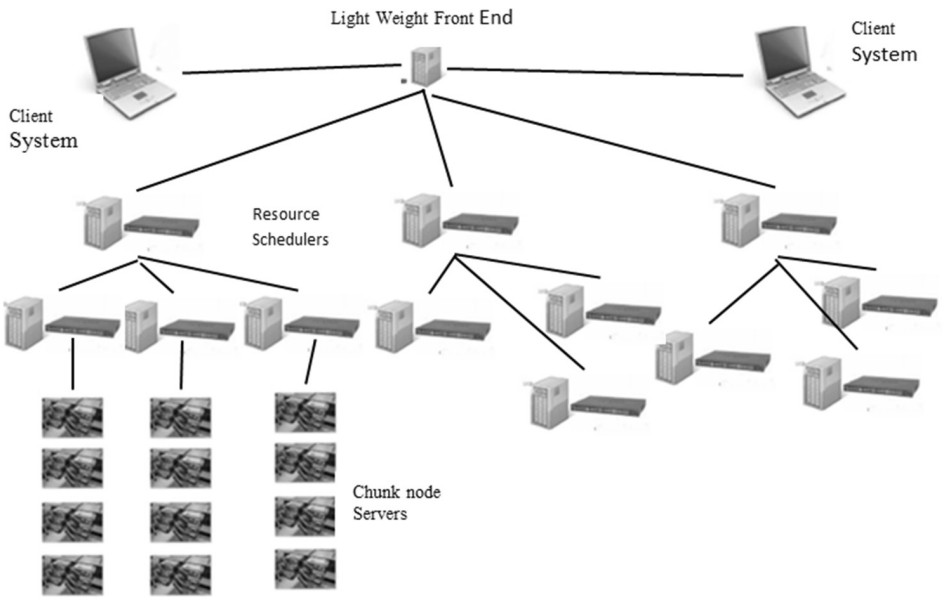


Fig 2: vertically distributed framework

Load Balancing Algorithm

In our proposed work we using two sharing approaches such as Time sharing and space sharing. Time sharing approach helps to balance the load in a number of jobs on process node servers and also helps to allocate that job to nodes can execute according to its time quantum which results in getting less weight time for the jobs. We use the parameter multiprogramming level (MPL) to control the number of tasks among which the processor is time shared at given time. Since each and every processors have different processing speed, the MPL [14] is determined as follows

$$MPL = \frac{CPU\ speed * Basic\ MPL}{slowest\ processor\ speed}$$

The space sharing technique also allows splitting the job on different processing nodes if one node is not able to full the requirements of the job then the job will be split into the different processing nodes which make the job to be executed in less time. The proposed system also uses the demand driven approach it makes the system more efficient. In the demand-driven approach if the node is in the idle state then it will demand their parent for the jobs and if the parent does not have the job then it will demand the job from its parent. it will make the system wait less for the jobs. In the workload model, all tasks of jobs have equal service demand. Job cumulative service demand is dividing into maximum jobs and each job will have a demand for minimum time. This workload shows the advantage of space sharing policy. The adaptive scheduling used for Heterogeneous Multi-cluster System can be framed using following steps:

I. **Job selection**: Job selection policy is used to select the jobs in the queue. The global scheduler consists the jobs in the queue. The aim of scheduling policy is to carry the job from the queue in some manner. So we use First Come First Serve policy. It is one of the simple policies and it has less overhead as compared to other policies. It implements just one queue which holds the tasks in the order they come in. The job is served in arrival order.it is done by our lightweight front end server. In case if it crashed also we won't face any data loss and it recovered instantly

II. **Selecting site**: The Site/Cluster is selected on the basis of where can our job perfectly runs. The Most-fit policy is used to select the cluster. The perfect policy is used to minimize the data that is divided into fragments by choosing the appropriate cluster which wasteless number of processing nodes and by taking care of the other jobs in the queue.
    - In our scenario, we define the tree structure in which the main front end server divided into three resource schedulers. After this three resource schedulers it further divided into three nodes each, our system contains total 9 name nodes. Each node connected to 8 chunk servers.
    - All the nodes find their BPU request according to the total sum of their BPUs range. In this system, all the processing nodes perform the task which is divide equally to them.
    - Multiprogramming Level (MPL) of all the node is also fixed. We have formula to calculate MPL of each processor. MPL= (Processor speed ×Basic MPL)/ (slowest processor speed). Or we can make it fixed (preferred) to simplify it. We can fix it to 2 i.e. each processing nodes can do time-sharing between two jobs. For time-sharing we use Round Robin algorithm.
    - Round Robin algorithm is used to find the waiting time and remaining time to finish the task of all nodes to allocate the jobs equally to them.

And after all scheduling and allocation of chunks the details will be stored by resource scheduler in their index table by means of the parameters such as BPU, starting time, ending time, node id .at each and every iterations this file table will be updated.

Numerical Result

Every processing node in a cluster consists of different number of BPUs (Intra cluster heterogeneity). Number of BPUs for each processor it is fixed to find the performance of our model. For example N33 = 1 BPU, N34 = 4 BPUs, N35 = 2 BPUs, N36 = 12 BPUs, N37 = 2 BPUs, N38 = 3 BPUs, N39 = 10 BPUs, N40 = 7 BPUs. In starting System will be in neutral state. There won't be any jobs in intermediate layers or any other processors. Jobs travel down in hierarchy order based on the demand from the client systems Multiprogramming Level (MPL) of all the processor is also fixed. We have formula to calculate MPL of each processor. Or we can make it fixed (preferred) to simplify it. We fixed it as 2 i.e. each processing nodes can do time-sharing between two jobs. And we got the result as follows:

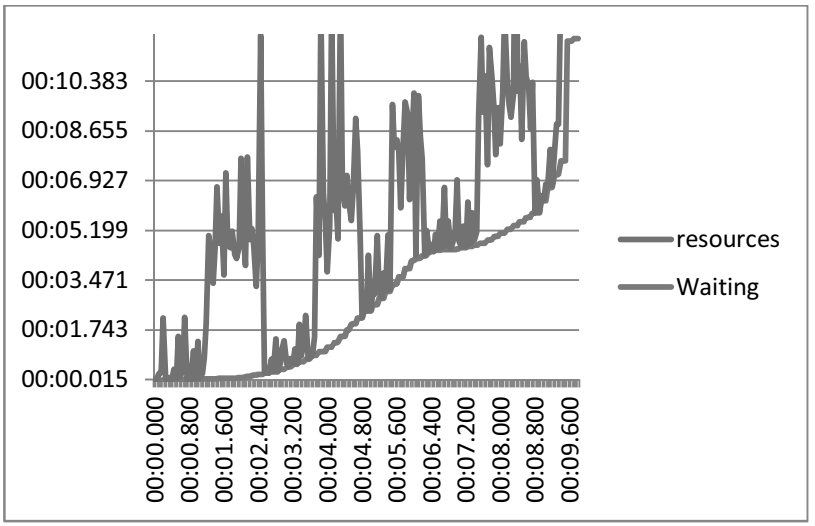


Fig 3: performance of the system graphically

Representing the values in the form of table as follows:

Table 1: Comparison of waiting times between our and existing system

resources	Average wait time new system	Average Waiting time old system
resources10% ar=6	00:02.840	00:03.117
resources 20% ar=12	00:03.112	00:03.384
resources30% ar=18	00:04.123	00:04.856
resources 40% ar=24	00:05.224	00:05.999
resources50% ar =30	00:06.111	00:06.455
resources 60% ar=36	00:07.222	00:07.888
resources70% ar=42	00:08.567	00:09.111
resources 80% ar=48	00:09.445	00:10.111
resources 90% ar=54	00:10.345	00:11.234

Conclusion

In this dissertation work, a new scalable and efficient scheduling algorithm in distributed file system is planned and then enforced in virtual cloud computing environment using Microsoft visual studio, in c# language. Our proposal is to balance the loads of nodes, to increase the processing speed in file system and also to reduce the cost as much as possible. This thesis presents design of a scalable and efficient distributed file system. The system uses a light weight front and back end server to

manage sessions and compute the storage and processing of data. This design solves the potential bottleneck scenario that the name node server of current systems by can be allocation the work load into further host. Our research work conjointly offers an adaptive and efficient resource allocation scheme which may lead to full link utilization and hence much reduced chunk transfer time. By visualizing the parameters in graphs and tables we can able to simply identify that the response time and data centre processing time is improved yet as well as cost is reduced in comparison to the existing scheduling parameters. Based on the numerical results presented, our algorithm will overcome standard existing distributed file systems .our model can be directly implemented in current distributed file systems.

Reference

[1] Dean, J.and S.Ghemawat. 2008." MapReduce:simplified data processing on large clusters.Commun.ACM 51(1):107-113".
[2] ASF.2009.Apache Hadoop Core. http://hadoop.apache.org/core
[3] Ali M. Alakeel, (2010),"A Guide to Dynamic Load Balancing in Distributed Computer Systems",International Journal of Computer Science and Network Security, VOL.10 No.6, June 2010.
[4] Hadoop Distributed File System, http://hadoop.apache.org/hdfs/, 2012
[5] Debessay Fesehaye, Rahul Malik, Klara Nahrstedt, A Scalable Distributed File System for Cloud Computing.(n.d)[online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.176.5236&rep=rep1&type=pdf
[6] Hadoop Distributed File System "Rebalancing Blocks", http://developer.yahoo.com/hadoopp/tutorial/module2.html#rebalancing, 2012
[7] Cloud computing principles, systems and applications NICK Antonopoulos (n.d.)[online].Available: http://mgitech.wordpress.com.
[8] C.Vecchiola, S.Pandey, and R.Buyya,"High-Performance Cloud Computing: A view of scientific applications,"CoRR,Vol.abs/0910.1979,2009
[9] Mladen A. Vouk, Cloud Computing Issues, Research and Implementations, Proceedings of the ITI 2008 30th International Conference on Information Technology Interfaces, 2008, June 23-26.
[10] lizhe wang, rajiv Ranjan. (n.d.). Cloud Computing Methodology, Systems and Applications. Available:http://www.unitiv.com.
[11] Luyang Dong, Bin Gong, (2012),"A Hierarchical Scheduling Policy for Large-Scale Rendering",IEEE International Conference on Systems, Man, and Cybernetics, 2012.
[12] Tejinder Sharma, Vijay Kumar Banga, (2013), "Efficient and Enhanced Algorithm in Cloud Computing", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-1.
[13] Jasmin James, Dr. Bhupendra Verma,(2012)," Efficient VM Load Balancing Algorithm For A Cloud Computing Environment", Jasmin James et al. International Journal on Computer Science and Engineering (IJCSE),Vol-4,Issue No.9.
[14] J. H. Abawajy and S. P. Dandamudi, "Parallel job scheduling on multicluster computing systems," in Proceedings of the IEEE International Conference on Cluster Computing,Hong Kong, China, 2003, pp. 11-18.
[15] Jaliya Ekanayake and Geoffrey Fox, High Performance Parallel Computing with Clouds and Cloud Technologies, Presented at Cloud Computing - First International Conference, CloudComp, Munich, Germany, 2009
[16] Jonthan Strickland. How Cloud Computing Works (n.d.) [online]. Available:http://www.howstuffworks.com/cloud-computing/cloud-computing1.html
[17] Abhisek Pan, John Paul Walters, Vijay S. Pai, Dong-In D. Kang, Stephen P. Crago, "Integrating High Performance File Systems in a Cloud Computing Environment", The International Workshop on Data-Intensive Scalable Computing Systems (DISCS), in conjunction with the 2012 ACM/IEEE Supercomputing Conference (SC'12), November 2012.
[18] Hung-Chang Hsiao, Member, Hsueh-Yi Chung, Haiying Shen,, and Yu-Chang Chao,(2013)," Load Rebalancing for Distributed File Systems in Clouds", IEEE Transactions On Parallel And Distributed Systems, Vol. 24, No. 5.
[19] Cong Wang, Qian Wang, and Kui Ren, Wenjing Lou, Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing, Presented at IEEE INFOCOM 2010.
[20] Satoshi Togawa, Kazuhide Kanenishi, Private Cloud Cooperation Framework of e-Learning Environment for Disaster Recovery , IEEE International Conference on Systems, Man, and Cybernetics, 2013.
[21] Sonal Guleria1, Dr. Sonia Vatta2, (2013) "To Enhance Multimedia Security In Cloud Computing Environment Using Crossbreed Algorithm",International Journal of Application or Innovation in Engineering and Management, Volume 2, Issue 6.
[22] CemOzdogan,(2011, Feb.14). Round robin scheduling. [Online]. Available: http://siber.cankaya.edu.tr/OperatingSystems/ceng328/node125.html.
[23] Amandeep Kaur Sidhu, Supriya Kinger, (2013)," Analysis of Load Balancing Techniques in Cloud Computing", International Journal of Computers & Technology,Volume 4 No. 2.
[24] Martin Randles, Enas Odat, David Lamb, Osama Abu- Rahmeh and A. Taleb-Bendiab, "A Comparative Experiment in Distributed Load Balancing", 2009 Second International Conference on Developments in eSystems Engineering.
[25] Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. (2009, Feb. 10)," Above the clouds: A Berkeley view of cloud computing", EECS Dept., Univ. California, Berkeley, No. UCB/EECS-2009-28.
[26] Anthony T.Velte, Toby J.Velte, Robert Elsenpeter,"Cloud Computing A Practical Approach", TATA McGRAW-HILL Edition 2010.

Fuzzy Logic Classification based Approach for Linear Time Series Analysis in Medical Data Set

Manish Pandey*, Meenu Talwar**, Sachin Chauhan*** and Gurinderjit Kaur****
* Principal, CGC-College of Engineering, Mohali, Punjab
principal.cgccoe@gmail.com
** Assistant Professor, CGC-College of Engineering, Mohali, Punjab
cgccoe.cse.meenu@gmail.com
***Assistant Professor, Chandigarh Engineering College, Mohali Punjab
Sachindiet@gmail.com
****Associate Professor, CGC-College of Engineering, Mohali, Punjab
cgccoe.appsc.gn@gmail.com

Abstract: Health-care management systems are of great relevance now days due to provision of an easy and quick management in all aspects of a patient, not necessarily medical. Furthermore, there are more and more cases of pathologies in which diagnosis and treatment can be only carried out by using medical imaging techniques. With an ever-increasing prevalence, medical images are directly acquired in or converted into digital form, for their storage as well as subsequent retrieval and processing. Data Mining is the process of extracting information from large data sets through using algorithms and techniques drawn from the field of Statistics, Machine Learning and Data Base Management Systems. Traditional data analysis methods often involve manual work and interpretation of data which is slow, expensive and highly subjective. Authors propose a robust ontology based multidimensional data warehousing and mining approach to address the issues of organizing, reporting and documenting diabetes cases including causalities. Data mining procedures, in which map and data views depicting similarity and comparison of attributes extracted from warehouses, are used in the present studies, for understanding the ailments based on gender, age, geography, food habits and hereditary traits. Time series forecasting takes the past values of a time series and uses them to forecast the future values. In this paper, we have proposed a new algorithm for multistep ahead time series forecasting. The original time series and differenced series are classified using Competitive Learning Neural Network.

Keywords: Neural Network, Time Serious Analysis, Regression Technique, Fuzzy Logic.

Introduction

Data Mining and Fuzzy Logic

Health care can be considered as the prevention, treatment, and management of illness and the preservation of mental and physical well being through the services offered by the medical, nursing, and allied health professions. Health care embraces all the goods and services designed to promote health, including preventive, curative and palliative interventions, whether directed to individuals or to populations. The organized provision of such services may constitute a health care system [1]. To implement the conceptions about the quality of the applied medical treatment the quality management system of the organism (medical center) must realize some characteristic activities, which are classified as following points:

- Identification of the basic elements (compartments) that necessary defines and realizes patient's global activity (treatment of the patients); Coordination of the activities of these elements;
- Definition and adaptation of criteria's and methods to make the control of the activity of the composing elements and the entire medical process;
- Make a permanent supervision, measurement and analyze of the process (based on the defined criteria's);
- Implement the proposed activity's to obtain the planed results, and realize a continuous amelioration of the process.

Data mining has been defined as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data" [9] and "the science of extracting useful information from large data sets or databases" [10]. It is the core principle of the knowledge discovery process, which also includes data selection, pre processing and cleaning, transformation and reduction, evaluation, and visualization. In the context of healthcare and biomedicine, data mining is often viewed as a potential mean to identify various biological, drug discoveries, and patient care knowledge embedded in the extensive data collected. Furthermore, data mining provides results that possibly highlight vaguely understood

doctrine and provide useful insights to help in decision making processes. Fuzzy Logic is usually known as an appropriate method for sorting and handling large amounts of data, but has also proven in recent years to be an excellent choice for many control system applications since it mimics human control logic. Fuzzy logic handles the concept of partial true that is true values between "completely true" and "completely false. It is very robust and forgiving of operator and data input and often works when first implemented with little or no tuning [7].

Linear Regression in Fuzzy Logic

A linear regression is a major factor evolved as an independent variable to explain the dependent variable change. In the real studies, the dependent variable changes often by several important factors. When more than one independent variable and dependent variables is in linear relationship, regression analysis is carried out by multiple linear regression Setting up y as the dependent variable, x1, x2, ... xk as independent variables, There is a linear relationship between the k independent variables and n the dependent variable, then multiple linear regression function is:

$y = b_0 + b_1x_1 + b_2x_2 \ldots + b_kx_k + e$

Where b_0 is a constant, b1, b2, ... bk for the regression coefficients. Multiple regression model parameter estimation is the same as linear regression equation.

Related Work

Castiglione et al. [1] has presented that in the clinical evaluation and treatment on a patient, there are mostly two relevant factors: the former is the training and experience of the physicians, the latter is the amount of information on which they may rely on. Such information may include the diagnosis of previous identical or similar cases, estimates on the possible healing times and any other information or comments considered clinically relevant. Ianosi [2] has presented in their paper regarding health management, life style modification which is the important solution against not just healthy life expectancy but medical crisis. Obviously, medical field plays a quite important role for prevention and treatment of disease. However, almost all patients consume very short time such as 3 to 5 minutes in medical treatment. On the other side, our ordinary life is at home, here the words of home is used as expanded meaning including our time for working and other activities. Lan Yo [3] has presented regarding Data Mining, which is the process of extracting information from large data sets through algorithms and techniques drawn from the field of Statistics, Machine Learning and Data Base Management Systems. Lan Yo [4] has presented in the paper that in practice, many data mining exercises using data drawn from patients with particular conditions are performed to provide medical researchers with some insight into the disease that could lead to a greater understanding of the condition and suggest possible interesting directions for research. Frawley [5] has presented that Healthcare organizations practicing evidence-based medicine strive to unite their data assets in order to achieve a wider knowledge base for more sophisticated research as well as to provide a matured decision support service for the care givers..

Time Series Forecasting

Time series occur in various domains in great number and heterogeneity. In general, a time series s can be described as a sequence (x1, x2, x3 ..., xn) containing n data points xi. These data points can consist of real numbers, for instance of the river level or the voltage of an EEG derivation [8] measured at certain typically equidistant points in time; or more complex, they can be highly multidimensional, for example in market basket analysis [9], where xi corresponds to a customer's transaction containing e.g. the time of the transaction, a customer ID and bought items. In recent work on model-free analyses, wavelet transform based methods (for example locally stationary wavelets and wavelet decomposed neural networks) have gained favor. Multi scale (often referred to as multi resolution) techniques decompose a given time series, attempting to illustrate time dependence at multiple scales.A number of different notations are in use for time-series analysis: $X = \{X_1, X_2, ...\}$

is a common notation which specifies a time series X which is indexed by the natural numbers. Another common notation is: $Y = \{Y_t : t \in T\}$.

Conditions

There are two sets of conditions under which much of the theory is built as: Stationary process and Ergodicity. In addition, time-series analysis can be applied where the series are seasonally stationary or non-stationary. Situations where the amplitudes of frequency components change with time can be dealt with in time-frequency analysis which makes use of a time–frequency representation of a time-series or signal.[4].

Autoregressive Models

The general representation of an autoregressive model, well-known as AR(p), is

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t$$

Where the term ε_t is the source of randomness and is called white noise. It is assumed to have the following characteristics:

1. $E[\varepsilon_t] = 0$
2. $E[\varepsilon_t^2] = \sigma^2$
3. $E[\varepsilon_t \varepsilon_s] = 0 \quad \forall t \neq s$

Forecasting Methods

One of the main goals of time series analysis is to forecast future values of the series. A trend is a regular, slowly evolving change in the series level. The use of intuitive methods usually precludes any quantitative measure of confidence in the resulting forecast. In the Single-Equation Regression Models the variable under study is explained by a single function (linear or nonlinear) of a number of explanatory variables. The equation will often be time-dependent (i.e., the time index will appear explicitly in the model), so that one can predict the response over time of the variable under study to changes in one or more of the explanatory variables.

Modelling the Causal Time Series

With multiple regressions, we can use more than one predictor. It is always best, however, to be parsimonious, that is to use as few variables as predictors as necessary to get a reasonably accurate forecast. The forecast takes the form:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$,

where β_0 is the intercept, β_1, β_2, . . . β_n are coefficients representing the contribution of the independent variables X_1, X_2,..., X_n.

Statistical control limits are calculated in a manner similar to other quality control limit charts, however, the residual standard deviation are used.

Modelling Seasonality and Trend

Seasonality is a pattern that repeats for each period. For example annual seasonal pattern has a cycle that is 12 periods long, if the periods are months, or 4 periods long if the periods are quarters. Seasonal index represents the extent of seasonal influence for a particular segment of the year. The calculation involves a comparison of the expected values of that period to the grand mean. A seasonal index is how much the average for that particular period tends to be above (or below) the grand average. Therefore, to get an accurate estimate for the seasonal index, we compute the average of the first period of the cycle, and the second period, etc, and divide each by the overall average. The formula for computing seasonal factors is:

$S_i = D_i/D$,

where:

S_i = the seasonal index for i^{th} period

D_i = the average values of i^{th} period

D = grand average

i = the i^{th} seasonal period of the cycle.

Result Analysis

Numerical Application

The table 6.1 provides number of patients on monthly basis at the hospital. The number of patient shows a seasonal pattern. The final step is represented in figure 6.1, shows the forecast is to use the seasonal index to adjust the trend projection. One simple way to forecast using a seasonal adjustment is to use a seasonal factor in combination with an appropriate underlying trend of total value of cycles.

Linear Regression Equations

If we expect a set of data to have a linear correlation, it is not necessary for us to plot the data in order to determine the constants *m* (slope) and *b* (y-intercept) of the equation y = mx + c. Instead, we can apply a statistical treatment known as linear regression to the data and determine these constants. Given a set of data (x_i, y_i) with *n* data points, the slope, y-intercept and correlation coefficient, *r*, can be determined using the following:

$$m = \frac{n\sum(xy) - \sum x \sum y}{n\sum(x^2) - (\sum x)^2}$$

Table 6.1: Number of patients in different months, for three years and forecast of third year using linear regression

	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec	Sum
2010	202	456	512	1167	1209	13	402	387	893	329	516	787	6873
2011	219	28	425	530	1398	1059	1544	747	654	266	523	769	8162
2012	442	659	629	363	1021	179	806	1637	2070	1313	1027	1110	11256
Mean	287.667	381	522	686.667	1209.33	417	917.333	923.667	1205.6	636	688.67	888.67	8763.67
Index	0.41	0.54	0.74	0.97	1.71	0.59	1.30	1.31	1.71	0.90	0.98	1.26	12.42
Expected 2013	180	356	465	353	1750	106	1048	2143	3537	1183	1002	1398	13521

$$b = \frac{\sum y - m\sum x}{n}$$

$$r = \frac{n\sum(xy) - \sum x \sum y}{\sqrt{\left[n\sum(x^2) - (\sum x)^2\right]\left[n\sum(y^2) - (\sum y)^2\right]}}$$

(Note that the limits of the summation, which are *i* to *n*, and the summation indices on *x* and *y* have been omitted.), implicitly applying regression to the sample data.

Application of linear regression technique for the calculation of line equation for the month 2010

Table 6.2: Calculation for the Summation of the Table Content for the Calculation

∑x	78
∑y1	6873
∑x*y1	45798
∑x*x	650
∑y1*y1	5437311
$(\sum x)^2$	6084
$(\sum y)^2$	47238129

After applying the following linear regression formula of slope and constant for the line equation, we will get the value of m and b. Where n = 12 (for 12 months in a year), after applying the formula for the slope of the line and the constant b for the line equation y = mx + b, we get: m = 7.856643 and b = 521.6818
Now for the plot of the graph for x = month of the year, y = variable from line equation after applying the value of slope (m) and constant (b), and y1 = number of patients in the month. We get:

Table 6.3: Calculation for the 12 months slope of the line after applying linear regression

x	y	y1
1	529.538	202
2	537.395	456
3	545.252	512
4	553.108	1167
5	560.965	1209
6	568.822	13
7	576.678	402
8	584.535	387
9	592.392	893
10	600.248	329
11	608.105	516
12	615.962	787

Application of linear regression technique for the calculation of line equation for the month 2011

Table 6.4: calculation for the summation of table content in year 2011

∑x	78
∑y2	8162
∑x*y2	57325
∑x*x	650
∑y2*y2	7891462
$(\sum x)^2$	6084
$(\sum y)^2$	66618244

After applying the following linear regression formula of slope and constant for the line equation, we will get the value of m and b: m = 29.87413 and b = 485.9848, Where n = 12 (for 12 months in a year) Now for the plot of the graph for x = month of the year, y = variable from line equation after applying the value of slope (m) and constant (b), and y2 = number of patients in the month. We get:

Table 6.5: Calculation for the 12 months slope of the line after applying linear regression in 2011

x	Y	y2
1	515.859	219
2	545.733	28
3	575.607	425
4	605.481	530
5	635.355	1398
6	665.23	1059
7	695.104	1544
8	724.978	747
9	754.852	654
10	784.726	266
11	814.6	523
12	844.474	769

Application of linear regression technique for the calculation of line equation for the month 2012

Table 6.6: Calculation for the Summation of the Table Content for the Calculation in 2012

∑x	78
∑y3	11256
∑x*y3	86393
∑x*x	650
∑y3*y3	13856640
$(\sum x)^2$	6084
$(\sum y)^2$	126697536

After applying the following linear regression formula of slope and constant for the line equation, we will get the value of m and b: m = 92.51049 and b = 336.6818, Where n = 12 (for 12 months in a year)
Now for the plot of the graph for x = month of the year, y = variable from line equation after applying the value of slope (m) and constant (b), and y3 = number of patients in the month. We get:

Table 6.7: Calculation for the 12 months slope of the line after applying linear regression in 2012

x	y	y3
1	429.192	442
2	521.703	659
3	614.213	629
4	706.724	363
5	799.234	1021
6	891.745	179
7	984.255	806
8	1076.77	1637
9	1169.28	2070
10	1261.79	1313
11	1354.3	1027
12	1446.81	1110

Application of linear regression technique for the calculation of line equation for the month 2013

Table 6.8: Calculation for the Summation of the Table Content for the Calculation

∑x	78
∑y4	13520.611
∑x*y4	109022.57
∑x*x	650
∑y4*y4	26129000
$(\sum x)^2$	6084
$(\sum y)^2$	182806926

After applying the following linear regression formula of slope and constant for the line equation, we will get the value of m and b. m = 147.8224 and b = 165.8723, Where n = 12 (for 12 months in a year):
Now for the plot of the graph for x = month of the year, y = variable from line equation after applying the value of slope (m) and constant (b), and y1 = number of patients in the month. We get:

Table 6.9: Calculation for the 12 months slope of the line after applying linear regression

x	y	y4
1	313.695	180
2	461.517	356
3	609.339	465
4	757.162	353
5	904.984	1750
6	1052.81	106
7	1200.63	1048
8	1348.45	2143
9	1496.27	3537
10	1644.1	1183
11	1791.92	1002
12	1939.74	1398

Comparison of slopes for four year

Table 6.10: Comparison of slops for four years

x	y1	y2	y3	y4
1	514	496	421	418
2	521	528	520	466
3	529	560	618	514
4	537	592	716	562
5	544	624	815	610
6	552	656	913	658
7	559	689	1012	706
8	567	721	1110	754
9	574	753	1208	802
10	582	785	1307	850
11	590	817	1405	898
12	597	849	1503	946

We get the different slope for different years

Table 6.11 Slope of four different years

m (2010)	7.85664
m (2011)	29.8741
m (2012)	92.5105
m (2013)	147.822

On comparing the four different slopes of four years we got that the predicted slope of the fourth year is about the average of the three year slopes hence we can say that the fourth year prediction is the good quality prediction for the patient data. The statistical reports in the following pages shows the various reports and analysis of patient data that can be represented using figure 6.6.

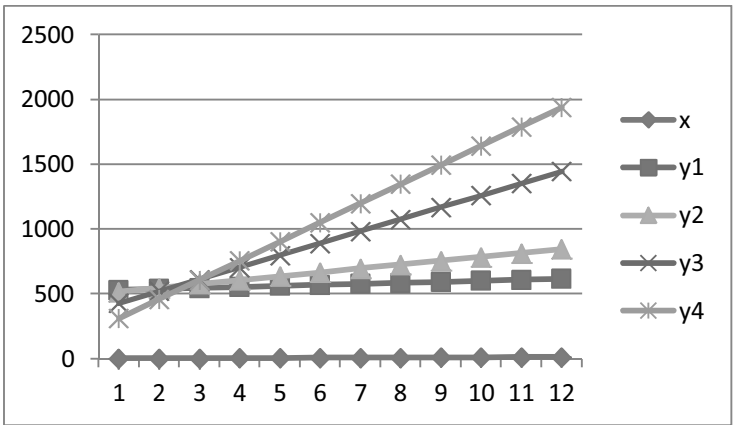


Figure 6.6: Comparison of slops for actual data of three years and the predicted value for fourth year

Conclusion

Inherent in the collection of data taken over time is some form of random variation. There exist methods for reducing of cancelling the effect due to random variation. Widely used techniques are smoothing. This technique, when properly applied, reveals more clearly the underlying trends. However, the data is not properly managed. As a result of this, majority of out-patients do not have full medical record. With this situation, the physician's time is wasted since they have to collect this information again and in addition, it becomes very difficult for them to keep track of the patients. This reduces the ability to carry out high quality clinical research in the hospitals, and compromises the continuity of healthcare as well as the quality of

healthcare delivery in the hospital. A Data Mart has been designed to collect, store, organize and retrieve the medical information of patients. A simple way of detecting trend in seasonal data is to take averages over a certain period. If these averages change with time we can say that there is evidence of a trend in the series.

References

[1] IANOSI ENDRE "Considerations about efficient health care management systems", Proceedings of the 3rd International Conference on E-Health and Bioengineering - EHB 2011, 24th-26th November, 2011, Iaşi, Romania
[2] Endre Ianosi, V. Vacarescu "Dialysis apparatus Technical and quality aspects (in Romanian)", Timisoara, Ed. Orizonturi Universitare, 2002, ISBN 973-8391-26-1.
[3] Lan Yu "Data Mining on Test Data of Physical Health Standard", 978-1-4244-3894-5/09/$25.00 ©2009 IEEE.
[4] Lan Yu" Association Rules based Data Mining on Test Data of Physical Health Standard", 2009 International Joint Conference on Computational Sciences and Optimization.
[5] W. J. Frawley, G. Piatetsky-Shapiro and C. J. Matheus, "Knowledge Discovery in Databases: An Overview", in G. Piatetsky-Shapiro and W. J. Frawley (eds.), Knowledge Discovery in Database. AAAI/MIT Press, pp.127, 1991.
[6] Sashi K Nimmagadda "Multidimensional Data Warehousing & Mining of Diabetes & Food-domain ontologies for e-Health", 978-1-4577-0434-5/11/$26.00 ©2011 IEEE.
[7] Stefan Kleinmann, Ralf Stetter, Praveen Kumar Kubendra Prasad" Optimization of a Pump Health Monitoring System using Fuzzy Logic", 2013 Conference on Control and Fault-Tolerant Systems (SysTol) October 9-11,2013. Nice, France.
[8] T. Schluter and S. Conrad, "An approach for automatic sleep stage ¨ scoring and apnea-hypopnea detection," in Proc. of the 10th IEEE Int. Conf. on Data Mining (ICDM), 2010, pp. 1007–1012
[9] T. Schluter and S. Conrad, "TEMPUS: A Prototype System for Time ¨ Series Analysis and Prediction," in IADIS European Conf. on Data Mining 2010. IADIS Press, 2010, pp. 11–1.
[10] A.S. Chen, M.T. Leung and H. Daouk, "Application of Neural Networks to an Emerging Financial Market: Forecasting and Trading the Taiwan Stock Index," Computers and Operations Research 30, 2003, 901-923.
[11] W. Kreesuradej, D. Wunsch, and M. Lane, "Time-delay neural network for small time series data sets," in World Cong. Neural Networks, San Diego, CA, June 1994.
[12] H. Tan, D. Prokhorov, and D. Wunsch, "Probabilistic and timedelay neural network techniques for conservative short-term stock trend prediction," in Proc. World Cong. Neural Networks, Washington, D.C., July 1995.

Design and Implementation of Cost Effective Controller for Solar PV Application

Pulkit Singh* and D K Palwalia**
* Poornima Institute of Engineering & Technology Jaipur
pulkit.singh@poornima.org
** Department of Electrical Engineering, RTU Kota
dheerajpalwalia@gmail.com

Abstract: Growing pollution and depleting fossil fuel reserves have encouraged exploration and exploitation of non-conventional energy sources since oil crisis faced in late 1970. Among available renewable energy sources, photovoltaic (PV) generation is supposed to play major role in future energy scenario. High installation cost and low efficiency has been major challenges for its wide spread usage. This work focuses on reducing cost of generation from PV and enhances power tracking with low cost controller design module. Maximum power point tracking (MPPT) highly depends on atmospheric conditions and exposure of PV panel surface to solar radiation. So MPPT technique should be good enough in dynamic atmospheric conditions. Different algorithms are design in which the Perturb and Observer (P & O) and Incremental conductance (INC) are widely used MPPT techniques.

Present work focus to design the cost effective PV controller module. In the PIC technology C/C^{++} Coding is used to design the algorithm but the drawback of the PIC technology is the complex coding and debugging is much more difficult. Emphasizing on DsPIC30f4011, in which automatic C/C^{++} code for the algorithm generated in the MATLAB using simulation and can easily be dumped on the respected pins and for the completion of the coding MP-Lab IDE software is used which interface between the DsPIC chip and DsPIC hardware kit. For the compilation process complier XC16 and C30 complier is used.

Efforts have been made to design the cost effective and reliable operation of PV system using IGBT based boost converter and rectifier circuit thereby ensuring the minimum switching losses, reducing size and cost of controller. The new topology is well suited for drives and renewable energy applications.

Keywords: Solar PV, MATLAB, MPLAB, DsPIC30f4011, P&O Algorithm.

Introduction

Energy plays an important role in our daily life. With rise in dependence on electrical energy, new sources of energy need tobe explored and exploited in order to meet the energy demand. Most common sources of energy currently utilized world-wide for generating electricity includes coal (39.3%), petroleum (0.7%), natural gas (27.6%), nuclear (19.5%), hydro power (6.7%), wind (4.2%) and other renewable (2.1%) covers mainly geothermal, biomass & PV energy [1]. Fossil fuel based energy sources cause emission of carbon particle and harmful gases, causing severe environmental concerns. This has encouraged power system researchers to increase dependence on renewable energy based generation. Renewable energy is promising and inexhaustible in nature. Amongst the available natural sources, wind, solar and hydro energy are omnipresent in abundance around the globe. Wind energy is intermittent in nature, involves high initial installation cost and gets affected by the geographical condition. For terrestrial applications, solar energy has gained great attraction due to easy installation, high reliability and simplicity in design.

PV cells are used to convert solar energy into electrical energy with the help of power electronics based converters. Solar panel can easily be mounted on roof of houses, multi storage buildings and complexes. Besides these advantages, the drawback is that solar energy is not available during the night hours. So, auxiliary backup unit is necessary to assure continuous supply of electrical energy. Moreover atmospheric conditions like cloud, partial shadowed zone, dust, and snow reduces the overall efficiency of the PV cells. Kumar and Palwalia [1] discussed various MPPT algorithms used to track maximum solar power like perturb & observe (P&O), hill climbing, incremental conductance (Inc), etc [1]. In this work, a new cost effective and reliable microcontroller based P&O tracking has been implemented as MPPT. A MATLAB and MP-lab software tool has been used to obtain simulation results. The obtained results have been tested on real time environment with the help of hardware assembly and software platforms in order to design low cost and user friendly system.

Aim and Objective

PV panel installation cost is quite high due to low efficiency of conversion equipment. In order to encourage its usage, cost efficient design of the system should be used. Presented the cost efficient MPPT hardware for PV module and its compatibility has been investigated on software as well as hardware platform.

The main objectives of this dissertation can be summarized as –

- Investigate working of PV module (solar cell, connections) and its application in different area by recent & relevant literature survey.
- Inspect different MPPT techniques, algorithm and converter topology depending on area of application according to their need and advancement e.g. DC- DC converter.
- Study PIC micro-controller and DsPIC, and examine pin configuration of DsPIC Kit. Mark their advantage over sensors and other methodology used for implementation.
- Analyze PV system with PIC and DsPIC Microcontroller.
- Establish software compatibility via simulation of PWM and design algorithm in MATLAB toolbox.
- Generate C language code in MATLAB and obtain hex-code with the help of MP-Lab software platform, i.e. code compilation in DsPIC.
- Incorporate obtained code in micro-controller to obtain PWM gate pulse for IGBT module connected to PV system via DsPIC kit.

Motivation

In a developing country like India, energy demand has increased exponentially in the last few decades. Majority of power is generated by conventional sources like coal and petroleum. This has increased pollution and depletion of fuel reserves. Conventional sources can deplit in short duration and takes very long time to recover. This energy generation and consumption patterns have called upon need to increase dependence on green energy sources like wind, solar, hydro, etc. High initial installation cost of renewable energy modules, energy constraints and increasing per capita energy consumption have been challenging issues for energy authorities. This need to be coped up by encouraging research programs all over the country. India has vast energy reserve in form of conventional and non-conventional sources. Energy scenario in India has been shown in table 1.

Table 1: Energy Scenarios in India

S. No	Source	% Contribution
1.	Thermal	64.75%
2.	Hydro	21.73%
3.	Nuclear	2.78%
4.	Renewable & other Sources	10.73%

Grid connected (Capacities in MW)

Table 2: Capacity of grid connected power plants

Wind Power	1234.11
Solar Power	827.22
Bio-Power	132.00
Waste to Power	12.00
Total	2311.88 MW

Off-Grid connected (Capacities in MW)

Table 3: Capacity off grid connected power plant

Waste to Energy	0.50
Biomass	10.50
Biomass Gasifiers-Rural-Industrial	0.20 8.67
Hybrid systems	0.13
SPV Systems	46.50
Total	66.50 MW

To overcome the energy crises in power sector in last few decades, planning commission of India has set a new ministry called ministry of new and renewable energy (MNRE) with a motive to encourage use of non-conventional energy sources. MNRE has taken initiatives like Jawaharlal Nehru national solar mission (JNNSM) with an aim to install 20GW of grid connected solar system and 2 GW of off grid by year 2022. MNRE, in venture with Indian government, provides subsidy on PV module installation and provides extra benefits to people using maximum power m renewable sources.

Desgin of Photovoltaic Cell

Solar energy is the form of energy which is taken by the sun in the form of solar radiation. In other words we can define photovoltaic cells are those cells which convert solar energy direct into the electricity, using semi conductor material silicon or germanium. If we describe the background the word photo voltaic it comes from the Greek word which means “light ” indicate to “photo” and “voltaic” indicate to “electrical”. Efficiency of the solar cell is mainly 30 - 40%. Due to the low efficiency in the PV system different techniques are used to extract maximum power from solar panel like PIC, DSP, and FPGA & DsPIC.

Single solar cell can give only about 0.5 volt, using the solar for the terrestrial application such as home lighting, water pumping etc power generated by single solar cell is not enough. To increase the rating number of cell are connected in the series or parallel which is known as PV module. For obtaining the higher power PV panel should be connected in the array. Here the PV panels are the combination of PV modules. Series connections are responsible for increasing the voltage of the array whereas the parallel connection is responsible for increasing the current in the array.

Fig 1: PV terms

The relationship between current and voltage may be determined from the diode characteristic equation that is:

$$I = I_{ph} - I_o(e^{qv/kt} - 1) \quad (3.1)$$

$$I = I_{ph} - I_d \quad (3.2)$$

Where q is the electron charge, k is Boltzmann constant, I_{ph} is the photo current, I_o is the reverse saturation current, I_d is the diode current and T is solar cell operating temperature (K).

An ideal source can be considered as a current source where the current produce by the solar cell is proportional to the solar irradiation falling on it. But the behavior of the ideal PV cell totally changes if we consider the practical circuit. Electrical

losses, optical losses are seen in the practical circuit. Modeling of the PV cell can be representation by the two type single diode model & double diode model .In the double diode model the optical losses are representation by the current sources .While the generated current I_L is directly proportional to the solar intensity. Two diode connected in the parallel representation the recombination losses. They are connected in the reverse because the recombination current flow opposite to the direction of the light generated current. Saturation current I_{S1} will flow in the diode J_{o1} due the diffusion and saturation current and Saturation current I_{S2} will flow in the diode J_{o2} due to the recombination of the space charge carrier. Electrical losses (ohmic loss) which occur due to series resistance Rs and shunt resistance R_{sh} .Series resistance offered the path to the current which is flowing in the solar cell. Shunt resistance indicate the leakage path of the current in a solar cell therefore it is represented in parallel with the current sources.
I-V equation of the solar cell is given by –

$$J = J_L - J_O\left(\exp\frac{qv}{KT} - 1\right) \qquad (3.3)$$

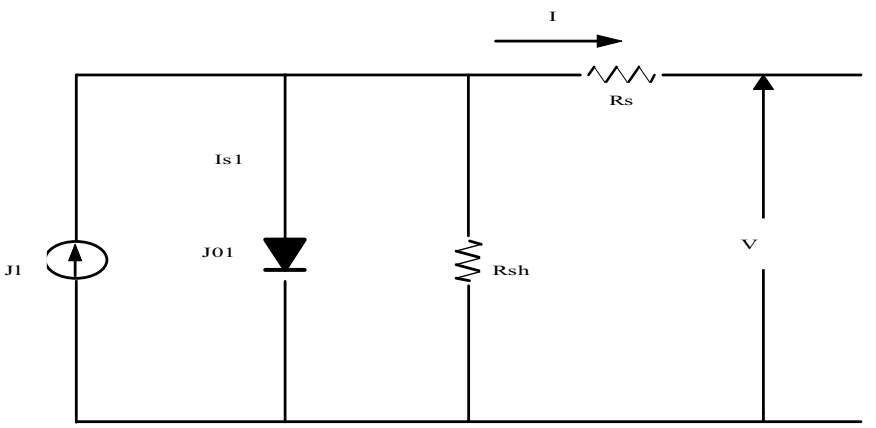


Fig 2: Single diode model

$$J = J_l - J_{O1}\exp\left(\frac{q(V+IR_s)}{KT}\right) - J_{02}\exp\left(\frac{q(V+IR_s)}{KT}\right) - \frac{V+IR_s}{R_{SH}} \qquad (3.4)$$

Equation (3.4) represented the two diode model equation. Term J_{01} represents the recombination in base and emitter region of cell &J_{02} represent the recombination of the space charge region.

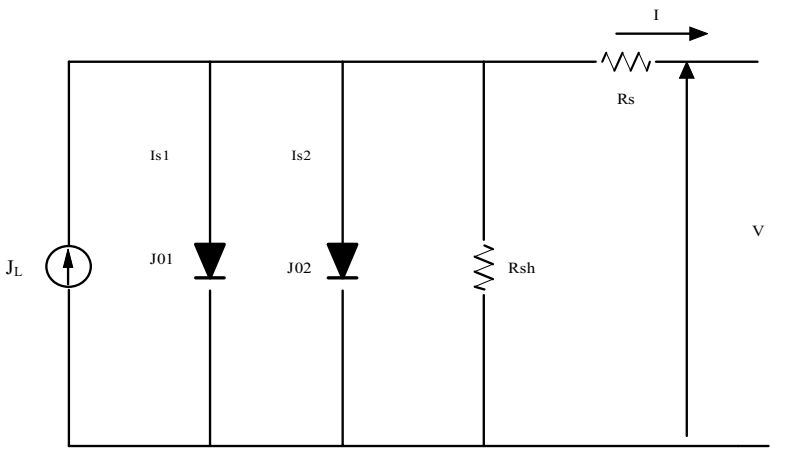


Fig 3: Double diode model

In the simple solar cell model $J_{02} = 0$.equation of the simple solar cell can be written as –

$$J = J_l - J_{O1} \exp(\frac{q(V + IR_s)}{KT}) - \frac{V + IR_s}{R_{SH}}$$

Here n is the diode ideality factor its value lie between 1 and 2 where 1 is for ideal diode and above equation represent the single diode model.

Ratings of 20w PV Module

The simulations are carried out using MATALB/SIMULINK package. The developed mathematical model of the PV array is used for the simulation studies. Various parameters of the PV array are determined and chosen. For the simulation work, we consider the solar panel model of rating 20 watt. Parameter ratings are taken from PV module datasheet

Table 4: Parameter of 20W Solar Module

Maximum Power	20 Watt
Open circuit voltage	1.2 V
Short circuit current	1.8 A
Ns	10
Np	2
Ideal factor	1.3
Band gap of semi conductor use in a cell	1.13 eV

Simulation results of 20W Solar PV Module

I-V curve of PV panel

I-V (Current-Voltage) curve originated from the equation (3.1) for particular value of the voltage .Current value we get and plot the curve this curve, shows what should be the current at a certain voltage. When the I_{pv}=0 we will get open circuit voltage (V_{oc}) of PV panel, when Voc=0we will get short circuit current (I_{sc}).In I-V curve represents the maximum power point corresponding voltage V_{mpp} and corresponding current I_{mpp}.

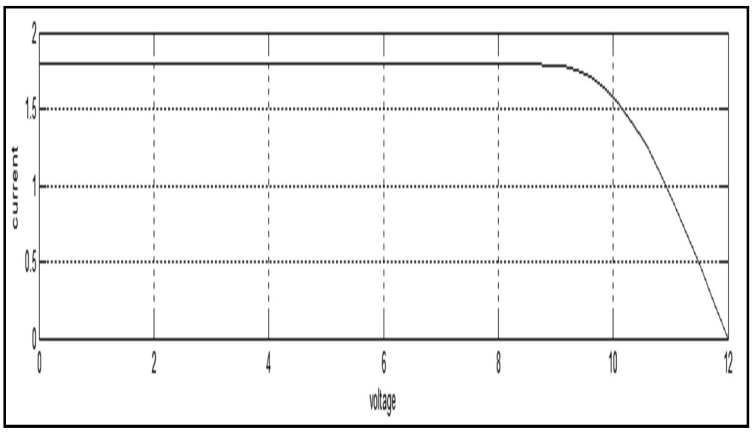


Fig 4: I-V cure at 1000W/m^2 irradiance

P-V curve of PV panel

Multiplication of output current and output voltage gives the output power, at particular value of Current (I_{mpp}) and voltage (V_{mpp}), will give maximum power P_{mpp}. Figure 5 shows P-V (Power-Voltage) curve of PV panel, point shows the maximum power point of the panel.

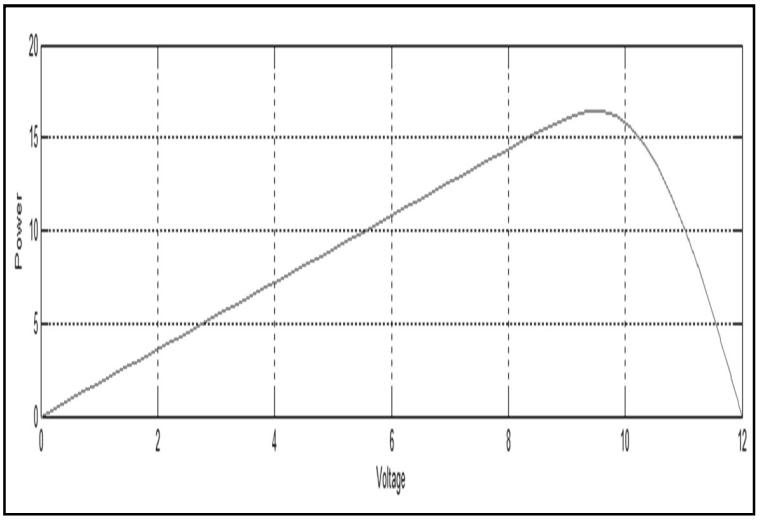


Fig 5: P-V cure at 1000W/m^2 irradiance

Varying insolation condition

Figure 6 shows the P-V curve in different insolation conditions

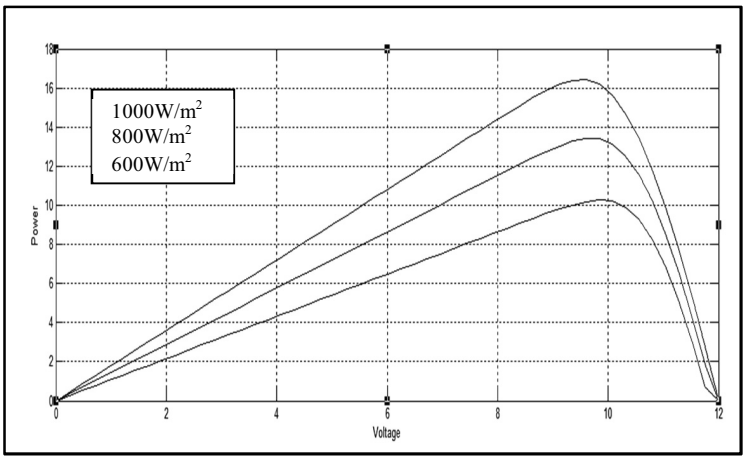


Fig 6: P-V curve at 1000W/m^2,800W/m^2,600W/m^2

Table 5: Power at different irradiance at constant temperature

Irradiance (W/m^2)	**Power (watt)**	**Temperature (^{o}C)**
1000	17.5	25
800	13	25
600	10	25

Varying temperature condition

Figure 7 & 8 shows the P-V and I-V curve in different temperature conditions, point shows peak power of each curve, as temperature increases peak power shifted downwards.

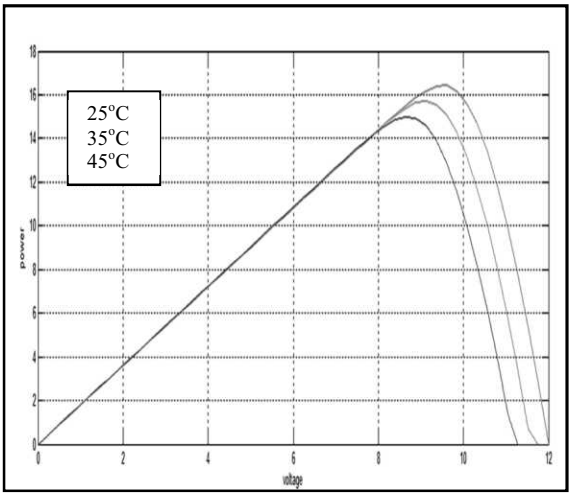


Fig 7: P-V Curve at temperature 25°C, 35°C, 45°C

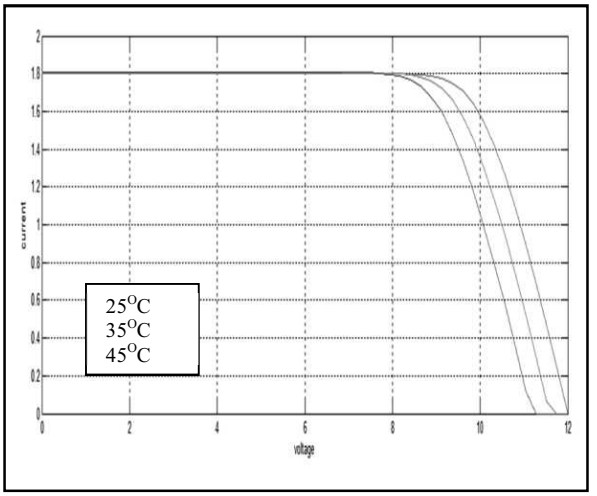


Fig 8: I-V Curve at temperature 25°C, 35°C, 45°C

Maximum Power Point Tracking

Output of the solar panel varies with respect to the sun position, temperature and insulation level. Out of these conditions there are two more conditions also which affect the output power of the PV system – cloudy day and partial shadings. As the efficiency of the PV panel is low i.e. ~ 10 – 25% and if the above condition occurs then the output power of the PV panel decreases. We cannot increase the efficiency of the PV panel but we can extract the maximum power from the panel and the point at which maximum power is extracted is known as MPPT. In other words, we can define controller which tracks the maximum power point locus of the PV array/panel & it is known as MPPT.

Methods of MPPT

Algorithms of MPPT are of various types and are implemented for obtaining the maximum power. Algorithms are used in the DsPIC to implement the maximum power tracking. Different MPPT techniques are given below –

- Hill climbing/ P & O Method
- Incremental conductance Method
- Fractional open circuit voltage/ short circuit current
- Fuzzy and neural network
- dP/dV or dP/dI feedback control

Hill climbing/ P & O Method

The P & O algorithm is widely accepted algorithm due to its simplicity and easy implementation. This algorithm is also known as Hill Climbing (HC) Algorithm. The difference between the two is that only P & O algorithm will work on the PV array voltage and PV array current & hill climbing algorithm work on the concept of Duty Ratio. Although Hill climbing and P & O method are used to obtain MPP but the concept is same. In the P&O algorithm, from the I-V characterizes and conclude that the operating on the left of MPP, when there is increase in voltage then there is increase in the power while on the right hand side if voltage decreases power also decreases. Therefore with increase in power perturbation should be same to reach the desired MPP and if the power is decreased then perturbation should be positive.

Modeling of PV Module with P & O and Boost Converter

The Simulink model of the required solar cell and boost converter system is as shown. This is for the P&O method. Here the solar cell is represented by a block named 'Photovoltaic cell'. Boost converter, which consists of a 0.001H inductor and a 1F capacitor. This boost converter is used to step up the voltage to the required value. The gating signal to the boost converter is generated by comparing the signal generated by the MPPT algorithm to a repeating sequence operating at a high frequency. The load is a 10 ohm resistance P&O algorithm is applied to track MPP. Data sheet is taken from- Energy PV module 285PC8.

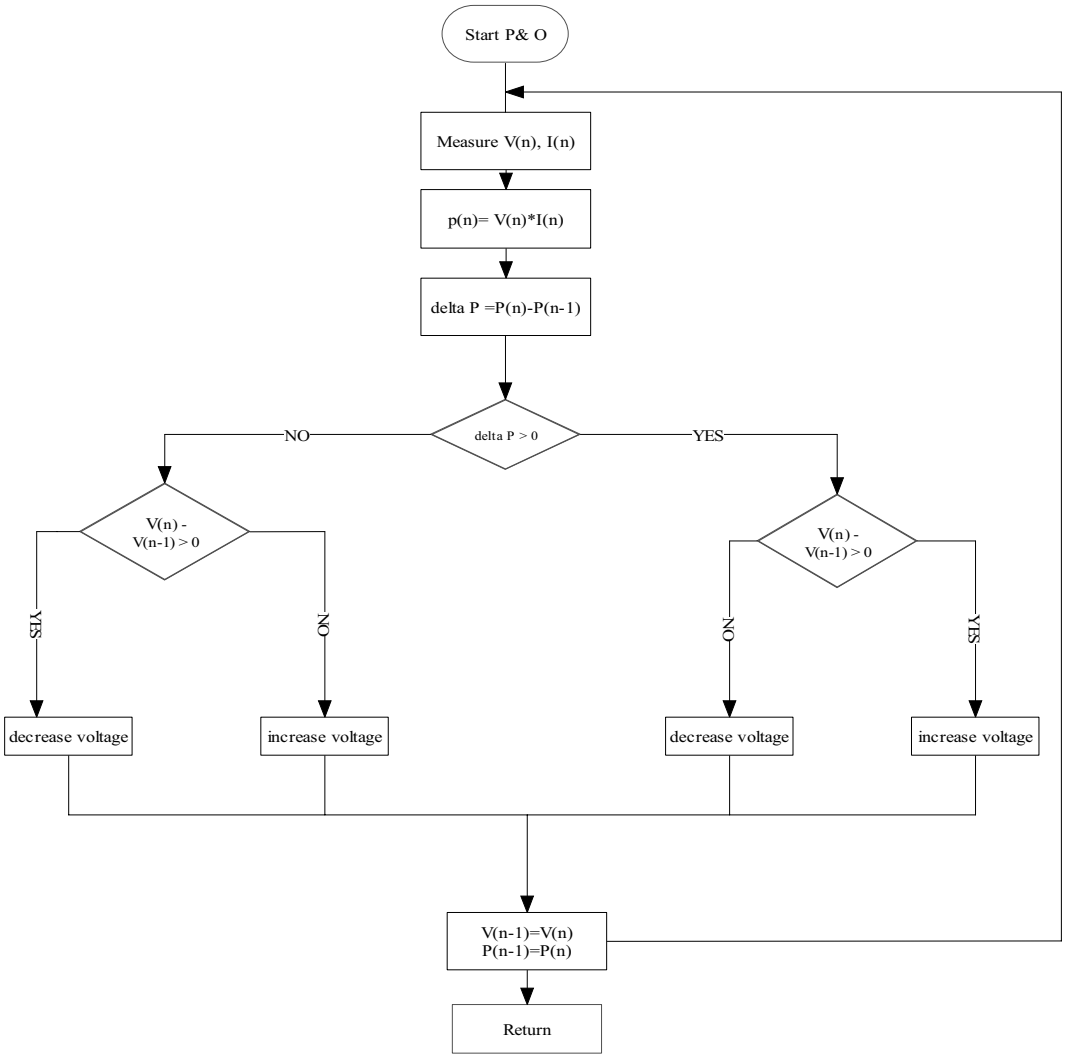


Fig 9: Flow chart of P&O Algorithm

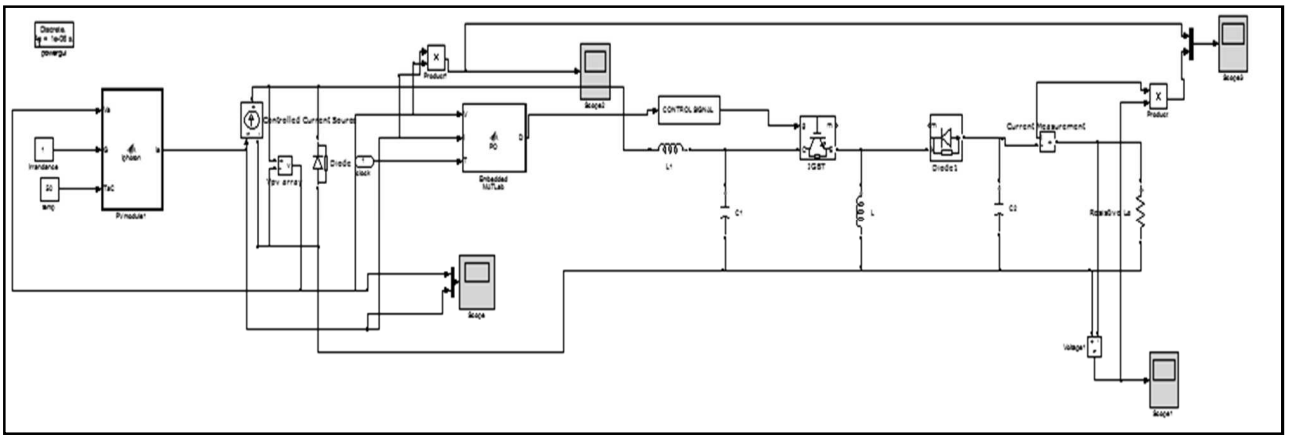


Fig 10: Simulation of PV system

Table 6: PV module 285PC8

Voltage at V_{max} (V_{mp})	35.9 V
Current at I_{max} (I_{mp})	7.95 A
Maximum Power	276 Watt
Open circuit voltage V_{oc}	44.5 V
Short circuit current I_{sc}	8.56 A
Module efficiency	14.37 %

Simulation result

PV module is simulated with MPPT with boost converter and output voltage, output power curve is plotted. Figure 11-16 shows input and output power.

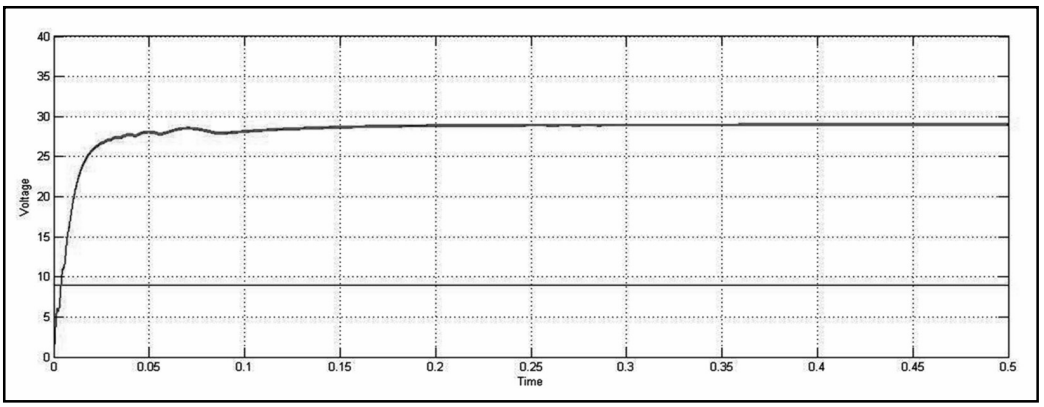


Fig 11: Input voltage 26V before boost converter

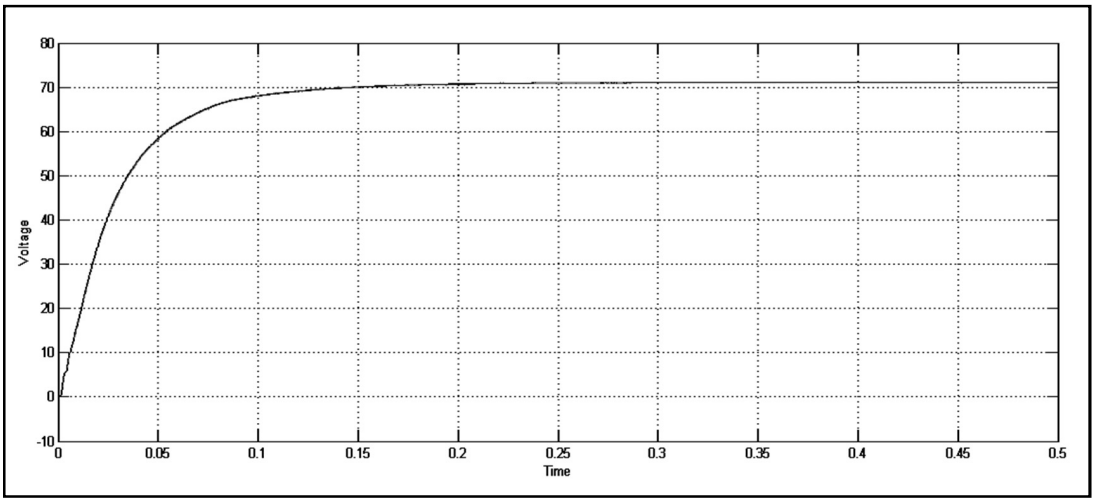


Fig 12: Output voltage 70V after boost converter

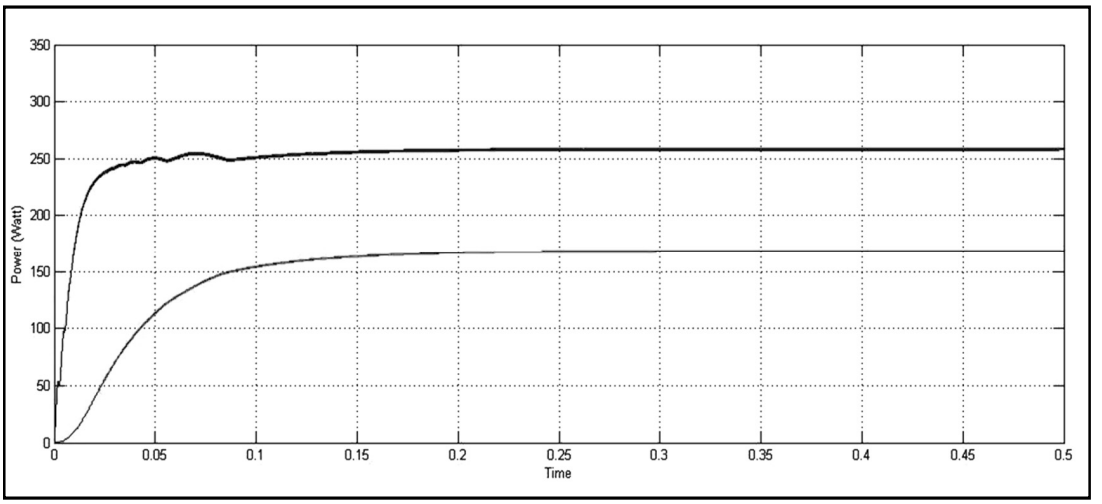


Fig 13: Output and input power at 1000W/m^2

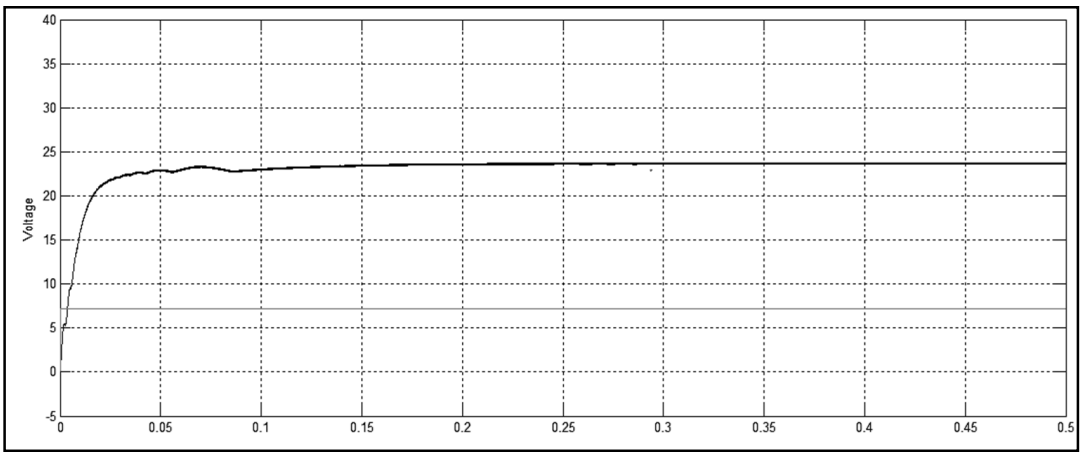


Fig 14: Input voltage 21V before boost converter

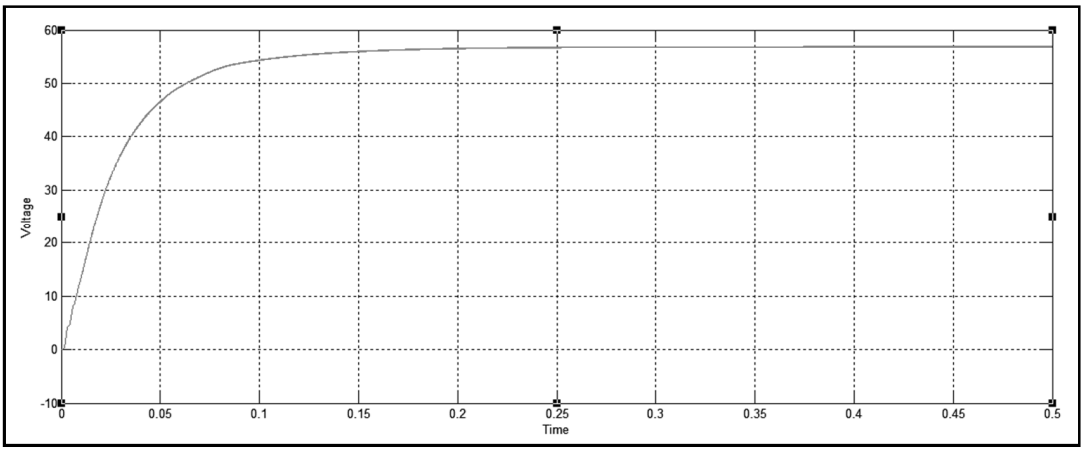


Fig 15: Output voltage 55V after boost converter

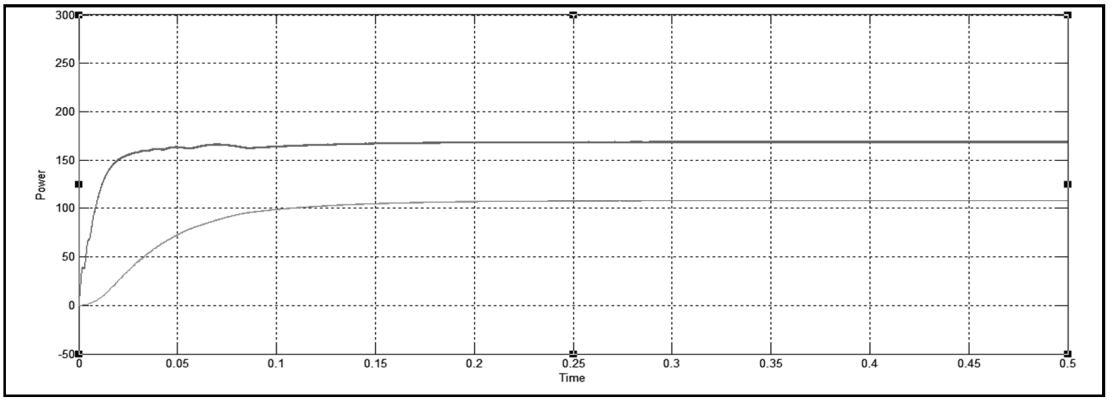


Fig 16: Input and output power at 800 W/m^2

Table 7: Simulation result

Irradiance	Input voltage	Output voltage	Input power before MPPT	Output power after MPPT
1000W/m^2	26 V	70V	150 W	250W
800W/m^2	21 V	55V	90W	150W

Hardware Simultion Results

With the advantage of automatic generated code its gives another benefit of a low price which is very economical to the industry purpose. MPP is tracked using P & O Algorithm, for tracking this algorithm different medology used to track MPP and to extract the maximum power form the PV module. DsPACE, FPGA, Sensors & PIC are the different mode to track the MPP but these modules have some disadvantage which is higher cost, complex, and not user friendly system. If we consider the DsPACE and FPGA module they are they are costly and difficult to implement on the large scale. In the PIC family 18, 24 c coding is to be done in this configuration it is very difficult to find the error and error line is not shown in the system we have to find the error by checking all the line which will make very complex and lengthy. Microchip has developed a unique configuration DsPIC30F4011 which contains all the parameter to make the system low cost, easy and reliable.DsPIC30F4011 has a six PWM channel which can be utilized in the converter and automatic C/C++ code can be generated and easily dumped on the chip using MPLab. In this error finding is very simple if the simulation has some error then built will not success and error in the line can be shown in red color. This makes system easier and reliable.

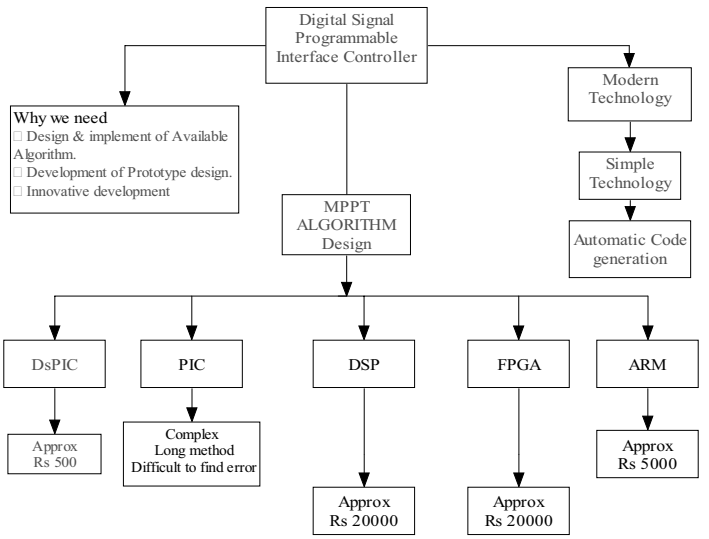


Fig 17: Flow chart showing advantage of DsPIC controller

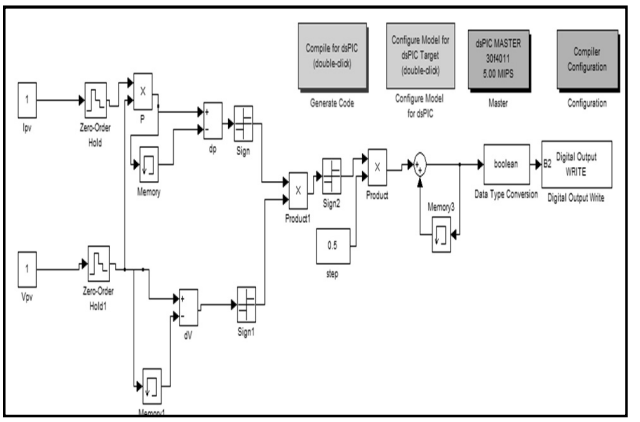


Fig 18: Simulation of P&O Algorithm using DsPIC Patch

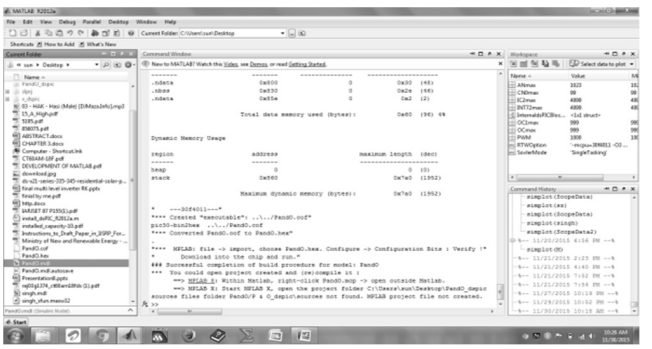


Fig 19: Successful build for P&O Algorithm

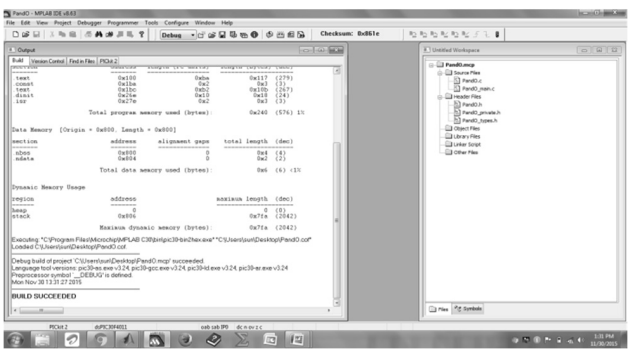


Fig 20: Successful build of code in MP Lab

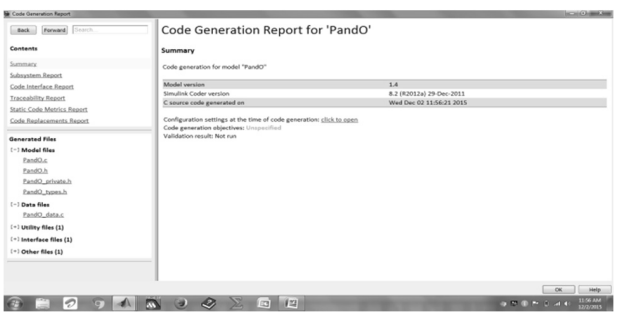


Fig 21: C Code generated report for P&O Algorithm

Hardware Implementation for Pv System

Hardware design of PV system is done using DsPIC30F4011. Flow chart describing the theme of the hardware.

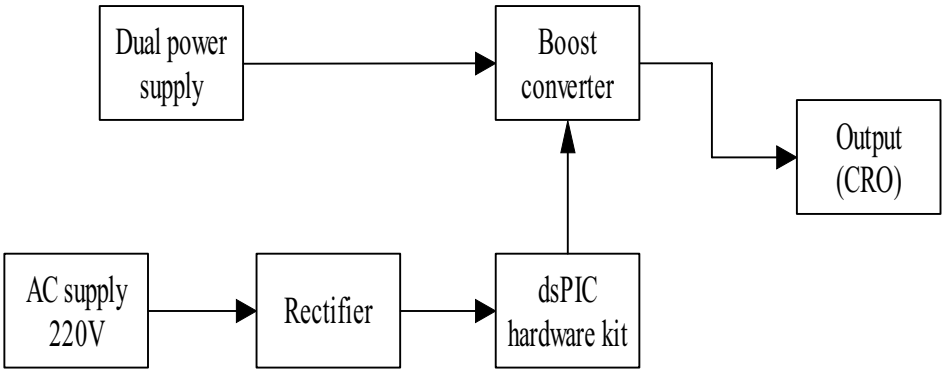


Fig 22: Block diagram showing PV system hardware

Dual power supply is used to vary the dc voltage and dual power supply act as PV module in this 9V is boost up to 24V dc using DsPIC30f4011. In the PV module as the voltage vary w.r.t to temperature and irradiances. Similarly experiment is performed by varying the voltage. Single integrated diagram is shown below-

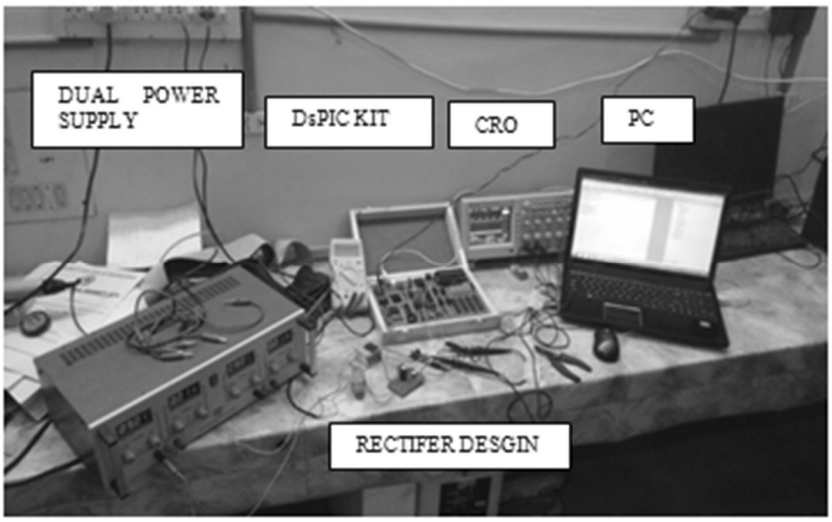


Fig 23: Hardware set up

Table 8: Components Rating

Components	Rating	Numbers
Capacitor	100uF	1
Transformer	220V-15V	1
IGBT	CT60AM-18F	1
TPL-250	3V/15V	1
IC-7825	15V	1
DsPIC30F	4011,6 Channels PWM	1
Boost converter	5V-35V	1
Resistor		2
Hardware kit		1
Diodes		4

Output waveform

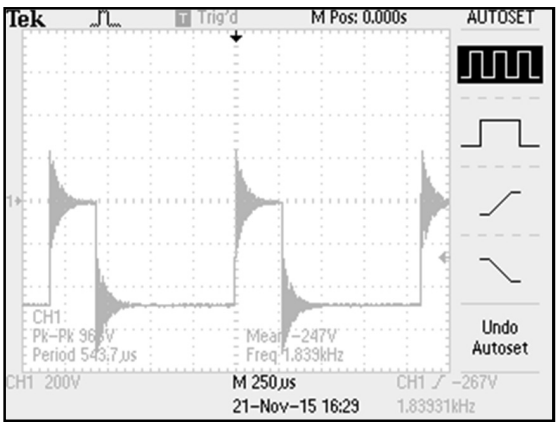


Fig 24: Boost voltage 9V to 24V using P&O Algorithm

Conclusion

Solar PV is a technology that offers a solution for a number of problems associated with fossil fuel. It is clean and continuously imports the energy from the sources. Round the globe India placed in the top which has highest solar irradiances. PV cell covert solar into electricity with a specified irradiance but due to partial shading irradiances goes lower and output power of the PV panel decreases. To get the better output Maximum Power Point (MPP) is track by designing different algorithm and we have designed controller based hardware using P&O algorithm of low cost and reduce complexity due to the automatic code generated in the MATLAB. Simulation and hardware result are successfully obtained.

Future Work

Designed low cost solar PV system can be installed in the small town, residential buildings etc.
The designed model can be used or implemented as a-

- Standalone unit.
- Design for the charge controller.
- Combination of the series and parallel connection can be explored.
- Cope up with the higher non linearity fuzzy algorithm to obtain MPP.
- For the larger PV system they can be implemented with the multi level inverter.
- Design of buck/boost algorithm can also be done to make the system reliable

References

[1] P.Kumar, D.K Palwalia "Decentralized Autonomous Hybrid Renewable Power Generation" "Hindawi Publishing Corporation" Volume 2015, Article ID 856075.
[2] R. Garg, A. Singh, S. Gupta "PV Models and Dynamic simulation of MPPT in MATLAB", IEEE Trans, 2014 pp 6-12.
[3] J.M. Blames, J.A. Carrasco, E. Avila "Maximum Power Point Estimator for Photovoltaic solar array" IEEE Mediterranean MELECON-2006 pp 889-892.
[4] J. Ahmad, "A Fractional open circuit voltage MPPT for PV Array" IEEE ,2nd ICSTE, 2010, Vol-1 pp 247-250.
[5] U. Kamran, S. Sirta, S. Yovsawat "A low cost Micro Controller based Maximum Power Point Tracking System with Multiple String connection for PV stand alone Application" IEEE 10th international conference, PEDS-13,pp 1343-1348.
[6] A. H. Niasar, Z. Zare, F.R. Far "A low cost P&O based MPPT ,combined with two degree sun tracking" IEEE, PEDSTC-2015,pp 119-124.
[7] H. A. Sher, AF Mautaza, A Noman "A new sensorless hybrid MPPT algorithm based on fractional short circuit current measurement and P&O MPPT" IEEE Transaction ,Vol 6,Issue 4 ,pp 1426-1434.
[8] B. Kuczeski, R. Philip, W. C Mensser "A platform for building PIC application for control and instrument" IEEE American conference Vol 7 pp 5162-5168.
[9] K. S. Phani Kiranmai, M. Veerachy "A single stage power conversion system for the PV MPPT application" IEEE international conference pp 2125-2130.
[10] G. Petrone, G. Spagavolo, M. Vitelli "An anolog techniques MPPT PV Application" IEEE Transaction Vol 59, Issue 12 PP 4713-4722.
[11] Kai Chen, Shulin Tain, Yahua Cheng, Libing Bai "An improved MPPT Controller for PV system under partial shading condition" IEEE Transaction Vol 5, Issue 3, PP 978-985.

[12] S. Bokswan, T. Benjanarasth "Code generation of fractional filter for DsPIC Micro controller" IEEE Conference, IENCON-2011 PP-1275-1279.
[13] F. Palmino, J. O Pinto, L. H Pereria "cost effective photovoltaic water pumping system for remote regions communities" IEEE conferences, ECCE-2014.PP-3426-3433.
[14] S. Ozturk, I.Cardirci "DSPIC Microcontroller based implementation of a fly back PV micro inverter using direct digital synthesis" IEEE conference, ECCE-2013, PP 3426-3433.
[15] A. Alnabusi, R. Dhauadi " Efficiency optimization of a DSP based standalone PV system using fuzzy logic an dual MPPT control" IEEE Transaction Vol 8, Issue 3, PP 573-584.
[16] laird, L. VDD "High step up dc dc topology and MPPT algorithm for use with a thermoelectric generator". IEEE Transaction PP 3147-3157.
[17] S.S. Valnjkar, S. D. Joshi, N. R. Kukarni "Implementation of MPPT charge controller for renewable energy" IEEE conference ICACCCT-2014, PP-255-259.
[18] Y. P. Siwakoti, B. B. Chhetri, D. Bista "Microcontroller based intelligent DC/DC converter to track maximum power point for solar photovoltaic module, IEEE conference CITRES 2010 PP 94-101.
[19] S. R. Osman, N. A. Rohin "Microcontroller based solar battery charging system with MPPT feature with low irradiance" IEEE conference CGAT 2013 PP 437-441.
[20] M. Killi, S. Samata "Modified perturb and observe MPPT algorithm for drift avoidance in photovoltaic system" IEEE Transactions Vol-62, Issue 9, PP 5549-5559.
[21] D. Sera, L. Mathe, T. Kerkes " on the P&O and INC MPPT methods for PV system" IEEE Journal Vol 3 Issue 3,PP 1070-1078.
[22] M. R. Islam, Guo Youguaugo, "Simulation of PV array characteristics and fabrication of Micro controller based MPPT" IEEE conference ICECE 2010, PP 155-158.
[23] M.S. Palli, F. M. Munshi, S. L. Medegar "Speed control of BLDC Motor with four switch three phase inverter using digital signal controller" IEEE conference, ICPACE-2015,PP-371-376.
[24] S. Z. Munji, R.A. Rahim, MHF Rahiman "Two controller interaction using C"IEEE conference 2010 PP 290-292.
[25] S. K. Kolimalla, M. K Mishra, "Variable perturbation size adaptive P&O ,MPPY Algorithm for sudden changes in irrandiance" IEEE transaction Vol 5 Issue 6 PP 718-728.
[26] H. Pate "MATLAB-Based Modeling to Study the Effects of Partial Shading on PV Array Characteristics" IEEE transaction Vol 23 Issue 1,2008 PP 302-310.
[27] T. Esram, P. L. Chapman "Comparison of Photovoltaic Array Maximum Power Point Tracking Techniques" IEEE transaction Vol 22 Issue 2, 2007 PP 439-449.
[28] P. Sanjeevi Kumar, G. Grandi "A Simple MPPT Algorithm for Novel PV Power Generation System by High Output Voltage DC-DC Boost Converter" IEEE Conference ISTE 2015 PP 214-220.
[29] Adithya Vangari, Divyanagalakshmi Haribabu "Modeling and Control of DC/DC Boost Converter using K-Factor Control for MPPT of Solar PV System" IEEE Conference ICEEE 2015 PP 1-6.
[30] Sarthak Jain, Anant Vaibhav "Comparative Analysis of MPPT Techniques for PV in Domestic Applications" IEEE Conference PIICON 2014, PP 1-6.
[31] Asim Datta, Dipankar Mukherjee, Sanjoy Debbarma "A DsPIC based Efficient Single-Stage Grid-Connected Photovoltaic System" IEEE conference TENCON 2014 PP 1-4.
[32] S. S. Valunjkar, S. D. Joshi "Implementation of Maximum Power Point Tracking Charge Controller for Renewable Energy" IEEE Conference ICACCCT 2014, PP 255-259.
[33] A. Nadjah, B. Kadri, "New Design of Dual Axis Sun Tracker with DSPIC Microcontroller" IEEE Conference PEMC-2014 PP 1030-1034.
[34] Mohammad B. Shadmand, Mostafa Mosa "An Improved MPPT Technique for High Gain DCDC Converter using Model Predictive Control for Photovoltaic Applications" IEEE Conference APEC 2014 PP 2993-2999.
[35] Ahmed Abdalrahman, Abdalhalim Zekry "Control of The Grid-connected inverter using DsPIC Microcontroller" IEEE conference JEC-ECC-2013, PP 159-164.
[36] Abdullah M. Noman, Khaled E. Addoweesh "Simulation and DSPACE Hardware Implementation of the MPPT Techniques Using Buck Boost Converter" IEEE conference CCECE 2014 PP 1-8.
[37] M. Elshaer, A. Mohamed "Smart Optimal Control of DC-DC Boost Converter in PV Systems" IEEE conference (T&DNA) -2015, PP 403-410.
[38] Lajos Török, Stig Munk-Nielsen "Simple Digital Country of a Two-Stage PFC Converter Using DSPIC30F Microprocessor" IEEE conference PEMD-2010 PP 1-4.
[39] Khanchai Tunlasakun, Krissanapong Kirtikara, Sirichai Thepa "Comparison Between FPGA-Based and DsPIC-Based Islanding Detection for Grid Connected Inverter in a Photovoltaic System" IEEE conferenceICCAS-2007 PP 416-419.
[40] T.-F. Wu, S.-A. Wang, C.-L. Kuo "Design and Implementation of a Push-Pull Phase-Shifted Bi-directional Inverter with a DsPIC Controller" IEEE conference PEDS-2009 PP 728-733.
[41] Nicola Femia, Giovanni Petrone "Optimization of Perturb and Observe Maximum Power Point Tracking Method" IEEE Trans. Vol 20 Issue 4 PP 963-973.
[42] Panom Petchjatupom, Wannaya Ngamkham "A Solar-powered Battery Charger with Neural. Network Maximum Power Point Tracking Implemented on a Low-Cost PlC-microcontroller" IEEE conference PEDS-2015 PP 507-510.
[43] K. S. Phani Kiranimai, M. Veerachary "A Single-Stage Power Conversion System for the PV MPPT Application" IEEE conference ICIT-2006 PP 2125-2130.

Code Generated For P&O Algorithm

```
 2  * PandO.c
 3  *
 4  * Code generation for model "PandO".
 5  *
 6  * Model version              : 1.4
 7  * Simulink Coder version: 8.2 (R2012a) 29-Dec-2011
 8  * C source code generated on: Mon Nov 30 13:30:34 2015
 9  *
10   * Target selection: dspic.tlc
11   * embedded hardware selection: 16-bit Generic
12   * Code generation objectives: Unspecified
13   * Validation result: Not run
14   */
15  #include "PandO.h"
16  #include "PandO_private.h"
17
18  /* Real-time model */
19  RT_MODEL_PandO PandO_M_;
20  RT_MODEL_PandO *const PandO_M = &PandO_M_;
21  static void rate_monotonic_scheduler(void);
22
23  /*
24   * Set which subrates need to run this base step (base rate
always runs).
25    * this function must be called prior to calling the model
step function
26   * in order to "remember" which rates need to run this
base step.  The
27   * buffering of events allows for overlapping preemption.
28   */
29  void PandO_SetEventsForThisBaseStep (boolean_T
*eventFlags)
30  {
31    /* Task runs when its counter is zero, computed via
rtmStepTask macro */
32    eventFlags[1] = ((boolean_T)rtmStepTask(PandO_M,
1));
33  }
34
35  /* rate_monotonic_scheduler */
36  static void rate_monotonic_scheduler (void)
37  {
38    /* Compute which subrates run during the next base time
step.  Subrates
39     * are an integer multiple of the base rate counter.
Therefore, the subtask
40     * counter is reset when it reaches its limit (zero means
run).
41     */
42    (PandO_M->Timing.TaskCounters.TID [1]) ++;
43    if ((PandO_M->Timing.TaskCounters.TID [1]) > 1) {/*
Sample time: [0.002s, 0.0s] */
44      PandO_M->Timing.TaskCounters.TID [1] = 0;
45    }
46  }
47
48  /* Model output function for TID0 */
49  void PandO_output0(void)             /* Sample time:
[0.001s, 0.0s] */
50  {
51    {                                  /* Sample time: [0.001s, 0.0s] */
52      rate_monotonic_scheduler();
53    }
54  }
55
56  /* Model update function for TID0 */
57  void PandO_update0(void)             /* Sample time:
[0.001s, 0.0s] */
58  {
59    /* (no update code required) */
60  }
61
62  /* Model output function for TID1 */
63  void PandO_output1(void)             /* Sample time:
[0.002s, 0.0s] */
64  {
65    /* (no output code required) */
66  }
67
68  /* Model update function for TID1 */
69   void PandO_update1 (void)                  /* Sample time:
[0.002s, 0.0s] */
70  {
71    /* (no update code required) */
72  }
73
74  void Pando output (int_T tid)
75  {
76    switch (tid) {
77     case 0:
78      PandO_output0 ();
79      break;
80
81     case 1:
82      PandO_output1();
83      break;
84
85     default:
86      break;
87    }
88  }
89
90  void Pando update (int_T tid)
91  {
92    switch (tid) {
93     case 0:
94      PandO_update0();
95      break;
96
97     case 1 :   98       PandO_update1();
99      break;
100
101     default:
102      break;
103    }
104  }
105
106  /* Model initialize function */
107 voids Pando initialize (void)
108  {
109    /* Registration code */
110
111    /* initialize real-time model */
112    (void) memset((void *)PandO_M, 0,
```

```
113                 sizeof(RT_MODEL_PandO));
114
115     /* S-Function "dsPIC_MASTER" initialization Block:
<S1>/Master */
116     /* Solver mode : Multitasking */
117     /* CONFIG TIMER 1 for scheduling steps             */
118     ConfigIntTimer1 (T1_INT_PRIOR_1 & T1_INT_ON);
119     T1CON = 0x8000;                 /* T1_PS_1_1       */
120     PR1 = 4999;
121
122     /* Configuration TRIS */
123     /* Configuration ADCHS */
124     ADPCFG = 0U;
125   }
126
```

A New Approach of Offline Parameters Estimation for Vector Controlled Induction Motor Drive

Krishna R More*, P N Kapil** and Hormaz Amrolia***
* Department of Electrical Engineering, Institute of Technology, Nirma University, Ahmedabad, India
14meep17@nirmauni.ac.in
** Department of Electrical Engineering, Institute of Technology, Nirma University, Ahmedabad, India
pnkapil@nirmauni.ac.in
***Department of Electrical Engineering, Institute of Technology, Nirma University, Ahmedabad, India
hormaz.amrolia@nirmauni.ac.in

Abstract: For efficient performance of any vector controlled drive, motor parameters are to be estimated with utmost accuracy. Offline and online parameter estimation techniques are generally used for estimating parameters. In this paper, the offline technique is adopted due to its advantage over latter technique. Under offline parameter estimation, the new strategy is simulated for estimating the machine parameters. Inverter test and single phase test are used for estimating parameters. Only three switches of the inverter are used for estimating parameters in inverter test and four switches in single phase test. These tests are easy to implement with no extra hardware, no mechanical job is required and equations are easy to understand and implement. These tests are simulated in MATLAB Simulink and parameters obtained are compared with actual parameters of machine model available in MATLAB Simulink.

Keywords: Offline parameter estimation, inverter test, single phase test, dc resistance test, open circuit test, locked rotor test.

Introduction

Induction motors are widely used the electric machines in high-performance drive applications. Vector control of induction motor requires a precise knowledge of the motor parameters for its efficient operation. Numerous methods for estimating induction machine parameters such as online and offline parameter estimation have been developed exclusively for application in high-performance drives. Any vector controlled induction motor drive is inverter fed, numerous tests based on an inverter supply have been developed in recent past for determination of the required parameter values. Such methods are called offline parameter identification methods. Numerous possibilities exist nowadays to update the parameter values during the drive operation, termed as online parameter estimation methods [1], [2], [3]. Offline parameter estimation technique is adopted, as parameters are estimated at the initial stage itself with the inverter and then the estimated parameters are used in the system. There is no need to sense the signals continuously in offline parameter estimation, in order to estimate parameters as it is mandatory in online parameter estimation. No standard tests like an open circuit, locked rotor test are required where the mechanical task is required to estimate the parameters. Moreover, for estimating parameters offline, no PI controller is required. So, hardware requirement reduces. The only inverter, sensors, and controller are required in this technique. Offline parameter estimation technique uses inverter test and single phase test for estimating machine parameters. Inverter test for R_s, L_s and single phase test for L_{ls}, L_m and R_r.

Parameter Estimation Tests

An equivalent circuit of induction motor valid for steady state is shown in Fig. 1.The circuit is composed of two resistors, the stator resistance (R_s) and the rotor resistance (R_r), and three ideal inductors, the stator leakage inductance (L_{ls}), the rotor leakage inductance (L_{lr}) and the magnetizing inductance (L_m). The stator voltage is represented by V_s and s is the slip frequency normalized with respect to the stator frequency. The purpose of this work is to measure the values of these five ideal components.

Traditional Parameter Estimation Tests

DC Resistance Measurement

DC Measurement test is used to determine the stator resistance (R_s). This test consists of applying a DC voltage to stator winding. The DC voltage (V_{dc}) and current (I_{dc}) are measured and R_s value is computed by dividing both measurements. Basically, a dc voltage is applied to the stator windings of an induction motor. Because the current is dc, there is no induced

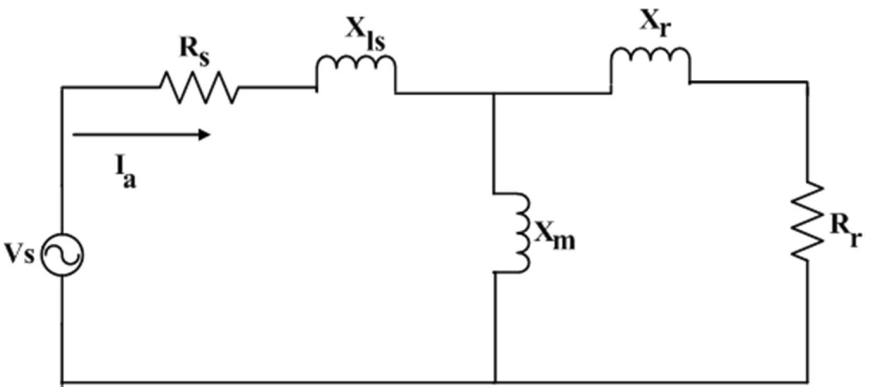


Figure 1: Steady State Induction Motor per phase equivalent circuit

voltage in the rotor circuit and no resulting rotor current. Also, the reactance of the motor is zero at direct current. Therefore, the only quantity limiting current now in the motor is the stator resistance, and that resistance can be determined. The basic circuit for the dc test is shown in Fig. 2 where, dc power supply connected to two of the three terminals of a Y-connected induction motor. [6]

$$R_s = \frac{V_{dc}}{2I_{dc}} \qquad (1)$$

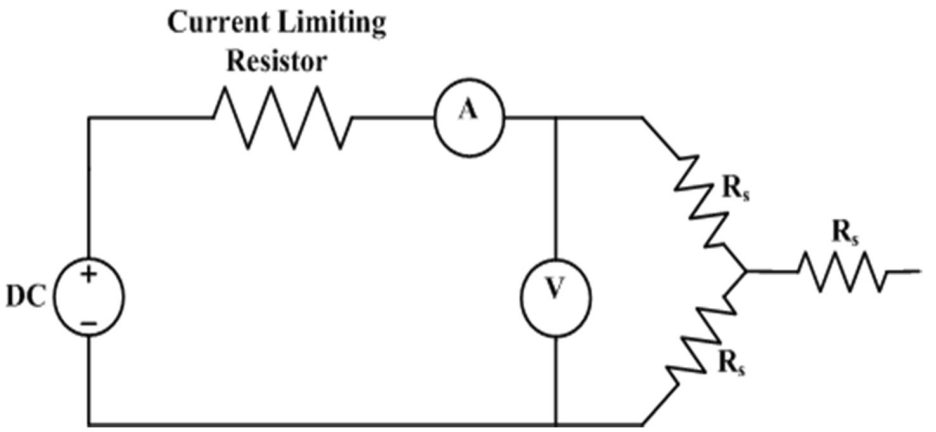


Figure 2: DC Resistance Test

No Load Test

Input power and motor current are measured at different voltage levels. Motor current and input power at rated voltage is called as no load current (I_{nl}) and no load power (P_{nl}) respectively. Motor no-load resistance, no load impedance, and no load reactance can be calculated. No load test is performed to calculate no load (stator) reactance (X_{nl}) or stator inductance (L_s). The motor is allowed to run at rated voltage and allowed to reach rated speed on no-load. The equation to measure (X_{nl}) is given below (equ.2). The equivalent circuit of induction motor under a no-load condition is shown below in Fig. 3. The simulation arrangement for open circuit test is also shown below in Fig. 4 [5].

$$X_{nl} = \frac{Q}{I^2} \qquad (2)$$

$$Q = \sqrt{(V_a I_a)^2 + (P_a)^2} \qquad (3)$$

where, Q = Reactive Power, V_a = Phase Voltage, I_a = Phase Current, P_a = Active power per phase, X_{nl} = No-Load Reactance

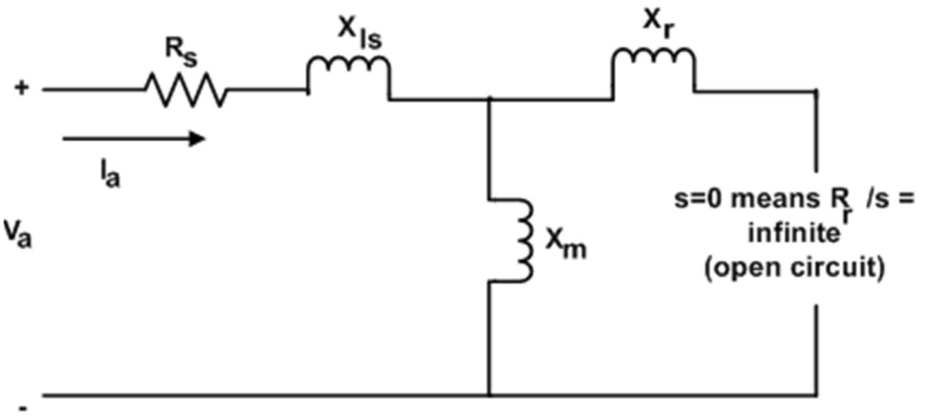


Figure 3: Three phase equivalent circuit of induction motor under no load condition

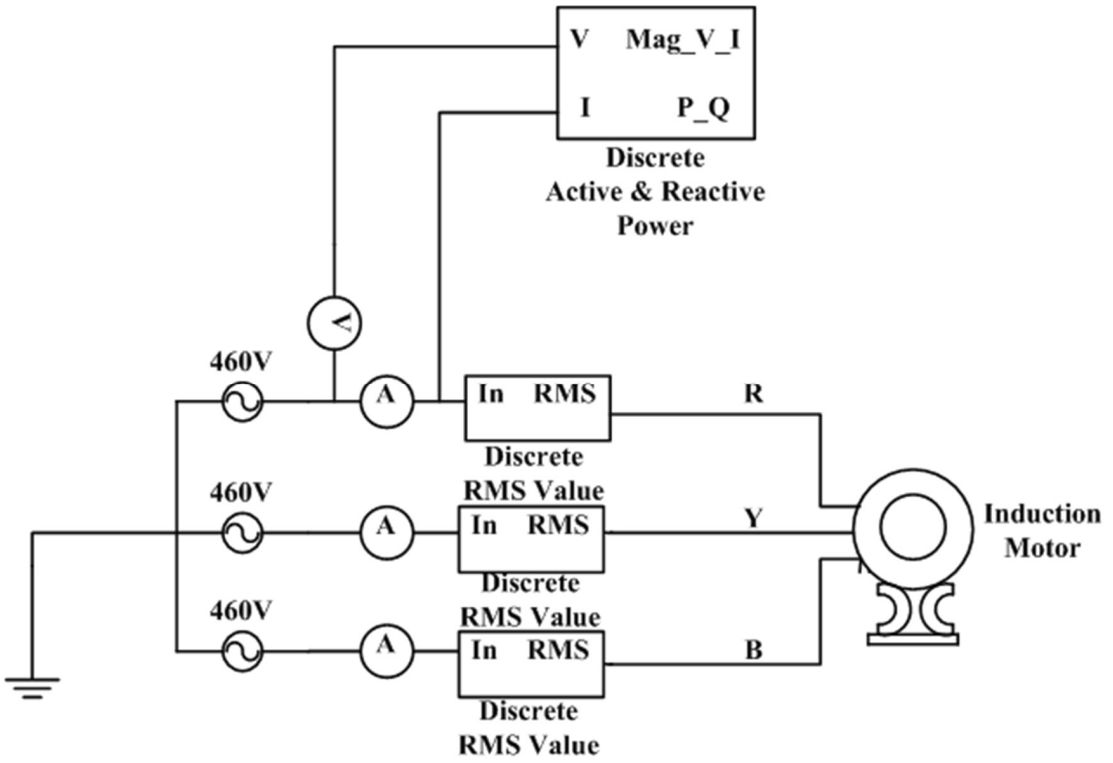


Figure 4: Simulation arrangement for induction motor under no load condition

Locked Rotor Test

Motor phase to phase voltage and input power at rated current are called as locked rotor voltage (V_{lr}) and locked rotor power (P_{lr}). Motor locked rotor resistance (R_r), locked rotor impedance (Z_r) and locked rotor reactance (X_r) can be calculated. Fig.5 shows the behavior of an induction motor under locked rotor test. Locked rotor test is performed to obtain stator leakage reactance (X_{ls}) (equ.4) basically stator leakage inductance (L_{ls}) is estimated (equ.5), magnetizing reactance (X_m) i.e. magnetizing inductance (L_m) can be estimated from (equ.6) and (equ.7) respectively. Rotor resistance (R_r) is estimated by using X_{ls}, X_m and R_2 (equ.11). In this test, the motor is allowed to run at a voltage less than the rated voltage till the rated current flows through the windings. Another way to estimate Rr is by using power factor (equ.12) and (equ.13). The simulation arrangement for locked rotor test is also shown below Fig.6 [5].

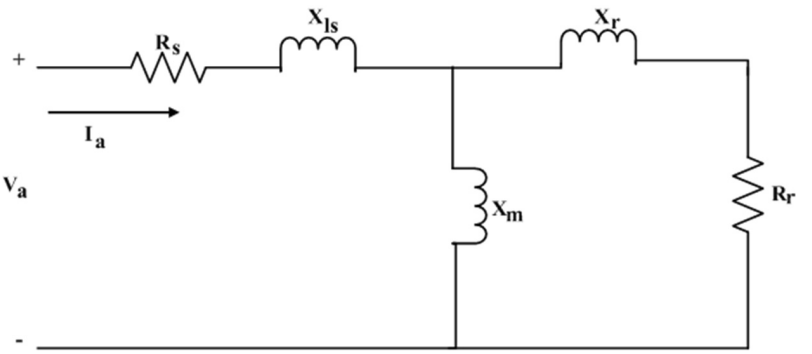


Figure 5: Simplified equivalent circuit of three phase induction motor under locked rotor test

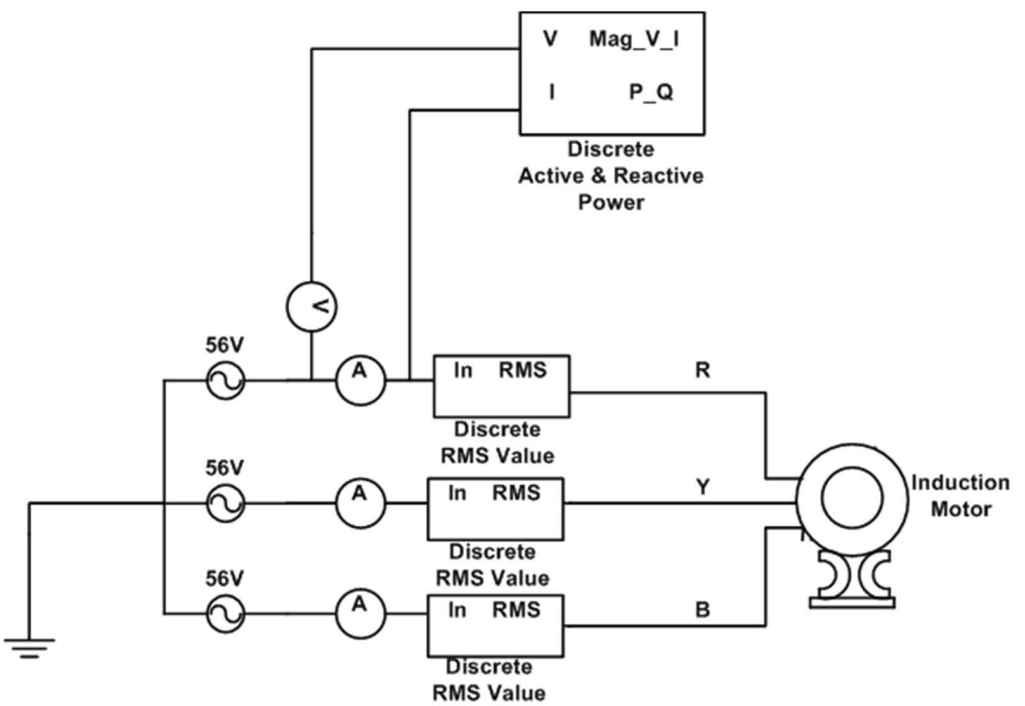


Figure 6: Simulation arrangement for induction motor under locked rotor condition

The equations for estimating various parameters are mentioned below:-

$$X_{ls} = 0.3 * \frac{Q}{I^2} \tag{4}$$

where, Q is Reactive Power
Here, 0.3 factor to be multiplied for the class C motor [4].

$$L_{ls} = L_{lr} = \frac{X_{ls}}{2\pi f} = \frac{X_{lr}}{2\pi f} \tag{5}$$

$$X_m = X_{nl} - X_{ls} \tag{6}$$

$$L_m = \frac{X_m}{2\pi f} \tag{7}$$

$$L_s = L_{ls} + L_m \tag{8}$$

$$R = \frac{P}{I^2} \tag{9}$$

where, P is Active Power

$$R_2 = R - R_s \tag{10}$$

$$R_r = R_2 * \left(\frac{X_{ls} + X_m}{X_m} \right)^2 \tag{11}$$

$$\cos\phi_s = \frac{P}{\sqrt{3} * V_a * I} \tag{12}$$

$$V_a = \left(I_{sc} \angle \phi_s \right) * \left(R_s + R_r + jX_{ls} \right) \tag{13}$$

where, I_{sc} is short circuit current and ϕ_s is lagging power factor angle.

The calculation results are compared and percentage error between the actual parameter of machine model available in MATLAB Simulink and calculated parameter through above equations is also calculated are shown in tables below.

Table 1. For 5 HP Machine Model available in MATLAB

Parameter	Calculated Value	Actual Value	Percentage Error (%)
R_s	0.923 Ω	1.115 Ω	17.21
L_s	0.20895 H	0.209674 H	0.345
L_{ls}	0.00355 H	0.005974 H	40.57
L_m	0.2054 H	0.2037 H	0.834
R_r	1.0689 Ω	1.083 Ω	1.3

Table 2. For 10 HP Machine Model available in MATLAB

Parameter	Calculated Value	Actual Value	Percentage Error (%)
R_s	0.5851 Ω	0.6837 Ω	14.42
L_s	0.15226 H	0.152752 H	0.3
L_{ls}	0.00246 H	0.004152 H	40.75
L_m	0.1498 H	0.1486 H	0.8
R_r	0.4956 Ω	0.451 Ω	1.2

Table 3. For Default HP Machine Model available in MATLAB

Parameter	Calculated Value	Actual Value	Percentage Error (%)
R_s	0.02873 Ω	0.029 Ω	0.93
L_s	0.035175 H	0.035489 H	0.88
L_{ls}	0.000355 H	0.000599 H	40.73
L_m	0.03482 H	0.03459 H	0.66
R_r	0.02052 Ω	0.22 Ω	6.81

Table 4. For 1 HP Machine Model available in MATLAB

Parameter	Calculated Value	Actual Value	Percentage Error (%)
R_s	9.589 Ω	12.625 Ω	24.05
L_s	0.803 H	0.8117 H	0.07
L_{ls}	0.0361 H	0.0617 H	41.49
L_m	0.7762 H	0.750 H	3.49
R_r	11.63 Ω	12.316 Ω	5.57

Proposed Parameter Estimation Methods

Buck Chopper

Fig.7 shows the circuit diagram of buck chopper. When the switch is ON energy is supplied to the load and is simultaneously stored in the inductor. During the ON period diode is in reverse bias condition. When the switch is OFF the energy stored in the inductor is fed to the load, which flows through the diode making it forward bias. Thus, the load is supplied with energy at every instant as long as input supplies the energy. Duty ratio decides the ON and OFF period of the switch.

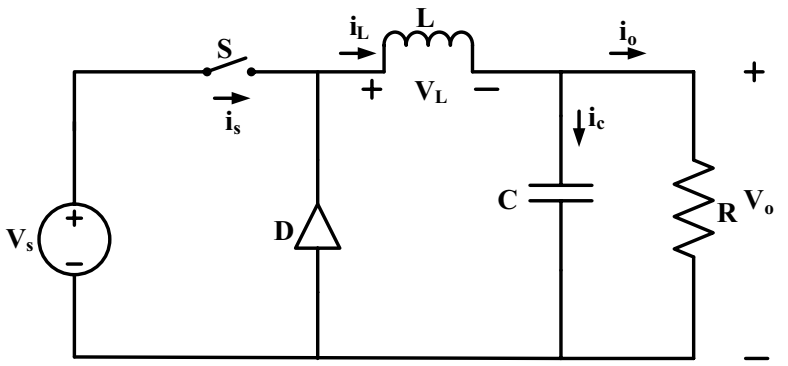


Figure 7: Buck Chopper

Here, in buck chopper 75% is set as duty ratio for 1 kHz frequency at 10 V. The value of L=10 mH and load is 10Ω. The voltage and current waveforms are shown in Fig.8 and Fig.9.

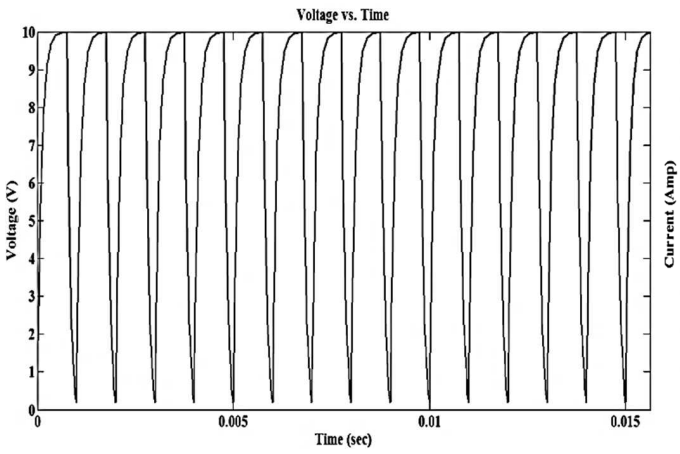


Figure 8: Voltage Waveform

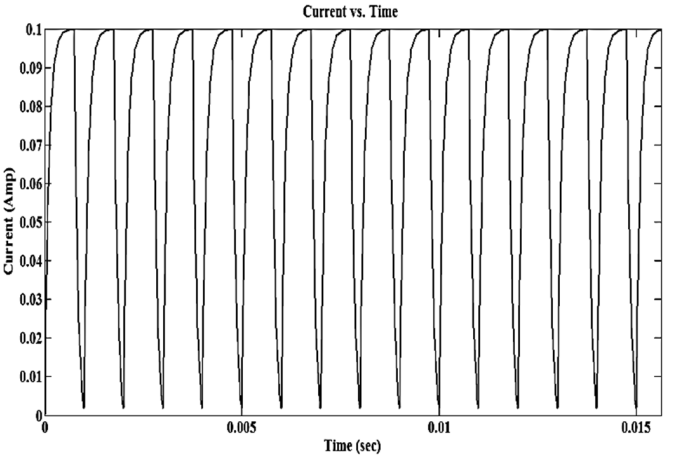


Figure 9: Current Waveform

Inverter as a Buck Chopper

A new strategy is proposed for measuring R_s and L_s using an inverter, which is operated as a buck chopper. Fig.10 shows inverter operating as a buck chopper. As shown in Fig. 10, when switch S1 and S6 are ON, current flows through motor windings through phase R and phase Y and circuit gets complete as S6 is ON. When S1 is OFF energy stored in motor windings flows through phase Y and phase R as S6 and S4 are ON now as shown in Fig. 11. The duty ratio is 75% for S1 and it is 99.9% for S6. Only three switches S1, S4, S6 are being turned on to estimate these parameters. The R-phase and Y-phase are only in action i.e. the machine is getting supply from R and Y phase only. The value of L_s is calculated by observing the value of time at 63.8% of I_{max} (definition of the time constant). The magnified image of the current waveform is shown in the Fig.14, which shows the value of 63.8% of I_{max} is obtained at 0.1087 sec.

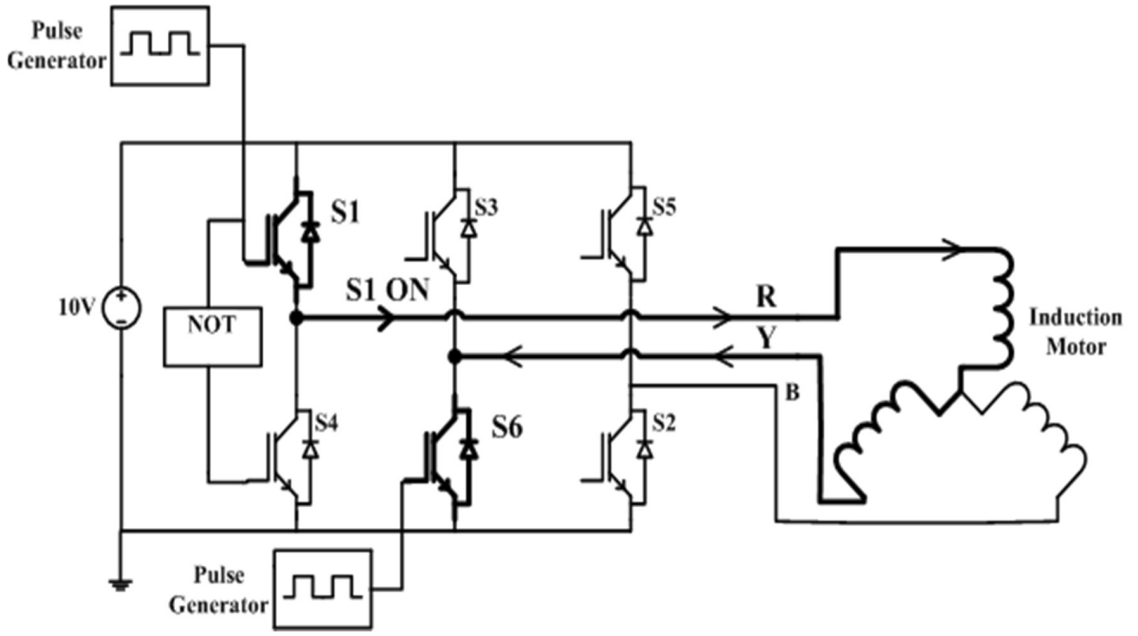


Figure 10: Inverter Circuit for estimating R_s & L_s - ON State

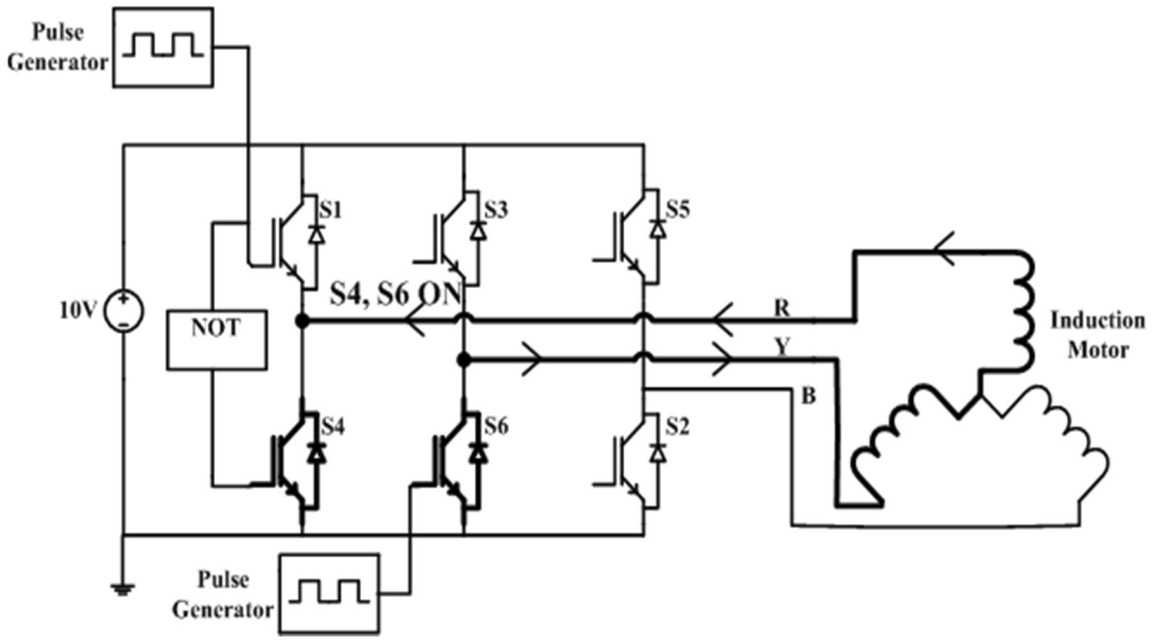


Figure 11: Inverter Circuit for estimating R_s & L_s - OFF State

The R_s is measured with the formula (equ.14) and the time constant is used to measure L_s (equ.15) with the help of current waveform.

$$R_s = \frac{DV_{dc}}{2I_{dc}} \tag{14}$$

where, D = Duty Ratio of the switch

$$\tau = \frac{L_s}{R_s} \tag{15}$$

where, τ = Time constant i.e. Time at 63.8% of Imax
The voltage and current waveform are shown in the figures below.

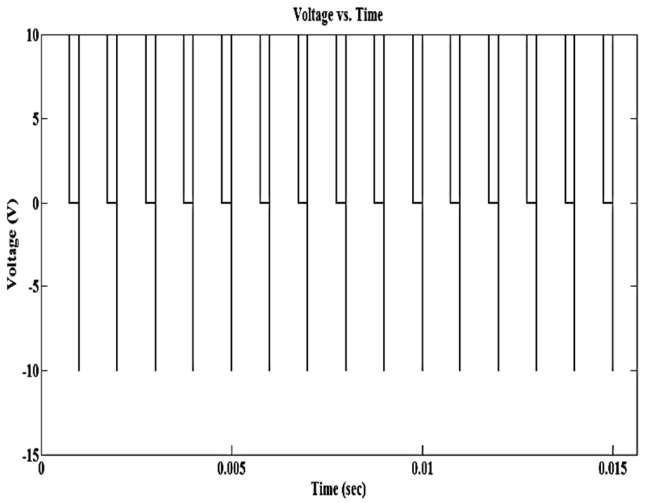


Figure 12: Voltage Waveform

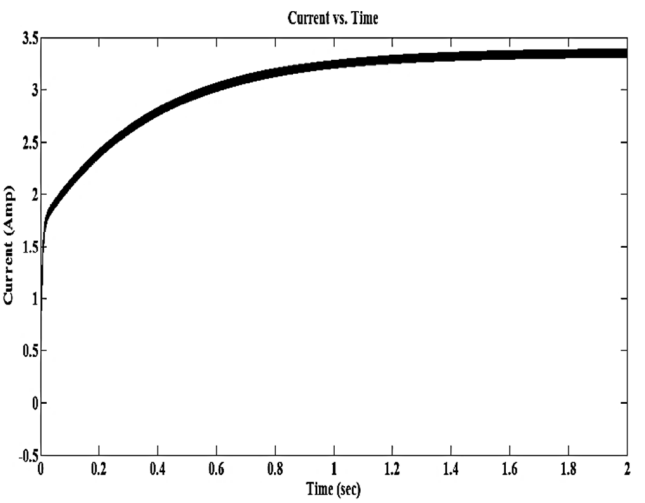


Figure 13: Current Waveform

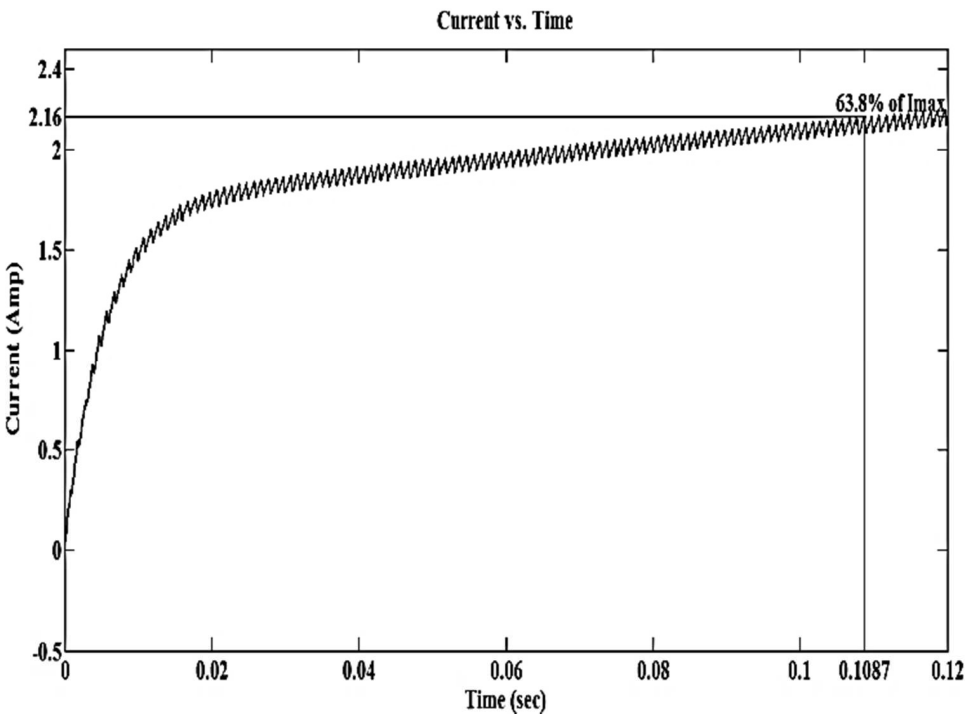


Figure 14: Magnified Current Waveform

Single Phase Test as Locked Rotor test

In single phase test motor is supplied with on only single phase by disconnecting one of the phases out of the three phases. Thus, the condition becomes like locked rotor test, where the rotor is blocked manually. So, equations used for calculating the parameter in locked rotor tests can be used to estimate parameters through single phase test. This test is used to estimate stator leakage inductance (L_{ls}), magnetizing inductance (L_m), rotor resistance (R_r). The circuit diagram for single phase test is shown in figures below. The current flows through S1, S6 for the positive half cycle Fig.15 and for negative half cycle it flows through S3, S4 Fig.16. The voltage and current waveform are shown in Fig.17 and Fig.18 respectively.

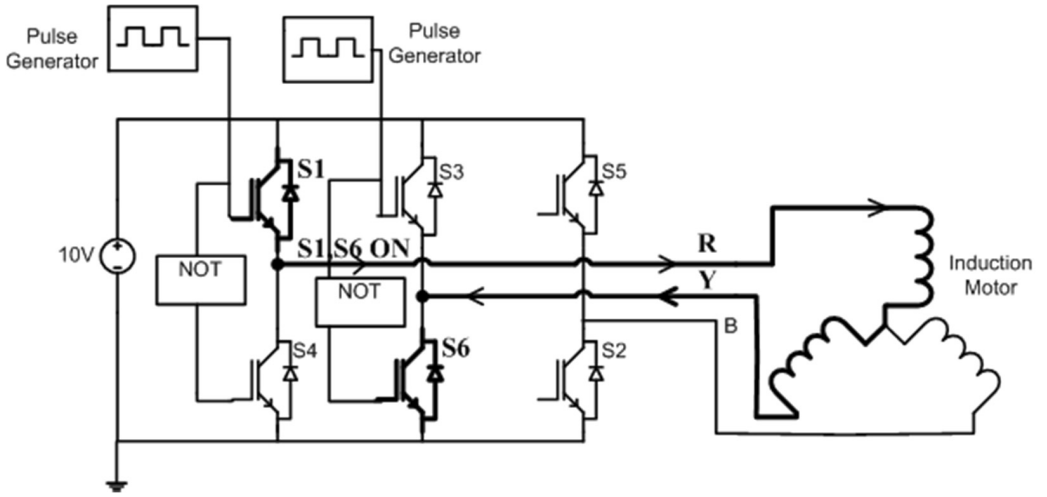


Figure 15: Inverter Circuit for Single Phase test - ON State

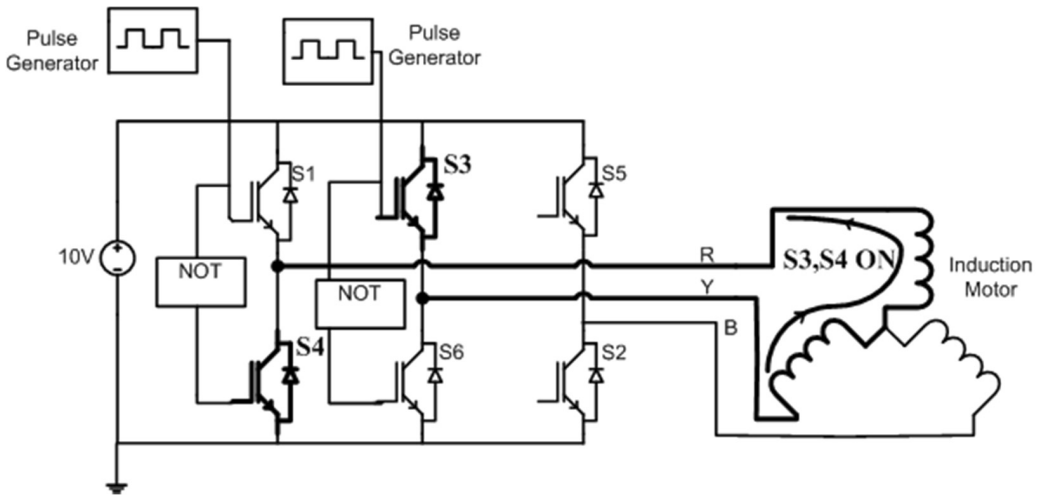


Figure 16: Inverter Circuit for Single Phase test - OFF state

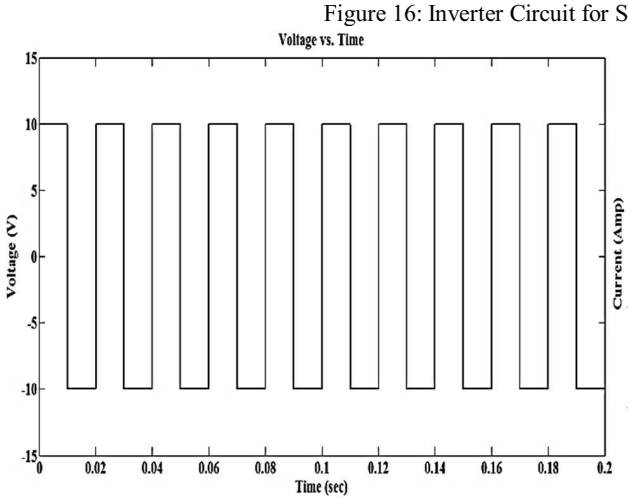


Figure 17: Voltage Waveform

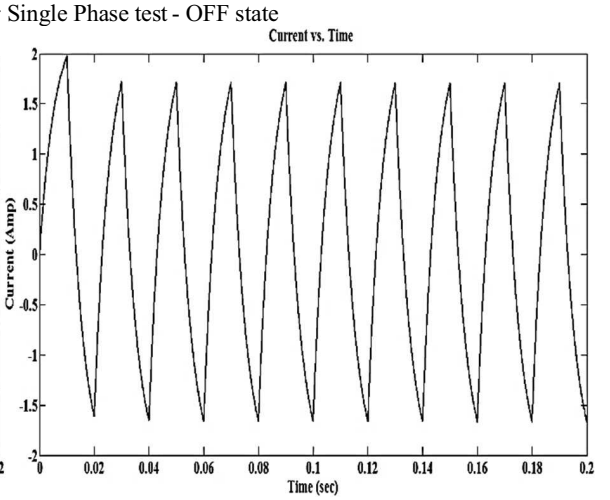


Figure 18: Current Waveform

The calculation results are compared and percentage error between the actual parameter of machine model available in MATLAB and calculated parameter through equations is also calculated are shown in tables below.

Table 5. For 5 HP Machine Model available in MATLAB

Parameter	Calculated Value	Actual Value	Percentage Error (%)
R_s	1.105 Ω	1.115 Ω	0.9
L_s	0.12 H	0.209674 H	42.768
L_{ls}	0.0050 H	0.005974 H	16.308
L_m	0.115 H	0.2037 H	43.54
R_r	2.115 Ω	1.083 Ω	95.3

Table 6. For 10 HP Machine Model available in MATLAB

Parameter	Calculated Value	Actual Value	Percentage Error (%)
R_s	0.6787 Ω	0.6837 Ω	0.73
L_s	0.02898 H	0..152752 H	81.03
L_{ls}	0.00333 H	0.004152 H	19.80
L_m	0.02565 H	0.1486 H	82.74
R_r	1.35 Ω	0.451 Ω	199.33

Table 7. For Default HP Machine Model available in MATLAB

Parameter	Calculated Value	Actual Value	Percentage Error (%)
R_s	0.0299 Ω	0.029 Ω	3.10
L_s	0.0123 H	0.035489 H	65.34
L_{ls}	0.000457 H	0.000599 H	23.71
L_m	0.011843 H	0.03459 H	65.76
R_r	0.0684 Ω	0.022 Ω	210.91

Table 8. For 1 HP Machine Model available in MATLAB

Parameter	Calculated Value	Actual Value	Percentage Error (%)
R_s	12.5 Ω	12.625 Ω	0.1
L_s	0.46 H	0.8117 H	43.33
L_{ls}	0.0516 H	0.0617 H	16.37
L_m	0.4084 H	0.750 H	45.55
R_r	21.82 Ω	12.316 Ω	77.17

Observations

i. With the increase in power of motor the percentage error in R_s decreases while it increases for L_s for the inverter test.
ii. For a specific motor at a constant frequency but varying duty ratio (increasing) percentage error in R_s decreases and also for L_s it decreases.
iii. For a specific motor at constant duty ratio but increasing frequency, percentage error decreases in Rs and Ls.
iv. From calculations, it is found that the value of R_r is less than R_s.
v. From calculation, it is found that the percentage error in L_{ls} is more but the combined percentage error of L_{ls} and L_m i.e. L_s is less i.e. actual L_s equivalent to calculated L_s.
vi. It is found that percentage error of L_s calculated from inverter test is more than that found from open circuit test.
vii. It is found that percentage error of R_r calculated considering power factor is more compared to that when power is not considered.
viii. It is concluded that the percentage error between the actual parameters and calculated parameters is more in single phase test as compared to the lock rotor test.

Conclusion

In offline parameter estimation technique, inverter test, and single phasing test are simulated in MATLAB Simulink for estimating induction machine parameters. Parameters are also estimated with standard tests. The estimated parameter value is compared with an actual parameter value of machine model from MATLAB Simulink. The percentage error is calculated for each parameter. It is observed that R_s is estimated more accurately with inverter test in comparison with DC resistance test. Estimation of L_s is quite accurately done with inverter test. L_{ls}, L_m, R_r are quite accurately estimated from single phase test. When compared with parameter estimation technique used in [1], parameters are precisely estimated with inverter test and single phase test. Hence, with less hardware requirement and complexity, parameters are precisely estimated with inverter test and single phase test.

References

[1] Jun Zheng, Yunkuan Wang, Xiaofei Qin and Xin Zhang, “An Offline Parameter Identification Method of Induction Motor”, Proceedings of the 7th World Congress on Intelligent Control and Automation, Year: 2008, Pages 8898-8901.
[2] Hamid Toliyat, Emil Levi, Mona Rana “A Review of RFO Induction Motor Parameter Estimation Techniques ", Power Engineering Review, IEEE, Volume: 22, Year: 2002, Pages: 271-283.
[3] Hubert Schierling, “Fast And Reliable Commissioning of AC Variable Speed Drives By Self-Commissioning ", Industry Applications Society Annual Meeting, 1988, Conference Record of the 1988 IEEE, Volume: 1, Year: 1988 Pages: 489 - 492.
[4] K.K.Pandey, P. H. Zope “Estimating Parameters of a Three-Phase induction motor using Matlab/Simulink ", International Journal of Scientific and Engineering Research, Volume: 4, Year: 2013, Pages: 425 – 431.
[5] Saffet Ayasun, Chika O. Nwankpa “Induction Motor Tests Using MATLAB/Simulink and Their Integration into Undergraduate Electric Machinery Courses ", IEEE Transactions on Education Volume: 48, Year: 2005, Pages: 37 - 48.
[6] http://iitg.vlab.co.in

A Comparative Study of Switching Strategies for Single Phase Matrix Converter

Mohammadamin Yusufji Khatri * and Hormaz Amrolia **
*Department of Electrical Engineering, Institute of Technology, Nirma University, Ahmedabad, India
aminykhatri@hotmail.com
** Department of Electrical Engineering, Institute of Technology, Nirma University, Ahmedabad, India
hormaz.amrolia@nirmauni.ac.in

Abstract: AC-AC power conversion particularly for speed control of AC drives is done with single phase cyclo-converters. In this work the single-phase matrix converter (SPMC) topology is used as a cyclo-converter. IGBT are used as power switches. The sinusoidal pulse width modulation has been used for generation of pulses. This paper presents three different switching strategies for SPMC. A comparison is made between these three strategies based on parameters such as type of load to which the strategy is restricted, number of switches kept ON during single time interval, output voltage waveform of SPMC and its THD.

Keywords: Sinusoidal Pulse Width Modulation (SPWM), Insulated Gate Bipolar Transistor (IGBT), Single Phase Matrix Converter (SPMC), Pulse Width Modulation (PWM), Complementary of PWM (cPWM).

Introduction

AC-AC conversion consist of converting fixed ac voltage with fixed frequency to variable ac voltage with variable frequency, which can be done by two methods: 1) Indirect method and 2) direct method. Indirect method is the most common approach for the ac-ac power conversion which consist of a rectifier at supply side and inverter at the load side. Such an arrangement would require energy storage element like capacitor or an inductor in the intermediate dc link. These element make the converter bulky and useless in application requiring regenerative operation. These limitation can be overcome by the direct method for ac-ac conversion without any intermediate dc link [1]. Such an operation is performed by a cyclo-converter which converts the ac power at one frequency directly to another frequency. The most desirable features for any power frequency changer are: 1) Simple and compact power circuit 2) Generation of load voltage with arbitrary amplitude and frequency 3) Sinusoidal input and output currents 4) Operation with unity power factor for any load 5) Regeneration capability. These characteristics are not fulfilled by the conventional cyclo-converters, which leads to the use of matrix converter topology as it fulfills the ideal features. The matrix converter provides an "all silicon" solution for direct ac-ac conversion without any intermediate dc link thus eliminating the use of reactive energy storage elements. It consists of bidirectional switches which allows any output phase to be connected to any input phase. The topology was first proposed by Gyugyi. In recent years, matrix converter has gained a lot of attention for the traction application. The commutation between the switches in a matrix converter results in current spikes which are their major drawback [2]. Development of three phase matrix converter started with the work of Venturini and Alesina published in 1980 [4]. The SPMC was first developed by Zuckerberger [5]. Study of other SPMC topology had been carried out by Hossieni [6] and Abdollah Khoei [7] and Saiful [8]. Due to the absence of natural free-wheeling paths commutation issues need to be resolved in any PWM type of converter [9]. When inductive loads are used there may be switching spikes [10]. Switching arrangements for safe commutation is proposed by Zahiruddin. [11]. Amongst previous development of the cycloconverter includes; work on improvements of harmonic spectrum in the output voltage with new control strategies [12], new topology [13] and study of the cyclo-converter behaviour [14]. In this work, three different switching strategies are studied for obtaining AC-AC conversion from 50Hz to 150Hz. A computer simulation model on SPMC for cyclo-converter operation using MATLAB/Simulink (MLS) software package is developed. The simulation results for all three strategies have been portrayed and studied in this paper with all their advantages and limitation.

The single phase matrix converter as a cyclo-converter

The SPMC requires four bi-directional switches capable of blocking voltage and conducting current in both directions for its cyclo-converter operation. The basic circuit diagram is shown in Figure. 1. Currently due to the unavailability of any such discrete semiconductor device which could fulfil the needs, hence Common Emitter anti-parallel IGBT, diode pair is used as shown in Figure. 2. Diodes are used to provide reverse blocking capability to the switch module. IGBT's are used due to its

high switching capabilities and high current carrying capacities for high power applications

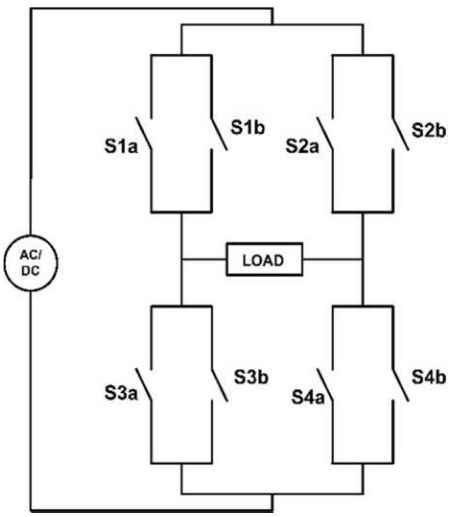


Figure 1: Basic circuit diagram of SPMC

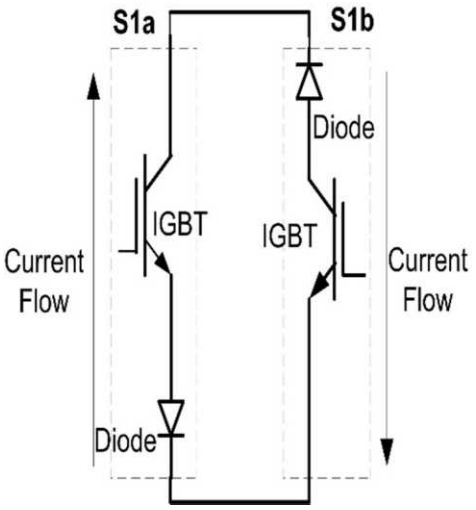


Figure 2: Bidirectional switch

Switching strategies for SPMC as cyclo-converter

Strategy I

In this strategy only two switches are kept on at a time. One of the two switch is provided with a PWM pulse and other switch is modulated with a continuous pulse. The switching sequence is shown in Table 1.

- At any time‘t’ only two switches S1a and S4a will be kept ON and conduct the current flow during the positive cycle of input source (state 1). (Ref Figure 3)
- At any time‘t’ only two switches S1b and S4b will be kept ON and conduct the current flow during the negative cycle of input source (state 2). (Ref Figure 4)
- At any time‘t’ only two switches S2a and S3a will be kept ON and conduct the current flow during the positive cycle of input source (state 3). (Ref Figure 5)
- At any time‘t’ only two switches S2b and S3b will be kept ON and conduct the current flow during the negative cycle of input source (state 4). (Ref Figure 6)

The switching signals according to Table 1 is shown in Figure 7. Due to the turn off characteristics of IGBT the practical realization of the switching sequence in the SPMC is not instantaneous and simultaneous. Here the tailing off of the collector current will create a short circuit with the next switch turn ON especially when the inductive loads are used, resulting in switching spikes [16]. This strategy leads to two damaging effects which would lead to undue stress and destruction of switches: 1) current spikes are generated in the short circuit paths 2) voltage spikes will be induced due to change in current across inductance of load.

Table 1: Switching sequence for strategy I

Input frequency	Target output frequency	Time Interval	State	Switch modulated	PWM switch
50 Hz	150 Hz	1	1	S1a	S4a
		2	3	S2a	S3a
		3	1	S1a	S4a
		4	2	S1b	S4b
		5	4	S2b	S3b
		6	2	S1b	S4b

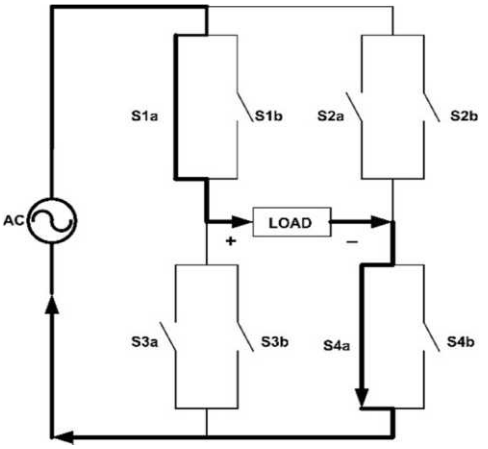


Figure 3: State 1 positive cycle

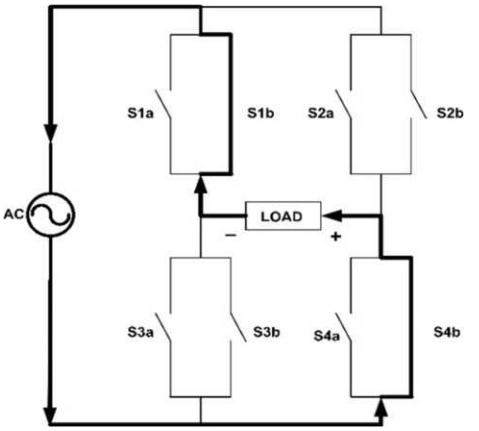


Figure 4: State 2 negative cycle

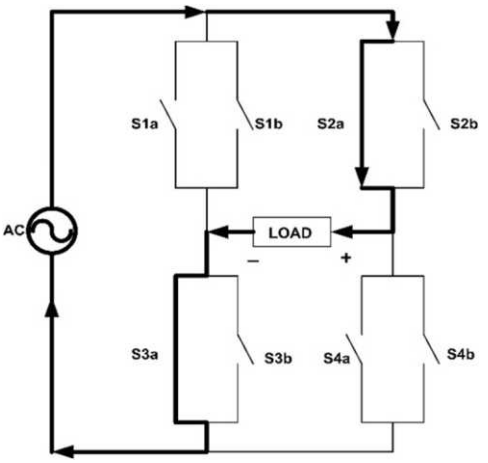


Figure 5: State 3 positive cycle

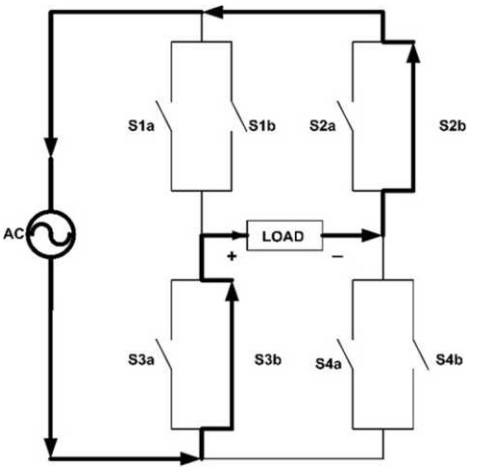


Figure 6: State 4 negative cycle

Both will subject the switches with undue stress leading to its destruction. Thus it cannot be used with RL load as it provides no path to discharge the energy across the inductor. Switching arrangements for safe-commutation as proposed by Zahirrudin can be used to eliminate the generation of switching spikes [17]. Refer Figure 8 to Figure 15 for the output voltage, current and their THD. Strategy II overcomes this limitation.

Strategy II

In this strategy three switches are used for a single interval out of which one is a PWM pulse and other two are modulated with a continuous pulse. The switching sequence are shown in the Table 2. Refer Figure 24 to Figure 27 respectively to understand the different operating states shown below.

- At any time‘t’, two switches S1a and S4a (PWM) will be kept ON and conduct the current flow during the positive cycle of input source, with S2b turn ‘ON’ for commutation purpose (state 1).
- At any time‘t’, two switches S4b and S1b (PWM) will be kept ON and conduct the current flow during the negative cycle of input source, with S3a turn ‘ON’ for commutation purpose (state 2)
- At any time‘t’, two switches S2a and S3a (PWM) will be kept ON and conduct the current flow during the positive cycle of input source, with S1b turn ‘ON’ for commutation purpose (state 3)
- At any time‘t’, two switches S3b and S2b (PWM) will be kept ON and conduct the current flow during the negative cycle of input source, with S4a turn ‘ON’ for commutation purpose (state 4)

This scheme can be used with R as well as RL load and is thus better than previous one. The THD can be further improved with proposed switching strategy. Figure 16 shows the switching signals according to Table 2. Refer Figure 17 – Figure 20 for the output voltage, current waveforms and their THD.

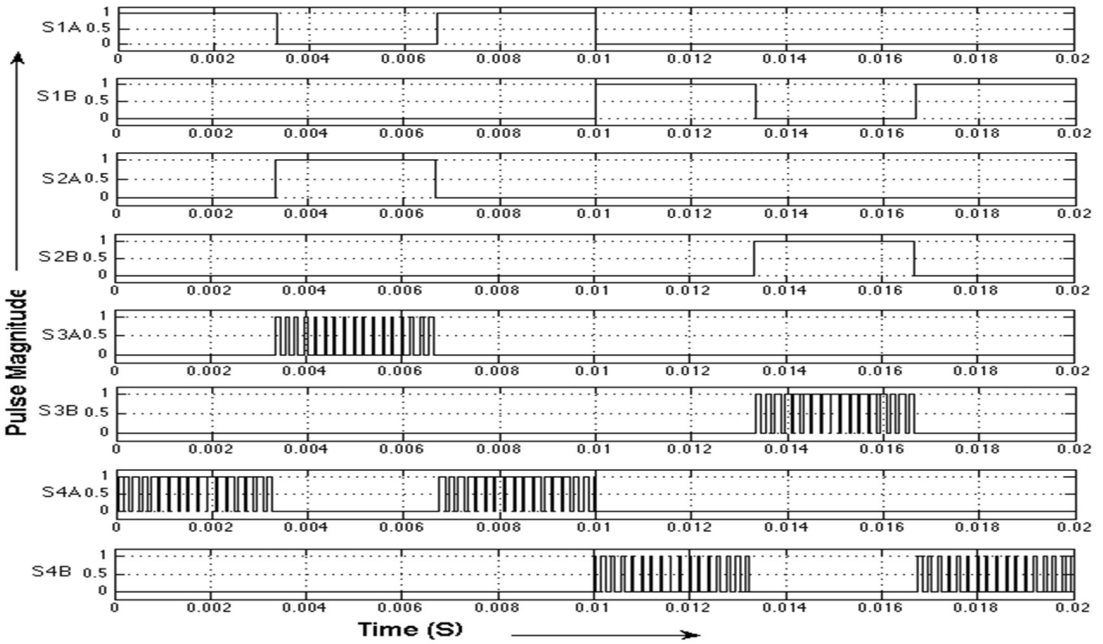


Figure 7: Switching signals for strategy I

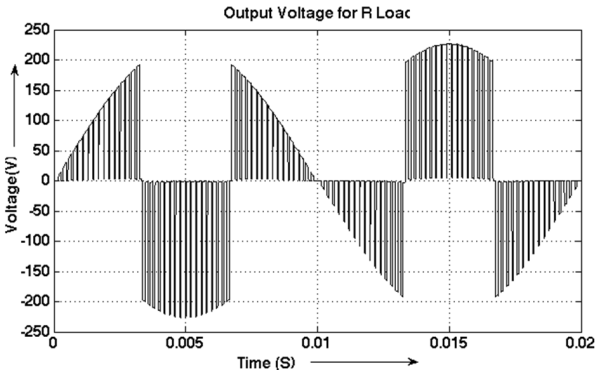


Figure 8: Output voltage for strategy I (R load)

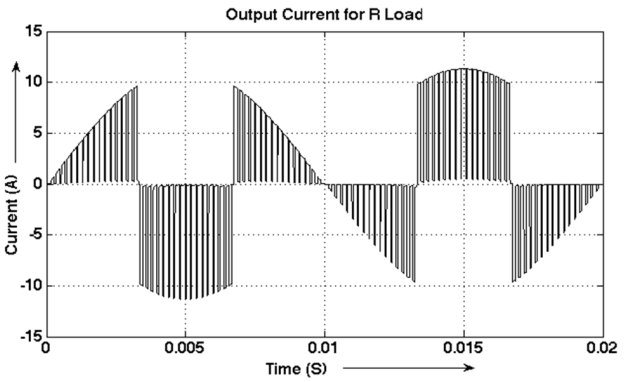


Figure 9: Output current for strategy I (R load)

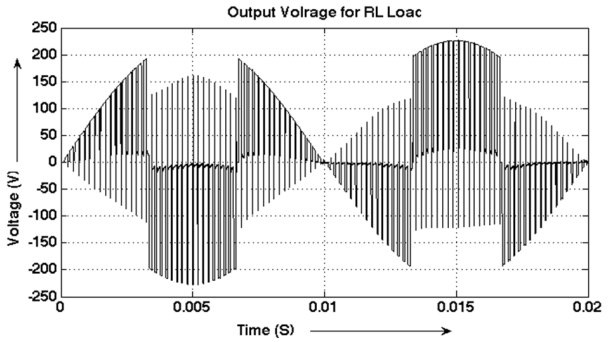


Figure 10: Output voltage for strategy I (RL load)

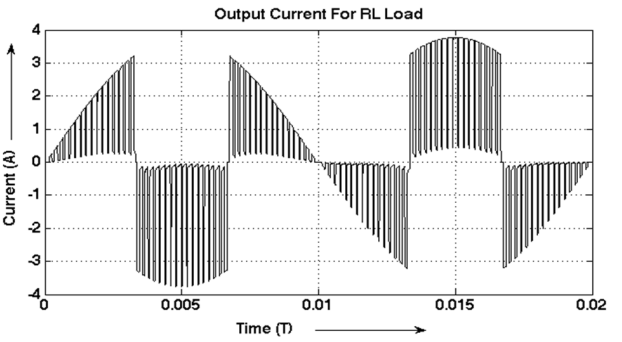


Figure 11: Output current for strategy I (RL load)

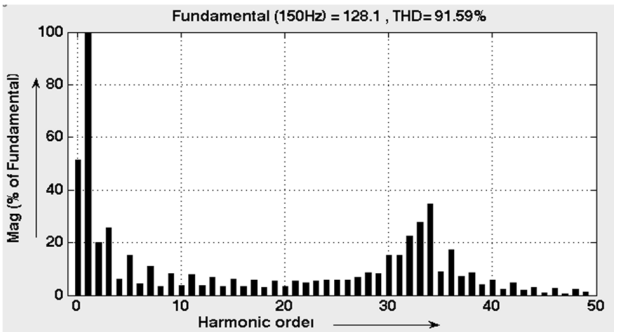


Figure 12: FFT analysis for output current with R load

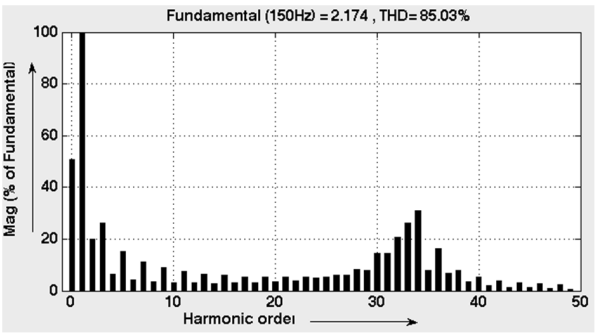


Figure 13: FFT analysis for output current with RL load

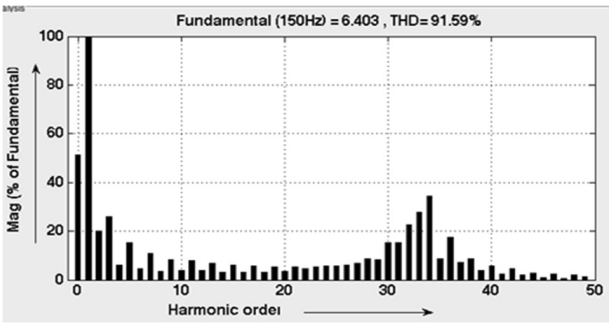


Figure 14 FFT analysis for output voltage with R load

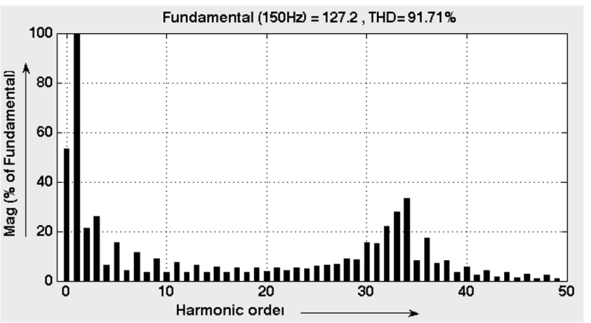


Figure 15 FFT analysis for output voltage with RL load

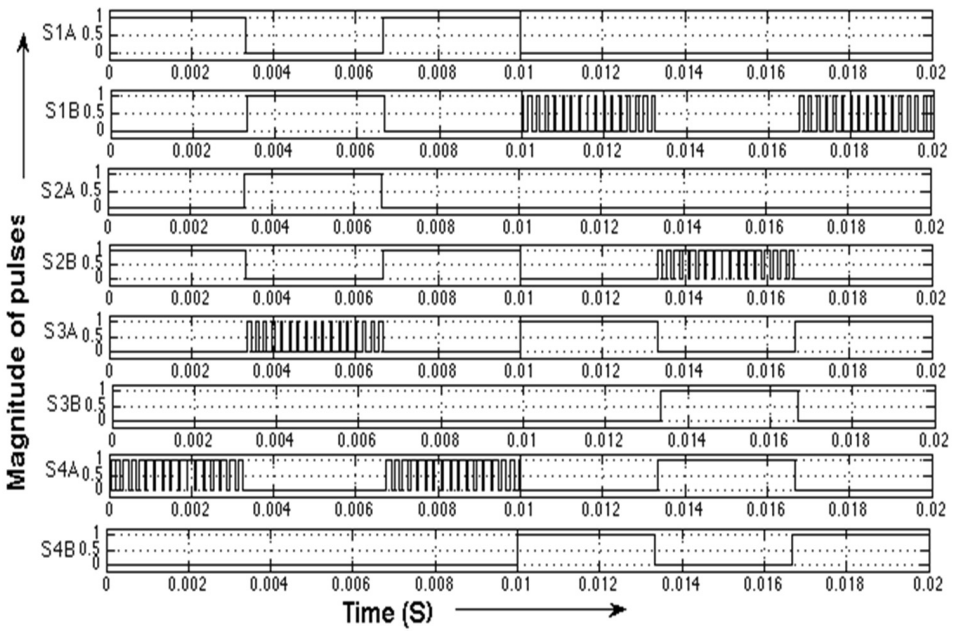


Figure 16 Switching signals for strategy II

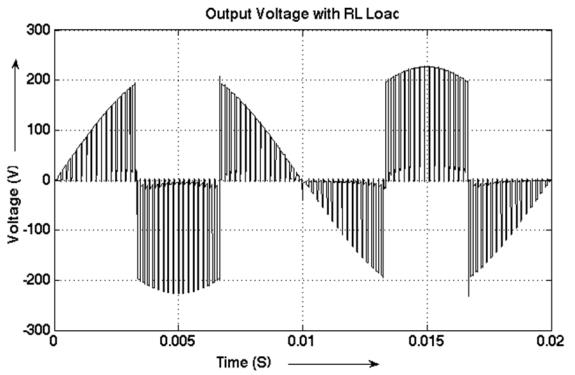


Figure 17 Output voltage for strategy II (RL load)

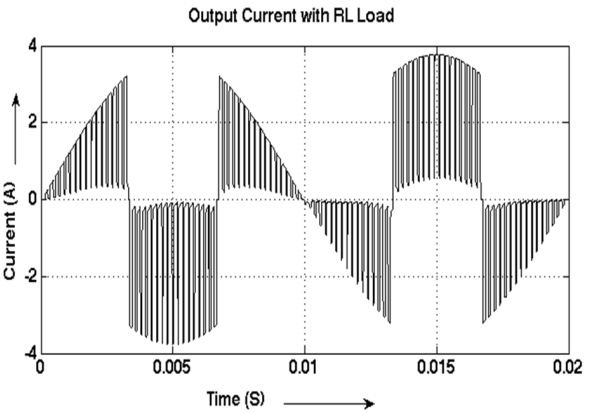


Figure 18 Output current for strategy II (RL load)

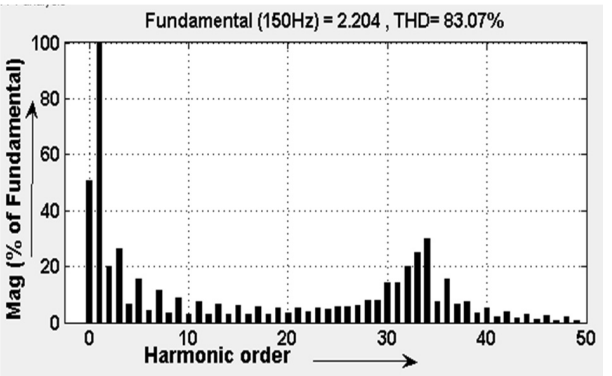


Figure 19 FFT analysis for output current with RL load

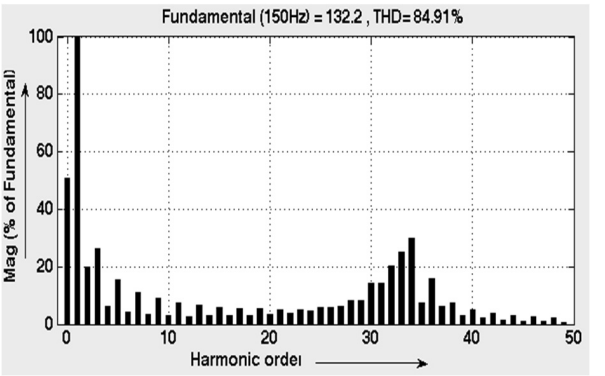


Figure 20 FFT analysis for output voltage with RL load

Strategy III - Proposed Strategy

In this strategy three switches will be used for a single time interval, out of which one switch will always be modulated with a continuous pulse and other two will be PWM switches. Thus one PWM signal is given to the switches used for free-wheeling (say PWM) and other PWM signal is given to the switch through which the load current would flow (say cPWM). These two PWM are generated such that one would be complementary of the other (only in that interval in which it is to be ON). Refer Figure 22, which clearly shows the three signals given to the switches of state 1. The two PWM signals in Figure 22 are zoomed in Figure 23, which shows that signal "cPWM" given to S2b is the complementary of signal "PWM" given to switch S4a, while S1a is modulated with a continuous pulse as shown in Figure 23. The switching sequence for proposed strategy is shown in Table 2.

- At any time't', two switches S1a and S4a (PWM) will be kept ON and conduct the current flow during the positive cycle of input source, with S2b (cPWM) turn 'ON' for commutation purpose (state 1). [ref fig 24]
- At any time't', two switches S4b and S1b (PWM) will be kept ON and conduct the current flow during the negative cycle of input source, with S3a (cPWM) turn 'ON' for commutation purpose (state 2). [ref fig 25]
- At any time't', two switches S2a and S3a (PWM) will be kept ON and conduct the current flow during the positive cycle of input source, with S1b (cPWM) turn 'ON' for commutation purpose (state 3). [ref fig 26]
- At any time't', two switches S3b and S2b (PWM) will be kept ON and conduct the current flow during the negative cycle of input source, with S4a (cPWM) turn 'ON' for commutation purpose (state 4). [ref fig 27]

As compared to strategy II we get an improved THD and also it can be used with both R load and RL load. Refer Figure 21 for the switching signals according to Table 2. The output voltage, current waveform and THD are shown in Figure 28 to Figure 31.

Table 2: Switching sequence for strategy II and proposed strategy

I/P Frequency	Target O/P Frequency	Time Interval	State	For Strategy II			For Strategy III (Proposed Strategy)		
				Switch modulated with continuous pulse		PWM Switch	Switch Modulated with continues pulse	PWM Switch	
					For Free Wheeling			PWM	cPWM
50 Hz	150 Hz	1	1	S1a	S2b	S4a	S1a	S4a	S2b
		2	3	S2a	S1b	S3a	S2a	S3a	S1b
		3	1	S1a	S2b	S4a	S1a	S4a	S2b
		4	2	S4b	S3a	S1b	S4b	S1b	S3a
		5	4	S3b	S4a	S2b	S3b	S2b	S4a
		6	2	S4b	S3a	S1b	S4b	S1b	S3a

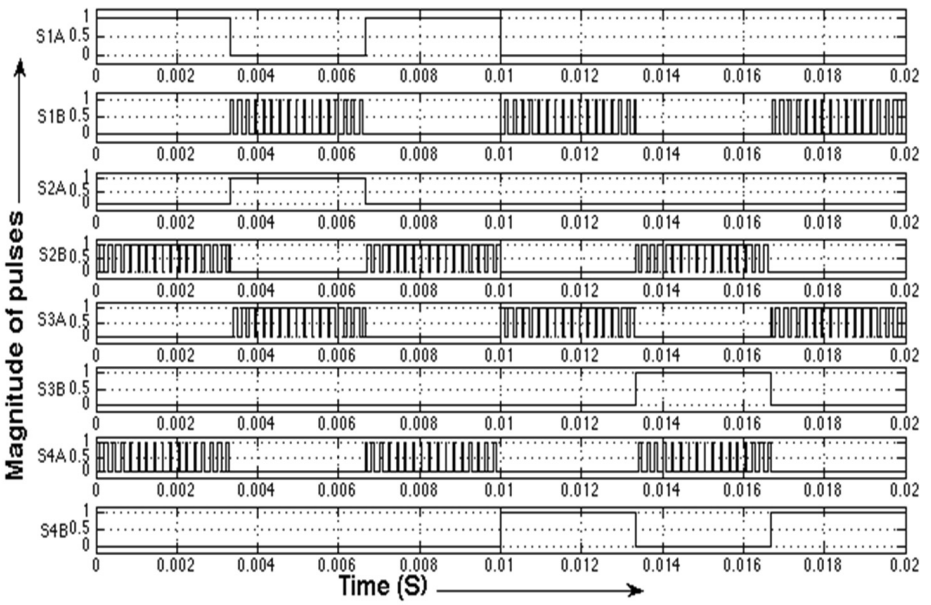


Figure 21: Switching signals for proposed strategy

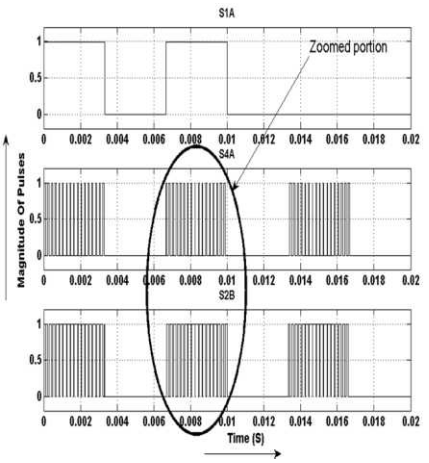


Figure 22 Switching signals for state 1 of proposed strategy

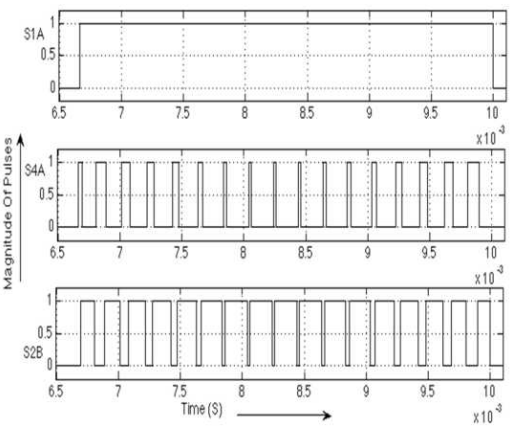


Figure 23 Zoomed view of Fig 22

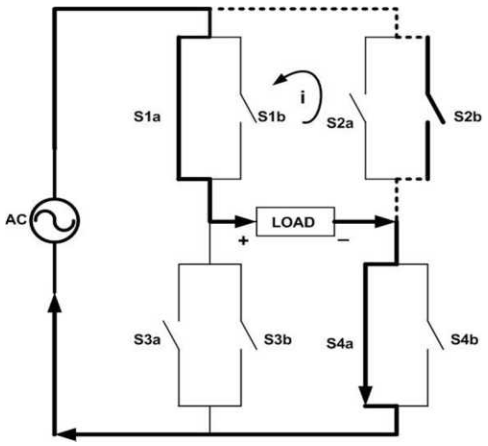


Figure 24 State 1 positive cycle

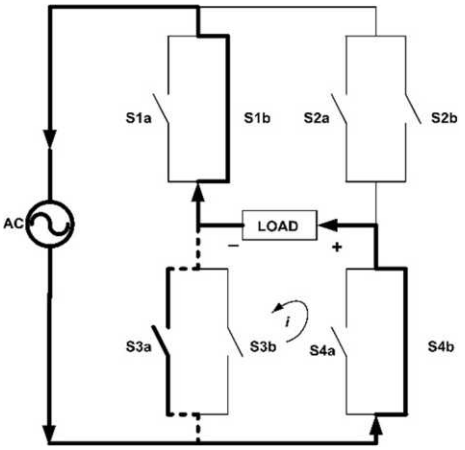


Figure 25 State 2 negative cycle

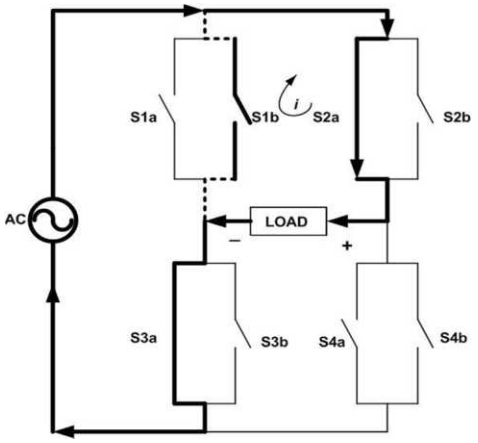


Figure 26 State 3 positive cycle

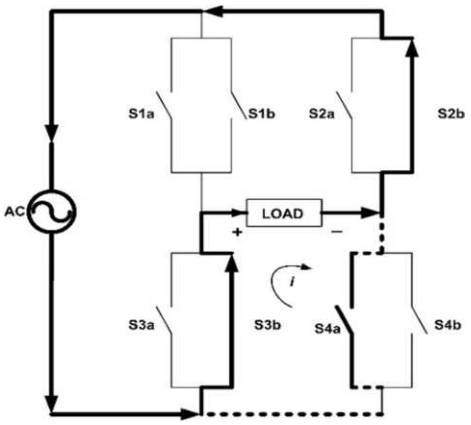


Figure 27 State 4 negative cycle

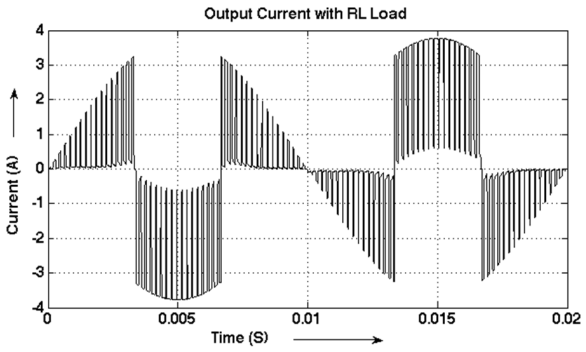


Figure 28 Output current for proposed strategy (RL load)

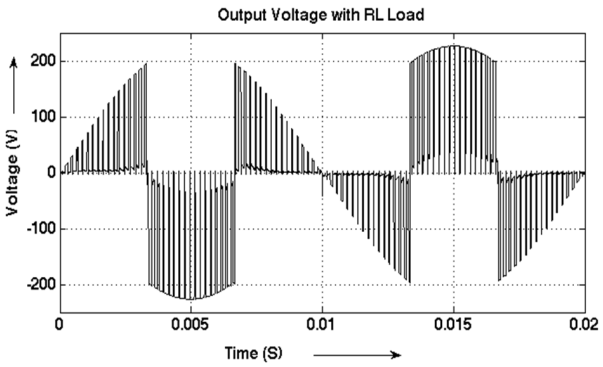


Figure 29 Output voltage proposed strategy (RL load)

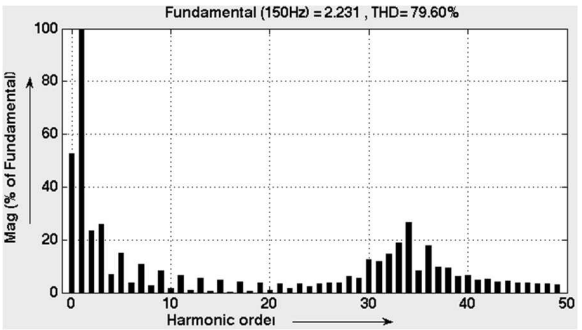


Figure 30 FFT analysis for output current

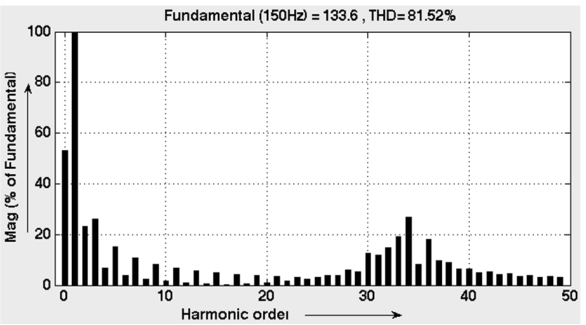


Figure 31 FFT analysis for output voltage

MATLAB Implementation

The matrix converter topology is developed in MATLAB/Simulink (MLS) software package as shown in Figure 32.

Table 3: THD results

Strategy	Current THD(%)	Voltage THD(%)
I (R Load)	91.59	91.59
I (RL Load)	85.05	91.71
II (RL Load)	83.08	84.91
III(RL Load) [Proposed Strategy]	79.60	81.52

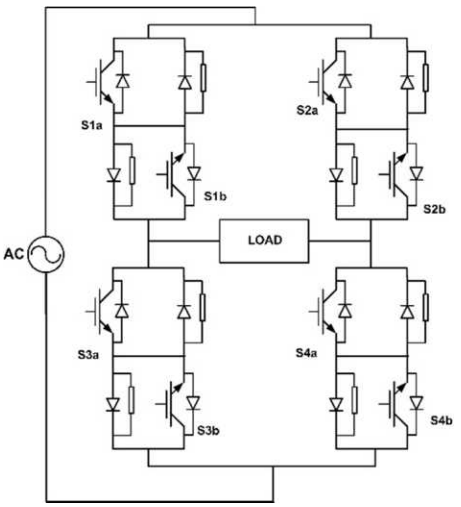


Figure 32 Simulation circuit diagram

The results in Table 3 shows that THD is improved with the proposed strategy. Moreover, it also provides us with a path for free-wheeling as three switches are kept ON for a single time interval.

Conclusion

Three different switching strategies have been presented in this paper. Comparison between all three strategies have been done and their advantages and limitations have been studied. The operational behavior is verified using MATLAB/Simulink with the SimPowerSystem Block Set. Results conclude that the strategy I can be used only for R load, strategy II can be used with R and RL load, proposed strategy (strategy III) provides better THD compared to strategy II.

References

[1] Anand kumar, Dr. P.R. Thakura, "Hardware Development and Implementation of Single Phase Matrix Converter as a Cycloconverter and as an Inverter," Power Electronics (IICPE), IEEE 6th India International Conference, 2014.

[2] V.V.Subrahmanya Kumar Bhajana, Pavel Drabek, Martin Jara, Bedrich Bednar, "A Novel ZCS Single Phase Matrix converter for Traction Applications," Power Electronics and Applications (EPE'14-ECCE Europe), 16th European Conference, 2014.

[3] Zahirrudin Idris, Mustafar Kamal Hamzah, Ngah Ramzi Hamzah, "Modelling & Simulation of a new Single-phase to Single-phase Cycloconverter based on Single-phase Matrix Converter Topology with Sinusoidal Pulse Width Modulation Using MATLAB/Simulink", University Teknologi MARA, 40450 Shah Alam, Malaysia.

[4] Venturini, M., and Alesina, A., "The Generalized transformer: a new bi-directional sinusoidal wave form frequency converter with continuosly adjustable input power factor," IEEE Power Electron. Spec. Conf. Rec., 1980, pp. 242-252.

[5] Zuckerberger, A., Weinstock, D., Alexandrovitz A., "Single-phase Matrix Converter," IEE Proc. Electric Power App, Vol.144 (4), Jul 1997 pp. 235-240.

[6] Hosseini, S.H.; Babaei, E, "A new generalized direct matrix converter," IEEE International Symposium of Industrial Electronics, 2001. Proc. ISIE 2001. Vol (2), pp.1071-1076.

[7] Abdollah Koei & Subbaraya Yuvarajan, "Single-Phase AC-AC Converter Using Power Mosfet's," IEEE Transaction on Industrial Electronics, Vol. 35, No.3, August 1988, pp.442-443.

[8] Firdaus, S., Hamzah, M.K., "Modelling and simulation of a single-phase AC-AC matrix converter using SPWM," Proceedings on Student Conference on Research and Development, (SCOReD2002), 16-17 July 2002, pp.286-289.

[9] Empringham, L., Wheeler, P.W., Clare, J.C., "Intelligent Commutation of Matrix Converter Bi-directional Switch Cells Using Novel Gate Drive Techniques," 29th Annual IEEE Power Electronics Specialists Conference, 1998. PESC 98 Record., Vol.1, 17-22 May 1998, pp.707- 713.

[10] Kwon, B.-H., Min, B.-D., Kim, J.-H., "Novel Commutation Technique of AC-AC Converters," Electric Power Application, IEE Proceedings-, Vol. 145 (4), July 1998, pp. 295-300.

[11] Zahiruddin Idris, Siti Zaliha Mohammad Noor & Mustafar Kamal Hamzah, "Safe Commutation Strategy in Single-phase Matrix Converter", IEEE Sixth International Conference PEDS 2005, Kuala Lumpur, Malaysia.

[12] Karamat, A., Thomson, T., Mehta, P., "A novel strategy for control of cycloconverters," 22nd Annual EEEE Power Electronics Specialists Conference, 1991. PESC '91, Record. 24-27 June 1991, pp. 819 - 824.

[13] Choudhury, M.A., Uddin, M.B., Bhuyia, A.R., Rahman, M.A., "New topology and analysis of a single phase delta modulated cycloconverter," Proceedings of the 1996 International Conference on Power Electronics, Drives and Energy Systems for Industrial Growth, 8-11 Jan. 1996, Vol. 2, pp.819-825

[14] Hill, W.A., Ho, E.Y.Y., Nuezil, LJ. "Dynamic behavior of cyclo-converter system," IEEE Transactions on Industry Applications, Vol. 27(4), July-Aug. 1991, pp.750-755

[15] K.V.S Bharath, Ankit Bhardwaj, "Implementing Single Phase Cyclo-Converter Using Single Phase Matrix Converter Topology with Sinusoidal Pulse Width Modulation," International Journal for Technological Research in Engineering Volume 2, Issue 6, February-2015.

[16] Wheeler, P.W., Rodriguez, J., Clare, J.C., Empringham, L., Weinstein, A., "Matrix converters: a technology review," Industrial Electronics, IEEE Transactions on, Volume: 49, Issue: 2, April 2002 Pages: 276 – 288

[17] Wheeler, P.W., Clare, J.C., Empringham, L., Bland, M., Kerris, K.G., "Matrix converters," Industry Applications Magazine, IEEE, Volume: 10, Issue: 1, Jan-Feb2004 Pages: 59 – 65.

[18] Cho, J.G., and Cho, G.H, "Soft-switched Matrix Converter for High Frequency direct AC-to-AC Power Conversion," Int. J. Electron., 1992, 72, (4), pp. 669-680.

[19] Idris, Z.; Hamzah, M.K.; Saidon, M.F., "Implementation of Single-Phase Matrix Converter as a Direct AC-AC Converter with Commutation Strategies," 37th IEEE Power Electronics Specialists Conference, 2006. PESC '06, Page(s):1-7.

[20] H.M.Hanafi, N.R. Hamzah, A.Saparon and M.K.Hamzah, "Improved Switching Strategy of Single-Phase Matrix Converter as a Direct AC-AC Converter", University Teknologi MARA, 40450 Shah Alam, Malaysia.

Author Index