1st International Conference on Computers and Electronics Engineering 2016

ICCEE-2016

9th JANUARY, 2016, HYDERABAD, INDIA

1st International Conference on Computers and Electronics Engineering 2016

ICCEE-2016

9th JANUARY, 2016, HYDERABAD, INDIA

ICCEE-2016



ICCEE-2016 COLLABORATED WITH B.I.T SINDRI

Venue

FTAPCCI, Federation House, Red Hills, Hyderabad 500004, Telangana, India



McGraw Hill Education (India) Private Limited

McGraw Hill Education Offices

New Delhi 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



on McGraw Hill Education (India) Private Limited

Published by McGraw Hill Education (India) Private Limited, P-24, Green Park Extension, New Delhi 110 016.

1st International Conference on Computers and Electronics Engineering 2016 (ICCEE-2016)

Copyright © 2016, by McGraw Hill Education (India) Private Limited.

No part of this publication may be reproduced or distributed in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise or stored in a database or retrieval system without the prior written permission of the publishers. The program listing (if any) may be entered, stored and executed in a computer system, but they may not be reproduced for publication.

This edition can be exported from India only by the publishers, McGraw Hill Education (India) Private Limited.

ISBN (13): 978-93-85880-97-1 ISBN (10): 93-85880-97-7

Managing Director: Kaushik Bellani

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

Senior Researcher-Product Development: Navneet Mehra

Head—Production (Higher Education and Professional): Satinder S Baveja Assistant Manager—Production: Atul Gupta

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

Information contained in this work has been obtained by McGraw Hill Education (India), from sources believed to be reliable. However, neither McGraw Hill Education (India) nor its authors guarantee the accuracy or completeness of any information published herein, and neither McGraw Hill Education (India) nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that McGraw Hill Education (India) and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

Typeset at Tej Composers, WZ 391, Madipur, New Delhi 110 063, and printed at

Cover Design: Kapil Gupta Cover Printer:

Contents

Key N	lote Speaker	vii
Gener Progr	ral Chair cam Committee	ix xi
1.08		
1.	Codon Adaptation Index Correlates Better with Gene Expression in Bacteria having Stronger Selection	1
2.	Digital Image Watermarking Techniques using Artificial Intelligence: New Trends in Information Security	6
3.	Green ICT: The New Era of Effective and Efficient Green Learning Based on ICT	13
4.	Concept Drift in Data Streams with Decision Trees	20
5.	Taxonomy of Security Attacks in Cognitive Radio Networks – A Survey	31
6.	A Survey on Association Rule Mining using Genetic Algorithm	
7.	Ingressing Embedded System through RC Cars Working in Wireless Network	44
8.	An Artificial Neural Network Based Classification Approach for Mining Spatial Database of Epidemiology in India	51
9.	Constraint Based Sequential Pattern Mining: Survey and Current Research Challenges	
10.	Design of Full Adder using Amorphous Silicon Thin Film Transistor	64
11.	Review on Classification and Quality Analysis of Fruits using Artificial Neural Network	70
12.	Identification of DoS Attack and its Countermeasure Based on Dynamic Threshold Value of Traffic Flow	77
13.	An Improved Task Scheduling Algorithm on e-learning using Cloud	82
14.	Dynamic Matrix Based Symmetric Key Cryptography	
15.	Next Generation Mobile Network Technology: 5G	95
16.	Analysis of Data Gathering Protocol and Security Issues during Data Aggregation in WSN	101

17.	Adaptive Threshold Memristor Model	105
18.	Modeling, Control and Simulation of Integrated Photovoltaic Cell-Fuel Cell-Battery Based Distributed Generation System	110
19.	Social, Ethical & Other Issues of Networked World	121
20.	A Survey on: PSO-GA Based Back Propagation Learning-MLP (PSO-GA-BP-MLP)	127
21.	Evaluation and Analysis of eLearning Models	133
22.	Thermohydraulic Performance of Solar Water Heater with Twisted Tapes and Boosted Mirrors	140
23.	The Software Reliability Assessment using Neuro-fuzzy Systems with Encoded Input	149
24.	Designing Display System using Led and Atmega-16 Controller with Low Power Consumption through Solar System	157

Key Note Speaker

Dr. Srinivas Pentyala

Director of Translational Research & Associate Professor of Anesthesiology Faculty Member in Urology, Health Sciences, Physiology & Biophysics NY, USA

Dr. Mohammed Zafar Ali Khan

IIT Hyderabad, India

Dr. M A Rizvi

National Institute of Technical Teachers' Training and Research Bhopal, India Govt. of India Institute (Under Ministry of HRD)

Title of Keynote: Enhancing Routing Efficiency using Agents

General Chair

Dr. M A Rizvi

National Institute of Technical Teachers' Training and Research Bhopal, India Govt. of India Institute (Under Ministry of HRD)

Professor (Dr.) Raj Kamal, Ph.D.

(IITD) Devi Ahilya University

Program Committee

M. N. Doja	A. Q. Ansari
Jamia Millia Islamia, India	Jamia Millia Islamia, India
Nur Izura Udzir	Manoj Kumar
University of Putra Malaysia, Malaysia	AIACTR, India
Rama Bhamidipathy	Vishal Bhatnagar
Open Text, India	AIACTR, India
Mamta Rani	Manju Khari
Central University of Rajasthan, India	AIACTR, India
Saurabh Agarwal	Abhishek Shukla
NIEC, India	R.D. Engineering College Technical Campus, Ghaziabad, India
Ashok De	Arup Kumar Pal
NIT, India	ISM, India
Rajiv Ranjan	Pankaj Raj
BIT, India	B.I.T. India
U.K. Dey	Mahesh Kumar H. Kolekar
BIT, India	IIT, India
G. P. Biswas	Kalyan Chatterjee
ISM, India	ISM, India
Ravi Shankar	R.K Singh
KIIT, India	IGIDTUW, India
Chun-Che Hung	Ibrahim El Rube
NCNUIM, Taiwan	Taif University, KSA
Dr. G KUMAR	S.K. Singh
BIT, India	CIT, India
Arun Sharma	Sudhanshu Kumar Jha
IGIDTUW, India	NIT, India
Salman Abdul Moiz	Maya Ingle
University of Hyderabad, India	D.A. University, India
Sanjay Tanwani	Abhay Kumar
D.A. University, India	DAUNIV, India

Md. Saiful Islam	Zainul Abdin Jafrey
King Saud University KSA	Jamia Millia Islamia, India
Sumant Katiyal DAUNIV, India	Mohammad Zarour Prince Sultan University, KSA
Zbigniew Leonowics	U B Desai
Wroclaw University of Technology, Poland	IIT Hyderabad, India
B.S. Shulaja	Asadullah Shaikh
Dr. Ambedkar University of Technology, India	Najran University, KSA
Mohammed Zafar Ali Khan	Kashif Minur
IIT Hyderabad, India	University of Hafr Al-Batin, KSA
Fai Ahmed Khan	Tamanna Siddiqui
IIT Hyderabad, India	Aligarh Muslim University, India
Chinthaka Premchandra	Mohamed Maher Ben Ismail
Tokyo University of Science, Tokyo	King Saud University, KSA
Jalal AL-Muhtadi	Md. AL-Amin Bhuiyan
King Saud University, Saudi Arabia	King Faisal University, KSA
Binod Kumar	Aradhana Gautam
Savitribai Phule Pune University, India	IES, INDIA
R. Balu	P V L Narayana Rao
Bharathiar University, India	Wolkite University, Ethiopia
Jagdish C. Bansal	Suresh Veluru
IIITM, India	United Technologies Research Centre, Ireland
K. V. Arya	Yogachandran Rahulamathavan
IIITM, India	City University London, UK
D. P. Agarwal	Selvakumar Manickam
University of Cincinnati, USA	National Advanced IPv6 Centre, USM, Malaysia
Shingo Yamaguchi	S. Karupayyah
Yamaguchi University, Japan	Technische Universitat Darmstadt, Germany
T. Perumal	Ms. Parama Bagchi
University Putra Malaysia, Malaysia	RCC Institute of Information Technology, India
Sebastian Garcia	Soumik Roy
CTU University, Prague, Czech Republic	Tezpur University, Assam
	Vijay Bhaskar Semwal IIIT-Allahabad

Codon Adaptation Index Correlates Better with Gene Expression in Bacteria Having Stronger Selection

Siddhartha Sankar Satapathy Department of Computer Science and Engineering Tezpur University Tezpur, Assam, India, 784028 ssankar@tezu.ernet.in

> Ajit Kumar Sahoo Department of ITM Revenshaw University Cuttack, Odisha, India, 753003

Suvendra Kumar Ray Department of Molecular Biology and Biotechnology Tezpur University Tezpur, Assam, India, 784028

Abstract—Codon Adaptation Index (CAI) is used for theoretically predicting gene expression from codon usage bias in organisms such as *Escherichia coli* and yeast. In this study we have compared gene expression data with CAI values in seventeen different bacteria using receiver operating characteristic (ROC) curve. CAI performs well in predicting the gene expression in five out of the seventeen bacteria that includes *E coli, Bifidobacterium longum, H. influenzae, Rhodopseudomonaspalustris,* and *Staphylococcus aureus*. Comparative analysis of the gene expression data between the high expression genes and the whole genome indicated that all the above five bacteria except *S. Aureus* were with strong selected codon usage bias. In general our observation suggests that CAI may be used to predict gene expression in bacteria with strong selection pressure on codon usage.

Keywords: codon usage bias; codon adaptation index; gene expression; selection

I. INTRODUCTION

The standard version of the genetic code table presents a translational mapping from 61 sense codons to 20 amino acids. In this degenerate table with more codons than amino acids, except Met and Trp other eighteen amino acids are encoded by two to six different codons. Different codons that encode the same amino acid are known as synonymous codons. Even though synonymous codons encode the same amino acids, it has been found that these codons are not used uniformly in organisms. Some synonymous codons are used more frequently than the other synonymous codons. This phenomenon is known as codon usage bias (CUB), which is variable across different organisms as well as among genes within a genome.

Genome composition (G+C%) that varies approximately from 13.0 to 75.0 in bacteria is a major determinant of codon usage bias [1,2]. Synonymous codons with G or C nucleotide at third codon position are often frequently used in genomes rich in G+C nucleotides, and reverse is also true in A+T rich organisms. Genome composition is considered to be due to various mutational factors [3] and therefore, mutation pressure is the primary factor for variation in codon usage across bacteria. Along with the mutational mechanisms, selective forces also influence synonymous codon usage though in variable strength [4]. Translational efficiency is considered as a primary selection factor in causing CUB [5].Certain codons are used more frequently in high expression genes for efficient translation. In support of selection on codon usage bias Gouy and Gautier (1982) reported codon usage difference between high and low expression genes in *Escherichia coli* [6].

Sharp and Li (1987) utilized this codon usage pattern in high expression genes to theoretically predict gene expression from codon usage bias [7]. They formulated a measure of CUB known as codonadaptation index (CAI). CAI measures – to what extent codon usage in a gene g is similar to that in a given reference set of high expression genes. CAI value ranges from 0.0 to 1.0. Higher is the similarity, larger is the CAI value of a gene. Higher is the CAI value more is the expression level. They showed that CAI performs fairly well in E. coli in terms of correlation with gene expression. CAI is commonly used methods for theoretically gene expression prediction.

Though the selection pressure on codon usage is ubiquitous [8], its strength varies among bacteria species [9]. In this manuscript, we considered 17 species of bacteria whose gene expression data were available in online databases to analyzed usefulness of CAI in predicting gene expression. We observed that CAI predicts gene expression fairly well in bacteria with strong selection pressure on codon usage.

II. MATERIALS AND METHODS

A. Nucleotide sequences and gene expression data of bacterial genomes

Gene sequences were taken from DDBJ site (www.gib.genes.nig.ac.jp). In total seventeen bacterial genomes belonging to seven bacterial groups covering a genomic G+C content from 32.84% to 72.00% and genome size from 1.83 mbp to 9.11 mbp (Table 1) were analyzed. Transcriptomics expression data for the bacterial strains are obtained from NCBI GEO website (ftp://ftp.ncbi.nih.gov/pub/geo/DATA/SeriesMatrix/).

B. Calculating Codon Adaptation Index (CAI)

CAI is a measurement of the relative adaptedness of the codon usage of a gene towards the codon usage of highly expressed genes[10]. The relative adaptedness of each codon (w) is the ratio of the usage of each codon, to that of the most abundant codon within the same synonymous family. The CAI index is defined as the geometric mean of these values.

$$CAI = exp\left(\frac{1}{L}\sum_{k=1}^{L}\ln\left(w_{k}\right)\right)$$

Where w_k is the relative adaptedness of the k^{th} codon and L is the number of synonymous codons in the gene. To generate CAI values, we used CodonW [8].

C. Generating Receiver Operating Characteristic (ROC) curve

Receiver operating characteristics (ROC) curve is useful for organizing classifiers and visualizing their performance [11]. ROC curve is a two-dimensional plot between true positive (tp) and false positive (fp) rates which are defined below. In this plot tp rate is plotted on the Y axis and fp rate is plotted on the X axis (Fig. 1). Diagonal line represents random classification. As the curve approaches towards the top left corner, performance of the classifier improves. For 100% accurate classification area under the curve (AUC) is 1.0 whereas for random classification AUC value is 0.5.



Fig. 1. Receiver operating characteristics (ROC) graph

In our analysis, we considered Positives as the genes with high expression values (more than the median value) and also with high CAI value. Negatives are the genes with high CAI but with low expression values. We consider gene expression/CAI value as high, when the value is more than the median value.

D. Measuring selection pressure on codon usage bias

The four-fold degenerate site (FDS) in the coding sequences has been used in the study of selection pressure on CUB [2][12-15] because nucleotide substitution per se is not under any selection pressure at the site due to the unaltered amino acid sequence in a protein. In the standard genetic code table there are 32 codons belonging to eight amino acids with such FDS. Satapathy et al. (2014) defined a measure Um(g) [Unevenness measure of a gene g] based on the frequency variation of nucleotides at the FDS across the eight family boxes (FBs) [4]. Range of the Um(g) value is between 0.0 and 1.0. This measure is an indicator of selection pressure on codon usage. We used online tool available at http://agnigarh.tezu.ernet.in/~ssankar/cub.php for calculating Um(g).

III. RESULT AND ANALYSIS

A. CAI values do not match with gene expression uniformly across organisms

We generated ROC curves in each organism for comparison between CAI and gene expression. In total seventeen organism were considered in this study (Table 1). The ROC curves are shown in Fig. 2. We also calculated area under curve (AUC) for each organism (Table 1). Higher is the AUC more is the similarity. In our analysis, we observed that AUC varies from organism to organism. In some organism, AUC is quite high suggesting CAI is comparable to experimentally calculated gene expression whereas in several other organisms, AUC was close to 0.5 suggesting CAI is not comparable with gene expression. AUC values were higher than 0.7 in several organisms such as *Escherichia coli, Haemophilus influenza, Bifidobacterium longum, Rhodobactersphaeroides, Staphylococcus aureus*, whereas in some other organism such as *Bradyrhizobiumjaponicum, Pseudomonas aeruginosa, Lactobacillus plantarum, Streptomycescoelicolor* AUC was very low.

Sl. No	Bacteria	Genome size (bp)	G+C%	Um(g) HEG	Um(g) WEG	Diff	AUC
1	Bradyrhizobium japonicum	9105828	64.06	0.22	0.19	0.03	0.61
2	Bifidobacterium longum	2260266	60.13	0.35	0.25	0.10	0.75
3	Bacillus subtilis	4214630	43.52	0.16	0.18	0.02	0.59
4	Desulfovibrio vulgaris Hildenborough	3773159	63.28	0.17	0.15	0.01	0.61
5	Escherichia coli	4639675	50	0.45	0.25	0.20	0.81
6	Haemophilus influenzae	1830069	38.15	0.33	0.20	0.12	0.76
7	Listeria monocytogenes	2944528	37.98	0.16	0.15	0.01	0.54
8	Lactobacillus plantarum	3348625	44.42	0.21	0.13	0.08	0.66
9	Nitrosomonas europaea	2812094	50.72	0.22	0.19	0.03	0.63
10	Pseudomonas aeruginosa	6264404	66.56	0.33	0.27	0.05	0.56
11	Pseudomonas syringae	6538260	58.34	0.25	0.26	0.01	0.67
12	Rhodopseudomonas palustris	5467640	65.03	0.23	0.15	0.08	0.74
13	Rhodobacter sphaeroides	4603060	68.79	0.28	0.23	0.06	0.61
14	Staphylococcus aureus	2903636	32.84	0.25	0.24	0.01	0.72
15	Streptomyces coelicolor	9054847	72	0.19	0.13	0.07	0.65
16	Streptococcus mutans	2030921	36.83	0.19	0.14	0.05	0.54
17	Thermus thermophilus	2116056	69.5	0.22	0.23	0.01	0.64

TABLE 1. COMPARISON BEETWEEN AUC AND STRENGTH OF SELECTION ON CUB IN SEVENTEEN BACTERIA

Note: HEG/WEG: High/Weak Expression Genes, Diff = (Um(g), HEG) – (Um(g) WEG). HEG: High expression gene set, WEG: Weak expression gene set. In each organism, we have considered 100 highly expression genes in HEG and all the genes in WEG as most of them are known to be weakly expressed. Higher is the difference value, more is the selection pressure difference between high and low expression genes. Higher is the AUC value, better is the performance of CAI in terms of prediction of gene expression from codon usage bias.

B. CAI performs better in organisms with stronger selection pressure on CUB

In order to understand mechanism behind similarity between CAI and gene expression we further analyzed strength of selection pressure on CUB. Though selection pressure on codon usage prevails in organisms for optimal translation, it's strength varies from organism to organism. Unevenness measure of a gene g, Um(g) is an indicator of selection pressure [4]. Higher is the difference between Um(g) values between high and weak expression genes, stronger is the selection pressure on codon usage. We calculated Um(g) values for 100 high expression genes (HEG) and weak expression genes (WEG), and difference between them (Table 1). Some of the organisms with high difference values (≥ 0.7) are *Rhodobactersphaeroides*, *Bifidobacterium longum*, *Haemophilus influenza*, E. coli. In these organisms AUC values are also high. Similarly in several other organisms such as *Lactobacillus plantarum*, *Bradyrhizobiumjaponicum*, *Bacillus subtilis*, *Desulfovibrio vulgaris*Hildenborough where difference values are low (≤ 0.6), AUC values are also low. These observations indicate that CAI may be used to theoretically predicting gene expression in organisms with stronger selection pressure. However there were also some exceptions to this observation. Organisms such as *Staphylococcus aureus* and *Pseudomonas syringae* with very low difference values were with high AUC values. In these organisms there might be some other mechanism playing important role on codon usage bias for which further analysis is required.



Fig.2. ROC curves in seventeen bacteria

In conclusion, our analysis suggest that detailed analysis on selection and other mechanisms is required before using computational models such as CAI for predicting gene expression in organisms.

ACKNOWLEDGEMENT

We acknowledge financial support under twining project from DBT, Govt. of India.

IV. REFERENCES

- [1] G.A. Palidwor et al.,"A general model of codon bias due to GC mutational bias,"PLoS ONE, 5, e13431, 2010.
- [2] E.P.C. Rocha and E.J. Feil, "Mutational patterns cannot explain genome composition: are there any neutral sites in the genomes of bacteria?" PLoS Genet, 6, e1001104, 2010.
- [3] S.L. Chen et al, "Codon usage between genomes is constrained by genome wide mutational processes," Proc Natl AcadSci USA, 101, 3480–3485, 2004.
- [4] S.S. Satapathy et al., "Selection on GGU and CGU codons in the high expression genes in bacteria," J MolEvol, 78, 13–23, 2014.
- [5] R. Hershberg and D.A. Petrov, "General rules for optimal codon choice," PLoS Genet, 5(7), e1000556, 2009.
- [6] M. Gouy and C. Gautier, "Codon usage in bacteria: correlation with gene expressivity," Nucleic Acids Res, 10, 7055– 7074, 1982.
- [7] P.M. Sharp and W.H. Li, "The codon adaptation index a measure of directional codon usage bias, and its potential application," Nucleic Acid Res, 15, 1281–1295, 1987.
- [8] F. Supek et al., "Translational selection is ubiquitous in prokaryotes," PLoS Genet, 6, e1001004, 2010.
- [9] P.M. Sharp et al., "Variation in the strength of selected codon usage bias among bacteria," Nucleic Acids Res, 33, 1141–1153, 2005.
- [10] J.F. Peden, CodonW, "PhD Thesis, University of Nottingham," 1999.
- [11] T. Fawcett,"An introduction to ROC analysis" (2006) J Pattern Recognition Letters. 27(8), 861–874, 2006.
- [12] N. Sueoka, "Intrastrand parity rules of DNA base composition and usage biases of synonymous codons," J MolEvol, 40, 318–325, 1995.
- [13] J.R. Lobry and N. Sueoka, "Asymmetric directional mutation pressures in bacteria," Genome Biol, 3, 1–14, 2002.
- [14] W. Jia and P.G. Higgs, "Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection," Mol Biol Evol, 25, 339–351, 2008.
- [15] R. Hershberg and D.A. Petrov, "Evidence that mutation is universally biased towards AT in bacteria," PLoS Genet, 6, e1001115, 2010.

Digital Image Watermarking Techniques using Artificial Intelligence: New Trends in Information Security

Aditi Zear

Department of Computer Science & Engineering, Jaypee University of Information Technology Waknaghat, Solan, Himachal Pradesh - India aditizear93@gmail.com

Amit Kumar Singh Department of Computer Science & Engineering, Jaypee University of Information Technology Waknaghat, Solan, Himachal Pradesh -India amit 245singh@yahoo.com

Pardeep Kumar

Department of Computer Science & Engineering, Jaypee University of Information Technology Waknaghat, Solan, Himachal Pradesh -India pardeepkumarkhokhar@gmail.com

Abstract—This paper discusses the basic concepts of digital watermarking techniques, important spatial and transform domain techniques and compare the performance of some reported transform domain based watermarking techniques using artificial intelligence (AI) such as Genetic Algorithm (GA), Differential Evolution (DE), Neural Networks (NN), Clonal Selection Algorithm (CSA) and Particle Swarm Optimizer (PSO) are used in different ways to optimize the embedding strength of watermark by analyzing host image. Finally, the current state-of-the-art in transform domain watermarking using AI techniques is also discussed. This paper will be more important for researchers to implement effective watermarking method using AI technique.

Keywords: Image watermarking ; AI techniques; robustness; transform and spatial domain techniques.

INTRODUCTION I.

From last few decades growth in technology of computers and computer networks provides quality of service and higher bandwidth for both wireless and wired networks. However, the representation of media in digital form and evolution of internet also made easy to transmit digital media such as image, audio, video in an effortless way. These advancements also raised some security related issues for the protection of multimedia data and require some data hiding techniques. Basically, the data hiding is described as the process of embedding data such as copyright information into digital media such as image, audio and video. The goal of data hiding is not to restrict the access to host signal, but to ensure that the embedded data should be inviolate and recoverable. There are two techniques for data hiding: Steganography and Watermarking. Steganography is defined as the technique of hiding communication, the hidden content is embedded in some cover media so that there will not be any eavesdropper's suspicion. Digital watermarking is one of the new methods that provide protection against various attacks, data authentication and security to digital media. The watermarking is the process of embedding secret information in the form of signal called watermark into digital media (i.e. image, audio and video) so that this secret information can detected and extracted out to check the real owner or identity of digital media [1-3].

The digital watermarks are playing an important role in Fingerprinting, copyright protection, healthcare, Source Tracking, Secured E-Voting Systems and Remote Education and Insurance Companies. The important characteristics of watermark are robustness, imperceptibility, capacity computational cost and security [4]. These characteristics are very important for any watermarking system. However, there exists some tradeoff between the performance factors robustness, imperceptibility and capacity. Robustness is achieved when modifications made to host image by watermark are significantly large. It also affects the imperceptibility of the image because these modifications can be seen by human eyes. Therefore, some optimization techniques are required to balance these performance factors. Recently, AI techniques have been used to balance robustness and imperceptibility. Artificial intelligence techniques [5-10] are used as an optimization technique to search optimal sub-bands and coefficients in transform domain to embed watermark with different scaling factors. In addition, the AI techniques can be used as optimization techniques to remove some round off errors when coefficients in transform domain are transformed to spatial domain. The watermarks can be classified on the basis of permanency, visibility, detection and domain. According to domain, watermarking techniques can be classified as spatial domain and transform domain techniques.

A. Spatial Domain Techniques

The spatial domain techniques [4, 11-12] works by embedding watermark by modifying values of pixels. It is simple technique and requires less time and computational complexity. However, this technique is less robust against attacks. Table I shows the comparison between some important spatial domain techniques. The important spatial domain techniques are least significant bit (LSB), Correlation-Based Techniques and Patchwork Techniques

	LSB	Correlation Based Tech.	Patchwork
Imperceptibility High perceptual imperceptibility		Imperceptibility depends upon gain factor i.e. low gain factor results in high imperceptibility and vice versa.	Better imperceptibility
Robustness	Lacks robustness, vulnerable to noise, cropping, scaling etc	Like imperceptibility robustness also depends upon gain factor i.e. high gain factor results in more robustness	Robust against most type of attacks
InformationAmount of informationAmountembedded depends onAmountchoosing the subset ofimage pixels		Amount of information embedded depends upon gain factor	Embed small amount of information

TABLE I. COMPARISON OF IMPORTANT SPATIAL DOMAIN TECHNIQUES

A. Transform Domain Techniques

Spatial domain watermarking techniques are simple and easy to implement however they are not robust and also they do not allow further processing in order to increase the robustness of watermark. Transform domain techniques [13-14] provide more robustness as compared to spatial domain techniques. In these techniques host image is converted into transform domain and then watermark is embedded in transform coefficients. To obtain watermarked image inverse transform is performed. Table II shows the comparison of some important transform domain techniques. The important transform domain techniques are Discrete Cosine Transform (DCT) [14-15], Discrete Wavelet Transform (DWT) [16-18], Discrete Fourier Transform (DFT) and Singular value decomposition [6,8,14].

TABLE II: COMPARISON OF IMPORTANT TRANSFORM DOMAIN	TECHNIQUES

	DCT	DWT	DFT
Complexity	Computational cost and time is more as compared to spatial domain techniques	Computational cost and time is more as compared to DCT	Computational cost and time is more as compared to both DCT and DWT
Representation	Frequency space representation of image	Spatial and frequency representation of image	Phase and magnitude representation of image
Blocking Artifact	Present in higher compression ratios	Not present	Not present
Energy Compaction	Present because most of visually significant information is concentrated in few coefficients	Not present	Not present
Multiresolution property	Not present	Present i.e image can be shown from low to high resolution	Not present
Strongest components (for Low frequency F _L band robustness) Description		Low frequency band (LL)	Central component (low frequency)

Resistance against Geometrical attacks	Not resistant	Not resistant	Resistant
Effect of change in transform coefficient	Affects entire image except it is implemented using block based approach	Affects image locally because of its property of spatial frequency locality	Full frame transform, affects entire image

II. REVIEW OF AVAILABLE TRANSFORM DOMAIN BASED WATERMARKING USING ARTIFICIAL INTELLIGENCE (AI) METHOD

Review of image watermarking using noticeable AI techniques is carried out and presented below.

Wei et al. [5] proposed a new approach for optimization in 8*8 DCT domains using genetic algorithm (GA) to improve the robustness and imperceptibility of image spread spectrum watermark algorithm. AC coefficients have been modified to embed spread spectrum watermark. These coefficients have been chose by GA in different bands to find best suitable AC coefficients for image with different characteristics. The experiment results show that the method is tested against signal processing attacks. The method is robust and achieved good image quality for different attacks. Nilchi et al. [9] presented a non-blind robust image watermarking technique based on DCT and Full Counter Propagation Neural Network (FCNN). FCNN has been used to simulate the perceptual characteristics of the original image. These perceptual features of the original image have been used to determine highest changeable threshold values of DCT coefficients and to embed watermark binary image. The watermarking algorithm has very good robustness to JPEG compression attacks as compared to other attacks. Aslantas et al. [6] introduced an optimal DWT-SVD based image watermarking technique using Particle Swarm Optimizer (PSO). In this scheme host image has been decomposed into sub bands and after that singular values of each sub band of host image have been modified by different scaling factors to embed watermark image. These modifications have been optimized using PSO to obtain highest robustness and transparency. Experimental results show that this scheme has given good results as compared to methods that use single scaling factor. Lai et al. [19] introduced an image watermarking scheme using SVD and micro-genetic algorithm. The values of scaling factors has been obtained and optimized by means of micro-genetic algorithm to embed watermark. Watermark has been embedded by modifying singular values of cover image by multiple scaling factors. The experimental results show that watermark is robust with highest NC value equal to 1. Aslantas et al. [7] proposed Intelligent Optimization Algorithms (IOA) namely Genetic Algorithm (GA), Differential Evolution (DE), Clonal Selection Algorithm (CSA), and Particle Swarm Optimization (PSO) to improve performance of fragile watermarking based on DCT. In embedding process after modifying the least significant bits of transformation coefficients, these coefficients when transformed from frequency domain to spatial domain produces some round off errors. These errors have been corrected by intelligent optimization algorithms. Experimental results show that GA, DE and PSO have similar performance. CSA has produced better PSNR results with the expense of increased computational complexity.

Kumsawat et al. [20] proposed a blind digital image watermarking scheme in multi wavelet domain. In this scheme embedding method has been based on quantization index modulation technique. Genetic algorithm has been used as an optimization technique and used to search for optimal quantization steps to improve robustness of watermark and quality of watermarked image. Experimental results show that watermark is robust to common attacks such as median filtering, Gaussian filtering, image rotation etc. Veysel Aslantas [8] presented a scheme for developing a robust image watermarking without losing the transparency using SVD watermarking method. The singular values of host image have been modified by multiple scaling factors. These modifications have been optimized using DE to achieve robustness and transparency. Experimental results obtained from this method show improvement in transparency and robustness for various signal processing attacks. This method has also been found superior over the use of constant scaling factor and even better than the results of using Genetic Algorithm. Poonam et al. [21] presented a method to improve robustness and imperceptibleness in watermarking. The singular values of 3rd level DWT approximation matrix of original image has been embedded with the singular values of watermark. Genetic algorithm has been used as an optimization algorithm to optimize scaling factors with which watermark should be embedded in host image. Fitness function in this algorithm takes two values PSNR and correlation. Experimental results show that increase in scaling factors results in better performance in terms of robustness by keeping image quality to reasonable levels. Lai et al.[22] introduced watermarking technique that uses SVD and GA. Singular values of cover image have been modified by multiple scaling factors to embed watermark. Genetic Algorithm has been used to search proper values to satisfy both imperceptibility and robustness requirements. Experimental results show that proposed scheme survived successfully after various signal processing attacks and also outperformed other similar works. Vafaei et al.[23] proposed a blind watermarking method in discrete wavelet transform domain based on neural networks. Artificial neural networks have been used to adjust the watermark strength. Binary image as watermark has been embedded repetitively into selected wavelet coefficients. Feed Forward Neural Networks satisfactorily maximize watermark strength using proper trainings by being adaptive based on knowledge of block features. Thus proposed method has high robustness and imperceptibility to different types of attacks such as cropping, filtering etc. Chaudhary et al.[24] presented a new method for adaptive watermark strength optimization in DCT domain.

Genetic Algorithm has been used to select DCT sub band and Particle Swarm Optimization (PSO) has been used as an optimization technique to intelligently select watermark strength. The proposed scheme shows better performance for imperceptibility and robustness tests. Ali et al.[25] proposed a technique using SVD and Differential Evolution in DCT domain. The original image has been partitioned into blocks and these blocks have been transformed into DCT domain. The ocefficients of each block have been collected and a low-resolution approximation image has been constructed. SVD has been applied on this approximation image and its singular values have been modified with singular values of watermark. DE algorithm identifies best multiple scaling factors for embedding to achieve robustness without compromising quality of image. The watermark has been scrambled by Arnold Transform before embedding. Experiment results show that proposed technique yields strong robustness to geometrical and image processing attacks and imperceptibility is also promising.

S.no	Author's name	Techniques Used	Watermark/ Cover Image	Results (Highest values)
1	Wei et al. [5]	DCT, Genetic Algorithm (GA)	Watermark of size 2048 bits Gold hill image (256*256)	PSNR = 51db Similarity value = 8.5
2	Nilchi et al. [9]	DCT, Full Counter Propagation Neural network (FCNN)	Watermark: binary image (64*64) Barbara gray image (512*512)	PSNR =46.15db Similarity percentage value =99.92%
3	Aslantas et al. [6]	DWT-SVD, Particle Swarm Optimizer (PSO)	Watermark: gray level image (256*256) Lena Image (512*512)	NC = 0.9997
4	Lai et al. [19]	SVD, micro-Genetic Algorithm (GA)	Watermark: binary image (32*32) Lena image (256*256)	NC = 1
5	Aslantas et al. [7]	DCT, Genetic algorithm (GA), Differential Evolution (DE), Particle Swarm Optimizer (PSO), Clonal Selection Algorithm (CSA)	Watermark: binary image (64*64) Lena, Baboon, F-16 and Peppers image (256*256)	PSNR = 56.2858db NC = 1
6	Kumsawat et al. [20]	Multiwavelet domain Genetic Algorithm (GA)	Watermark of size 512 bits Lena, Baboon, Gold hill and Peppers gray level image (512*512)	PSNR = 46.70db NC = 1
7	VeyselAslantas [8]	SVD Differential evolution (DE)	Watermark: grey level image (32*32) Lena image (256*256)	PSNR = 38.023db NC = 1
8	Poonam et al. [21]	DWT Genetic Algorithm (GA)	Watermark: original watermark image (64*64) Original dog image (512*512)	PSNR = 47.9997db NC = 0.9702
9	Lai et al. [22]	SVD Genetic Algorithm (GA)	Watermark: binary image (32*32) Lena Image (512*512)	Proposed scheme survive after attacks and outperforms other similar works.
10	Vafaei et al. [23]	DWT Feed Forward Neural Network (FFNN)	Watermark: binary image 32*32 Lena, Baboon, Airplane, Barbara gray scale images (512*512)	PSNR = 48.25db NC = 1
11	Chaudhary et al. [24]	DCT Particle Swarm Optimizer (PSO)	Lena, boat, baboon, couple images (512*512)	PSNR = 48.10db NC = 0.99
12	Ali et al. [25]	DCT-SVD Arnold Transform Differential evolution (DE)	Watermark: gray scale image (64*64) Airplane, baboon image (512*512)	PSNR = 36.3848db NC = 0.9998
13	Ali et al. [26]	DWT-SVD Differential evolution (DE)	Watermark: binary image (64*64) Couple image (512*512)	PSNR = 35.2357db
14	Han et al. [27]	SVD Genetic Algorithm (GA)	Lena, Baboon, Barbara, Peppers, Bird, Cameraman, Gold hilland Airplane (256*256)	PSNR = 45.82db NC = 1
15	Yen et al. [10]	DCT, Back Propagation Neural Network (BPNN)	Watermark : original image (32*32) Lena image (256*256)	NC = 1

Fable III. SUMMA	ARY OF SOME TR	ANSFORM DOMAIN V	VATERMARKING	USING AI TECHNIQUES
-------------------------	----------------	------------------	--------------	---------------------

This method has not been good for attacks such as histogram equalization, gamma correction etc. Ali et al. [26] proposed an innovative watermarking scheme based on Differential Evolution and DWT-SVD transform domain. During embedding process host image has been transformed into sub bands of different frequencies by 3rd level DWT then watermarked image has been scaled down with multiple scaling factors and embedded in singular value matrix of LL and HH sub bands of host image to make watermark invisible and robust. DE has been used to search these optimal scaling factors. To overcome false positive problem watermark has been embedded in lossless manner. Watermark has been used to generate secret key instead of embedding it to host image. Experimental results shows that this method not only guarantees imperceptibility but also provide strong robustness. Han et al. [27] proposed a new watermarking method based on host image analysis using SVD and Genetic Algorithm. This method has enhanced embedding capacity of watermark in order to embed more watermarks in host image and having same visual effect. Genetic algorithm has been used in embedding process to improve the image quality, security and robustness of watermarked image. Algorithm also satisfies an optimal compromise between robustness and image quality. Threshold T has been used to control embedding rate. Experiments ensured that this scheme has achieved imperceptibility and good robustness to general signal processing attacks. Yen et al.[10] proposed a scheme to suppress the interference affecting watermarks. Watermarks have been embedded in host image using DCT technique. After that various common interferences such as salt-and-pepper noise, Gaussian noise, clipping and rotation have been used to destroy the watermarked image. Back Propagation Neural Network algorithm when combined with DCT watermarking scheme has been used to suppress these interferences that affect watermarks. The simulation results demonstrate the effectiveness of this technique in restoring watermarked image.

Based on the literature survey, AI techniques efficiently remove the tradeoff between the robustness and imperceptibility of the watermarked images. The results can be improved further by finding the best possible combinations of watermarking techniques and AI techniques. The above mentioned techniques are summarized in Table III. The current state-of-the-art in transform domain based image watermarking using AI techniques as available in the reviews is given below:

П. In [5], the method is focuses on the robustness of the watermark. Genetic Algorithm has been used to choose different bands in DCT domain to find best suitable AC coefficients to embed watermark. The method [9] achieved robustness by simulating HVS model of image using intelligent neural network so that image quality do not damage. Full Counter Propagation Neural Network (FCNN) used to determine best changeable threshold values of DCT coefficients, which are used to embed watermark binary image. The methods [6, 8, 19 and 22] focuses on both the robustness and imperceptibility of watermarked image. Different AI techniques such as Particle Swarm Optimizer, micro-Genetic algorithm, Genetic Algorithm, Differential evolution have been used in modification of singular values by different scaling factors in DCT domain. The method [7] has achieved higher transparency and robustness. The method has been based on removing round off errors occurred while transforming coefficients from frequency domain to spatial domain using DCT watermarking technique. These errors have been corrected using Intelligent Optimization Algorithms (IOA). This method [20] is based on method to search optimal quantization steps to improve robustness and also by providing good watermarked image quality. Genetic Algorithm has been used as an optimization technique in wavelet domain. A watermarking method is to improved the robustness and imperceptible in [21]. Genetic algorithm has been used as an optimization algorithm to optimize scaling factors to embed watermark in singular values of 3rd level DWT. The method discussed in [23] focused on adjusting tradeoff between robustness and imperceptibility of watermarked images. Artificial Neural Networks (ANN) has been used to adjust the watermark strength using DWT watermarking technique. The method discussed in [24] is based on improving the robustness and imperceptibility of an image. Genetic Algorithm (GA) has been used to search optimal sub-band. Particle Swarm Optimizer has been used to select the watermarking strength using DCT watermarking technique. The method proposed in [25] achieved better performance in terms of robustness without compromising image quality. Hybrid watermarking techniques has been used where DCT coefficients to be changed collected into low approximation image and then SVD has been applied on this image. Differential Evolution has been used to identify best scaling factors. Arnold transform has also been applied to scramble the watermark. In [26], the method focuses on both robustness and imperceptibility and based on combination of SVD and DWT. DE has been used to search optimal scaling factors to embed watermark in LL and HH sub bands. The method proposed in [27] focuses mainly on image quality but it also enhances the security and robustness. GA has been used to enhance embedding capacity of watermark. The method proposed in [10] improved the quality of watermarked image for different signal processing attacks. Back Propagation Neural Network (BPNN) is combined with DCT has been used to suppress interference that affects watermark.

IV. Conclusion

This paper discussed the basic concepts of digital watermarking techniques, important spatial and transform domain techniques and compare the performance of some reported transform domain based watermarking techniques using artificial intelligence (AI) are used in different ways to optimize the embedding strength of watermark by analyzing host image. Finally, the current state-of-the-art in transform domain watermarking using AI techniques is also discussed. Following points are to be concluded by us 1) AI based watermarking methods have very good performance against signal processing attacks. However, the computational complexity of these methods is too high. 2) Many image watermarking methods are proposed by various authors/scientist but there is no single technique that can provide satisfactory performance against robustness, capacity, imperceptibility and security requirements simultaneously. 3) Also, researches and applications in AI based watermarking are blooming area that can further be explored.

V. References

- [1] W. BENDER, D. GRUHL, N. MORIMOTO AND A. LU (1996), TECHNIQUES FOR DATA HIDING. IBM SYSTEMS JOURNAL, VOL. 35, NOS 3&4, PP. 313-336.
- [2] NIELS PROVOS AND PETER HONEYMAN (2003), HIDE AND SEEK: AN INTRODUCTION TO STEGANOGRAPHY. IEEE COMPUTER SOCIETY, IEEE SECURITY & PRIVACY, 1540-7993/03, PP. 32-44.
- [3] A. CHEDDAD, J. CONDELL, K. CURRAN AND P. MCKEVITT (2010), DIGITAL IMAGE STEGANOGRAPHY : SURVEY AND ANALYSES OF CURRENT METHODS. SIGNAL PROCESSING, VOL. 90, PP. 727-752.
- [4] P. Singh And R S Chadha (2013), A Survey Of Digital Watermarking Techniques, Applications And Attacks. International Journal Of Engineering And Innovative Technology (Ijeit) Vol. 2, Issue 9, Pp. 165-175.
- [5] W. Zhicheng, L. Hao, Dai Jufeng And W. Sashuang (2006), Image Watermarking Based On Genetic Algorithm. Ieee International Conference On Multimedia And Expo Icme 1-4244-0367-7/06, Pp. 1117-1120.
- [6] V. ASLANTAS, A. L. DOGAN AND SERKAN OZTURK (2008), DWT-SVD BASED IMAGE WATERMARKING USING PARTICLE SWARM OPTIMIZER.IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO (ICME) 978-1-4244-2571-6/08, PP. 241-244.
- [7] V. ASLANTAS, S. OZER AND S. OZTURK (2009), IMPROVING THE PERFORMANCE OF DCT-BASED FRAGILE WATERMARKING USING INTELLIGENT OPTIMIZATION ALGORITHMS. OPTICS COMMUNICATIONS 282, PP. 2806–2817.
- [8] V. ASLANTAS (2009), AN OPTIMAL ROBUST DIGITAL IMAGE WATERMARKING BASED ON SVD USING DIFFERENTIAL EVOLUTION ALGORITHM. OPTICS COMMUNICATIONS 282, PP. 769–777.
- [9] N. Nilchi Ahmad R And T. Ayoub (2008), A New Robust Digital Image Watermarking Technique Based On The Discrete Cosine Transformation And Neural Network. International Symposium on Biometrics And Security Technologies, Islamabad, Pp. 1-7.
- [10] C-T YEN AND Y-J HUANG (2015), FREQUENCY DOMAIN DIGITAL WATERMARK RECOGNITION USING IMAGE CODE SEQUENCES WITH A BACK-PROPAGATION NEURAL NETWORK. MULTIMED TOOLS APPL. DOI 10.1007/S11042-015-2718-Y.
- [11] T. S Tagare and S. Kuri (2015), Digital Image Watermarking -An Overview. International Journal of Computer Science Information Engineering Technologies, Issue 3 Volume – 1 – Series-4.
- [12] R. Singh (2015), Digital Image Watermarking: An Overview. International Journal of Research (IJR), Vol. 2, Issue 05, pp. 1087-1094.
- [13] V. M. POTDAR, S. HAN AND E. CHANG (2005), A SURVEY OF DIGITAL IMAGE WATERMARKING TECHNIQUES. 3RD IEEE INTERNATIONAL CONFERENCE ON INDUSTRIAL INFORMATICS (INDIN), 0-7803-9094-6/05, pp. 709-716.
- [14] P. Parashar and R. K. Singh (2014), A Survey: Digital Image Watermarking Techniques. International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 7, No. 6, pp. 111-124
- [15] N. Ahmidi. and Reza Safabakhsh (2004), A Novel DCT-based Approach for Secure Color Image Watermarking. IEEE Proceedings of the International Conference on Information Technology: Coding and Computing, pp. 709-713.
- [16] M. Barni and F Bartolini (2001), Improved Wavelet-Based Watermarking Through Pixel-Wise Masking. IEEE Transactions On Image Processing, Vol. 10, No. 5, pp. 783-791.
- [17] E. Ganic and Ahmet M. Eskicioglu (2004), Robust DWT-SVD Domain Image Watermarking: Embedding Data in All Frequencies. ACM 1-58113-854-7/04/0009.
- [18] Nikita Kashyap and G. R. Sinha (2012), Image Watermarking Using 3-Level Discrete Wavelet Transform (DWT). I. J. Modern Education and Computer Science, DOI 10.5815/ijmecs.2012.03.07, pp. 50-56.

- [19]C-C Lai, H-C Huang and C-C Tsai (2008), Image watermarking scheme using singular value decomposition and micro-genetic algorithm. IEEE International conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), DOI: 10.5815/IIH-MSP.2008.168, pp. 469-472.
- [20] P. Kumsawat, K. Attakitmongcol and A. Srikaew (2009), An Optimal Robust Digital Image Watermarking Based on Genetic Algorithms in Multiwavelet Domain. WSEAS Transactions On Signal Processing, Issue 1, Volume 5, pp. 42-51.
- [21] POONAM, S. KUNDU, S. KUMAR AND K. CHANDER (2012), EFFICIENT GENETIC ALGORITHM BASED IMAGE WATERMARKING USING DWT-SVD TECHNIQUES. IEEE, INTERNATIONAL CONFERENCE ON COMPUTING SCIENCES (ICCS) DOI 10.1109/ICCS.2012.32, pp. 82-87.
- [22] C-C LAI, C-H YEH, C-H KO AND C-Y CHIANG (2012), IMAGE WATERMARKING SCHEME USING GENETIC ALGORITHM. IEEE, INTERNATIONAL CONFERENCE ON GENETIC AND EVOLUTIONARY COMPUTING, DOI 10.1109/ICGEC.2012.103, PP. 476-489.
- [23] M. Vafaei, H. Mahdavi-Nasab and H. Pourghassem (2013), A New Robust Blind Watermarking Method Based on Neural Networks in Wavelet Transform Domain. World Applied Sciences Journal, DOI: 10.5829/idosi.wasj.2013.22.11.2828, pp. 1572-1580.
- [24] A. Chaudhry, B. Sikandar, M. Ishtiaq, A. Jaffar, J. Y. Kim, M Ishtiaq and T. A. Tuan (2013), Genetic Swarm Based Robust Image Watermarking. Proceedings of the 7th International Conference on Ubiquitous Information Management and Communication.
- [25] M. Ali, C. Wook Ahn and M. Pant (2014), A robust image watermarking technique using SVD and differential evolution in DCT domain. Optik 125, pp. 428–434.
- [26] Musrat Ali, Chang WookAhn and Patrick Siarry (2014), Differential evolution algorithm for the selection of optimal scaling factors in image watermarking. Engineering Applications of Artificial Intelligence pp. 15–26
- [27] J. Han, X. Zhao and C. Qiu (2015), A digital image watermarking method based on host image analysis and genetic algorithm. Journal of Ambient Intelligence and Humanized Computing, DOI 10.1007/s12652-015-0298-3.

Green ICT: The New Era of Effective and Efficient Green Learning Based on ICT

Monika Sahu RIE, NCERT Bhopal (M.P.) India monikasahu7@gmail.com

Madhup Shrivastava NITTTR, Bhopal (M.P.) India 08madhup@gmail.com

M.A. Rizvi Associate Professor, NITTTR Bhopal (M.P.) India <u>marizvi@nitttrbpl.ac.in</u>

Abstract—As information and communication technologies are rising and the organizations are reaching globally, the possibility and influence of the ICT in daily life appear unparalleled. Mostly the ways to make sense, most of our information exchange with other people, are not directly human and expressive, but interfacing through interactive and non interactive like TV, radio, magazines, music etc. Computers open new forms of learning which eliminates the restrictions of older approaches. Globalization is driven in fragment by incredible and rapid ICT developments, and young people are often among the first to take advantage of new developments in this area. ICT plays major role in development in learning, development in the world of work and changing the nature of social interaction but it also affects the nature in adverse ways like "Global Warming". In this paper an effort is put to reduce those bad outcomes by adapting the "Green computing" over the network.

Keywords: ICT (Information and Communication Technology), IT (Information Technology), Green Computing, Global Warming, BST (Binary Search Tree)

I. INTRODUCTION

In today's world youth are distinguished by their cultural, economical, social and educational dissimilarities; particular circumstances are somehow depends on where a person is grown up. More than 800 million youth are at a standstill lack basic literacy skills and at another side, the use of information and communication technologies (ICT) is moving upward irrespective of the differences in living environments. An unprecedented and joining global media culture has made that confronts and often outshines such established structures of socialization as family, colleges and educational places.

In the difficult enriching situation where young people are fighting to choose the direction in their life or simply to continue to exist, to enhance their living conditions, and to extend their identities, they are choosing diverse, unique and technical ways. So some call it the era of Information, some others use the word technical culture or techno-capitalism, global media culture, or simply globalization, which refers to the process in which the all survive as "united and equally associating principles" signifies obeying the similar rules with different extends.

Two most important theories motivate the role of ICT: the first is that the proliferation of the expertise is resulting rapid revolutions in all regions of life; the next is that ICT function to combined and normalize culture [1]. Information and communications technology (ICT) delivers all the expertise used to access the information related to the communication through different channels in different ways. So ICT is often regarded as an enhanced means for information technology (IT) and its scope is broader than IT. ICT has more newly been used to explain the convergence of several technologies and the use of common transmission lines carrying very assorted data and communication types and formats [2].

ICT is used as an extended mean for information technology (IT), but is a specific term that stresses the role of unified communications and the conjunction of telecommunications, computers as well as important software, interface, memory, and audio-visual systems, which allow users to create, access, store, transmit, and manipulate information [3].

A. ICT In Education:

An advance world with old classrooms; If we analyze current situation with the one hundred years ago, we feel astonishing advances in science, math, commerce, medical, transportation, and lot other areas. But if we compare the classrooms the conditions are still very much similar as before, that student energetically copying all the content as said by the teacher, expecting to memorize the facts and write it out on an exam. Information and communication technologies are of dominant importance to the upcoming of education. ICT in education initiatives are basically focusing on the following areas:

- Escalating access through distance learning
- · Opening a knowledge network for students
- · Training teachers

- Expansion of the availability of quality educational accessories
- Enhancing the efficiency and effectiveness of educational administration [4]

Worldwide research has shown that ICT can direct to enhance student learning and better teaching methods. ICT through integrated curriculum has a considerable and affirmative impact on student achievement, especially in terms of "Knowledge, "Comprehension", "Practical skill" and "Presentation skill" in respective subjects. "ICT in Education" means "Teaching and Learning with intervention of Technology". Educational ICT tools can be divided into 3 categories: Input source, Output source and others [5] refer Figure 1. Below:



Fig 1. ICT Tools

For the teacher, ICT is an initiation into:

- · Exploring educational possibilities with technology
- · Need analysis for right choices of hardware, software and ICT interactions, and
- Become a critical user of ICT

For the student, it is the beginning into:

- Creativeness and problem solving,
- An introduction to the world of information and technologies, and
- An opportunity to shape career pursuits [6]

In recent years we tried to use the computer and technology to enhance the efficiency and effectiveness of education in student's life. Moreover, different technologies are typically used in combination rather than as the sole delivery mechanism. There are various ICT tools used for education; that can be interactive or non-interactive. The way we categorize the tools is based on its use and function and mostly used in education as well. These are:

1. Interactive Whiteboard 2. Educational games 3. Intranet 4. E-Books [7]

B. Types of ICT based of teaching learning:

- E-learning: It is now-a-days commonly integrated with higher education and corporate training. It includes learning at all levels, both formally and informally, it also uses an information network known as online learning. Web-based learning is a part of e-learning and refers to learning using the Internet, mainly using a browser. It can also be preferred as learning through the electronic devices.
- Blended learning: This is also the type that is gaining currency is blended learning. This refers to learning models that combine traditional classroom work with e-learning modules.
- Open and distance learning: Open and distance learning is defined as "a way of providing learning opportunities that is characterized by the disconnection of teacher and learner in time or place, or both time and place; learning that is certified in some way by an institution or agency".
- Learner-centered environment: The National Research Council of the U.S. defines learner-centered environments as those that "pay careful attention to the knowledge, skills, attitudes, and beliefs that learners bring with them to the classroom". It is best done in a group setting with the teacher as facilitator or guide [8].

Advantages of ICT

- · It improves the speed of sending the information or communication
- · It provides globalization over the information exchange

- It provides a voice to those who normally would not speak up in groups
- It is cost effective than older mediums of communication
- It provides great availability due to use of internet
- Creation of new jobs as Computer programmers, Systems analysts, Hardware and Software developers, and Web designers

• It also helps in Education: Information is freely available to anybody and all with an internet connection only [5][9][10] *Disadvantages of ICT*

- Poor substitution for face-to-face (ftf) communication and expensive to afford
- · Hard for teachers to use with a lack of experience using ICT tools
- · Reliance on Technology making the memory weak
- Threats like viruses, worms etc. can attack [5][9][10]
- CO₂ emission which causes pollution and one of the reason of global warming

Due to the global warming expansion, the new concept is introduced which focuses on the technology which saves energy and try to prevent the pollution due to modern IT i.e. called "Green Computing".

C. Green Computing:

Population of world are facing the biggest environmental threat today is global warming, caused by carbon emissions into environment. About 98% of CO_2 emissions can be directly accredited to energy expenditure (according to a research by the Energy Information Administration). Many associations are now speaking openly about an aspiration to operate in a "green" manner, launching standards for environmental practices and sustainability on their commercial web and many industries are submitting/ paying of carbon tax also for the assets they are using and the environmental impact of the inventions and services they produce. Green computing aspires to accomplish economic feasibility and improve the way computing devices are used. Green IT preparations contain the energy efficient computing, disposal and recycling to develop the environmentally sustainable products [11].

To promote green computing concepts at all possible echelons, the following four corresponding approaches are introduced:

- Green use: try to minimize the electricity consumption of computers and the surrounding devices
- Green disposal: try to re-use an existing computer or appropriately disposing or recycling of discarded electronic apparatus
- Green design: scheming energy-efficient computers, servers, printers, projectors and other digital devices
- Green manufacturing: Minimizing waste during the manufacturing of computing devices and other subsystems to reduce the environmental impact of these activities [12]

There are many ways to reduce the impact of ICT over environment which somehow control the side effects of ICT over nature. These include:

- Just try to repair ICT equipment when they disturbing before replacing
- Safe disposal and recycling of ICT equipment
- Turn off the device when it's not being used
- When purchasing choose energy-saving devices
- Try to switch over the online services that are using green energy resources [13]

Many IT producers and traders are continuously investing in designing energy proficient computing devices, reducing the use of hazardous materials and encouraging the recyclability of digital devices and papers used for printing. It is also known as Green Information Technology (Green IT) [12].

The green computing implementation can also be categorized in another way. It can be as level wise:

1. Level 0: At Home Level - In this level we tried to visualize the concept of green computing at home means try to reduce the use of computer or use energy efficient devices.

2. Level 1: At Sharing Level - In this level we tried to visualize the concept of green computing at sharing level that means try to share the devices between users to reduce the no of devices used.

3. Level 2: At Infrastructure Level - In this level we tried visualize the concept of green computing at infra level means try to design the energy efficient devices and energy efficient network.

4. Level 3: At Cloud Computing level - In this level we tried visualize the concept of green computing at cloud level means at storage as well as network level.

There is a convincing need of appliances which take environmental factors into account at the time of design, driven by the need to align with organizational environmental policies, reduce power and infrastructure costs and to reduce current or future carbon costs [11].

In this paper an attempt is made to identify the problem with ICT environment and try to solve that problem too with the help of Green Networking using the concept of Green Computing.

II. PROBLEM FORMULATED

To keep balance between technology and environment "Green" has to be adopted. In ICT computer and related device which is largely used in education because technology based learning is used. This use of computers decreases the use of print material which reduces the cutting of tree; this will help in reducing global warming which is still a concern and needs to be further control.

There is a lot of energy emission due to large and inefficient networking which falls under the class of Level-2 of Green Computing. We have to reduce the emission in this level to achieve the goals of Green Computing at our infrastructure level. In this paper a trial is made to suggest a concept of networking to achieve the Green Computing at infra level.

PROPOSED METHODOLOGY

To achieve the goal of Green computing means to reduce the emission of energy while transmission of information, so there is need to design energy efficient network. To design a network, topologies are used. Consider wireless or wired network in tree form i.e. "Binary Search tree" form results as the following advantages in network:

- It allows fast lookup
- Addition and removal of nodes in network is faster
- Implementation is simple and dynamic in this form.

These advantages reduce the emission of energy and resulting as "Green Networking" *Functionality of the network*

To design the network in Binary Search Tree follow some properties of the data structure which are as follows:

- Elements in a binary tree are denoted by nodes.
- Nodes store the information in a tree.
- Each node in the tree must contain at least 3 fields containing:

a) IP Address at mid b) Pointer to left subtree (IP of left node) c) Pointer to right subtree (IP of right node)

- For every node with the IP address x in the tree
- Value of x is greater than all values in x's left subtree.
- Value of x is less than all values in x's right subtree.
- Both left subtree and right subtree are also binary search trees

Formation of network

To design the physical connection between nodes we have to follow the properties of BST. For example initially we have 7 nodes in our network with the IP Address like 172.16.2.39, 172.16.2.30, 172.16.2.46, 172.16.2.19, 172.16.2.32, 172.16.2.41 and 172.16.2.48. The initial structure of the network will be as shown in Fig 2.:



Fig 2. Tree structured Network

Insertion of node in network

Insertion operation will insert a node into the network and will become leaf node.

- Before the insertion of node in BST, the position of the new node must be determined. This is to ensure that after the insertion, the BST characteristics should be maintained.
- Steps to insert a new node in BST

1. First find out the position of the new node in the tree by comparing 2. Set NULL value to left and right pointer

3. Assign the IP address to be stored in the tree 4. Finally place the node in tree

For Example: If we place the node with IP Address 172.16.2.42 then the structure of the network will look like as shown in Refer fig 3:



Fig 3. Insertion of a node into Network

Deletion of node in network

When a node is deleted from the network, the linked node of the deleted node must be taken care of to ensure that the property of the Binary search tree should be maintained.

• There are 3 possible cases to delete a node in a tree:

1. Delete a leaf 2. Delete a node with one child 3. Delete a node that has two children

Delete a Leaf Node

If the node to be deleted is a leaf: Set the pointer in x's parent to NULL and delete it immediately

For example if we want to delete 172.16.2.42 from the tree (Fig 3) then the network will look like which is shown Refer Fig. 4:



Fig 4. Network after the deletion of leaf node

Delete a Node with One Child

Adjust a pointer from the parent to bypass that node i.e. the pointer of parent node now point to the child node of the deleted node. For example if we want to delete the node with IP Address 172.16.2.41 from the network (Fig 3) then the network will look like which is shown in fig. 5:



Fig 5. Network after the deletion of node with one child

Delete a Node with 2 Children

To delete a node x that has two children

- Replace the IP of that node with the minimum element at the right subtree.

- Delete the minimum element that has either no child or only right child because if it has a left child, that left child would be smaller and would have to be chosen. For example if we want to delete 172.16.2.46 from network (Fig 3) then the resultant network will look like as shown in fig 6:



Fig 6. Network after the deletion of node with two children

Searching of node in network

Searching in network is depends on the comparison between the address values of node of network and the node which we have to find.

- First compare with the root if it is found then ok
- If it is greater then go to right sub-tree otherwise go to left sub-tree
- Repeat the steps until the node is not found [14]
- •

III. CONCLUSION

The world is grooming and the changes are reflecting in our lives as dependency on technology. To be aware with these technologies, it should be exemplified in curriculum also. This technology can be demarcated as ICT. ICT is the umbrella term

which includes all the technologies which are used for communication in sense of information exchanging. Through the ICT we are contending with the external world but some complications which are produced as effect of excess use of ICT are the main causes of "Global Warming". In this paper we tried to decrease the emission of energy from the networking i.e. "Green Computing". We arrange the network in a Binary Search Tree manner to reduce its side effects on environment. Through this technique we get some advantages like:

- Searching will be easy which saves energy (Unicasting / Multicasting can be done instead of Broadcasting for finding the node)
- Insertion and deletion of nodes in network will be easy which also saves energy
- Better and easy understanding of network to work over that, at time of fault.

If we are competent to diminish even around fractional points in global warming that will be beneficiary for our next generations. So try to "Go Green".

IV. FUTURE WORK

To improve the concept of Green Computing and implementing it in our technology many changes can be done, like implementation of green concept at computer level, infrastructure level, networking and sharing level and cloud level. There is lots of scope to implement green over above levels using different concepts at different level. This will finally control the "Global Warming".

V. REFERENCES

- [1] Youth and Information and Communication Technologies (ICT) World YOUTH Report, 2003 Page 311-333
- [2] http://www.techopedia.com/definition/24152/information-and-communications-technology-ict
- [3] https://en.wikipedia.org/wiki/Information_and_communications_technology
- [4] Information and Communication Technology (ICT) and Education ,The Challenge and the Opportunity: A Changed World with Unchanged Classrooms Harvard University Report, 2002
- [5] http://www.elmoglobal.com/en/html/ict/01.aspx
- [6] http://ictcurriculum.gov.in/
- [7] https://sites.google.com/a/myport.ac.uk/using-ict-in-education/types-of-ict-tools
- [8] https://en.wikibooks.org/wiki/ICT_in_Education/Definition_of_Terms
- [9] http://www.useoftechnology.com/advantages-disadvantages-communication-technology-organization/

[10] https://storify.com/RoryP1/some-of-the-advantages-and-disadvantages-of-commun

- [11] research.microsoft.com/pubs/78813/aj18_en.pdf, The Architecture Journal page 1-44
- [12] http://www.techopedia.com/definition/14753/green-computing
- [13] http://www.ictchampions.org.uk/2009/05/what-can-i-do-about-green-computing/
- [14] http://ocw.utm.my/file.php/31/Module/ocwSearchTreeDec2011.pdf, Binary Search Tree by Nor Bahiah Hj Ahm

Concept Drift in Data Streams with Decision Trees

E.Padmalatha CSE Dept, CBIT, Osmania University padmalatha@cbit.ac.in

C.R.K.Reddy CSE Dept, CBIT, Osmania University Padmaja Rani CSE Dept, JNTUH

Abstract—Large amount data is generated from various institutions like hospitals which generate data to determine disease from symptoms. A single instance of data may belong to a previously defined class (earlier identified disease) or a new class (a new disease like ebola). This data is continuously generated, and is called data stream. Data stream is a new era in data mining. Numerous algorithms are used for processing & classifying data streams. Traditional algorithms are not appropriate to process data stream which in cause generate problems regarding classification. A model which is developed from stream data for classification must update incrementally after the fresh arrival of new data. Data stream classification performance can be measured by various factors such as accuracy, computational speed, memory and time taken for processing. Proposes to implement and use the very fast decision tree (VFDT) algorithm can effectively perform a test-and-train process with a limited segment of data.

Distribution of data stream is always changing in the real world. This problem is usually defined as concept drift. The stateof-the-art decision tree classification method Concept adaptive Very Fast Decision Tree (CVFDT) can solve the concept drift problem well, but the efficiency is reduced because of its general method of handling instances in CVFDT without considering the types of concept drift. An algorithm called Efficient CVFDT (E-CVFDT) is proposed to improve the efficiency of CVFDT. E-CVFDT introduces cache mechanism and treats the instances in three kinds of concept drift respectively, i.e. accidental concept drift, gradual concept drift and instantaneous concept drift higher accuracy than CVFDT algorithm.

Keywords: VFDT, CVFDT, E-CVFDT, Concept drift.

I. INTRODUCTION

Traditional data mining methods, that require learning from the whole set of historical data, do not suit the fast changes in high speed data streams in real-time applications. So we use Very Fast Decision Tree (VFDT)[1], a well-known decision tree algorithm for DSM. Its underlying principle is a dynamic decision tree building process that uses a Hoeffding Bound (HB) to determine the conversion of a tree-leaf to a tree node by accumulating sufficient statistics of the new samples. VFDT is I/O bound in the sense that it mines examples in less time than it takes to input them from disk. It does not store any examples (or parts thereof) in main memory, requiring only space proportional to the size of the tree and associated sufficient statistics. It can learn by seeing each example only once, and therefore does not require examples from an online stream to ever be stored. It is an anytime algorithm in the sense that a ready-to-use model is available at any time after the first few examples are seen, and its quality increases smoothly with time. Data streams are dynamic and infinite in size ,in the sense data is continuously generated and changing and live streams may have no upper limit .A live stream, can be read only once.

("Just One Look"). The stream rate may be very high. Regular data mining is based on finite data collections stored in files or databases. Also data mining is done off-line in batch:

read (access) data in collection several times
analyze accessed data
store results in database (or file)

Currently, the most efficient algorithms available concentrate on making it possible to mine databases that do not fit in main memory by only requiring sequential scans of the disk. But even these algorithms have only been tested on up to a few million examples. In many applications this is less than a day"s worth of data. For example, every day retail chains record millions of transactions, telecommunications companies connect millions of calls, large banks process millions of ATM and credit card operations, and popular Web sites log millions of hits. As the expansion of the Internet continues and ubiquitous computing becomes a reality, we can expect that such data volumes will become the rule rather than the exception. Current data mining systems are not equipped to cope with them. When new examples arrive at a higher rate than they can be mined, the quantity of unused data grows without bounds as time progresses. Even simply preserving the examples for future use can be a problem when they need to be sent to tertiary storage, are easily lost or corrupted, or become unusable when the relevant contextual information is no longer avail-able. When the source of examples is an open-ended data stream, the notion of mining a database of fixed size itself becomes questionable. So continuous stream of information has challenged our storage, computation and communication capabilities in computing systems. And for effective processing of stream data, new data structure, techniques, and algorithms are needed. Because we do not have finite amount of space to store stream data, we often trade-off between accuracy and storage ^[5]. So we can classify into five category [2][3][4] so in table 1.

S.NO	ISSUES	CHALLENGES	DIFFERENT APPROACHES FOR THAT
	MEMORY	FLUCTUATED AND	SUMMARIZATION
1.			TECHNIQUES
		IRREGULAR DATA	
		ARRIVAL RATE	
		AND VARIANT	
		DATA	
		ARRIVAL RATE	
	MANAGEMENT	OVER TIME	
2.	DATA	QUALITY OF	LIGHT-WEIGHT
	PREPROCESSING	MINING RESULT	PREPROCESSING
		AND	TECHNIQUES
		AUTOMATION OF	
		PREPROCESSING	
		TECHNIQUES	
		~	
3.	COMPACT DATA	LIMITED	INCREMENTAL
	STRUCTURE	MEMORY SIZE	MAINTAINING OF
		AND LARGE	DATA
		VOLUME OF	STRUCTURE, NOVEL
		DATA STREAM	INDEXING, STORAGE
			AND QUERYING
			TECHNIQUES
			~
4.	RESOURCE AWARE	LIMITED	
		RESOURCE LIKE	AOG
		STORAGE AND	
		COMPUTATION	
		CAPABILITIES	
5.	VISUALIZATION	PROBLEM IN	STILL IS A RESEARCH
	OF RESULTS	DATA ANALYSIS	ISSUE (ONE OF THE
		AND QUICK	PROPOSED
		DECISION	APPROACH IS:
		MAKING BY	INTELLIGENT
		USER	MONITORING)

Tabel1: Classification of challenges in five categories

II. VERY FAST DECISION TREE (VFDT)

VFDT allows the use of either information gain or the Gini index as the attribute evaluation measure. It includes a number of refinements to the algorithm:

Ties. When two or more attributes have very similar G's, potentially many examples will be required to decide between them with high confidence. This is presumably wasteful, because in this case it makes little difference which attribute is chosen. Thus VFDT[5] can optionally decide that there is effectively a tie and split on the current best attribute if $\Delta < \varepsilon < T$, where T is a user-specified threshold[28].

G computation. The most significant part of the time cost per example is recomputing G. It is inefficient to recompute G for every new example, because it is unlikely that the decision to split will be made at that specific point. Thus VFDT allows the user to specify a minimum number of new examples n_{min} that must be accumulated at a leaf before G is recomputed. This effectively reduces the global time spent on G computations by a factor of n_{min} , and can make learning with VFDT nearly as fast as simply classifying the training examples. Notice, however, that it will have the effect of implementing a smaller

 δ than the one specified by the user, because examples will be accumulated beyond the strict minimum required to choose the correct attribute with confidence 1 - δ. (This increases the time required to build a node, but our experiments show that the net effect is still a large speedup.) Because δ shrinks exponentially fast with the number of examples, the difference could be large, and the δ input to VFDT should be correspondingly larger than the target[28].

Memory. As long as VFDT processes examples faster than they arrive, which will be the case in all but the most demanding applications, the sole obstacle to learning arbitrarily complex models will be the finite RAM available. VFDT's memory use is dominated by the memory required to keep counts for all growing leaves. If the maximum available memory is ever reached, VFDT deactivates the least promising leaves in order to make room for new ones. If p_1 is the probability that an arbitrary example will fall into leaf l, and el is the observed error rate at that leaf, then P_1e_1 is an upper bound on the error reduction achievable by refining the leaf. P_1e_1 for a new leaf is estimated using the counts at the parent for the corresponding attribute value. The least promising leaves are considered to be the ones with the lowest values of P_1e_1 . When a leaf is deactivated, its memory is freed, except for a single number required to keep track of P_1e_1 . A leaf can then be reactivated if it becomes more promising than currently active leaves. This is accomplished by, at regular intervals, scanning through all the active and inactive leaves, and replacing the least promising active leaves with the inactive ones that dominate them [28].

Poor attributes. Memory usage is also minimized by dropping early on attributes that do not look promising. As soon as the difference between an attribute's G and the best one's becomes greater than ε , the attribute can be dropped from consideration, and the memory used to store the corresponding counts can be freed [28].

Initialization. **VFDT** can be initialized with the tree produced by a conventional RAM-based learner on a small subset of the data. This tree can either be input as is, or over-pruned to contain only those nodes that VFDT would have accepted given the number of examples at them. This can give VFDT a head start that will allow it to reach the same accuracies at smaller numbers of examples throughout the learning curve [28].

Rescans. VFDT can rescan previously-seen examples. This option can be activated if either the data arrives slowly enough that there is time for it, or if the dataset is finite and small enough that it is feasible to scan it multiple times. This means that VFDT need never grow a smaller (and potentially less accurate) tree than other algorithms because of using each example only once[28].

3. Concept adapting Very Fast Decision Tree (CVFDT)

VFDT is included in knowledge data discovery assume training data is a sample drawn from stationary distribution.

Data stream not considered this assumption because of concept drift. So to continuously change data stream is our only goal. CVFDT[27] is an extended version of VFDT which provides same speed and accuracy advantages but if any changes occur in example generating process provide the ability to detect and respond.

Various systems with this capability (Widmer and Kubat, 1996, Ganti et al., 2000), CVFDT[8] uses sliding window of various dataset to keep its model consistent. In Most of systems, it needs to learn a new model from scratch after arrival of new data. Instead, CVFDT[18] continuous monitors the quality of new data and adjusts those that are no longer correct. Whenever new data arrives, CVFDT[16] incrementing counts for new data and decrements counts for oldest data in the window. The concept is stationary than there is no statically effect. If the concept is changing, however, some splits examples that will no longer appear best because new data provides more gain than previous one. Whenever this thing occurs, CVFDT create alternative sub-tree to find best attribute at root. Each time new best tree replaces old sub tree and it is more accurate on new data. It is generally occurs in VFDT algorithm, but CVFDT continuously monitors the validity of its old decisions, by maintaining more than sufficient statistics at every node in Decision tree. As decision tree has grown Forgetting an old example is slightly complicated by the fact that DT may have grown or changed since the example was initially incorporated. To avoid forgetting an example from a node that has never seen it, nodes are assigned a unique, monotonically increasing ID as they are created. After addition of an example to W[8], the maximum ID of the leaves it reaches in DT and all alternate trees is recorded with it. An example's effects are forgotten if the example whose ID is less than or equal to stored ID by decrementing the counts in the sufficient statistics.

CVFDTGrow: In CVFDTGrow[12,13] for each node reached by the example in Hoeffding Tree Increment the corresponding statistics at the node. If sufficient examples seen at the leaf in HT which the example reaches, Choose the attribute that has the highest average value of the attribute evaluation measure (information gain or gini index). If the best attribute is not the "null" attribute, create a node for each possible value of this attribute.[11,17]Forget Old Example: Maintain the sufficient statistics at every node in Hoeffding tree to monitor the validity of its previous decisions. VFDT only maintain such statistics at leaves. Than HT might have grown or changed since the example was initially incorporated. It will assign unique increasing ID as they are created. After each node reached by the old example with node ID no larger than the max leave ID the example reaches, Decrement the corresponding statistics at the node and For each alternate tree Talt of the node, forget (Talt, example, maxID).

VI. EFFICIENT CVFDT (E-CVFDT)

4.1 Accidental Concept Drift

In some cases the concept drift in data stream occurs accidently. For example, adds n examples with A, and a few examples with B, set S is $\{s_{a1}, s_{a2}, .s_{b1}, s_{b2}..., s_{an}\}$. It is means that, the stream S is flow into classification model, the original examples s_a keep classifying for a while, after times, new examples s_b involve in classifying. But s_b disappears after a short time, and there are only s_a left in the data stream. Defines this scene is accidental concept drift. Traditional CVFDT[15] algorithm addresses concept drift in this scene is not perfect. The examples with new concepts always have little amount and appears probability in data stream. This project considers these examples as fate or owns a special concept. Although these examples appear probability is small, and it is not so important for creating the sprite decision, traditional CVFDT always uses it to participate in the information gain calculation for best attribute selection. In this way, performance efficiency is reduced and makes the decision selection more complex. To solve this concept drift problem in this scene, E-CVFDT[30] algorithm uses a caching mechanism. when the number of examples in window $|W| \le w$ and $nw \ge n*P$ -discard, algorithm puts all examples of W to an array discard (). The value P-discard is a user-supplied threshold, means the percentage of the maximal discard number for total number of examples which flow into the classification model. This is an experience value, always valued 3%. This threshold is very important, if a comfortable value chosen, the performance of the algorithm improved distinctly. Suppose |W|<w, there are n examples, after a while, not have any new examples was sent in. Thus, these examples are considered as unexpected examples. If data pre-process is very credible, these examples are believed as the accidental occurs example. If after t times, some examples arrive. In order to reduce the complex of the information gain calculation, E-CVFDT deletes the examples in primary process. When |discard ()|>w, algorithm send all examples back to evolving the process of creating decision tree. Experiments show if the value P-discard is properly, the classification accuracy and performance efficiency is not only affected, but also have a better improved.

4.2 Gradual Concept Drift

In some cases concept drift occurs in evolving style in data stream. For example, involves n examples with A with r examples with B, then the S becomes $\{s_{a1}, s_{a2}, s_{b1}, s_{a3}, s_{b2}, \dots, s_{ar}, s_{br}, s_{br+1}, \dots, s_{bn}\}$. It is means that, the S flow into classification model, the original examples s_a keep classifying for a while, after t times, new example s_b occurs and keep the style of these forms for a while, after t times, sa disappears, and only sb maintain. This is consider as gradual concept drift. This form of concept drift is the commonly seen in real world stream. Different evolving concepts flow into the decision tree model to classify. Because of the characteristic of evolving data, the E-CVFDT algorithm will find the best split attribute by computing information gain at a very fast frequency. It's good for obtain a better accuracy of classification, but the performance efficiency needed to reconsider. The complex information gain calculation with evolving data distribution affects the performance efficiency. So E-CVFDT method regroups the data in memory, delays the new concept examples, and lets the original concept evolve in information gain calculation first, it must be useful for improving the performance efficiency. When the new concept example occurs, algorithm puts these examples into W, to cache them for a while. Then algorithm check the |W|, if |W| < w, the algorithm deletes the examples from the W in primary process of algorithm. It means that, if new concept examples arrive, it's not used for information gain calculation immediately, but kept in W. When |W|>w means W is full, all the examples in W are sent to evolve in information gain calculation. In this way, the same concept examples are classified first. And avoid the complex distribution of examples in data stream affects the performance efficiency. In this way, E-CVFDT[17] method regroups the data distribution of data stream, and to reduce the computing cost in unit time. So that, the method E-CVFDT performances efficiently.

4.3 Instantaneous Concept drift

In some cases concept drift appears instantaneously. For example, adds n examples to B, then stream S becomes $\{s_{a1}, s_{a2}, ..., s_{ar}, s_{b1}, s_{b2}, ..., s_{bn}\}$. It is means that, after the S flow into classification model, the original examples s_a keep classifying for a while. After t times, new examples s_b occurs. At the same time, examples s_a disappears and never occurs again, and there is only s_b maintain. This it as instantaneously concept drift. This style of concept drift in data stream is very easy to deal with. Because, after t times, the new concept occurs instantaneously in data stream, at the same time, the old concept disappear, and it will never occurs again. The system maintains only the new ones.

The traditional CVFDT algorithms addresses concepts drift in this scene quit well, it can achieve good accuracy of classification and comfortable performance efficiency. The E-CVFDT [27]adds a mechanism into traditional CVFDT algorithm, which just only regroups data distribution, and its time complexity is linear. There is no effect for process of creating decision tree of traditional CVFDT, thus, the E-CVFDT algorithm's performance as well as the CVFDT algorithm.

V. IMPLEMENTATION

The statistical details for the data that is given in. It shows the number of classes, examples and the majority class label and the class distribution. A pie chart representing the proportion of class labels is also shown. Next, the tree information and the percentage of classified and misclassified data is shown. At the bottom, a confusion matrix is displayed, representing the correctly classified and incorrectly classified data. For implementation of VFDT thousand.arff [29] dataset is used ,the statistics developed by the program is mentioned in figure 5.1.

Further it is compared with J48 in time accuracy.

A comparison is made by giving the same input to a traditional data mining model, J48. The output shows that J48 generated the results in more time compared to the proposed system, VFDT. But the constraint is that the accuracy achieved by J48 is relatively higher than VFDT.
claime Tree Pres	Section Apply Tex	t Data Statistics Details	
Statistics For V	er.		
			DONN
Number C	f Classes 12		
20 unitser a	E-margles ::32		
Magnetty C	See Label: UP		
Class Diet	allestien ::	-	
Classes	Label	Proparties	
	C.P.	30.27%	
	DOM:N	2017.9	
Tree	all consultions		
Mumber a	126 oden : 85		UP
20 uniber o	Loud 21 adea : 2	2	
Waanhoor a	Levels : 25		
Contraction	States :	0-141	26 sandress of Text Observations 15
	Predictie	A Claim Label	% Sdawlasseder det.orenell
		-	% Classified: 50.300002
UP	8 7		
Actual DOW	N 0 1		
Clam			

Fig. 5.1: Statistical details of the data

vfdt (run) 🛙	vfdt (run) #3 II	
U		
0		
0		
0		
Decision 7	ree	
before dec	isom	
Decision 7	ree	
after deci	som	
heloo		
Time taken	for decision tree construction:	Oms 21541n

Fig. 5.2. Time taken by VFDT system

For the Implementation of CVFDT and E-CVFDT three artificial datasets are considered. Initially E-CVFDT has higher accuracy values compared to CVFDT algorithm but as we increase the size of the window the accuracy of CVFDT increases and overcomes E-CVFDT accuracy in Figure 5.3 But E-CVFDT produces a constant accuracy thus giving us guarantee of giving a fixed accuracy value independent of window size.

Though CVFDT shows greater value of accuracy, it does not work in real time scenarios. It slowly betrays the window concept. As discussed in earlier, one cannot store instances of data stream due to infinite size of data. When the size of the window is increased, the number of instances available in window increases thus it results in classification of similar instances directly without any miss in the tree which eventually reduces the overhead of adding new concept to the tree. This causes drawback of having old concept reside in the tree, causing increase in the storage cost.

÷

S.No.	w	cvfdt	ecvfdt	30	1900	71.35	71.45
1	50	66.72	71.45	31	2000	71.48	71.45
2	70	70.08	71.45	32	2100	71.55	71.45
3	90	67.8	71.45	33	2300	71.42	71.45
4	110	70.08	71.45	34	2400	71.43	71.45
5	130	71.36	71.45	35	2500	71.43	71.45
6	150	71.34	71.45	36	2600	71.56	71.45
7	170	71.34	71.45	37	2700	71.6	71.45
8	190	70.08	71.45	38	2800	71.6	71.45
9	210	69.16	71.45	39	2900	71.6	71.45
10	230	69.16	71.45	40	3000	71.6	71.45
11	250	70.08	71.45	41	3100	71.6	71.45
12	270	70.08	71.45	42	3200	71.46	71.45
13	290	71.27	71.45	43	3300	71.46	71.45
14	300	71.27	71.45	44	3400	71.6	71.45
15	400	71.27	71.45	45	3500	71.46	71.45
16	500	71.27	71.45	46	3600	71.46	71.45
17	600	71.14	71.45	47	3700	71.46	71.45
18	700	71.14	71.45	48	3800	71.45	71.45
19	800	71.14	71.45	49	4000	71.45	71.45
20	900	71.14	71.45	50	4100	71.45	71.45
21	1000	71.14	71.45	51	4200	71.46	71.45
22	1100	71.14	71.45	52	5000	71.36	71.45
23	1200	71.27	71.45	53	6000	71.53	71.45
24	1300	71.14	71.45	54	7000	71.52	71.45
25	1400	71.2	71.45	55	8000	71.52	71.45
26	1500	71.45	71.45	56	9000	71.42	71.45
27	1600	71.45	71.45	57	10000	71.6	71.45
28	1700	71.45	71.45	58	15000	71.99	71.45
29	1800	71.32	71.45	59	20000	71.77	71.45

Fable 5.1	accuracy	values	using	dataset	1



Fig. 5.3. Accuracy-w graph using dataset1

S.no	w	cvfdt	ecvfdt	30	1900	67.72	71.45
1	50	65.02	71.45	31	2000	67.8	71.45
2	70	66.11	71.45	32	2100	68.1	71.45
3	90	64.19	71.45	33	2300	68.24	71.45
4	110	63.68	71.45	34	2400	68.2	71.45
5	130	64.32	71.45	35	2500	68	71.45
6	150	64.32	71.45	36	2600	68.73	71.45
7	170	66.29	71.45	37	2700	66.76	71.45
8	190	65.05	71.45	38	2800	68.65	71.45
9	210	65.09	71.45	39	2900	68.65	71.45
10	230	61.9	71.45	40	3000	69.97	71.45
11	250	64.18	71.45	41	3100	69.64	71.45
12	270	64.18	71.45	42	3200	69.63	71.45
13	290	63.91	71.45	43	3300	69.78	71.45
14	300	63.75	71.45	44	3400	69.67	71.45
15	400	64.22	71.45	45	3500	69.62	71.45
16	500	68.16	71.45	46	3600	69.79	71.45
17	600	68.18	71.45	47	3700	69.79	71.45
18	700	67.33	71.45	48	3800	70.03	71.45
19	800	67.87	71.45	49	4000	69.99	71.45
20	900	68.05	71.45	50	4100	70.02	71.45
21	1000	68.12	71.45	51	4200	70.06	71.45
22	1100	68.28	71.45	52	5000	69.91	71.45
23	1200	68.466	71.45	53	6000	70.08	71.45
24	1300	67.05	71.45	54	7000	68.71	71.45
25	1400	68.94	71.45	55	8000	69.9	71.45
26	1500	67.7	71.45	56	9000	70.35	71.45
27	1600	68	71.45	57	10000	68.18	71.45
28	1700	68.07	71.45	58	15000	70.24	71.45
29	1800	67.54	71.45	59	20000	69.51	71.45

Table 5.2 accuracy values using dataset 2



Fig. 5.4. Accuracy-w graph using dataset1

S.no	w	cvfdt	ecvfdt	S.no	w	cvfdt	ecvfdt
1	50	55.64	71.45	31	2000	69.4	71.45
2	70	60.95	71.45	32	2100	70.23	71.45
3	90	65.87	71.45	33	2300	69.96	71.45
4	110	65.87	71.45	34	2400	69.6	71.45
5	130	65.87	71.45	35	2500	69.6	71.45
6	150	68.92	71.45	36	2600	69.6	71.45
7	170	64.68	71.45	37	2700	69.93	71.45
8	190	58.44	71.45	38	2800	71.38	71.45
9	210	68.95	71.45	39	2900	70.8	71.45
10	230	68.6	71.45	40	3000	70.87	71.45
11	250	68.6	71.45	41	3100	70.87	71.45
12	270	68.6	71.45	42	3200	70.87	71.45
13	290	68.6	71.45	43	3300	70.87	71.45
14	300	68.6	71.45	44	3400	70.87	71.45
15	400	68.45	71.45	45	3500	70.87	71.45
16	500	69.4	71.45	46	3600	70.87	71.45
17	600	68.77	71.45	47	3700	70.87	71.45
18	700	68.45	71.45	48	3800	70.87	71.45
19	800	68.7	71.45	49	4000	70.87	71.45
20	900	68.98	71.45	50	4100	70.87	71.45
21	1000	69.34	71.45	51	4200	70.87	71.45
22	1100	69.48	71.45	52	5000	70.87	71.45
23	1200	69.48	71.45	53	6000	70.87	71.45
24	1300	69.48	71.45	54	7000	69.88	71.45
25	1400	68.98	71.45	55	8000	70.87	71.45
26	1500	68.7	71.45	56	9000	70.87	71.45
27	1600	68.7	71.45	57	10000	70.87	71.45
28	1700	68.7	71.45	58	15000	70.55	71.45
29	1800	68.7	71.45	59	20000	70.81	71.45
30	1900	68.98	71.45	60	20000	71.37	71.45

Table 5.3 accuracy values using dataset 3



Fig. 5.5. Accuracy-w graph using dataset3

VI. CONCLUSION

The decision tree algorithms Hoeffding tree, VFDT, CVFDT, E-CVFDT are used for the classification of data. The Hoeffding tree has a drawback of using large memory and stacking up the old instances. VFDT algorithm overcomes some of the drawbacks of Hoeffding tree because it can differentiate between ideal attribute and wasteful attribute but it still cannot handle concept drift. CVFDT algorithm handles concept drift by using window mechanism and also increases efficiency by generating alternate sub-tree but the efficiency is reduced because of its general method of handling instances in CVFDT without considering the types of concept drift.

E-CVFDT algorithm handles all these problems as it considers the types of concept drift. The E-CVFDT algorithm provides better accuracy than CVFDT algorithm. E-CVFDT algorithm guarantees constant accuracy for a given data set with respect to changing window size. CVFDT algorithm increases its accuracy when the size of the window is increased but only crosses E-CVFDT accuracy when window size is considerable compared to the size of data set which is not economical in real time situations where data stream size is infinite.

VII. REFERENCES

- [1] Pedro Domingos, Geoff Hulten '*Mining High Speed Data Streams*' In the ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD) 2000.
- [2] Kiran Kumar Patel, *Review on data stream classification Algorithm*, International Journal of Conceptions on Electrical and Electronics Engineering, Vol. 1, Issue 1, Oct 2013.
- [3] Vishal, Dr Saurabh Gupta, An Overview of data mining, International Journal of Computer Science and Information Technology, vol.5 (2), 2014.
- [4] Tusharkumar Trambadiya, Praveen Bhanodia, A Comparative study of Stream Data mining Algorithms,
- [5] International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 3, September 2012.
- [6] Tatsuya Minegishi, Masayuki Ise, Ayahiko Niimi, Osamu Konishi . Extension of Decision Tree
- [7] Algorithm for Stream Data Mining Using Real Data.
- [8] Jaiwei Han, Micheline Kamber, Jain Pei, Data Mining concepts and techniques, third edition, Elsevier Publications.
- [9] Ch.S.K.V.R.Naidu, S. Devanam Priya Sudam Sekhar Panda, High-Speed Data Stream Mining using VFDT,
- [10] International Journal of Electronics Communication and Computer Engineering, Volume 3, Issue (1).
- [11] Albert Bifet and Ricard Gavald'a. Learning from time-changing data with adaptive windowing. In SDM. SIAM, 2007.
- [12] Bifet, G. Holmes, B. Pfahringer, and E.Frank. "Fast perceptron decision tree learning from evolving data streams". In Proceedings of the 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining, pages 299-310, 2010.
- [13] Brifet, G. Holmes, R. Kirkby, and B. Pfahringer. "MOA: Massive online Analysis", http:// moa.cs.waikato.ac.nz/, Journal of Machine Learning Research (JMLR), 2010.
- [14] Domingos P, Hulten G. "Mining High-speed Data Streams" [C]//Proc. Of the 6th International Conference on
- [15] Knowledge Discovery and Data Mining. Boston, USA: [s.n.], 2000: 71-80
- [16] Edith Cohen and Martin J. Strauss. Maintaining time-decaying stream aggregates. J. Algorithms, 59(1):19–36, 2006.
- [17] Francisco J. Ferrer-Troyano, Jes'us S. Aguilar-Ruiz, and Jos'e Cristóbal Riquelme Santos. Discovering decision rules from numerical data streams. In Hisham Haddad, Andrea Omicini, Roger L. Wainwright, and Lorie M. Liebrock, editors, SAC, pages 649–653. ACM, 2004.

- [18]G. Widmer and M. Kubat. "Learning in the presence of concept drift and hidden contexts". Machine Learning, 23(1): 69-101, 1996.
- [19] Gerhard Widmer and M. Kubat. Learning in the presence of concept drift and hidden contexts. In *Machine Learning*, pages 69–101, 1996.
- [20] Hulten G, Spencer L, Domingos P. "Mining Time-changing Data Stream" [C]//Proc. Of the 7th International Conference on Knowledge Discovery and Data Mining. San Francisco, USA: [s.n.], 2001: 97-106.
- [21]I.H. Witten and E. Frank. "Data Mining: Practical Machine Learning Tools and Techniques". Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, second edition, June 2005.
- [22] J. Gama, Pedro Medas, Gladys Castillo, and Pedro Rodrigues, "Learning with drift detection", Symposium on Artificial Intelligence, in SBIA, pages 286-295, 2004
- [23] J. Gama, R. Rocha, and P. Medas. "Accurate decision trees for mining high-speed data streams". In KDD"03, pages 523-528, August 2003.
- [24] Jo^ao Gama and Mohamed Medhat Gaber, editors. Learning from Data Streams: Processing techniques in Sensor Networks. Springer, 2007.
- [25] Jo^ao Gama and Pedro Pereira Rodrigues. Stream-based electricity load forecast. In Joost N. Kok, Jacek Koronacki, Ramon López de M'antaras, Stan Matwin, Dunja Mladenic, and Andrzej Skowron, editors, *PKDD*, volume 4702 of *Lecture Notes in Computer Science*, pages 446–453. Springer, 2007.
- [26] P. L. Bartlett, S. Ben-David, and S.R. Kulkarni. "Learning changing concepts by exploiting the structure of change". Machine Learning, 41: 153-174, 2000.
- [27] R.Agrawal and G.Psaila. "Active data mining. In Proceedings of the First International Conference on Knowledge Discovery and Data Mining, pages 3-8, Montreal, Canada, 1995.
- [28] Remco R. Bouckaert. Voting massive collections of bayesian network classifiers for data streams. In Abdul Sattar and Byeong Ho Kang, editors, *Australian Conference on Artificial Intelligence*, volume 4304 of *Lecture Notes in Computer Science*, pages 243–252. Springer, 2006.
- [29] R.Kirkby. "Improving Hoeffding trees". Ph.D. thesis, University of Waikato, November 2007.
- [30] Tsymbal. "The problem of concept drift: definitions and related work", Technical Report TCD-CS-2004-15, Computer Science Department, Trinity College Dublin, Ireland. 2004.
- [31] An improvement CVFDT method for Concept Drift Data stream. Gang Liu, Cai-xia Liu, Chengdu. IEEE 2013.
- [32] Pedro Domingos, Geoff Hulten, Dept. of Computer Science Engineering University of Washington." Mining High-Speed Data Streams".
- [33] D. Patil, Parag Kulkarni, PhD "Adaptive Classification Algorithm for Concept Drifting Electricity Pricing
- [34] Data Streams." International Journal of Computer Applications (0975 8887) Volume 63-No.2, February 2013.
- [35] Hong-rong Cheng ; Zhi-guang Qin ; Qiao Liu "An improving CVFDT method for concept drift data stream"

[36] Communications, Circuits and Systems (ICCCAS), IEEE 2013

Taxonomy of Security Attacks in Cognitive Radio Networks - A Survey

Muzzammil Hussain Assistant Professor and Corresponding author Department of Computer Science and Engineering Central University of Rajasthan Kishangarh 305817, India <u>mhussain@curaj.ac.in</u>

Suditi Choudhary Research Scholar, Assistant Professor and Corresponding author Central University of Rajasthan Kishangarh 305817, India <u>suditichoudhary@gmail.com</u>

Abstract—Cognitive Radio (CR) is an emerging radio technology in which the available spectrum is intelligently shared between the licensed and unlicensed users. In Cognitive Radio Networks (CRN), Secondary Users (SU) are allocated free channels and may have to vacate them when the density of Primary Users (PU) increases. In many instances the SU misbehaves and tries to either retain the channels or disrupt the communication. In this paper, we have identified and classified various attacks possible in CRN. We have observed that attacks on CRN are possible during two important phases (Sensing and Accessing) of its working process. The attacks during spectrum sensing phase are known as spectrum sensing attacks which force the Cognitive Engine to make false decisions and end up in allocating the spectrum wrongly, whereas attacks during spectrum access phase are known as spectrum access attacks and are intended to either gain the channels forcefully or to disrupt the ongoing communications. The impact of security attacks on CRN is studied and analyzed in this paper. Most potential attack (PUEA) on CRN is simulated and its impact is studied and analyzed. It has been found that in the presence of PUEA, the spectrum allocation to PUs decreases rapidly and unlicensed users utilize the spectrum illegitimately.

Keywords: Cognitive Radio; licensed user; unlicensed user; primary user; secondary user

I. INTRODUCTION

Due to the increasing wireless devices all over the world the demand for spectrum is increasing enormously but the spectrum is fixed for the different wireless networks and they can't operate in some other frequency hence quality of service (QOS) is degrading, as the availability for particular wireless networks becomes less say cellular networks, on the other hand in some networks the spectrum remains unused due to less traffic specially TV transmission. Therefore, to make proper utilization of unused spectrum FCC gave permission to use the white space for all the other networks as that frequency was earlier reserved for TV transmission only which has nearly zero utilization due to digitalization of TVs. In that frequency TV will be the Primary User (PU) and any other device using that spectrum will be Secondary User (SU) and can use the empty spectrum for its transmission if there is no enough PUs occupying the complete spectrum without any interference to the Pus [11]. Mitola (2000) pioneered Cognitive Radio (CR) as a radio mechanism that has been derived from term Software Defined Radio (SDR), which can change its transmitter parameters dynamically according to the environment by adapting to it in order to communicate with each other which are intelligently controlled by software. The primary objective of a CR is to dynamically improve spectrum utility but they are neither field programmable nor have any software. Cognitive Radio Network (CRN) is a network of cognitive radios deployed to utilize the available spectrum at its best, incorporate reliability in communication and provide QOS. CRs are aware, adjustable, automatic operation, and adaptive to the environment. They follow spectrum management which includes properly managing spectrum sensing, decision, sharing and mobility of the users to other vacant spectrum bands. CRNs can be equipped and improve many areas like emergency management or disaster relief, search and rescue, mining industry, road traffic control, medicine, weather forecasting, military, etc [8].

II. SECURITY THREATS IN CRN

Threats [3][4][10] in CRN are the most possible instances that might exploit the vulnerabilities to violate security policies and degrade its performance. Here, we discuss various possible security threats in CRN shown in Fig 1.

D. Self-Propagating Behaviour

CRN have self-propagating nature, due to which the radios cooperatively function at same states through propagation of any change in state. This can lead to a serious attack, where an attacker can induce an attack state on one radio, due to self-propagation that attack state will be induced to its neighboring radios and so on. Finally, all the radios in the network end up in attack state and will malfunction.

E. Policy Radio Threats

In policy radio threats, the attacker manipulates the sensing signals, causing the radio to select a suboptimal configuration which may degrade its performance adversely.



Fig.1. Security Threats in CRN

F. Parameter Threats

In parameter threats, an attacker by emulating the signals of legitimate users tries to either access the operation parameters in CRN or forces cognitive engine to alter the parameters. These changes in operation parameters in CR may cause CR to work in suboptimal environment.

D. Learning Radio Threats

In this type of attack, the attacker easily convinces a radio to believe that counterfeited information is correct and this seriously impacts the behavior of radio in particular situations, leading to its malfunctioning.

E. Dynamic Spectrum Threats

Dynamic spectrum threats arise to the dynamic nature of CR in selecting, vacating, sensing, adapting and handoff of spectrum. In these attacks, an attacker can emulates the signals of PU and force the CR to wrongly identify spectrum holes and allocate them to illegitimate users. An attacker can also jam the common control channel and disrupt the sensing information.

An attacker can also seriously affect the spectrum decision and mislead the CR in the process of selecting the best available spectrum band to meet the QoS requirements. This may impact the communication or may be the CR ends up in selecting suboptimal band.

An intelligent attacker may simply jam the frequency of a PU and force it to either frequently do spectrum handoff or may also cause communication failure during spectrum handoff and this will adversely affect the communication.

III. VULNERABILITIES AND ATTACKS IN CRN

CRN works in a cyclic process first it senses the spectrum, then it understands the spectrum to detect unused bands, then it decides which frequency holes (unused bands) to be allocated and then it adapts to the environment by changing its parameters. Here, sensing the spectrum and adapting to it are different tasks and are done at different times. This time gap between these two operations gives ample scope to malicious nodes to attack the network [9].

Most of the attacks on CRN are possible and are executed during following two important phases of working of CRN: Spectrum Sensing and Spectrum Access defined further as shown in Fig. 2.

F. Spectrum Sensing Attacks

In this category of attacks, the attackers try to fool the base station during spectrum sensing phase. Generally, the attacker nodes propagate false signals that are sensed by the BS as legitimate signals these signals affect the decision phase and the BS ends up in wrongly allocating the free bands to illegitimate nodes. These attacks are generally executed to mislead the BS regarding available free frequency bands and identification of legitimate stakeholders of such bands. These attacks are executed by attackers with different objectives in different ways. Here, are few instances of this type of attack.

1) Primary User Emulation Attack (PUEA): In this attack a malicious SU emulates the signal features of a PU. The SU generates the signals which have similar parameters to that of a PU; thereby the spectrum is allocated to the attacker node. Here, a SU executes PUEA either to gain spectrum (known as selfish PUEA) or to disrupt other SUs from accessing spectrum (known as malicious PUEA) [2][5][12].



Fig. 2. Taxonomy of Security Attacks in CRN.

2) Objective Function Attack: In this attack a malicious node disturbs the channel during learning/sensing phase. So, that the optimized objective function of the radio is affected. Here, the attacker prevents the CR from adapting to change in its environment. The cognitive engine cannot feed that input to its objective function, which gives the best result for its application [4][5].

3) Spectrum Sensing Data Falsification Attack (SSDF): In this attack a malicious SU alters the local spectrum sensing data/reports and sends to the cognitive engine. Falsified local spectrum sensing reports adversely affects the decision phase and either the spectrum is allocated to illegitimate users or is left empty [2][3][5].

4) Control Channel Saturation DoS Attack (CCSD): In thus attack, performance of network is brought down by generating forged MAC control frames and saturating the control channel. When many MAC. Control frames are exchanged, then the channel becomes a bottleneck and it affects the channel reservation in negotiation phase. Thereby, the data channels remain unallocated wasting the spectrum [5].

5) Spectrum Sensing Data attack: This attack is similar to SSDF attack, where the spectrum sensing data is counterfeited and sent to Cognitive Engine which affects decision phase and incorrect allocation of bands to PU and SU. This may lead to further attacks to be executed on CR [2].

6) Incumbent Emulation Attack: In this attack a malicious SU emulates the signals of an incumbent to get priority over other SUs [3].

G. Spectrum Access Attacks

These type attacks are executed when the legitimate users are accessing their allocated channels to transmit their data. In these attacks the illegitimate nodes try to disrupt the ongoing communication process by either introducing Denial of Service or by capturing the channels is usage. Here are few instances of this type of attacks.

1) Cross Layer Attack: Network designers have come up with a concept known as Cross-Layer design, in which information is exchanged among different layers so as to enhance its efficiency and utilization. But, this concept is also being intelligently used by smart attackers, who are designing attacks that simultaneously attack protocols at multiple layers thereby bringing down the network completely and also all the layers are attacked it becomes a tedious task to detect and recover from such attacks [5].

2) Jamming Attack: In CR, a malicious node can transmit signals to the base station with the same frequency as that of a legitimate mode. The attacker competes with the PUs to access the same channels. This may lead to interference and the signals of legitimate PU get disrupted [5].

3) Hello Flood Attack: In this attack, the attacker node floods the networks with messages and claims to be neighbor of many of the nodes. The attacker encourages nodes to send its packets intended for a specific destination through it and drops them [2][5].

4) Lion Attack: Lion attack is executed to reduce the performance of network especially TCP. It is done by forcing frequency handoffs often frequently. The attacker frequently jams the signal thereby forcing the network to change its frequency. This brings down the throughput of CP as the nodes need to change their frequencies often frequently [5].

5) Sinkhole & Sybil Attack: In this attack, an attacker advertises itself as having the shortest route by tampering route reply packet. This causes a neighbouring node to forward its packets to the attacker node. The attacker node now can simply drop all the packets or can forward only selective packets. The attacker can also modify the packets passing through it. Sybil attack is the attack in which an attacker copies the entity of a legitimate user. Usually it does this by emulating the signals of a PU [2][5].

6) *Wormhole Attack:* Wormhole Attack is executed by two nodes where they form an insecure tunnel between themselves and convince legitimate nodes to send packets through them. They advertise themselves as having best route between a pair of nodes and causing the packets to be transmitted through them [2].

SIMULATION OF PUEA

The PUEA was introduced by changing the signal properties (Signal Strength, AoA) of few selected SUs to that some random PUs. The PUEA was simulated in Matlab, simulation environment is:

Area	1000m X 1000m
No. Of PUs	15
No. Of SUs	30
Signal Strength	0.5db
Node Distribution	Random
Node Movement	Random Waypoint

TABLE 1. SIMULATION ENVIROMENT

RESULTS AND ANALYSIS



Fig. 3. Time Vs Number of PUs/SUs.

Fig. 3 shows the graph plotted for number of PUs and SUs allocated spectrum over period of time. From Fig. 3 it is clear that initially as there are no attacker nodes the number of PUs allocated spectrum are more but over the time as the selected SUs start emulating PUs, the number of PUs allocated the spectrum is decreasing rapidly and at the same rate the SUs being allocated spectrum are increasing rapidly. It is because the SUs are emulating PUs and are getting allocated spectrum.

IV. CONCLUSION

CR is an emerging radio technology designed with the primary objective of minimizing the utilization of spectrum. The allocation of available free spectrum to the SUs may lead to many security attacks due to selfish or malicious behavior of SUs. It has been found that mostly the SUs execute attacks during two important phases (sense and access) of working of CRN. The security attacks during spectrum sensing phase are intended to force the cognitive engine to make false decisions and end up in allocating the spectrum wrongly, whereas security attacks during spectrum access phase are executed to either gain the channels forcefully or to disrupt the ongoing communication. We have simulated and studied the impact of PUEA on CRN, as PUEA is most potential attack on CRN, it has been found that when an attacker node executes PUEA on CRN, the spectrum allocation to PUs decreases rapidly and thereby most of PUs are kept waiting for the spectrum, which is being used illegitimately by unlicensed users.

V. REFERENCES

- K.-C. Chen, Y.-J. Peng, N. R. Prasad, Y.-C. Liang and S. Sun, "Cognitive radio network architecture: Part I-General structure," Proc. ACM ICUIMC, pp.114 -119, 2008.
- [2] H. Idoudi, K. Daimi and M. Saed., "Security Challenges in Cognitive Radio Networks," In Proc of the WCE, vol. 1, July 2014.
- [3] M. A. Butt, M. Zaman, "Cognitive Radio Network: Security Enhancements," Journal of Global Research in Computer Science, vol. 4(2), pp. 36-41, Feb. 2013.
- [4] T. C. Clancy and N. Goergen, "Security in cognitive radio networks: Threats and mitigation," IEEE CrownCom,vol. 3, pp. 1-8. IEEE, 2008.
- [5] S. Bhattacharjee, R. Rajkumari and N. Marchang, "Cognitive Radio Networks Security Threats and Attacks: A Review," Journal of Computer Applications, ICICT, pp. 16-19, 2014.
- [6] P. P. Bhattacharya, R. Khandelwal, R. Gera, and A. Agarwal, "Smart Radio Spectrum Management for Cognitive Radio," IJDPS, vol. 2(4), pp. 12-24, Aug. 2011.
- [7] H. C. van Tilborg, Encyclopedia of Cryptography and Security. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.

- [8] S. Parvin, F. K. Hussain, O. K. Hussain, S. Han, B. Tian, and E. Chang, "Cognitive radio network security: A survey," In Journal of Network and Computer Applications, Elsevier, vol. 35, pp. 1691-1708, 2012.
- [9] S. Bhattacharjee, S. Sengupta, and M. Chatterjee, "Vulnerabilities in Cognitive Radio Networks: A survey," *In Computer Communications, Elsevier*, vol. 36(13), pp. 1387-1398, 2013.
- [10] Y. Zhang, G. Xu, X. Geng, "Security threats in cognitive radio networks," In High Performance Computing and Communications (HPCC) IEEE, vol. 10, pp. 1036-1041, Sept. 2008.
 [11] A. Popescu, "Cognitive radio networks," IEEE COMM, vol. 9, pp. 11-15, June 2012.
- [12] S. Choudhary, M. Hussain. "Primary User Authentication in Cognitive Radio Network using Signal Properties." In Proceedings of the Third International Symposium on Women in Computing and Informatics, pp. 290-296. ACM, 2015.

A Survey on Association Rule Mining Using Genetic Algorithm

Shailza Chaudhary Department of Computer Science and Engineering Jaypee University of Information Technology, Waknaghat. HP, India <u>shailza1101@gmail.com</u>

Pardeep Kumar

Department of Computer Science and Engineering Jaypee University of Information Technology, Waknaghat. HP, India pardeepkumarkhokhar@gmail.com

Abstract—Data mining technology has emerged as a method of distinctive hidden patterns and trends from massive volume of information. The general goal of data mining method is to extract information from an information set and remodel it into a noticeable structure for more use Association rule (AR) mining could be a data processing task that discovers interesting relations between variables in information. It's supposed to spot robust rules discovered in databases mistreatment different measures of interest. This paper presented a survey of assorted association rule mining approaches. Many factors are thought of for creating comparison and additionally description concerning varied knowledge is given which will give raised potency also as correct results.

Keywords: Survey, Data mining, association rule mining, genetic algorithm

I. INTRODUCTION

From recent number of year explosive growth in quantity of knowledge or data get noticed. Knowledge can be easy numerical figures, multimedia system knowledge, web data, text knowledge and abstraction knowledge, is additionally being hold on in files knowledge bases and data repositories. Finding omnipresent model in a very great deal of knowledge is one amongst the key issues. For this explicit reason knowledge mining is attracted by info business and therefore the world and is needed to show knowledge into helpful info and data.

Data mining is the method of winning the specified info from giant databases. Extracted info is employed for various areas like business analysis, consumer maintenance, characteristic the frauds, and scientific discoveries [1]. There exist various models of knowledge mining like classification, clustering, call tree and neural networks from that association rule mining is additionally a vital model. Association rules area unit accustomed extract the frequent patterns or casual structure among the set of things from given information. The pattern and rule discovered area unit supported the bulk of ordinarily continual items in dataset. Today Association rule mining is generally utilized in many alternative areas like telecommunication networks, market and risk management, internal control mobile mining, graph mining, academic mining, etc. the standard application of association rule mining is market basket analysis that considers the shopping for habits of consumers. Typical example of super market with sizable amount of transactions for association rule mining is: bread \rightarrow jam [sup=10%, conf=80%], 100 percent support states that of client purchase bread and jam at the same time, and 80% confidence means that eightieth customers purchased bread additionally purchase the jam. Support and confidence are 2 basic parameters of association rules to come up with the attention-grabbing association rules. To discover the attention-grabbing association rules domain specialists specifies the minimum support (minsup) and minimum confidence (minconf) from given set of transactions item sets are called interesting if have greater support and confidence from minsup and minconf. To mine the association rules firstly find all the item sets having specified threshold support, secondly generate association rules from these item sets.

Various works have been proposed so far that promises to extract information in form of association rules. A rigorous survey of various such algorithm and comparison between them is summarized in the next section.

II. LITERATURE SUVEY

The study of various algorithm available for association rule mining has been divided in 3 categories i.e. classical approaches for association rule mining, association rule mining using genetic algorithm for categorical data and association rule mining using genetic algorithm for numerical data.

H. CLASSICAL ALGORITHM FOR ASSOCIATION RULE MINING

1) AIS

The AIS algorithm was the very first published algorithm for mining association rules by generating all large itemsets in a transaction dataset [1]. This technique only generates rules with only one item in the consequent part. That is, the association rules generated by this process are in the form of $X \Rightarrow I_j \mid \alpha$, where X is a set of attributes and I_j is a single attribute in the set I, and α is the confidence parameter of the generated rule.

2) SETM

SETM algorithm was proposed by [5] with aim of using SQL for obtaining frequent itemsets. In this algorithm each member of the large itemset keeps information of TID i.e. unique ID of the record as <TID, itemset>. Further process of this algorithm is similar to that of AIS.

But shortcoming of this algorithm is also the same as AIS i.e. it also generates large number of candidate sets and require multiple read scans of dataset and arranging of data again and again.

3) APRIORI

The apriori algorithm presented by [2] was the major milestone in association rule mining area. The work presented in this uses the axiom that any subset of a frequent itemset must be a frequent itemset. Hence it generates the candidate itemsets of next pass by merging the large itemsets of the previous pass and eliminating all small subset of previous pass as well. Therefore it can be concluded that apriori performs much better than the previous two algorithm as it works with much less candidate sets and is much faster and efficient.

4) CARMA

CARMA stands for Continuous Association Rule Mining Algorithm [6] works with the concept of computation of frequent itemsets online. Bring them online, CARMA directly shows the present association rules to the user of dataset in real time which also authorize the user to make relevant change in the parameters as per the need such as minimum support or minimum confidence, for any transaction at the first scan of the dataset. It requires at most 2 dataset scans. It generates the itemsets in the prior scan and does the counting work for all the itemsets in the next scan.

All these algorithms and many others published in the past, no doubt perform well in generating association rules which are accurate but they are computationally intensive and most of them generate much more rules than needed. These algorithms also don't consider numeric values of the attributes so, to make things easier we thought of studying mining algorithm which uses genetic algorithm as there basis. Because genetic algorithm is meant for doing the computationally intensive task fast and more efficiently which is the need for rule mining in large databases. Some of earlier algorithm which uses this concept are discussed further.

I. ASSOCIATION RULE MINING USING GENETIC ALGORITHM

Various algorithms has been proposed so far for association rule mining from databases using genetic algorithm. Based on type of association rules generated, these algorithm can be categorized as categorical association rule mining algorithm and numeric association rule mining algorithm.

1) Categorical Association Rule Mining Algorithm

Categorical association rules are the IF-THEN rules generated from a categorical or a binary dataset.

In a binary dataset, a rule like $A,B,C \Rightarrow D,E$ signifies that:

If items A, B, and C are bought together, then items D and E will also be purchased. But, these rules do not indicate anything about the number of items that can be purchased; these rules simply indicate that whether the attribute is present or absent from the rule.

For applying GA for finding such rules, four main designing components are needed i.e. Chromosome encoding for representing rules in form of GA understandable chromosomes, Crossover operator for generating new child chromosomes, Mutation operators for introducing diversity in the population and fitness function which will evaluate fitness measure of each rule. After deciding these factors only thing left is to apply basic genetic algorithm which is defined in introduction part above, using all these components. Various algorithms discussed based on such bifurcation:

1.1) Chromosome Representation:

Broadly there are two chromosome representation techniques are present in literature for categorical ARM. In the first approach (Pittsburgh approach), a batch of possible association rules are encoded in a chromosome. This encoding approach is well suitable for classification rule mining, where the goal is to generate a good quality set of rules. However, in ARM, the main goal is to find a set of rules were each rule is good in itself. Therefore, for ARM case, the Michigan approach in [3] is mostly used, in Michigan approach each chromosome represents exactly one rule in itself, is more suitable. Most of the ARM techniques use Michigan chromosome representation.

In an early work in [7], the author uses the Michigan approach as follows: each chromosome has length 2k, where k is the number of items. The chromosomes in this case were binary strings where each attribute is given two bits. If these two bits are 00 or 11, then the attribute is present in the antecedent part or consequent part of the rule, respectively, else the attribute is not present in the rule. Likewise in paper [9], each chromosome has two parts. The first part indicates the location of the attribute in the rule and the other part indicates the categorical value it has. The prior part consists of two bits where the attribute appears in the antecedent part or the consequent part of the rule, if the bits are 10 and 11, respectively; else, it is meant to be absent from the rule. The other part represents the categorical values carried by attributes in binary form. However, the authors in the paper did not given any justification of how the binary value of the attribute in the second part will appear and how categorical state will be managed if the number of states for an attribute is not an exact power of two.

The main demerit of choosing a binary encoding scheme is that the length of the chromosome is large when the number of attributes increases, because at least two bits are required for each attribute representation. An integer encoding can be used as a solution to this problem.

An integer encoding scheme has been used in association rule mining using multi objective genetic algorithm (ARMGA) in [9], in this work the chromosomes encode the index of the attributes. A chromosome represented as encoding a *k*-rule, *k* is the total number of attributes in the antecedent part and the consequent part of the rule, and this chromosome will have k + 1 genes. The first gene position represents the differentiating position of the chromosome where the attributes of antecedent and the consequent are separated. For example, if *Ai* represents the *i*th item, then the chromosome $\{3 \mid 5 4 2 1 3\}$ represents the rule *A2*, *A5*, *A4* \Rightarrow *A1*, *A3*. This kind of representation reduces length of the chromosome significantly.

1.2) Objective Functions:

Even though support count and confidence are two well-known objectives that are generally to be maximized, there are several other parameters available to measure the interestingness or strength of association rules. Some of the parameters, which can be used by different algorithms for optimization in a multi-objective scenario are comprehensibility, conviction, interestingness, performance, lift, coverage, precision etc. Each function has its own significance and combination of best of these functions gives best result in the end.

1.3) Evolutionary Operators:

Mostly, when binary encoding scheme is used, some of the standard crossover and mutation operators are used. For example, in in [7], bit-flip mutation and multipoint crossover have been used. In [10] mutation operator, bit flip has been adopted, however, the authors did not specifically mention which crossover operator should be used.

In [11], where integer encoding for the chromosomes is used, an order-1 crossover technique is taken in consideration. In this technique, first a segment is selected from any two parent chromosomes and these are copied to the two child chromosomes. Next, starting from the right side of the segment, the values of the genes that didn't appear in the selected segment of the first parent are copied to the first child. The same procedure is repeated for the second child as well. The mutation operator introduces diversity by replacing a selected attribute value from the chromosome with a random attribute value not present in the chromosome currently.

1.4) Final Selection:

After appropriate iteration of the algorithm the set of rules generated are treated as final selected rules. On the basis of all the above parameters the interestingness and importance of rules depends.

2) Numeric Association Rule Mining Algorithm

Numeric association rule can be represented mathematically in the form: $(l1 \le A1 \le h1) \land (l2 \le A2 \le h2) \Rightarrow (l3 \le A3 \le h3).$

Here *Ai* represents the *i*th attribute. *li* and *hi* represent the lower and upper bound of the attribute values, respectively. Thus, [*li*, *hi*] defines an interval of values for the attribute *Ai* with lower bound and upper bound.

2.1) Chromosome Representation:

The chromosomes are needed to encode the lower and upper bounds of the intervals of the attributes participating in a rule representing for numeric or quantitative association rules. In [8], proposed work MODENAR used chromosomes where each attribute has three part. The first part tells whether the attribute is present or not in the rule, and if present, in which part of the rule (antecedent or consequent) it is present. The second and third part tells the lower and upper bounds of the value of attribute. The first part can have integer numbers such as 0, 1, or 2, which represents the occurrence of the attribute in the antecedent part of the rule, the occurrence of the attribute in the consequent of the rule, and the absence of the attribute from the given rule, respectively. Next, the second and the third part can take real numbers from the related attribute range. Further it is to be noted that as MODENAR uses differential evolution as an optimization technique and works on real-valued chromosomes, the authors considered a round off operator for handling the integer value part of the chromosome. A similar encoding scheme is adopted in [12]. The only difference is that in this case, the prior part of the chromosome, instead of taking the values as 0, 1, 2, adopts the values 0, 1, and -1, respectively, to denote the same meaning.

In [14] also a three part chromosome is present where first number represents lower bound and second number represents upper bound of the attribute. With this there is a tag bit which represents position of attribute in rule i.e. if two tag bits are 00 then the attribute present in antecedent part of rule, if it is 11 then the corresponding attribute is present in the consequent part and if tag bits are 10 or 01 that represents the absence of attribute in rule. For example, A(00)(2.1)(3.5) B(11)(1.6)(4.5) C(01)(73)(73) represents if(A is between 2.1 and 3.5) then (B is between 1.6 and 4.5) and C is not present in rule.

2.2) Objective Functions:

MODENAR optimizes four criteria of the rules in [8]: support, confidence, amplitude of the intervals and comprehensibility, which make up the subparts of the rule. Comprehensibility is the measure that is used to forces the search procedure to get shorter rules, this is done due to the assumption that relatively shorter rules generates more non redundant information. They also suggested that the amplitude of the intervals should be smaller for interesting rules, but the reason for this is not explained.

In [12], three major interestingness parameters are simultaneously optimized: lift, comprehensibility, and performance. Where Performance is calculated by the product of confidence and support measure. Lift is calculated as the ratio of support count of the rule to the product of the supports of the antecedent and the consequent of the rule. A higher value of the lift measure signifies that the rule is more interesting, since its support is high with respect to the supports of its antecedent and its confidence. The comprehensibility can be defined simply as the reciprocal of the number of attributes in the rule. In [14], again three objective functions are discussed that are confidence, comprehensibility, and interestingness. First two objective was similar to above and third objective that is interestingness considered to find rules that are more interesting or useful for the user, rather than all possible rules.

2.3) Evolutionary Operators:

MODENAR in [7] uses the standard crossover and mutation operators introduced by the approach of differential evolution. Moreover, this approach also introduced a rounding operator which round's off the first part of the attribute which should have an integer value (0, 1, 2) for evaluating the objective function values. In [12], multipoint crossover is used as crossover operator. Further, the two parts of the chromosome uses two different mutation operators for mutation. In the first part, where the value of chromosome are -1, 0, or 1, a random value is choosen from the set $\{-1, 0, 1\}$ and it replaces the existing value. For the other part of the chromosome which consist of the lower and upper bounds of the chromosome, a random mutation is applied.

In [14], the attribute part of the rule i.e. (A, B....) remains fixed for the crossover or mutation step. K – Point crossover is used in this paper and for mutation, swapping of two tag bits is taken by using bit-flip mutation. This means that the existing bits pairs (00, 01, 10, 11) are randomly transformed into each other. Same is taken for the lower and upper value they are also changed randomly.

2.4) Obtaining Final Solution:

All in [8], [12] and [14] use a Michigan approach of rule mining by encoding one rule in one chromosome. Thus, the final nondominated set gives a set of numeric rules. Thus, there is no need to select any particular solution from the final non-dominate set. All the solutions will serve as the final selected rule set.

Many other algorithms have also been proposed so far in literature. Comparison of various such proposed works by different authors so far are summarized here in table 1 and table 2. Table 1 compares the various algorithms on basis of design components that are used in respective algorithm for generating association rules. Table 2 describes the specifications of each algorithm i.e. the key concept of the algorithms, shortcoming as observed by us and the remarks i.e. improvement that can be thought of for these algorithms

FITNESS FUNCTION	Confidence, Comprehensibilty, intersetinoness	Support,Confidence, Comprehensibility, amplitude of	Relative Confidence	3 at a time among Support,Confidence, Interest, cosine, Comprehensibility.	Support , Confiden0ce	Lift, Comprehensibility, Performance	Precision, Coverage, Generality	Confidence, Comprehensibility, Interestingness	Confidence*Support	Comprehensibility ,Interestingness,
MUTATION	Bit flip	DE/Rand/I	Probability based Random	Bit-Flip	Random Replacement	Random increase /decrease	Bit flip	Random tag mutation	Bit-flip	Random mutation
<u>CROSSO</u> <u>VER</u>	Multi-point	Single point	Two point	Not Mentioned	Order-I	Multi-point	Fixed Single Point	k-point crossover	Single Point	Multi point
ENCODIG	Binary (Michigan)	Mixed (Integer +Real)	Integer (Michigan)	Binary (Michigan)	Integer (Michigan)	Real valı (Michigan)	Binary (Michigan)	Real valued (Michigan)	Binary (Michigan)	Positional (Michigan)
TYPE	Categor ical	Numeri c	Categor ical	Categor ical	Categor ical	Numeri c	Categor ical	Numeri c	Categor ical	Numeri c
	Ghosh and Nath(2004) [7]	Alatas et. al. (2008) MODENAR[Xiaowei an et. al. (2009)ARMGA [9]	Anand et. al. (2009) [10]	Qodmanan et. al. (2011) ARMMGA	Martin et. al. (2011) NSGA-II-OAR	Vashishtha et. al. (2011) [13]	B. Minaei et. al. (2013) MOGAR [14]	Didel et. al. (2013) [15]	Diana Martin et.al.

Table 1: Comparison between various works on the basis of design

REMARKS	More automatic setup is required.		Strict fitness function required	·	To be generalized for finding variant length rule.	Can be generalized for negative association rule		Better fitness function should be used	
SHORTCOMINGS0	Require user's Involvement	Increase in no. of attributes and there distinct values may cause problem	No.of rules generated are more	Increase in no. attributes could problematic	I	Increase in no. of attributes would be problematic	Rough Intervals	More no. of rules generated . Increase in no. of attributes could be problematic.	Increase in no. of attributes could be problematic.
SPECIFICATION	Less rules generated	Simple and easy to implement. Intervals are selected on the go. Database independent.	No Minimum Support required. Fast in execution time. System Automation. Database independent	Better if rule length is smaller	No min. support or confidence is required from user. Flexible on changing fitness. Best population is generated in each generation.	Good trade-off between Interpretability and accuracy. Evolutionary Learning of Intervals of attributes	Steps are taken to remove weak association rules.	Easy Concept	Mine Positive and Negative QAR. Low computational cost. Good Scalability
ALGORITHM	Ghosh and Nath(2004) [7]	Alatas et. al. (2008) MODENAR [8]	Xiaowei an et. al. (2009)ARMGA [9]	Anand et. al. (2009) [10]	Qodmanan et. al. (2011) ARMMGA [11]	Martin et. al. (2011) NSGA-II-QAR [12]	B. Minaci et. al. (2013) MOGAR [14]	Didel et. al. (2013) [15]	Diana Martin et.al. (2014)MOPNAR[17]

Table 2: Specification ad shortcoming of various algorithms

III. CONCLUSION AND FUTURE SCOPE

This paper reviewed the progress of research on association rule mining. In this paper we have presented the review of traditional algorithm used for association rule mining later we have discussed about various algorithms developed for association rule mining using genetic algorithm. After survey it is found that genetic algorithm is a suitable machine learning algorithm which can be used for association rule mining. Still there is a scope of improvement in these algorithm regarding scalability, speed and better rule generation.

VI. References

- R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," in Proc. of the Twentieth International Conference on Very Large Databases, Santiago de Chile, Chile, 1994.
- [2] R. Agrawal, T. Imielinski and A. N. Swami, "Mining Association Rules Between Set of Items in Large Databases," in ACM SIGMOD International Conference on Management of Data, Washington D.C., 1993.
- [3] D. E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison- Wesley Publishing Co., Inc., 1989.
- [4] C. Borgelt, "Efficient implementations of Apriori and Eclat," in Proc. Workshop Freq. Itemset Mining Implement, Melbourne, FL, 2003.
- [5] M. Houtsma and A. Swami, "Set-Oriented Mining for Association Rules in Relational Databases," in Proc. of the 11th IEEE International Conference on Data Engineering, Taipei, Taiwan, March 1995.
- [6] C. Hidber, "Online Association Rule Mining," in Proc. ACM SIGMOD International Conference on Management of Data, Philadelphia, PA, USA, June 1999.
- [7] A. Ghosh and B. Nath, "Multi-objective rule mining using genetic algorithms," Journal of Information Science, Elsevier, vol. 163, pp. 123-133, June 2004.
- [8] B. Alatas, E. Akin and A. Karci, "MODENAR: Multi-objective differential evolution algorithm for mining numeric association rules," *Journal of Applied Soft Computing, Elsevier*, vol. 8, no. 1, p. 646–656, January 2008.
- [9] X. Yan, C. Zhang and S. Zhang, "Genetic algorithm-based stratergy for identifying association rules without specifying actual minimum support," *Journal of Expert System Application, Elsevier*, vol. 36, no. 2, p. 3066–3076, January 2009.
- [10] R. Anand, A. Vaid and P. K. Singh, "Association rule mining using multiobjective evolutionary algorithms: Strengths and challenges," in *Proc. of Nature & Biologically Inspired Computing*, Coimbatore, India, December 2009.
- [11] H. R. Qodmanan, M. Nasiri and B. Minaei-Bidgoli, "Multiobjective association rule mining with genetic algorithm without specifying minimum support and minimum confidence," *Journal of Expert System Application, Elsevier*, vol. 38, no. 1, p. 288-298, January 2011.
- [12] D. Martin, A. Rosete, J. Alcala-Fdez and F. Herrera, "A multiobjective evolutionary algorithm for mining quantitative association rules," in *Proc. of 11th International Conference on ntelligent Systems Design and Applications*, Cordoba, pp. 1397–1402, November 2011.
- [13] J.Vashishtha, D. Kumar, S. Ratnoo and K. Kundu, "Mining comprehensible and interesting rules : A genetic algorithm approach," *International Journal of Computer Applications*, vol. 31, no. 1, pp. 39-47, Oct 2011.
- [14] B.Minaei-Bidgoli, R.Barmaki and M. Nasir, "Mining numerical association rules via multiobjective genetic algorithms," Journal of Information Science, Elsevier, vol. 233, pp. 15-24, June 2013.
- [15] A. Didel and Poonam, "Association rule mining using genetic algorithm," International Journal of Science Technology & Management, vol. 1, no. 5, pp. 61-67, 2013.
- [16] B. Al-Maqaleh, "Discovering Interesting association rules: A multi-objective Genetic algorithm approach," International Journal of Application in Information Science, vol. 5, no. 3, pp. 47-52, February 2013.
- [17] D. Martin, A. Rosete and J. Alcala, "A new multiobjective evolutionary algorithm for mining a reduced set of interesting positive and negative quantitative association rules," *IEEE Transaction on evolutionary computation*, vol. 18, no. 1, pp.

Ingressing Embedded System through RC Cars Working in Wireless Network

Prachi Jaiswal Acropolis Institute of Technology and Research. Indore (M.P) Email: <u>prachijaiswal497@gmail.com</u> Sarita Rathore Acropolis Institute of Technology and Research. Indore (M.P)

Shubhangi Jain Acropolis Institute of Technology and Research. Indore (M.P)

Abstract—"Intelligence Inside" means the master is performing, guiding the plan from Inside. The most relevant feature is to develop any such technology which are least depended under human conveyance. Thus embedded systems are inleted to technology and become one of the prominent feature in today's faster environment. Such system are mainly consist of software, hardware and network, all three comprise of their individual specifications. In this paper, we give an overview of such systems and also briefly describes real-time application with RC car working in wireless network. The analogous term comes with embedded system is automation ,dedicated to operate or control a process by highly automatic means mainly through smart system reducing human intervention to minimum.

These paper outlets a application of embedded system through RC cars working in wireless networks. This shows synergies between hardware (Audino, RF circuit), network (Wi-Fi), and Software through this car. The paper highlights the future scope of Real-world cars that could achieve the concept of automated driving system. The future from these 21st century stands to be more Hi-Tech where 90% resource would be invested in virtual-sovereignty.

Keywords: Microcontroller (Arduino), Embedded java, TCP/IP protocols.

I. INTRODUCTION

Intel introduced the world's first microprocessor where a Japanese company Busicom designed an integrated chip that could sustain an embedded system [6]. The embedded systems are sudden evolution and its usage, has been seen from last few decades. As name suggests these system are stew of both hardware and software components, where interaction between then makes a real embedded system with most prominent real life application.

We have till many such system implemented on very minute scale .Thus there is a need to scout embedded system more deeply as, Dashboard navigation system, personal digital assistance (PDA). Certainly those consisting more potential, creativity would only be able to progress this system more to new extends.

II. LITERATURE REVIEW

Embedded system are still apprised show in paper [7] as at this stage a creative approach is required to upgrade Embedded system level. Published work by[1][18] on Infrared Controlled RC cars has drawback of distance as, it works only in line of sight. Similar working with varies technologies to provide efficient and distance covering feature is show in paper [2][5], they highlights implementation of RC Robot car through GSM using DTMF signals with the usage of ATmega8 microcontroller. In this work the signal is processed by ATmega8 and decoded by DTMF MT8870. The emerging trends in ride to make embedded system more versatile home automation system is delivered in script[9]. Whereas [8] uses Bluetooth/WLAN mode of communication to transfer signal, [11] provides consistent usage of E-Controller having(8 channel, 32 ports). Emphasis, a special mode of communication through GPRS(TCP/IP embedded) between Modem and server[10]. While journal[4] describes Zigbee and Arduino communication module. Meanwhile [7] provides the futures huge aspects that will overrides present system with technologies like Zigbee.

III. DEEP INTO THEME

Embedded system are mainly dwell of processing radio signals, through embedded hardware and software. They bring imagination to reality; make hardware to work accordance with human intervention, senses and knowledge. The main disparity in this paper is to take embedded system to new pitch, solstice the technology to new high edges

A. NETWORKING

NEED OF COMMON MODE OF TRANSMISSION

For interaction between hardware and software, it is necessary that they both should have a common medium to interact. So it is compulsory that they must follow same set of protocols else the flow of data would be more complicated which would reduce the system's efficiency. The instruction generated by the software's that implements TCP\IP protocol, the hardware must catch on TCP\IP protocols. Hence the hardware programming would be based on TCP\IP protocol. The primary purpose behind following same set of protocols on both hardware and software side is to have efficient data communication. The major advantage of this decision is that the dataflow would be controlled and probability of getting error would be reduced.

WIRELESS NETWORK

Embedded system are the most imposing facet for future which tends such system ratify wireless mode of communication to follow ,Today we are mostly concerned with Radio frequency as mode of data transfer. Thus an analysis is required to asset best of them, below shows the comparison among them.

INFRARED TECHNOLOGY

- Only support touch-touch connectivity.
- Slow speed
- Low range of data is support
- No interface with hardware
- Only peer to peer connectivity.

BLUETOOTH

- Only line of sight connectivity.
- More battery consumption.
- Less interface with hardware.

Thus, found souped up technology for embedded system, and that leads to inclusion of RADIO FREQUENCY (RF) in embedded system with overcomes the precedent challenges as by providing better communication over large network, less battery consumption, more interface with hardware etc.

a) B LANGUAGES

C has become the prominent language while implementing any embedded system. Great software needs more perfect language, it is startling knowing that one of such language is C that has proven itself appropriate for both 8- bit and 64- bit processors[6]. It's snap and small approach to program an embedded system is addition in C , It has the benefit of processor , which grant programmers to peruse on algorithms , rather than on the details of a language.

OTHER EMBEDDED LANGUAGES

There are also languages, which can be used for embedded such as – assembly, C++, and Ada – are worth mentioning in greater details[6].In last few decades, embedded software were written in the assembly language to spot processor. The assembly provides programmers total curb to processor and hardware components, all these with a cost. Assembly consist of many disadvantages, as programmer has to birr upto the processor level to instruct .C++ is an object-oriented superset of C that is increasingly popular among embedded programmers[6].Its features are dominated over C and very supportive to software developers.

A software engineer who spends their time in software development would learn nothing about hardware. Thus software engineers must often understand the hardware in order to develop embedded system. Embedded systems are built with semiconductor parts called chips that are purchased from manufacturers [6]. Chips are microprocessors or micro controllers, which are clock-driven semiconductor devices uses electronic l circuits[19]. It performs read binary instructions from a storage device, accept input, process data, instructions and provide results as output[6].

C. EMBEDDED HARDWARE AND SOFTWARE

EMBEDDED HARDWARE

Embedded systems need a microcontroller, and kinds of microprocessors that are used quite divers. Lists of some of the common microprocessor families are: the Zilog Z8 family, Intel 8051/80188/x86 family, Motorola 68k family and the PowerPC family [6].

Embedded system needs memory for storing data as well as to store program .As normal execution store program and data at same place, the Embedded system store them at different spaces ,just because Embedded system does not have hard drives, and programs always executed from memory itself. Thus uses ROM or Read Only Memory. This system fortunately consists of such memory that can programmed as needed. Main problem persists as data cannot be stored in such diverse memory, thus it needs a formal storage

New concerns in such system are relevant on the physical component it is controlling. Very often these systems have a standard serial port, a network interface, input/output (I/O) interfaces, or hardware to interact.



Figure 1.FUNCTIONAL DIAGRAM [6]

Address bus is used by processor to refer a specific memory allocation within ROM more over to a concerned peripheral chip shows fig1. Data buses are used for transferring data between the processor and memory or I/Ol devices. Control bus is for timing the signals (synchronize the flow) between the ROM or peripheral devices.

EMBEDDED SOFTWARE

Embedded software developer needs bizarre level of hardware curb without enduring the profits of high orientation languages. Compilers and cross compilers are basic needs to plays a prominent role in such systems. Source code provided to embedded system needed to be convert into an executable file that in term can be loaded onto a ROM chip. This loading process consists of three specific steps, and the source code runner is called the host device shows in fig (2).

- Firstly the source file that is designed as embedded software must be complied or assembled into system understandable object file.
- Second, all such object file produced after compiling or assembling are called as relocatable program.
- Third, it consist a code that contain a executable file that pouted into a ROM chip. These ROM,I/O devices and software inclusion makes an embedded system.



Fig. 2. SOFTWARE WORKING [6]

D. DEBUGGING

The debugging feature of software and hardware is only detectable in execution of an Embedded system. Mostly it's not possible to audit the errors. Thus desired tools are needed to get inspected inside the device and spot the register values, data, memory locations and instruction flow.

Logic analyzers help out the hardware engineers by testing the robustness, Electrical connections and circuitry [3]. On the software side, emulators enable software engineers to deal with bugs.

E. EMBEDDED DEVELOPMENT ENVIRONMENT

Embedded system hardly consists of any keyboard, screen, disk drives or any I/O devices that are basic of designing any application. Therefore programming for such systems is done on a complete computer with all programming tools called a host. Embedded Software that's programmed, compiled, assembled and linked is it send to the purpose hardware.

Software must be compiled, linked, relocated and porting the executable object file into the ROM, needs to debug the Software.

IV. REAL TIME APPLICATION THROUGH RC CARS

This section focuses on the design, analysis of a RF-based remote controlled Robotic Car that can be run in a WI-FI environment .As a demo model that could be implemented in real world car with more powerful and specific hardware and software design .These show an experimental approach . In these connected world. It's rare to see people without their smart phones. So, it shows a way to automakers to get inside to launch a new era of connected automobiles that can serve up their own WiFi hot-spots[14].

Deep seated, an addition antenna uses 4G, that improves signal as compared to a portable hot-spot inside a vehicle [15]. This section is dedicated to future vision of automobile industries which mainly focuses on indulging new technologies with automotive parts.

A. HARDWARE

These parts of script described the basic hardware needs while implementing these demo models.

•	Multimeter	Pliers
•	Arduino	Soldering Iron
•	WI –FI shield	Solder Wick
•	RJ45 ends	Screwdrivers

- Motor control H-Bridge
- Caps[16]

PCB etching supplies Copper Clad Boards

Arduino is an automotive hardware and software company [19] that analyses and manufactures microcontroller for embedded systems and others. Arduino have sets of digital and analog I/O pins that can be interfaced to shields and circuits. It features serial communications interfaces, including USB, for loading programs from personal computers [17]. For programming such kits, the Arduino platform provides an integrated development environment (IDE) based on the Real time applications, which includes C++ and Java programming languages[19].

B. SOFTWARE HARDWARE ORIENTED

Arduino platform comes with multiple features, both to maintain its hardware, software add ups. Thus this platform comes with a dynamic IDE called (ARDUINO IDE) which is the most grounded feature to program such circuits. This IDE provides very efficient interface to program a hardware device with many built in function and schedules. That makes hardware to work in accordance with software.

CONTROLLER SOFTWARE

In accordance to control such system, specific segment need to be designed and this are desktops applications, smart phones (android, iPhones) application which have a unique program regarding their embedded system requirement. The controller segment is the only that includes all stipulation about system that includes following term:

- Transfer Protocols.
- Port Connection.
- Accomplishment task.
- Debugging.

The controller must define Transfer Protocols that is how instruction is transfer from controller to Arduino in Wi-Fi, because this is the only mode of communication and there should be proper instruction that transfer data without losing them or without generating errors. Java is such programming tool which provides a user to let its data flow.

The next challenge is all about data exchange port. Ardunio need to get specified by controller from where it has to accept data (a lot data port also exist), and report while interacting with controller. Thus needs a facet that defines the specific port that will only be dedicated to arduino for its instruction to transfer and feedback reporting. Arduino is an automotive hardware and software company [19], that analyse and manufactures microcontroller for embedded systems and others. Arduino have sets of digital and analog I/O pins that can be interfaced to shields and circuits. It features serial communications interfaces, including USB, for loading programs from personal computers [17]. For programming such kits, the Arduino platform provides an integrated development environment (IDE) based on the Real time applications, which includes C++ and Java programming languages [19].

C.ALGORITHM

This Script consists of working algorithm of an:

- Assign port1 and port3 as input and port2 as output.
- Load port1 into register and load port3 into another register [12].
- Enter selection routine, under the routine input and test are done using routines. If 1 is pressed right motor starts. If 2 is pressed on the keypad, it leads to forward direction.
- If 3 is pressed key, left motor starts. If 4 is pressed on the keypad, left motor stops and right motor moves forward.
- If 5 is pressed motors stops.[12] If 6 is pressed on the keypad, right motor stops and left motor moves forward. If 8 is pressed on the keypad, both motor moves backwards.
- If any other key is pressed the motors, the motors will remain in their initial position[12].



Fig. 3. ALGORITHM FOR RC CARS [13]

D. COURSE OF ACTION

This script of Embedded java programming provides feature for programming a network with TCP/IP [12]. For more details on programming refer to:

 [1]http://www.instructables.com/id/How-to-build-an-Arduino-WiFi-4x4-with-Android Cont/?ALLSTEPS/CarControlerWindows.zip
 [2]http://www.instructables.com/id/How-to-build-an-Arduino-WiFi-4x4-with-Android-Cont/?ALLSTEPS/CarControllerAndroidSetup.zip
 The following shows the preceding to design real RF cars using arduino in WI-FI.

STEP 1: DESIGN MOTOR CONTROL CIRCUIT...

- This circuit design provides power control and Motor regulator for arduino.
- This has 7.2 power consumption for motor regulator, steering and arduino.
- This Script uses SN754410 dual h bridge for motor and steering, and this pin is direct soldering at end of RJ 45 into the shield.

STEP 2: DESIGN ARDUINO SHIELD CIRCUIT...

- A Arduino shield that will pass the pins to the WiFi shield and connect to the motor controller shield via the RJ45 Connector and 2 screw terminals[16.1].
- The pin RJ 45 is critical, if any wire gets crossed it will debugs.

STEP 3: TEST SOFTWARE WITH IP/PORT.

- This is the GUI that sends instruction to arduino which moves the system.
- The interface is designed mostly using Embeddd java, matlab.
- As all goes well, this embedded system would move according to controller in Wi-fi range

V. CONCLUSION

Embedded systems recrement differs significantly via as detailing and software designing. Particularly every system has specific hardware and software demands that they cannot relay in other system. Novel enthusiasms, a step in the right direction,

as it provides combination to hardware and software that are important for cost-potent embedded systems. However for successful future it may well need to increase scope, more practical approaches. Thus, an out of box approach is needed to increase the habitude of Embedded system i.e. more to imagination, fewer the recent

Thus, an out of box approach is needed to increase the habitude of Embedded system i.e. more to imagination, fewer the recent application.

Fewer application that are aspected to see in future are:

- ZigBee technology for home automation and security.
- Tumor recoginisation using Mat lab and PIC.

IV. REFFERNCES

[1]Dharmain c.c,"Design with microcontroller":IR Remote Controlled Car(PWM motor using ATmega8).

• JOURNAL PAPAER

BOOKS

- [2]Gupta,Sabuj s,Arman R.O Mohammad, S.H,Nahid H.S" Design and implementation of Mobile operated Toy Car by DTMF".
- [3] Kiran kranthi chikka, Reddy gv nagajuna "ELECTROCARDIOGRAPHY MONITERING SYSTEM"
- [4]Kuang-YowLian,Sung-Jung Hsiao and Wen-Tsai Sung:"Mobile Monitoring and Embedded Control System for Factory Environment".
- [5]Ranu kaushik and Renuka Singh"GSM Mobile control Robo car"(2013).
- [6]S.mythili, C.sunilkumar "FUTURE OF EMBEDDED SYSTEM FOR JACOFEAST 2K2" Tiruchy.
- STAND ALONE
- [7]. TECHINAL REPORT: Chan Chung Hoong: "Wi-fi Remote Control car via mobile device(2013)".
- CONFERENCE PAPER
- [8] Alheraish, A.; "Design and implementation of home automation system," Consumer Electronics, IEEE Transactions on , vol.50, no.4, pp. 1087-1092, Nov. 2004.
- [9] Alkar, A.Z.; Buhur, U.; , "An Internet based wireless home automation system for multifunctional devices," Consumer Electronics, IEEE Transactions on , vol.51, no.4, pp. 1169- 1174, Nov. 2005
- [10] Chen Chao; Meng Keqilao; Gao Muyu; , "The Remote Wireless Monitoring System Based on GPRS for Off-Grid Wind Turbine," New Trends in Information and Service Science, 2009. NISS '09. International Conference
- [11]. Ximin Zhang; Junding Sun; Lihua Zhou; , "Development of an Internet Home Automation System using Java and Dynamic DNS Service," Parallel and Distributed Computing, Applications and Technologies, 2005. PDCAT 2005. Sixth International Conference
- ONLINE RESOURCE
- [12] Algorithm:http://www.iiste.org
- [13]Algorithm_flowchart:_www.codeproject.com/Article/126859/rc_cars_control_programming
- [14]Automaker usher in new era of connected cars :http://www.12voltnews.com/2014/05/13/automakers-new-era-connectedcars.
- [15]Connectivity cars :a new generation with built in wi -fi hotspot:http://www.nbcnews.com/business
- [16] Design and Model:http://www.instructables.com/id/How-to-build-an-Arduino-WiFi-4x4-with-Android-Cont/?ALLSTEPS
- [17]Details of starterkit::http://arduinostarterkitonline.blogspot.com/2015/09/arduino-starter-kit-online.html
- [18]Robolab Technology: IR Controlled Robot:http://www.robolabtechnology.com/robots.
- [19]Working and programming arduino: https://en.m.wikipedia.org/wiki/Arduino

An Artificial Neural Network based Classification Approach for Mining Spatial Database of Epidemiology in India

Priyanka Tyagi SV Subhrti University Meerut (U.P)

Amit Asthana, S V Subharti University Meerut (UP)

Abstract—The rapidly growing data creates the necessity of knowledge discovery from data which leads to promising emerging field, called the data mining or knowledge discovery from database (KDD). Spatial Data Mining describes as a process of discovering previously unknown, but potentially useful patterns from spatial database. The process of data mining could be the integration of many things including machine learning, database system, statistics, and information theory. There are many studies available of data mining in relational and transactional database; the concept is in high demand to apply it in many other applicative areas like spatial database, temporal database, multimedia database, object-oriented database etc.

The proliferation of spatial information during last several years has attracted the policy makers to extract useful patterns from large spatial datasets[1][2][3]. The heterogeneous geographical distribution of HIV/AIDS epidemiology and varied spatial socio-economic factors led us to work out an intelligent model which is able to correlate the trend of spread of this disease in India. In this paper we empirically study the role of socio economic factors like migration rate[MGRT], ratio of female to male literacy[LTFM], average distance traveled by migrant/ bridge population[AIMD], human development index [HDI], gender development index [GDI], in populating the disease in India. An Artificial Neural Network based model has been developed which has correlated these spatial and non-spatial factors with the various spread pattern of the disease. The result of the model reveals an interesting pattern which in agreement with the report published by the government on the basis of the physical survey of various geographical locations.

Data Source for the Analysis:

The sources of data for the case study work are: National AIDS Control Organization (NACO), Ministry Of Statistics and Program Implementation (Govt. of India), National Health Profile 2010 by Ministry of Health and Family welfare (Govt. of India).

Key Words: Spatial Data Base, Spatial Data Mining, ANN-BPN Model, Oracle10g.

I. INTRODUCTION

Spatial data mining has become a new and powerful tool for efficient and complex analysis of very large geospatial database [4][5][6]. It puts emphasis on extraction of interesting and implicit knowledge such as the spatial pattern or other significant mode not explicitly stored in the spatial database [7][8]. The geographical attributes involved in spatial database is an important aspect for many applications [4][9][1]. The HIV sentinel surveillance obtains HIV prevalence data from antenatal clinics (ANC) and sexually transmitted disease (STD) clinics as well as from high-risk groups. The information contained in the sentinel surveillance database reveals that India has a heterogeneous HIV epidemic. The main objective of this research is to develop an intelligent model which is able to take into account the important socio economic factors and forecast the prevalence, growth or declining trend of the epidemic like HIV/AIDS at various geographical locations[4][5]. The knowledge given by the model can be used to perform spatial prediction that could help the policy makers to plan and monitor the impact of HIV prevention and care intervention program.

The conventional analysis techniques have been based on traditional statistics and multidimensional data analysis. The traditional analysis is performed by diverse method like basic statistics (average, variance, histogram etc.), regression and correlation [10][11][12]. Those methods apply to quantitative or qualitative analysis of data. There was no effort to develop any

learning model which is able to learn from the accumulated data and environment factors and predict the most vulnerable geographical location for the epidemic like HIV/AIDS[13][14].

II. METHODOLOGY

2.1 Hypothesis: Migrants, particularly the bridge population (unskilled labors, truckers etc), bear a height risk of spreading HIV infection, which results from the condition and structure of the migration process. Available evidence suggests that migration could be fuelling the spread of HIV epidemic in high-out migration states. The analysis of the recent sentinel surveillance data (2008-09) shows that out of the 0.12 million estimated new infections in 2009, the six high prevalence states accounted for only 39% of the cases, while it is accounted for 41% in the states having high out migration rate. Now this HIV prevalence data is combined with the thematic data relating to the location of states. It is hypothesized that location wise Human Development Index [HDI], Literacy Ratio female to male[LTFM], Average Migration Distance of migrant population [AIMD] and migration rate[MGRT] influences the HIV growth rate [IDIF] in a state. The statistical calculation about the correlation of these factors is shown in figure [1]

Correlations							
		HDI	LTFM	AIMD	MGRT	IDIF	
HDI	Pearson Correlation	1	456	980**	927**	.665	
	Sig. (2-tailed)		.256	.000	.001	.072	
	Ν	8	8	8	8	8	
LTFM	Pearson Correlation	456	1	.553	.493	191	
	Sig. (2-tailed)	.256		.155	.214	.650	
	Ν	8	8	8	8	8	
AIMD	Pearson Correlation	980**	.553	1	.939**	608	
	Sig. (2-tailed)	.000	.155		.001	.110	
<u>.</u>	Ν	8	8	8	8	8	
MGRT	Pearson Correlation	927**	.493	.939**	1	705	
	Sig. (2-tailed)	.001	.214	.001		.051	
<u>,</u>	Ν	8	8	8	8	8	
IDIF	Pearson Correlation	.665	191	608	705	1	
	Sig. (2-tailed)	.072	.650	.110	.051	1	
	Ν	8	8	8	8	8	
	**. Correlatio	on is significa	nt at the 0.01	level (2-taile	d).		
	F	ig. 1: Corr	elation Tal	ole			

Now the conventional statistical approach of developing a predictive model would not work properly as the relationship between the dependent variable (IDIF) and the independent variables HDI,LTFM,AIMD,MGRT in quite non linear. The Auto Correlation factor (ACF) graph of the dependent variable is shown in figure[2] and the tabular value is shown in figure[3].



Fig. 2: ACF Graph of IDIF

	Autocorrelations									
		Series:I	DIF							
	Box-Ljung Statistic									
Lag	Autocorrelation	Std. Error ^a	Value	df	Sig. ^b					
1	027	.296	.008	1	.927					
2	224	.274	.676	2	.713					
3	232	.250	1.534	3	.675					
4	066	.224	1.621	4	.805					
5	.016	.194	1.627	5	.898					
6	.025	.158	1.653	6	.949					
a.	The underlying pro	cess assumed	is independe	ence (white n	oise).					
	b. Based on th	e asymptotic c	hi-square ap	proximation.						
		Fig. 3: ACF Ta	abular value of	fIDIF						

2.2 Model: The Artificial Neural Network [ANN] model used in the study is the multi layer perceptron (MLP). For machine learning, the model requires a desired output which correctly maps input to output. It has got a three layer of architecture (Input, hidden and output layer) as shown in figure 4.



Fig. 4: Back Propagation Artificial Neural Network Architecture

Inputs are: 11: HDI, 12: LTFM, 13: AIMD, 14: MGRT Output Y : HIV prevalence [INFDIFF] Parameters of ANN : Nodes in input layer: 4 No. of hidden layer: 1 Nodes in hidden layer: 1 Nodes in hidden layer: 3 Weight assigned in input to hidden layer: 12 weights Weight assigned in each hidden to output layer: 3 weights Learning rate: 0.6

III. Model Implementation

The relevant data on various state wise non-spatial and spatial attributes have been collected from National AIDS Control Organization[NACO], Planning Commission, UNDP Human Development Report 2009, and Ministry of Statistics and Program Implementation Report October 2011 and stored in ORACLE 10g spatial database.

An overview of state wise data reveals three patterns of HIV growth rate e.g., increasing, stable and decreasing. Accordingly the states with three categories are shown in map of India in different color for different categories [Fig 5].



Fig 5: The three categories of states having different growth rate of HIV

Based on the outcome of the correlation analysis only statistically significant data on four inputs e.g., HDI, LTFM, AIMD and MGRT have been considered for machine learning and other socio economic data which were not so significant are ignored.

3.1 Computational Results:

Data have been filtered using sql query in three categories of states. Computer program for running ANN back propagation algorithm has been developed in PL/SQL and stored as a procedure. The SDO_GEOM function in the spatial database of ORACLE 10g computes the average distance of the migrants as one of the significant inputs to the ANN model. Artificial Neural Network is trained with input – output data for each category of states.

For the three categories of states, the mean square [MSE] error graph of the trained ANN model is plotted as shown in figures [6-8].







. .

IV. CONCLUSION

This research successfully demonstrates the application of spatial data mining for modeling the prevalence of HIV in India which is found to be significantly influenced by the four parameters e.g., Human Development Index[HDI], Literacy Ratio female to male[LTFM], Average Migration Distance of migrant population [AIMD] and migration rate[MGRT]. The potential of spatial database of ORACLE 10g has been fully explored in classifying the states on the basis of HIV prevalence rate using sql query. The machine learning process using back propagation algorithm of Artificial Neural Network has been successfully implemented with PL/SQL procedure developed on ORACLE database.

The mean square error of machine learning process is found to be reasonably low for three types of data sets as defined in section 3.1 above. The computational result corroborates our hypothesis set forth in the section 2.1. This will help the researchers and planners to work on large spatial dataset and help the policy makers in prediction process and to plan and monitor the impact of HIV prevention and care intervention program.

Waight s of Nodes	Weight for increasing class	Weight for decreasing class	Weight for stable class				
				I1W1	.1005744	.1062326	.1245625
				I1W2	.4015816	.4153782	.4738058
I1W3	.6002843	.6034905	.6108007				
I2W1	1264225	0582184	0946846				
I2W2	.4447993	.5708079	.5212602				
I2W3	2763843	2251907	2530306				
I3W1	.4824482	.56627	.5134122				
I3W2	0252463	.1361366	.0465647				
I3W3	.5263539	.5874757	.5506734				
I4W1	.2274652	.246863	.2250029				
I4W2	0086566	.0232134	0232614				
I4W3	.4088744	.4245377	.4111825				
H1W1	7913804	-1.050212	8845848				
H2W1	-1.4159216	-1.6709461	-1.5484145				
H3W1	5991215	8575985	718109				
ERR_SQ R	.00044	.00244	.00083				

Table:1

V. REFERENCES

- [1] S. Shekhar and S. Chawla, A Tour of Spatial Databases. Englewood Cliffs, NJ: Prentice-Hall, 2002. ISBN 0-7484-0064-6.
- [2] S. Schönfelder. Some notes on space, location and travel behaviour. In Swiss Transport Research Conference, Monte Verita, Ascona, 2001.
- [3] Fisher, J.B.; Kelly, M.; Romm, J. Scales of environmental justice: Combining GIS and spatial analysis for air toxics in West Oakland, California. Health Place 2006, 12, 701–714.
- [4] U. Ozesmi and W. Mitsch, "A spatial habitat model for the marsh-breeding red-winged blackbird (agelaius phoeniceus 1.)," in Coastal Lake Erie Wetlands Ecological Modeling. Amsterdam, The Netherlands: Elsevier, 1997, vol. 101, pp. 139–152.
- [5] R. Pace and R. Barry, "Quick computation of regressions with a spatially autoregressive dependent variable," Geograph. Anal., 1997.
- [6] Zeitouni K., "A Survey on Spatial Data Mining Methods Temporal, Spatial and Spatio-Temporal Data Mining, Lecture Databases and Statistics Point of Views", Int. Conf. IRMA, Notes in Artificial Intelligence no2007, Springer, September 12-16,2000, Lyon, France, pp102-114.
- [7] R. Agrawal, "Tutorial on databse mining," in Proc. 13th ACM Symp. Principles of Databases Systems, Minneapolis, MN, 1994, pp. 75–76.
- [8] K. Koperski, J. Adhikary, and J. Han, "Spatial data mining: Progress and challenges," in Proc. Workshop Research Issues on Data Mining and Knowledge Discovery, Montreal, Canada, 1996, pp. 1–10.
- [9] S. Shekhar, S. Chawla, S. Ravada, A. Fetterer, X. Liu, and C. T. Lu, "Spatial databases: Accomplishments and research needs," IEEE Trans. Knowledge Data Eng., vol. 11, Jan./Feb. 1999.
- [10] "Sparse spatial autoregressions," in Statistics and Probability Letters. Amsterdam, The Netherlands: Elsevier, 1997, pp. 291–297.
- [11] J. P. LeSage, "Bayesian estimation of spatial autoregressive models," Int. Reg. Sci. Rev., vol. 20, pp. 113–129, 1997.
- [12] R. S. Bivand, E. J. Pebesma, and V. Gomez-Rubio. Applied Spatial Data Analysis with R. Springer Series in Statistics. Springer, New York, 1st edition, 2008.
- [13] P. Elliot and D. Wartenberg. Spatial epidemiology: Current approaches and future challenges. Environmental Health Perspectives, 112(9):998{1006, Jun. 2004.
- [14] P. Elliott, J.Wake_eld, N. Best, and D. Briggs. Spatial Epidemiology: Methods and Applications. Oxford Medical Publications. Oxford University Press, Oxford, 2nd edition, 2000.

Constraint Based Sequential Pattern Mining: Survey and Current Research Challenges

Pisal T.B. Solapur University, Solapur: Department of Computer Science Sinhgad Institute of Computer Sciences Pandharpur(MS),India pisaltbresearch@gmail.com

Dr. Arjun P. Ghatule Solapur University,Solapur: Department of Computer Science Sinhgad Institute of Computer Sciences Pandharpur(MS),India arjun1671@gmail.com

Abstract—In the present scenario almost every system and working is computerized. All information and data are stored in computer. The capacity of data storage is increasing tremendously. Retrieval of data, hidden data and information are quiet difficult tasks. In data mining, sequential pattern mining is a great technological solution for extracting data, hidden data and information from database. In recent years, sequential pattern mining has been broadly applied to several application domains such as E-commerce, Medical Science, Information Technology, Telecommunication, Educational System etc. Since last two decades a numbers of algorithms have been developed for efficient mining. These algorithms generate large number of patterns for large database but users are only interested in small subset of such patterns. Mining complete pattern from database is difficult and also user is not interested in all patterns. Constraints are essential for finding user centered pattern in various applications. Constraint based sequential pattern mining. In this paper, a systematic survey of the constraint based sequential pattern mining algorithms is performed and current research challenges in data mining are discussed. At the end of paper, comparative analysis is done on the basis of constraints used in various algorithms.

Keywords: Sequential Pattern Mining, Constraint based Sequential Pattern Mining, Regular Expression, and Context Free Grammar.

I. INTRODUCTION

Constraint based sequential pattern mining finds only interesting sequential patterns from large databases. It finds out only interesting subsequences from a sequence database as patterns. A huge amount of data is continuously being collected and stored in the database. Many industries are interested in mining sequential patterns from their databases. Constraint based sequential pattern mining is one of the most well-known methods and has number of applications in the field of data mining such as e-commerce for customer analysis, medical record for disease analysis etc. In the trade business, sequential patterns can be used for mining patterns from the transaction records of customers. If a customer procures a laptop, the customer comes back to buy a printer next time. The retailer can use such information for understanding customer's interests and demands, in order to predict their needs. In the education field to extract patterns from source code and students teamwork data. Pattern mining can be used in the field of telecommunications for mining group patterns from mobile users, for customer behavior prediction, for predicting location of mobile in future for location based services and for mining useful patterns for mobile companies[1][2].

II. BASIC CONCEPT

The primary goal of constraint based sequential pattern mining is to determine only interesting sequences in the given database. Algorithms for this problem are mined sequentially where datasets are ordered. The problem was first introduced by Agrawal and Srikant[3] for market basket analysis to finding association between the different items that customers put in their shopping basket. From the last two decades, many sequential pattern mining algorithms were proposed, but not all having the same constraints. The use of some constraints has a great impact on mining, only interesting sequential pattern for efficiency. Some basic concepts are needed to introduce the problem.

J. Items

Items are the collection of items from database.

K. Itemset

An itemset is a non-empty subset of elements from given set.

L. Sequence

A list of item sets in sorted order is known as Sequence. If sequence is not enclosed in any other sequence then it is known as maximal sequence. If sequence has K items then sequence is called a k-sequence. If the sequence is S then the length of the sequence S is number of elements in sequence and it is denoted by |S|. If user-specified minimum support threshold of sequence is F, the number of possible sequences is S and the database density is d, then density is calculated as d=F/S.

M. Frequent Sequence

If the given database sequence is D, a few user-specified minimum support threshold σ and constraint K, then the sequence is frequent if it satisfy the constraint and support at least σ .

III. CLASSIFICATION OF CONSTRAINTS BASED ON SEQUENTIAL PATTERN MINING

The efficiency of mining the complete set of sequential patterns has been improved; still mining faces tough challenges in both effectiveness and efficiency. Large number of sequential patterns in large databases, users is interested in a small subset of patterns. Presenting the complete set of sequential patterns in large databases is difficult to understand and usage. To overcome this problem, Jian Pei, Jiawei Han and Wei Wang[8] have systematically presented the various strong conditions for solving these problems. In 2006 they mentioned seven categories of constraints:

N. Item constraint

An item constraint specifies subset of items that should or should not be present in the patterns.

O. Length constraint

A length constraint specifies the condition on the length of the patterns, where the length can be either the number of occurrences of items or the number of transactions.

P. Super-pattern constraint

Super-patterns are those which contain at least one of a particular set of patterns as sub-patterns.

Q. Regular Expression constraint

A regular expression constraint is specified over the set of items using the set of regular expression operators such as parallel (+), series (.) and closure (*).

R. Aggregate constraint

An aggregate constraint is the constraint on an aggregate of items in a pattern, where the aggregate function can be avg, sum, max, min, standard deviation etc.

S. Duration constraint

A duration constraint is applied only in sequence databases where each transaction has a time-stamp in every transition. It requires that the sequential patterns in the sequence database must have the property such that the time-stamp difference between the first and the last transactions in a sequential pattern must be longer or shorter than a given period.

T. Gap constraint

A gap constraint set is defined only in sequence databases where each transaction in every sequence has a timestamp. It requires that the sequential patterns in the sequence database must have the property such that the timestamp difference between every two adjacent transactions must be longer or shorter than the given gap.

A Out of these seven constraint some another constraints are also used for constraint based sequential pattern mining. These constraints are:

A. RFM (Recent, Frequent and Monetary)

The time span between the first and the last purchase in a customer sequence must be within a user-specified threshold[12]. This constraint can assure that the purchasing behavior implied by a sequential pattern must occur in a reasonable period.

B. Target-Oriented

A target-oriented sequential pattern is a sequential pattern with a concerned itemset at the end of pattern. For most decision makers, when they want to make efficient marketing strategies, they usually interested only in the happening order of a concerned itemsets. Most sequential patterns revealed by using traditional algorithms, which are extraneous and inoperable.

C. Context Free Grammar

For pattern mining, regular expression and context free grammar have been proposed to be used as a constraint to reduce the number of discovered patterns. Finite Automata (FA) accepted language is represented by a simple expression known as regular

expression. Regular expression is used to specify the constraint according to the specific syntax rules and it is called a pattern. Regular expression is used in many text based applications such as search and modifying text in text editors and utilities. It is also used in various programming languages like C, C++, and Java etc. In Java for string searching and modification regular expression is used as package. Some constraints are not regular and we cannot construct equivalent FA for such type of pattern like aⁿbⁿ. For handling such type of language, Context Free Grammar (CFG) is used. CFG can also be used to mining sequential pattern of any type of language.

IV. LITERATURE REVIEW

During the last two decades, people have implemented and compared numerous algorithms that try to solve the sequential pattern mining problem as efficiently as possible. Out of these numerous algorithms which have made a significant contribution to improve the efficiency of sequential pattern mining using constraints are discussed in this section.

The Novel idea of the sequential pattern mining with regular expression constraints (SPIRIT) algorithm was proposed by Minos N. Garofalakis, Rajeev Rastogi and Kyuseok Shim[4]. They used Regular Expressions (RE) as flexible constraint specification tool. It prunes the search space of patterns during computation. It involves a generic user-specified RE constraints on the mined patterns, thus enabling considerably versatile and powerful restrictions.

PrefixGrowth is a pattern-growth method which was developed by Jian Pei, Jiawei Han and Wei Wang[5]. Like PrefixSpan that generates only those sequences which satisfies the given constraint and according to the constraint sequences that are prefixes of valid sequences. PrefixGrowth represents a new and promising methodology as effective and efficient mining sequential patterns with various constraints such as length, aggregate, duration etc.

Regular Expression Highly adaptive constrained local extractor (RE-Hacklealgorithm) is proposed by H. Albert-Lorincz, and J. Boulicaut[6]. It is a highly adaptive algorithm, which represents the tree structure of a RE. An Abstract Syntax Tree (AST) which encodes the structure of the canonical form of a RE constraints called Hackle-tree. Each intermediate node represents operator and leaf node contains sequences, which are atomic. These atomic sequences are accumulated by the union, concatenations, and Keene closures to the RE constraints.

Sandra de Amo, Waldecir P. Junior and Arnaud Giacometti[7] have developed, a constraint based algorithm for Mining Temporal Relational Patterns (MILPRIT). MILPRIT allows a comprehensive spectrum of constraints to be combined into the mining process. An extensive set of experiments of MILPRIT implemented over synthetic and real data. The experimental results show its efficiency for mining temporal relational patterns.

Jian Pei, Jiawei Han and Wei Wang [8] have developed a new algorithm constraint based sequential pattern mining: the pattern-growth (PG) methods. PG represents a new and capable methodology as effective and efficient mining sequential patterns with constraints. It is interesting to extend it towards mining sequential patterns with other more complicated constraints, and mining other kinds of time related knowledge with various constraints.

Yu Hirate and Hayato Yamana[9] have proposed a novel algorithm for generalizing sequential pattern mining with time intervals. Its mining is based on gap and time interval. This algorithm has capability to handle extended sequences and implementing item interval constraint. It gives efficient pattern for the user.

Novel algorithm called sequential frequent Patterns mining with Tough Aggregate Constraints (PTAC) was proposed by Enhong Chen, Huanhuan Cao, Qing Li and Tieyun Qian[10]. This algorithm was developed to reduce the cost of mining using tough aggregate constraints through incorporating two effective strategies. One to avoid checking data items one by one and second to avoid unnecessary projected databases. They conducted experiments to evaluate the performance of PTAC and its optimization on the datasets is generated by the IBM sequence generator. It is shows that PTAC and its optimization is better than PrefixGrowth.

Ming-yen Lin, Sue-chen Hsueh and Chia-wen Chang [11] came with an algorithm called Closed Time-constrained Sequential Pattern Mining (CTSP) for mining closed sequential patterns with minimum or maximum gap and sliding window constraints. First they defined the closure of closed time-patterns and find the prefix with super count. The definition of closure is based on the contiguous super sequence. The CTSP algorithm effectively mines closed time-constrained sequential patterns. CTSP uses memory-indexes and the time constraints to shrinks the search-space efficiently within the pattern-growth basis. The closure glance is bi-directionally completed. The investigational results display that CTSP has good performance with gap constraints, both for synthetic and real datasets. CTSP can mine databases; even though the size of the database is larger than the main memory.

Recently Sequence Mining Automata (SMA) was proposed by R. Trasarti, F. Bonchi, and B. Goethals [13]. SMA algorithm uses regular expression as a constraint and develops new Petri net type automata. It generates only those sequences which are satisfied by automata and whose support count are greater than or equal to minimum support threshold value.
Mansi Vithalani, Gordhan Jethava and Amit Ganatra[14] have presented the mining of sequential patterns using the PrefixSpan (Prefix-projected Sequential pattern mining) algorithm. It can reduce the useless candidate generation and generates only user centered patterns. This algorithm is used to find sequential patterns which satisfy the various constraints. This algorithm generates only interesting patterns and saving computational cost. They applied multiple constraints in sequential pattern mining to enhance the performance.

To overcome the constraint of algorithms of RE, Pattern Mining using Formal Language Tool (PMFLT) was introduced by Sunil Joshi, Dr. R. S. Jadon and Dr. R. C. Jain[15]. The PMFLT uses Chomsky Hierarchy for applying constraints. He specified four types of grammars namely unrestricted grammar, Context Sensitive grammar, Context Free grammar and Regular grammar. It can reduce the search space by categorizing constraint type and generate only those sequences which are frequent and satisfied given constraints.

The novel algorithm for Sequential Pattern Mining using time-constraint was proposed by Anita Zala and Mehul Barot[16]. First it constructs a 1-sequence for item using time set and recursively forms time-constrained sequential patterns of time-consuming durations. The time set is a set of data-sequence pointer and time pairs. For each item time sequence is unique. By applying time attributes extremely into the mining process, it gives the user interested patterns.

The Depth-First Spelling algorithm for mining sequential patterns in biological sequences with Gap constraints (DFSG) was proposed by Vance Chiang-Chi Liao and Ming-Syan Chen[17]. Mining pattern in biological sequences is difficult task, because the sequence length is long and alphabets are small. Finding biological sequence using gap constraint is important. DFSG is more efficient than GenPrefixSpan.

Bhawna Mallick, Deepak Garg and Preetam Singh Grover[18] have presented a robust Compactness, Frequency, Monetary Prefix Span (CFML-PrefixSpan) algorithm for mining all CFML sequential patterns from the customer transaction database. The CFML-PrefixSpan algorithm has developed a pattern-growth methodology which discovers sequential patterns by a divide-and-conquer approach. They have applied two advanced concepts namely, monetary and compactness. Those are derived from the aggregate and time constraints, in addition to the frequency for mining. The mined CFML sequential patterns have provided the valuable information regarding the customer purchasing behavior and ensure that all patterns have practical time spans with the decent turnover. The experimental results have confirmed that the effectiveness of sequential pattern mining algorithms can be improved extensively by integrating the monetary and compactness concepts into the mining process.

M. Kalaivany and V. Uma [19] suggested constraints based on PrefixSpan algorithm (CTPrefixSpan), which finds the related frequent patterns from the given data sequences by representing interval-based events. This is interval based algorithm which is efficient among the sequential pattern mining algorithms. In this algorithm multiple constraints are added such as item, aggregate and length. It increases the scalability, accuracy and hence increases the efficiency of the mining sequential pattern.

V. RESEARCH CHALLENGES

Today several algorithms are available for efficiently discovering sequential patterns with respect to constraint based mining. Mined patterns are extensively used in large number of applications. But still there are numerous research challenges in the field of data mining. These research challenges are:

- To find the complete set of patterns, when probable and satisfying the minimum support threshold.
- To find efficient pattern only necessary for a small number of database scans.
- To find a user-specific constraints.
- To handle the large search space for finding pattern.
- Repeated scanning of database for mining pattern.
- To handle other constraint such as Recency, Frequency and Monetary (RFM) constraints and check the efficiency of this constraint.
- To study target oriented sequential pattern mining and its application in various domain.
- To find the pattern in object oriented databases.
- There are number of issues that need to be studied for the development of particular Sequential Pattern Mining methods for particular applications.
- To study the distributed sequential pattern mining for large sequence databases having distributed sequences.
- To find multi variant time related pattern in various applications.

VI. COMPARATIVE STUDY OF CONSTRAINT-BASED SEQUENTIAL PATTERN MINING ALGORITHM

A Comparative study of constraint based sequential pattern mining algorithm is done on the basis of their various constraints.

SPIRIT: Sequential Pattern mIning with Regular expression constraints.

CFML-PrefixSpan: Compactness, Frequency, Monetary Prefix Span.

Time-Constraint: The Sequential Pattern Mining time-constraint.

PTAC: sequential frequent Patterns mining with Tough Aggregate Constraints.

DFSG: The Depth-First Spelling algorithm for mining sequential patterns in biological sequences with Gap constraints.

CTPrefixSpan: Constraints based Prefix-projected Sequential pattern mining.

GSPM: generalized sequential pattern mining with item intervals.

RE-Hackle: Regular Expression Highly Adaptive Constrained Local Extractor.

MILPRIT: Mining Interval Logic Patterns with Regular expressions constraints.

PrefixGrowth: Prefix Growth is a pattern-growth.	Prefix Span: Prefix-projected Sequential pattern mining.
PG: Sequential Pattern Mining: the pattern-growth.	CTSP: Closed Time-constrained Sequential Pattern mining
SMA: Sequence Mining Automata.	PMFLT: Pattern Mining using Formal Language Tools.
RE: Regular Expression.	CFG: Context Free Grammar.

TABLE 1: COMPARATIVE STUDY OF CONSTRAINT BASED SEQUENTIAL PATTERN MINING ALGORITHMS

	Constraints							
Algorithms	Item	Length	Supper- Pattern	Aggregate	Duration	Gap	RE	CFG
SPIRIT							YES	
PG				YES				
CFML- PrefixSpan				YES	YES			
Time-Constraint		YES			YES			
PrefixGrowth		YES		YES	YES		YES	
PrefixSpan		YES		YES	YES	YES		
PTAC				YES				
DFSG						YES		
CTPrefixSpan	YES	YES						
CTSP			YES	YES	YES	YES		
GSPM	YES				YES	YES		
RE-Hackle							YES	
MILPRIT	YES						YES	
SMA							YES	
PMFLT							YES	YES

VII. CONCLUSION

Today several algorithms are available for discovering sequential patterns with respect to constraint based sequential pattern mining. Constraint based sequential pattern mining improves the efficiency of mining process. There are number of constraints used to discover the user specific patterns. Context Free Grammar (CFG) as a flexible constraint specification for algorithms that enables user controlled focus to integrate into the pattern mining process. Based on the analysis of above sequence data, it is clear that if we use CFG as constraint in sequential pattern mining, then we can reduce the explore space and generate only those sequences which are frequent and user centric. Use of CFG enhances performance of constraint based algorithms and this technique is well improved.

VIII. REFERENCES

- J Neelamadhab Padhy, Dr. Pragnyaban Mishra and Rasmita Panigrahi, "The Survey of Data Mining Applications and Feature Scope", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.3, June 2012, pp.43-58.
- [2] Manish Gupta and Jiawei Han, "Applications of Pattern Discovery Using Sequential Data Mining", "unpublished".
- [3] Agrawal R., Imielinski T. and Swami A., "Mining association rules between sets of items in large Databases", Proceeding 1993 ACM-SIGMOD International Conference Management of Data (SIGMOD'93), New York: ACM, 1993, pp. 207–216.
- [4] Minos N. Garofalakis, Rajeev Rastogi and Kyuseok Shim, "SPIRIT: Sequential Pattern Mining with Regular Expression Constraints", Proceedings of the 25th VLDB Conference, Edinburgh, Scotland, 1999, pp.223-234.
- [5] Jian Pei, Jiawei Han and Wei Wang, "Mining Sequential Patterns with Constraints in Large Databases", CIKM'02, McLean, Virginia, USA, November 2002.
- [6] H. Albert-Lorincz, and J. Boulicaut, "Mining Frequent Sequential Patterns under Regular Expressions: A Highly Adaptive Strategy for Pushing Constraints", Proceeding of 3rd SIAM International Conference on Data Mining, May 2003, pp.316-320.
- [7] Sandra de Amo, Waldecir P. Junior and Arnaud Giacometti, "MILPRIT: A Constraint-Based Algorithm for Mining Temporal Relational Patterns", This Research has been Supported by CNPq-Brazil, Project, 2004.
- [8] Jian Pei, Jiawei Han and Wei Wang, "Constraint Based Sequential Pattern Mining: The Pattern-growth Methods", Springer Science + Business Media, LLC, 2006.
- [9] Yu Hirate and Hayato Yamana, "Generalized Sequential Pattern Mining with Item Intervals", Journal of Computers, VOL. 1, NO. 3, June 2006, pp.51-60.
- [10] Enhong Chen, Huanhuan Cao, Qing Li and Tieyun Qian, "Efficient Strategies for Tough Aggregate Constraint-based Sequential Pattern Mining", Elsevier- Information Sciences 178, 2008, pp. 1498–1518.
- [11] Ming-yen Lin, Sue-chen Hsueh and Chia-wen Chang, "Mining Closed Sequential Patterns with Time Constraints", Journal of Information Science and Engineering 24, 2008, pp.33-46.
- [12]C.K. Bhensdadia and Y. P. Kosta, "Discovering Active and Profitable Patterns with RFM (Recency, Frequency and Monetary) Sequential Pattern Mining A Constraint based Approach", International Journal of Information Technology and Knowledge Management, Volume 4, No. 1, January-June 2011, pp. 27-32.
- [13] R. Trasarti, F. Bonchi, and B. Goethals, "Sequence Mining Automata: a New Technique for Mining Frequent Sequences under Regular Expressions," Proceeding Eight International Conference on Data Mining, December 2008, pp.1061-1066.
- [14] Mansi Vithalani, Gordhan Jethava and Amit Ganatra, "Sequential Pattern Mining with Various Constraints: An Enhanced Approach", International Journals of Computer Technology & Applications, Vol 3 (3), May-Jun 2012, pp.1125-1131.
- [15] Sunil Joshi, Dr. R. S. Jadon and Dr. R. C. Jain, "Sequential Pattern Mining Using Formal language Tools", IJCSI International Journal of Computer Science Issues, September 2012, pp.316-325.
- [16] Anita Zala and Mehul Barot, "Mining Sequential Pattern with Time-Constraint", International Journal of Engineering and Innovative Technology (IJEIT), Volume 2, Issue 7, January 2013, pp.94-97.
- [17] Vance Chiang-Chi Liao and Ming-Syan Chen, "Efficient Mining Gapped Sequential Patterns for Motifs in Biological Sequences", Liao and Chen BMC Systems Biology, 2013, pp.1-13.
- [18] Bhawna Mallick, Deepak Garg, and Preetam Singh Grover, "Constraint-Based Sequential Pattern Mining: A Pattern Growth Algorithm Incorporating Compactness, Length and Monetary", The International Arab Journal of Information Technology, Vol. 11, No. 1, January 2014, pp.33-42.
- [19] M. Kalaivany and V. Uma, "Mining Sequential Patterns for Interval Based Events by Applying Multiple Constraints", International Journal on Computational Sciences & Applications (IJCSA) Vol.4, No.4, August 2014, pp.59-66.

Design of Full Adder using Amorphous Silicon Thin Film Transistor

Srikanth G Student, M tech, VLSI and Embedded Systems, R V College of Engineering, Bangalore, India srikanthg2007@rediffmail.com

Kariyappa B S Professor, Dept. of Electronics and Communications R V College of Engineering, Bangalore, India <u>kariyappabs@rvce.edu.in</u>

Ramavenkateshwaran N Asst. Professor, Dept. of Electronics and Communications R V College of Engineering, Bangalore, India <u>ramavenkateswarann@rvce.edu.in</u>

Uma B V HOD, Dept. of Electronics and Communications R V College of Engineering, Bangalore, India <u>umabv@rvce.edu.in</u>

Abstract—Over the years thin film transistors have attracted the attention of researchers from various domains. They have the potential advantages of low cost and low temperature fabrication process. They have found applications in flexible and wearable electronics, sensors and also in displays. This paper is focused on the design of a full adder using thin film transistors with amorphous silicon as the active material. The simulations of the devices are carried out using Silvaco ATLAS and that of the circuit is carried out using Silvaco GATEWAY.

Keywords: Thin film transistors, bottom contact structures, Top contact structures, ATLAS, UTMOST, GATEWAY

I. INTRODUCTION

Thin Film Transistors (TFTs) have found applications in displays, sensors and many other electronic applications. Their potential advantage of low cost, low temperature fabrication process makes them an enticing alternative for Metal Oxide Semiconductor Field Effect Transistors (MOSFETs). A TFT is similar to a MOSFET in basic design and function. It is a three-terminal device, in which a voltage applied to a gate electrode controls current flow between a source and drain electrode under an imposed bias. TFTs can be fabricated using a wide range of materials. The most common candidates are amorphous silicon, and zinc oxide.

TFT differs from the MOSFET in that the conducting channel is induced in the accumulation regime rather than through the formation of an inversion layer.

Amorphous silicon began to be used as a semiconductor in 1960s. The first amorphous silicon was prepared by sputtering a silicon target or by thermal evaporation and it had poor characteristics. Plasma Enhanced Chemical Vapour Deposition (PECVD) is usually used to grow hydrogenated amorphous silicon (a-Si:H) and amorphous silicon nitride (a-SiN) from silane (SiH₄). Amorphous silicon suffers from low electrical performance compared to polycrystalline silicon. Their lower mobility is a major disadvantage. Thin-film transistors (TFTs) based on amorphous oxide semiconductors (AOSs) are being broadly investigated as a robust candidate for replacing traditional amorphous silicon TFTs in the field of flat-panel displays (FPDs) due to the advantages of high mobility, optical transparency, low-temperature process, and low-cost [1]. However it has several advantages. It can be made thinner and it produces saving in the material cost, and it can be deposited at a much lower temperatures.

A TFT can have top contact or bottom contact structure. It is found that the intrinsic characteristics of top contact devices are superior to those of the bottom contact ones, which is usually believed to be due to poor contact characteristics and poor semiconductor quality of bottom contact TFTs. The mechanism behind the intrinsic characteristics differences is the deficiency of carriers at the source-channel interface, resulting to a very high potential drop, which the bottom contact devices suffer more. Remarkable improvements in drain current are expected by only inserting high carrier concentration region around the source/drain contact area, which totally eliminates the potential drop [2].

In this work we have designed and simulated a thin film transistor with amorphous silicon as the semiconductor layer. Despite the advantages of the top contact device, bottom contact structures can be fabricated much more easily compared to the top contact structures. The device structure for the design was chosen to be bottom contact. The device is simulated in Silvaco ATLAS tool. This device so designed device is used in the design a full adder circuit. Here the TFTs are used as pass transistors. The circuit simulation was carried out using Silvaco Gateway tool. While simulating the circuit it was assumed that both the signal and its complement are available thus reducing the number of transistors in the circuit and avoiding any multiple threshold voltage drops that is typical in pass transistor logic.

The rest of the paper is organized as follows. The second section focuses on the design of the thin film transistor and the full adder circuit schematic design. This section is followed by the discussion of the simulation results, followed by the conclusion.

I. DESIGN

TFT device can be made with a wide variety of structures, of which the four basic configurations are staggered, inverted staggered, coplanar and inverted coplanar. Depending on the material used as the semiconductor, one of these configurations is preferred. Inverted configurations are referred to as the bottom gate, while non-inverted ones are referred to as the top gate.

The thin film transistor that is designed has a bottom contact structure. The semiconductor used is amorphous silicon which has a thickness of $0.3\mu m$. The dielectric layers used are silicon dioxide and silicon nitride each of thickness $0.05\mu m$, aluminium is used for source, drain and gate contacts. The structure of the device that was designed and simulated is as shown in Fig 1.

A passivation layer is required so as to prevent the oxidation of the underlying amorphous silicon layer. As can be seen in Fig 1, it is easy to fabricate the device as less number of masks are needed, thus reducing the steps required in the fabrication procedure.



Fig 1. Structure of the TFT

Fig 2. Schematic of the Full Adder

The TFT structure shown in Fig 1 is used to design the circuit of a full adder. A full adder is a circuit that adds to numbers along with a carry. Full adders are usually used as components of parallel adders, which makes up a significant part of an Arithmetic Logic Unit (ALU), which in turn is used in Central Processing Unit (CPU)of computers. Full adder is designed using TFTs as pass transistors. In the design it is assumed that both the signal and its complements are available. The size of the each transistor is chosen as $L=4\mu m$ and $W=16\mu m$. The schematic of the circuit is as shown in Fig 2. The supply voltage of 5 V was given and the input to the circuit was a square wave of amplitude 5 V.

II. SIMULATION RESULTS

Thin film transistor is designed using Silvaco ATLAS tool. ATLAS is a physically-based two and three dimensional device simulator. It predicts the electrical behaviour of specified semiconductor structures and provides insight into the internal physical mechanisms associated with device operation. It can be used as standalone or as a core tool in SILVACO's VIRTUAL WAFER FAB simulation environment. In the sequence of predicting the impact of process variables on circuit performance, device simulation fits between process simulation and SPICE model extraction.

The Id/Vd characteristics of the device and the Id/Vg characteristics of the device are as shown in Fig 3 and 4 respectively. From Fig 3, we can see that the device behaves almost like a MOSFET, except that the current has a low value compared to the latter, as the amorphous silicon is not as good a conductor as polycrystalline silicon. The various regions of operation, cutoff, linear and saturation regions analogous to a MOSET transfer characteristic can be easily observed in this figure.



Simulations were run for different values of a parameter, while all other parameters were kept constant. During the simulations the following properties about the device current Id was observed. It was found that the current increases:

- 1. with decreasing dielectric thickness
- 2. with increasing semiconductor layer thickness
- 3. with both the gate and the drain voltage
- 4. with decreasing distance between the source and the drain(channel length)

The circuit for full adder is designed using the above TFT on Silvaco Gateway tool. Gateway is a schematic capture tool and is the entry point for the IC design flow. Gateway consists of a GUI for schematic entry and is integrated with the following simulators: Silvaco SmartSpice (Analog), Silvaco SmartSpice RF (RF), Silvaco Silos (Verilog), Silvaco Atlas (TCAD MixedMode), and Synopsys HSPICE (Analog).

The transient plot for the above full adder is shown in Fig 5 and 6 respectively. The output, sum of the circuit is about 3.6 volt as connecting the devices as pass transistors will reduce the output by one threshold voltage. And the carry output of the circuit goes as high as 5 volt for all input being one, this is suspected due to the capacitance effect.



Fig 5. Sum output for the Full Adder

Fig 6. Carry Output for the Full Adder

During the simulation of the circuit it is found that the circuit fails to operate at high speeds, in the GHz range, while it performs well in the MHz range. As can be seen from the Fig 5 and 6 the output voltage is less by 1.6 volt (approx), due to the threshold voltage drop across the transistors when connected as pass transistors.

VI. CONCLUSION

The Id-Vd and Id-Vg characteristics reveal that the device is almost similar to MOSFETs in operation. However, their oncurrent is very much lower compared to the MOSFETs. This is due to the lower mobility of amorphous silicon layer. The threshold voltage from the simulations is found to be close to 1.6 V. The results of circuit simulation have shown that TFTs can be very efficiently used in digital circuits, if the threshold voltage can be lowered. In this work (amorphous silicon) a-Si based TFT is used in the design of full adder and verified for its function which proves that flexible digital circuits can be designed using a-Si TFT.

V. REFERENCES

- K. Nomura, H. Ohta, A. Takagi, T. Kamiya, M. Hirano, and H. Hosono, "Room-temperature fabrication of transparent flexible thin-film transistors using amorphous oxide semiconductors," Nature, vol. 432, no. 7016, pp. 488–492, Nov. 2004.
- [2] Yoshinori Ishikawa, Yasuo Wada, and Toru Toyabe, "Origin of characteristics differences between top and bottom contact organic thin film transistors", Journal of Applied physics, 04 March 2010
- [3] Poornima Mittal, Brijesh Kumar, Y. S. Negi, B. K. Kaushik and R. K. Singh, "Organic Thin Film Transistor Architecture, Parameters and their Applications", 2011 International Conference on Communication Systems and Network Technologies
- [4] Dawen Li and Jay Guo, "Organic thin film transistors and polymer light emitting diodes patterned by polymer inking and stamping," J. Phys. D: Appl. vol. 41, pp. 105115 (1-7), 2008.
- [5] M. G. Kane, J. Campi, M. S. Hammond, F. P. Cuomo, B. Greening, C. D. Sheraw, J. A. Nichols, D. J. Gundlach, J. R. Huang, C. C. Kuo, L. Jia, H. Klauk, and T. N. Jackson, "Analog and Digital Circuits Using Organic Thin-Film Transistors on Polyester Substrates", IEEE ELECTRON DEVICE LETTERS, VOL. 21, NO. 11, NOVEMBER 2000
- [6] Yu-Chi Chang, Chia-Yu Wei, Yen-Yu Chang, Tsung-Yu Yang, and Yeong-Her Wang, "High-Mobility Pentacene-Based Thin-Film Transistors With Synthesized Strontium Zirconate Nickelate Gate Insulators", IEEE TRANSACTIONS ON ELECTRON DEVICES, VOL. 60, NO. 12, DECEMBER 2013
- [7] A. R. Khan, Dr. Javed Siddique, Dr. S. S. K. Iyer, "Advanced Technique of Vertical Organic Thin Film Transistor", IMPACT-2013
- [8] Kymissis, Organic Field Effect Transistors: Theory, Fabrication and Characterization. New York, NY, USA: Springer-Verlag, 2009, pp. 1-3.
- [9] B. Stadlober, M. Zirkl, M. Beutl, and G. Leising, "High-mobility pentacene organic field-effect transistors with a highdielectric-constant fluorinated polymer film gate dielectric," Appl. Phys. Lett., vol. 86, no. 24, pp. 242902-1–242902-3, Jun. 2005
- [10] Feng Yan and Hao Tang, "Application of thin-film transistors in label free DNA biosensors," Expert Rev. Mol. Diagn. vol. 10 no.5, pp. 547–549, 2010
- [11] H. Klauk, "Organic thin film transistor," Chem. Soc. Rev., 39, pp. 2643-2666, 2010.
- [12] S. M. Sze, Physics of Semiconductor Devices. New York, NY, USA:

Review on Classification and Quality Analysis of Fruits Using Artificial Neural Network

Mr. Lawate P.M. Research Student, Sinhgad Institute of Computer Sciences, Korti, Pandharpur, Solapur University, Solapur, arun1671@gmail.com

Dr. Ghatule A.P. Controller of Examination, Yashwantrav Chavan Maharashtra Open University, Nashik. arjun1671@gmail.com

Abstract—In the Present work the Pomegranate images are pre-processed to highlight the discriminating features of the fruit varieties. Thus, the feature vector is obtained and is subjected to categorization process. Automatic fruit grading and sorting requires the implementation of computer vision systems. It also uses the artificial neural network to enhance the accuracy of the result

I. INTRODUCTION:

The decision-making capabilities of human-inspection are being affected by external influences such as tiredness, penalty, partiality etc. Hence, development of a Machine Vision Systems (MVS). so as to overcome aforesaid influences. An artificial neural network based categorizer is developed which is trained using feed forward rule. Fruit processing industry contributes a major part in nation's growth, but there has been a decrease in production of good quality fruits, due to improper cultivation, lack of maintenance, very high post-harvest losses in handling and processing, manual inspection, lack of knowledge of preservation and quick quality evaluation techniques. Also, rising labour costs, shortage of skilled workers, and the need to improve production processes have all put pressure on producers and processors for the demand of a rapid, economic, consistent and nod-destructive inspection method [3]. In such a scenario, automation can reduce the costs by promoting production efficiency.

II. OBJECTIVES

- 1) The aim of this work is to develop a prototype for the automatic Computer vision based quality sorting of pomegranate including the development of new real-time computer vision algorithms.
- 2) In order to offer a quality product to the consumers as per requirement.

LITERATURE REVIEW

Devrim Unay, Bernard Gosselin [2] demonstrated that computer vision based system is introduced to automatically sort apple fruits.

Tejal Deshpande, Sharmila Sengupta, K. S. Raghuvanshi [1] Introduced image processing methodology to deal with one of the main issues of plant pathology i.e. disease grading. It specify automatically grade the disease on the Pomegranate plant leaves and fruit and also provide efficient module that identifies the Bacterial Blight disease on pomegranate plant.

Khoje S. A., Bodhe S. K. and Adsul A. [3] demonstrated new approach to classify the quality grade of guava and lemon fruit was developed using the SVM classifier based on curvelet transform based features.

Khoje S. and Bodhe S. [4] studied various size estimation metrics which are used as feature vectors for two classifiers namely

Feed Forward Neural network (FFNN) and Support Vector Machines.

Teoh C. C. and Syaifudin A. R. M. [5] presented a study on measuring the weight of Chokeman mangoes using image processing and analysis techniques.

Mahendran R, Jayashree G. C., Alagusundaram K. [6] presented the recent development and application of image analysis and computer vision system in quality evaluation of products in the field of agriculture Color features of fruits and vegetables

included mean, variance, ranges of the red (R), green (G), and blue (B) color primaries and the derived hue (H), saturation (S), and intensity (I) values.

Meenakshi M. Pawar, Sanman Bhusari, Akshay Gundewar [7] implemented for automatic grading and sorting system for quality control of pomegranates.

Sanjay Chaudhary, Bhavesh Prajapati [8] proposed technique of digital image processing to classify the banana in group or bunch. They also proved that Color and texture features combination gives best result for classification process.

Dayanand Savakar [9] proposes an algorithm for fruits classification based on the shape, color and texture proposes that when color and texture features are amalgamated, it gives better result over the all other previous method such as shape, color and texture. From the result we can find that shape based classification gives 83.2% accuracy, Color basis gives 90%, Texture basis give 89.60% and results are improved to 96 % when the color and texture features are amalgamated.

S. Arivazhagan [10] the recognition is done by the minimum distance classifier based upon the statistical and co-occurrence features derived from the Wavelet transformed sub- bands.

Hetal Patel [11] designed the algorithm aiming at calculation of different weights for features like intensity, color, orientation and edge of the test image.

Basvaraj. S. Anami, J. D. Pujari, Rajesh. Yakkundimath [12] Demonstrated color and texture features are used to recognize and classify different agriculture/horticulture produce into normal and affected using neural network classifier the experimental results indicated that approach significantly supports accuracy in automatic detection of normal and affected produce.

Sawsana Al-Rahbi and Annamalai Manickavasagan [13] Proposed Back propagation neural network (BPNN) was implemented to obtain the classification accuracies of the developed algorithm. The accuracy was improved to 90%

A. Khoshroo, A. Keyhani, R. A. Zoroofi, S. Rafiee, Z. Zamani, M. R. Alsharif [14] They used (MRI)Magnetic resonance imaging techniques to visualize internal structure of pomegranate.

III. METHODOLOGY

Proposed method wills Grade and Identify the defect in pomegranate fruit. The size is determined by the weight or maximum diameter of the equatorial section of the fruit, in accordance with the following table:

Sr. No.	Size code/Class	Weight in grams (Minimum)	Diameter in mm (Minimum)
1	Α	400	90
2	В	350	80
3	С	300	70
4	D	250	60
5	Е	200	50

The system is divided into the following steps:

1) Image Acquisition 2)Image Pre-processing 3)Segmentation 4)feature extraction 5) Calculating affected area and Normal area 6)Quality Grading (Color, Size, % of disease Infection)

System Architecture:

The System architecture of the whole system as shown in the following points.

1. Image Acquisition:

Image set will consist of a total 40 image samples of Pomegranate fruit. This image set was further divided into four subsets (10 partially affected, 10 moderately affected, 10 severely affected and the rest 10 which are normal). Images are taken from the digital camera having resolution of 13 Mega Pixels by keeping distance of 20 centimeters from the image sample.

2.Image Pre-processing:

The captured image in RGB color space for extraction of texture features, the image was converted to grayscale. Grayscale image are used to detect defected portion on the surface of pomegranate fruit. An original RGB image is used to Classify by using color of mature pomegranate fruit. This image also used to identify the healthy region on the surface of pomegranate fruit. To avoid computational delays associated with further image processing analysis, the images were resized in fixed resolution images.

3. Image Segmentation:

In computer vision, segmentation refers to the process of clustering the pixels with certain properties into salient regions and these regions correspond to individual surfaces, objects or natural parts of the objects. More precisely, image segmentation is the process of assigning a label to every pixel in an image so that pixels with the same label share certain visual characteristics.

We propose the following techniques to segment the affected area.



Fig: System Flow

4. Feature Extraction:

In this step we extract the desired features from the sample image for the analysis of pathogenic affected region of the fruit. Main goal of our project

is to provide a fast algorithm for fruit classification. Therefore, average, standard deviation, and median values are calculated over the segmented area of each fruit from all filter images [2].

A computer vision based fruit blemish inspection system should start with segmentation, followed by extraction of texture features and then lead to achieve correct grading of fruits into corresponding quality categories [3]. Here Color feature extraction and Texture Feature Extraction is employed.

4.1 Color Feature Extraction:

There are four color models namely RGB, HIS, CMY, YIQ. We have used RGB color model for analysis. Red, Green and Blue component of the image sample is extracted. We also extract the total count of red, green and blue pixels in the Pomegranate fruit. We also extract the total count of pixels occupied by entire Pomegranate, which gives the total area of Pomegranate.

4.2 Texture Features Extraction:

Important texture features like correlation, homogeneity, entropy and contrast are extracted along with color features. Haralick features calculation is done in two phases, i) Calculation of the Co-occurrence Matrices, ii) Calculation of the features based on the Co occurrence Matrix. A. Contrast: Contrast is the difference in luminance and/or color that makes an object. (or its representation in an image or display).

Different extraction equation as follow:

1) Correlation:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

-

Where:

- X_i is the pixel intensity or the gray scale value at x_i point in the unreformed image. y_i is the gray scale value at a point y_i in the deformed image
- x bar and y bar are mean value of the intensity matrix x and y.

2) Entropy:

Entropy = -
$$\sum_{i=0}^{n} P_i Log2 p_i$$

Where:

- P is the probability that the difference between 2 adjacent pixels is equal to i,
- \log_2 is the base 2 logarithm.

3) Contrast :

Where:

Contrast =

$$rac{I_{ ext{max}} - I_{ ext{min}}}{I_{ ext{max}} + I_{ ext{min}}},$$

• Imax and Imin representing the highest and lowest luminance.

5. Estimation of percentage of affected area:

The percentage of affected area is calculated using equation

1. Percentage of healthy area = (healthy area / total area) * 100 Where,

```
Healthy area = (Total area of fruit – Defected Area of fruit) /100,
Total area = total size of fruit.
```

2. Percentage of affected area = 100 - % of healthy area.



Fig: Defect Identification Flow Diagram

6. Defect Identification Flow Diagram:

The defect identification is required for quality classification. The flow of defect detection is shown in the following flow diagram.

7. Back Propagation Artificial Neural Network (BPANN) Model:

A Back propagation network (BPN) is best suited classifier and thus is the most popular choice for classification of agriculture product.

. The algorithm can be decomposed in the following four steps:

i) Feed-forward computation

ii) Back propagation to the output layer

iii) Back propagation to the hidden layer

iv) Weight updates

The algorithm is stopped when the value of the error function has become sufficiently small which explained in the following Fig.

The multilayer feed forward network ANN model used has adopted back propagation algorithm for training.



Fig: Back Propagation Artificial Neural Network

The number of neurons in the input layer is equal to the number of input features. The number of neurons in the output layer is equal to the category of affected part depending on % of affected area in fruit samples considered four types i.e. 1) Normal, 2) Partially affected, 3) Moderately affected, 4) Unhealthy. Oⁱ if the ith output of the network is "high" while all other outputs are "low".

IV. RESULT & CONCLUSION

The main purpose of this paper is to overcome the Manual classification of Pomegranate fruit. It also motivate to farmer accurate Quality and Reduce the Fording in the billing system of farmer and intermediate chain who provide the commutation of customer and retailer. It work also classification and grading on the basis of quality and obtain maximum accuracy and reducing human interface. Using this combined framework the Proper % of defected part of pomegranate fruit is estimated. This may achieved using K-Means Clustering technique. To predict result more accurately we use the BPNN so here we train the neural network on the basis of features like 1 Color 2 Texture and 3 Color & Texture, so we have got High accuracy in combination of both the feature like Color & Texture.

V. REFERENCES

- [1] Tejal Deshpande, Sharmila Sengupta, K. S. Raghuvanshi "Grading and and Fruit", International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014, 4638-4645.
- [2] Unay D. and Bernard Gosselin, "Artificial Neural Network-based Segmentation and Apple Grading by Machine Vision" IEEE, 2005.
- [3] Khoje S. A., Bodhe S. K. and Adsul A. "Automated Skin Defect Identification System for Fruit Grading Based on Discrete Curvelet Transform", International Journal of Engineering and Technology, vol. 5, no. 4, pp. 3251-3256, 2013.
- [4] Khoje S. and Bodhe S. "Comparative Performance Evaluation of Size Metrics and Classifiers in Computer Vision based Automatic Mango Grading", International Journal of Computer Applications, vol. 61, no. 9, pp. 1-7, 2013.
- [5] Teoh C. C. and Syaifudin A. R. M. "Image Processing and Analysis Techniques for Estimating Weight of Chokanan Mangoes", J. Trop. Agric. and Fd. Sc., vol. 35, no. 1, pp. 183–190, 2007.
- [6] Mahendran, R., Jayashree, G. C. and Alagusundaram, K. "Application of Grading of Fruits and Vegetables", Journal of Food Processing and Technology, 2011.
- [7] Meenakshi M. Pawar, Sanman Bhusari, Akshay Gundewar "Identification of Infected Pomegranates using Color Texture Feature Analysis" International Journal of Computer Applications (0975 – 8887) Volume 43–No.22, April 2012.
- [8] Sanjay Chaudhary, Bhavesh Prajapati "Quality Analysis and Classification of Bananas" Volume 4, Issue 1, January 2014 ISSN: 2277 128X, International Journal of Advanced Research in Computer Science and Software Engineering.
- [9] Dayanand Savakar "Identification and Classification of Bulk Fruits Images using Artificial Neural Networks" IJEIT Volume 1, Issue 3, March 2012.
- [10] S.Arivazhagan, R. Newlin Shebiah, S. Selva Nidhyanandhan, L. Ganesan "Fruit Recognition using Color and Texture Features". JETCIS, vol. 1 no. 2, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India, 2010.
- [11] Hetal N. Patel, Dr. R. K. Jain, Dr. M. V. Joshi "Fruit Detection using Improved Multiple Features based Algorithm". IJCA vol. 13, A. D. Patel Institute of Technology, New V. V. Nagar, Gujarat, India, 2011.

- [12] Basvaraj S. Anami1, J. D. Pujari, Rajesh.Yakkundimath "Identification and Classification of Normal and Affected Agriculture/horticulture Produce Based on Combined Color and Texture Feature Extraction" International Journal of Computer Applications in Engineering Sciences [ISSN: 2231-4946].
- [13] Sawsana Al-Rahbi and Annamalai Manickavasagan "Back Propagation Neural Network (BPNN) to Detect Surface Crack on Dates Using RGB Images" Journal of Medical and Bioengineering Vol. 4, No. 1, February 2015.
- [14] A. Khoshroo, A. Keyhani, R. A. Zoroofi, S. Rafiee, Z. Zamani, M. R. Alsharif "Classification of Pomegranate Fruit using Texture Analysis of MR Images" Agricultural Engineering International: the CIGR Ejournal. Manuscript 1182. Vol. XI., March 2009.

Identification of DoS Attack and Its Countermeasure Based on Dynamic Threshold Value of Traffic Flow

Preksha Nema Computer Technology and Applications National Institute of Technical Teachers' Training and Research Bhopal (M.P), India preksha.nema@gmail.com

Yogesh Sahu Computer Technology and Applications National Institute of Technical Teachers' Training and Research Bhopal (M.P), India yogesh.sahu22@gmail.com

M.A.Rizvi Dept. of Computer Engineering and Applications National Institute of Technical Teachers' Training and Research Bhopal (M.P), India marizvi@nitttrbpl.ac.in

Abstract—Recent generation is cyber craving. Everyone is flourishing with the enormously growing digital technologies. Today every job is done online, that's why the probability of being attacked by attackers is continuously increasing with increased online task therefore security of sensitive data that is being transmitted is one of the important issue to be preferably considered in order to prevent it from intruders. Cyber security is provided by employing various available security mechanisms. In this paper a brief overview of cyber security is presented with the analysis of various cyber attacks and their countermeasures. Among all the described attacks, Denial of service (DoS) attack is then analyzed in detail with its identification mechanism based on varying threshold value of traffic flow.

Keywords: Cyber security; Cyber space; Cyber attacks; Dos;DDos

I. INTRODUCTION

In current cyber era, it is relatively a very easy task of obtaining, communicating, monitoring and controlling information in computer systems over vast distances. Huge Network connectivity and its global access are prominent to cyber world of information technology. Besides all its benefits, extensive connectivity brings high risk of theft, fraud, and abuse. Networked systems are in great threats such as: denial of service attacks, virus, spyware, socially engineered Trojans, malware, phishing, password cracking and hacking are extensively rising. [1]

Cyber security is preventing systems, networks and data in cyberspace. The primary goal is to minimize the risks, including mitigation of cyber security attacks. The general security objectives are Availability, Integrity which also includes authenticity, non-repudiation and Confidentiality.

Cyber security techniques can be used to ensure system availability, integrity, authenticity, confidentiality, non-repudiation and to establish the user's trustworthiness. Almost every entity that is directly or indirectly connected to internet or using it are the part of cyber space. Cyber security aims at securing the cyber environment in following way:

The collection of policies and actions that are used to protect connected networks (including, computers, devices, hardware, stored information and information in transit) from unauthorized access, modification, theft, disruption, interruption or other threats.[2]

II. CURRENT SCENENARIO OF CYBER SECURITY

Cyber crime being committed over internet it is very deep matter. Cyberspace and its underlying infrastructure are vulnerable to a wide range of risk stemming from both physical and cyber threats and hazards. Sophisticated cyber actors and nation-states exploit vulnerabilities to steal information and money and are developing capabilities to disrupt, destroy, or threaten the delivery of essential services. [3]

A. Cyber Risks

Cyber risks can be divided into three distinct areas:

- **Cyber crime**: It is a malicious practice by group of individuals to steal money, data and cause disruption in cyber space. It can be carried out in various ways, such as obtaining the debit/credit card information and damaging the primary functions of a website.
- Cyber war: When the nation is conducting vandalism against another nation to cause destruction or to extract data.
- Cyber terror: Terrorist activities conducted in cyberspace by an independent organization of a nation.

B. Threats To Cyber Security

The foremost threat in cyber environment is the risk involved in ensuring security. The type of cyber attacks, attackers and their methods of attack and the consequences of successful attacks are very well included in threat analysis. Risk assessment combined with threat analysis allows an organization to evaluate security. [4] Some of the most common threats to cyber security include the following:

- Destruction of information
- Corruption or modification of information
- Information Theft or loss
- Intervention of services.

Threats are mainly of four types named as: accidental, intentional, active and passive. [5]

Accidental threats: These threats are those that come with no prior intention. For example operational blunders, system malfunctions.

Intentional threats: These threats are those that come with some prior intention. For example casual examination, using monitoring tools which are easily available.

Passive threats: These threats will not result in any modification to information contained in the systems.

Active threats: These threats involve the alteration of information contained in the system and to the state or operation of the system.

III. CYBER ATTACKS

Attacks that are generated in the cyber environment via worms, malware or by direct attack on critical infrastructure, such as telecommunications cables are known as cyber attacks. A combination of these attacks may also possible. The level of risk varies as low, medium and high among different components of the cyber space. So risk management is essential part of cyber security. In order to facilitate the risk analysis, it is beneficial to consider attacks as belonging to the following categories:

- Service Interruption Attacks: These types of attacks disable user access to the targeted services either temporary or permanently. Examples include lack of access to a web site, or the inability to conduct a financial transaction, or the ability to initiate a voice call. Several types of attacks can lead to service disruption. For example, denial of service (DoS), distributed denial of service attacks (DDoS), or damaging of buildings that host critical infrastructure could result in preventing users from accessing a service.[6]
- **Phishing**: It is the process of acquiring sensitive and personal information such as passwords, usernames and debit/credit card details by masquerading as a trustworthy entity in electronic communication.
- **Password Cracking**: It is the technique of recovering passwords from information which is being stored in or transmitted by a computer system.
- **Spyware:** It is a kind of malware whose purpose is to gather the information about user like their browsing habit or about their computer secretly. Beyond just simple monitoring one can extend the function of Spyware, after that they are able to gather several kind of information. [7]
 - History of surfing
 - List out frequently access sites by user.

Such program can also capable to interfere with user control on the computer like

- Redirect web browser activity
- Install additional software
- Change computer setting.
- Socially engineered Trojans:- In the context of cyber security social engineered Trojans to psychological manipulation of people into performing actions or divulging confidential information like, user get message through pop up or via email that they are infected by virus or your device is very slow, you should need such particular software to run.[8]
- **Backdoor Attack:** In such kind of cyber threat an attacker make effort to find a back door and use that to install a key logging software for allowing an illegal access to the system and try to modify the files, stealing information, installing harmful software.[9]

S.No.	ATTACKS	SYMPTOMS	COUNTERMEASURES
1.	Service Interruption Attacks	 Network performance becomes slow usually in opening files or accessing websites. Amount of spam received will increase excessively 	 Use Infiltrator Identification system. Use Strong updated anti-virus software on machine. Once a DoS attack identified one can reduce its impact by execution filters to block unwanted traffic.
2.	Phishing	 Spelling and grammatical mistakes in the received emails. Emails received that your account would be closed if you didn't respond to an email message. Spoofing popular websites or companies. 	 Avoid configuring a system or applications with unreliable proxy server settings. Never click unknown links in trustless websites. Firewall software used should be best available in market.
3.	Password Cracking	 Files unexpectedly appear or grow in size. Frequent crashes 	 Make sure that the password that is being used is very strong. Password for two different sites must not be same.
4.	Spyware	 pop-up advertisements will appear again and again. My settings have changed and one can't change them back to the way they were. My computer seems sluggish 	 Use anti-Spyware software, keep virus definition files updated, and scan your system for Spyware. Keep your computer up-to-date. Be careful about suspicious sites and emails. Enhance the security level of your computer. In case of emergency, back up important files
5.	Socially engineered Trojans	 Fake emails regarding request for financial information and password. Fraud calls with intension of taking information related to any bank account, ATM etc. 	 End-user education is one of the best countermeasures for socially engineered Trojans. Enterprises can further protect themselves by not allowing elevated users to surf the Web or answer email.
6.	Backdoor Attack	 Machine will take more time to start. Certain undesirable programs are installed without your permission. Certain program may take major CPU or memory. OS of Machine may crash suddenly. 	 Anti-virus and malware code scanning For malicious user-installed backdoors, use access control management and controlled software deployment. For developer-installed backdoors, disable them, change the defaults, or block access.

Table 1: Analysis of Various Cyber Attacks along with their symptoms and countermeasures

IDENTIFICATION MECHANISM FOR DOS ATTACK

In DoS attack, attacker makes memory resources too busy by creating lot of requests at the same time, because of that server cannot respond to the entire requests, thus legitimate user also denying accessing a machine. Broadly there are two types of Dos attack namely, DoS that is Performed by a single host and DDoS that is performed by a number of compromised machine that all targets the same victim. In DDOS attack hacker will fire huge number of fake messages to the websites and mail servers, so that the system performance will go slow which ends up with system crash. [10] Hackers use malicious codes and zombie machines in order to increase the effect of DDoS attack and thereby send such a huge quantity of fake messages on to the servers.

Step for Identify the attack and Strategy to Handle Situation:-

For identifying the DoS attack, there are various steps to be considered:-

Step 1:- Monitor the flow of traffic, if it is greater than the threshold value of traffic flow.

Step 2:- Check the traffic flow rate, if exceeds a certain threshold or not.

Step 3:- If threshold is greater than the set limit then try to identify the IP address from where the traffic is generated.

Step 4:- When the IP address is identified, send notification about the attack with IP address of traffic creator to all the connected routers in the network.

Step5:- Then Block all the traffic that is coming from the identified IP address.

Flow Chart:



After doing the in depth study of almost all the prominent cyber attacks, it has been concluded that DoS attack and its countermeasures are the prime area to be explore in the field of tremendously growing information technology. One can verify an attack is Dos attack, based on examine the threshold value of traffic flow. Continuous advancement in the technology however leads to the growth of more powerful attacks and as a result hackers and attacker are becoming more intelligent. This directly affects the cyber security of sensitive data that is the target of hackers. So to prevent this sensitive data it is foremost important to take countermeasures before it has been attacked by malicious attackers. No single countermeasure is sufficient to protect by intruder so huge number of countermeasures should be employed in order to improve the present situation of cyber security.

IV. REFERENCES

- [1] http://faculty.nps.edu/ncrowe/edg_attacks.htm
- [2] http://www.ukessays.com/essays/computer-science/study-on-current-cyber-attacks-and-countermeasures-computerscience-essay.php
- [3] http://www.cybersecuritycrimes.com/types-of-cyber-attacks/
- [4] http://www.itgovernance.co.uk/what-is-cybersecurity.aspx
- [5] http://www.dhs.gov/topic/combating-cyber-crime
- [6] http://www.infoworld.com/article/2616316/security/the-5-cyber-attacks-you-re-most-likely-to-face.html?page=2
- [7] http://searchsecurity.techtarget.com/definition/distributed-denial-of-service-attack
- [8] https://www.iplocation.net/denial-of-service.
- [9] http://www.cisco.com/en/US/products/ps6134/products_configuration_guide_chapter09186a00808498a8.html
- [10] Sherif M. Khattab, C. Sangpachatanaruk, D. Mosse, R. Melhem, and T. Znati, "Roaming honeypots for mitigating servicelevel denial-of-service attacks," in Proceedings of the 24th International Confer-ence on Distributed Computing Systems (ICDCS04),pp. 328–337. IEEE, IEEE Computer Society, Mar. 2004.
- [11] N. Ye, "A markov chain model of temporal behavior for anomaly detection," in Proceedings of the 2000 IEEE Workshop on Information Assurance and Security, United States Military Academy, West Point, NY, June 2000.
- [12] "A Way and Means of Intrusion Detection in Mobile Ad-Hoc Network" National Seminar on Recent Trends in Network Security March 2015, Sponsored by UGC
- [13] International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), ISSN: 2321-8169, PP: 4301 – 4306 June 15 Volume 3 Issue 6 entitled "Critical Analysis of Various Symmetric Key Cryptographic Algorithms".
- [14] Rahman, Md Saifur, Hemanshu R. Pota, and Md Jahangir Hossain. "Cyber vulnerabilities on agent-based smart grid protection system." PES General Meeting Conference & Exposition, 2014 IEEE. IEEE, 2014.
- [15] Kumar. V. Ananda. Krishan K. Pandev. and Devendra Kumar Punia. "Cvber security threats in the power sector: Need for a domain specific regulatory framework in India." Energy Policy 65 (2014): 126-133.
- [16] Richardson, Robert. "CSI computer crime and security survey." Computer Security Institute 1 (2008): 1-30.
- [17] Jackson, William D., Mark Jickling, and Baird Webel. "The economic impact of cyber-attacks." Congressional Research Service, Library of Congress, 2004.
- [18] Lazarevic, Aleksandar, Vipin Kumar, and Jaideep Srivastava. "Intrusion detection: A survey." Managing Cyber Threats. Springer US, 2005. 19-78.
- [19] Lewis, James Andrew. Assessing the risks of cyber terrorism, cyber war and other cyber threats. Washington, DC: Center for Strategic & International Studies, 2002.
- [20] Ahmad, Ateeq. "Type of Security Threats and It's Prevention." International Journal of Computer Technology and Applications 3.2 (2012).
- [21] http://www.ukessays.com/essays/computer-science/study-on-current-cyber-attacks-and-countermeasures-computer-science-essay.php#ixzz3qQEo1pSu

An Improved Task Scheduling Algorithm on e-learning Using Cloud

Sonal Dubey Regional Institute of Education, Bhopal, India ersonaldubey@gmail.com

Kratika Verma Regional Institute of Education, Bhopal, India vermakritika011@gmail.com

M.A. Rizvi Dept. of Computer Engineering and Applications National Institute of Technical Teachers' Training and Research Bhopal (M.P), India <u>marizvi@nitttrbpl.ac.in</u>

Khaleel Ahmad School of CS & IT, Maulana Azad National Urdu University, Hyderabad, India <u>khaleelamna@gmail.com</u>

Abstract—Cloud computing offers a dynamic provision of elasticity, scalability, virtualized resources, pays per use and measured service with the dynamic allocation of the computing resources as needed. Cloud computing is preferably suited for e-learning Systems. Resource optimization and QoS improvement are of main concern in case of scheduling in cloud environment. In this paper, we address the "Qualities of Service (QoS) improvement problem" in cloud on task scheduling, which is to allocate and schedule computing resources in a way that the QoS parameters i.e. latency and send/received (bytes/sec) over network are improved and users meet their applications' performance requirement. A new task scheduling algorithm using Dynamic Parallel Schedule (DPS) framework, in cloud computing environment, i.e. Quality of Service Improvement Simple Compute Farm (QISCF) is devised.

Keywords: E-Learning; Cloud Computing; Scheduling Algorithm; QoS Parameter; Task Scheduling

I. INTRODUCTION

One of the important problems that must be addressed to realize the advantages of cloud computing, and the focus of paper is improving the Quality of Service (QoS) parameters. QoS identifies a set of functional and non-functional attributes that can be used to evaluate the behavior of a service from different perspectives. These could be performance metrics, such as response time or security attributes, latency, data transmission speed over network, transactional integrity, reliability, scalability, availability etc.

A. E-Learning On Cloud

E-Learning is related to the virtualized distance learning by means of electronic communication specifically the internet. The E-Learning is internet enabled learning. The e-learning includes several advantages some of them are the flexibility, convenience, accessibility, consistency and repeatability. With the rapid growth of Cloud Computing, it is rising as a platform for delivering E-learning services to the users with great opportunities. The infrastructure buildup, the software updating and licenses managements, the hardware configurations, the infrastructure space, maintenance, and power consumption, and many other issues will be either eliminated or minimized using the Cloud technology.

B. Task Scheduling In Cloud Computing

A scheduling is a set of rules that determine the jobs to be executed at a particular time. The Task management is the key role in cloud computing systems. Task scheduling problems are premier which relate to the efficiency of the whole cloud computing facilities. Because of different QoS parameters such as CPU speed, CPU utilization, turnaround time, throughput, waiting time etc., task scheduling in cloud computing is different from conventional distributed computing environment. The demand for resources fluctuates dynamically so scheduling of resources is a difficult task. Task scheduling based on QoS parameters is necessary for efficient resource utilization and for satisfying user requirement.

II. DYNAMIC PARALLEL SCHEDULES (DPS)

In the thesis, DPS is being used for improving the QoS parameters i.e. memory usage and the average number of bytes per second accessed by the process in cloud environment.DPS applications are defined as directed acyclic graphs of operations. The fundamental types of operations are leaf, split, merge and stream operations. The inputs and outputs of the operations are strongly typed data objects. Fig 1 illustrates the flow graph of a simple parallel application, describing the asynchronous flow of data between operations



Fig. 1. Flow graph describing data distribution (split), parallel processing, and collection of results (merge)

The Split-Process-Merge construct shows all the basic building blocks for parallel schedules, and is suitable for simple compute farm applications.

III. THE NEW DEVELOPED ALGORITHM

In the proposed algorithm, three major operations are performed, they are: Split, Process and Merge operations. The concept used in this algorithm is: for a task to be executed, it is there in the ready queue. The task waits for its turn to be executed on a master node, and then a task splits into many sub-tasks in the split operation. Then these sub-tasks are executed in parallel as leafs in processing operation on the working nodes. Then the results of the sub-tasks or leafs are combined in merge operation to obtained the final desired result.

Quality of Service Improvement Simple Compute Farm (QISCF) Algorithm:

- 1. Select a Task file.
- 2. Read Data from the task file.
- 3. Read the Resources.
- 4. Categorise the Resources.
- 5. Split the Task into the sub-tasks.
- 6. Read common Resources.
- 7. Execute the Sub Tasks and Process the data.
- 8. Separate it in a file.
- 9. Evaluate the results of the sub tasks after execution.
- 10. Merge the results of the sub-tasks
- 11. Obtain the final results.

IV. SIMULATOR: ANEKA

Aneka is a platform and a framework for developing distributed applications on the Cloud. It harnesses the spare CPU cycles of a heterogeneous network of desktop PCs and servers or datacenters on demand. Aneka provides developers with a rich set of APIs for transparently exploiting such resources and expressing the business logic of applications by using the preferred programming abstractions. Aneka is a Cloud Application Development Platform (CADP) for developing and running compute and data intensive applications. As a platform it provides users with both a runtime environment for executing applications developed using any of the three supported programming models, and a set of APIs and tools that allow the users to build new applications or run existing legacy code.

V. EXPERIMENTAL SETUP

The set up used in the paper work for creating the Aneka cloud infrastructure and for the development of the algorithm, to obtain the desired result is described below in the Table 1 as:

Hardware/ Software	Aneka Master Node	Aneka Slave Nodes
Operating system	Windows 7 Home Basic Service	Windows 7 Home Basic Service
	Pack 1	Pack 1
Processor	Intel [®] Core [™] i5-3317u CPU @	Intel [®] Core [™] i3-2330M CPU @
	1.70 GHz	2.20 GHz
RAM	4 GB	2 GB
System type	64bit Operating system	64 -bit Operating system
Hard Disk Capacity	500 GB	320 GB
.NET Framework	.NET Framework 2, 3.5, 4	.NET Framework 2, 3.5,
Simulator (Aneka)	Aneka 2.0	Aneka 2.0
Programming Environment	Microsoft Visual Studio 2008	Microsoft Visual Studio 2008
Database (optional, if database	Microsoft SQL Server 2005	Microsoft SQL Server 2005
support is required)		

VI. RESULTS

The results showing the comparison in the values of parameters taken for analysis i.e. Latency (KB) and Network's read, write and total (B/sec), obtained by using the normal system method and the QISCF algorithm are shown below. These results can be monitored by the resource monitor of the task manager.

6.1 Parameters for Analyzing the Performance

In general terms, performance is the measure of how quickly a computer completes application and system tasks. Overall system performance might be limited by the latency in the network, average number of bytes per second transferred by the process etc.

• Send/Received(bytes/sec) over network

The number of bytes (bytes/sec) read in input/output operations generated by a process, including device I/Os. I/O Read Bytes directed to CONSOLE (console input object) handles are not counted.

The number of bytes (bytes/sec) written in input/output operations generated by a process, including device I/Os. I/O Write Bytes directed to CONSOLE (console input object) handles are not counted.

• Latency

Latency is a measure of time delay experienced in a system, the precise definition of which depends on the system and the time being measured. The Round trip latency excludes the amount of time that a destination system spends processing the packet. In our paper work we are working for Network Latency in Round Trip time, and the new designed QISCF Algorithm will minimize the latency. The comparison graph for both the methods is being displayed below in Graph 1. The comparison made here is for the network application Google Chrome.

The diagrams here shows both the parameters i.e. the LATENCY and the SEND/RECEIVED (bytes/sec).



Graph 1. Graph representing the Latency comparison (Chrome)



Graph 2. Graph representing the Send/Received (bytes/sec) comparison (Chrome)

The next comparison has been made for the network application Drop Box. The system method and then with the algorithm method and as it is clearly visible in Graph 3 and Graph 4, that both the parameters has been reduced with the algorithm method.



Graph 3. Graph representing the Latency comparison (Drop Box)



Graph 4. Graph representing the Send/Received (bytes/sec) comparison (Drop Box)

The next comparison has been made for the network application Firefox. The system method and then with the algorithm method and as it is clearly visible that both the parameters has been reduced with the algorithm method.



Graph 5. Graph representing the Latency comparison (Firefox)



Graph 6. Graph representing the Send/Received (bytes/sec) comparison (Firefox)

VII. CONCLUSION

In the paper, the cloud QoS parameters improvement problem is being addressed, which is to analyse and improve the quality of service parameters by scheduling the computing resources in a way that providers achieve high resource utilization and users meet their applications' performance requirements. The paper work has been done by developing the QISCF (Quality of Service Improvement Simple Compute Farm) algorithm based on the dynamic parallel scheduling algorithm. This newly developed algorithm will boost the performance of e-learning environment using cloud. The above comparisons in real time described with the help of graphs shows the improvement applied with the

help of the newly developed algorithm and also it opens up several options for the future work.

VIII. FUTURE WORK

There are several avenues of future works that could be pursued as the extensions of this research. The new developed and implemented task scheduling algorithm i.e. QISCF can be refined by including more QoS parameters to ensure that high-priority requests are served first in case of intense competition for resources and limited resource availability. This algorithm is useful in many ways for several purposes and parameters. This can be used for quality of service improvement for many others parameters like CPU usage, response time etc. Here it is used for task scheduling but it can be used for thread scheduling also.

REFERENCES

- [1] A. Fern'andez1, D. Peralta2, F. Herrera2, and J.M. Ben'itez2, L. Uden et al. (Eds.), An Overview of E -Learning in Cloud Computing: Workshop on LTEC 2012, AISC 173, pp. 35–46. springerlink.com c Springer-Verlag Berlin Heidelberg 2012
- [2] Rodrigo N. et al., "The Aneka platform and QoS-driven resource provisioning for elastic applications on hybrid Clouds", Future Generation Computer Systems, Elsevier B.V., 2011, pp. 861–870.
- [3] Aneka 2.0 Developing Task Model Applications by Manjrasoft Pvt. Ltd., 2010.
- [4] Vecchiola C., Chu X., and Buyya R. "Aneka: A Software Platform for .NET-based Cloud Computing", High Speed and Large Scale Scientific Computing, IOS Press, Amsterdam, Netherlands, 2009, pp. 267-295
- [5] Huang Lu, Chen Hai-shan, Hu Ting-ting. "Survey on Resource Allocation Policy and Job Scheduling Algorithms of Cloud Computing", Journal Of Software, Vol. 8: (2), February 2013, Page no: 480-487.
- [6] Buyya Rajkumar, Vecchiola Christian and Selvi Thamarai S. Mastering Cloud Computing, Tata McGraw Hill, New Delhi, 2013.
- [7] Gerlach Sebastian et al. "Fault-Tolerant Parallel Applications with Dynamic Parallel Schedules: A Programmer's Perspective: Dependable Systems", LNCS 4028, Springer-Verlag Berlin Heidelberg pp. 195–210, 2006.

Dynamic Matrix Based Symmetric Key Cryptography

Neetu Yadav Department of Computer Engineering and Applications National Institute of Technical Teachers' Training and Research Bhopal, India <u>neetuy596@gmail.com</u>

M.A. Rizvi Department of Computer Engineering and Applications National Institute of Technical Teachers' Training and Research Bhopal, India <u>marizvi@nitttrbpl.ac.in</u>

R.K. Kapoor Department of Computer Engineering and Applications National Institute of Technical Teachers' Training and Research Bhopal, India rkkapoor@nitttrbpl.ac.in

Abstract—In the current scenario of internet technology, internet has become a defacto medium to communicate with each other. A huge amount of data is shared over the internet security of that data is the biggest concern. To share data in secure form cryptography techniques are used. There are two types of cryptography techniques called symmetric key cryptography technique and asymmetric key cryptography technique. In symmetric key cryptography a single key is used to encrypt and decrypt data.On the other hand, in asymmetric key cryptography two keys: a private key and a public key are used for encryption and decryption purpose. In this paper a dynamic matrix based symmetric key encryption technique is proposed which reduces the drawbacks of the other encryption techniques. In this technique first a matrix is used to store the plain text and then convert that plain text characters into its ASCII form. Then, a random key which has the length equal to the order of the matrix is used to XNOR that text, after that encoding technique is used to encode that matrix and then transpose that matrix and send it to the receiver's end. Reverse operation of the encryption is performed at receiver's end to decrypt data. This technique is secure and also uses a small key as compare to other methods so it takes a small span of time to encrypt and decrypt data.

Keywords: Cryptography, Dynamic matrix, ASCII, XNOR, Transpose.

I. INTRODUCTION

In the last decades, ICT world has witnessed a significant change in the way communication is embedded by the internet. With the evolution of the technology there are mediums like social media, smart phones, etc. widely used to share information. A tremendous amount of data is introduced by these medium and that data also contains private information of the use. The Internet is also used for many military applications to share information. That information contains confidential data. To share the data in secure form a technique is required. Cryptography techniques are widely used to send data in encrypted form. There are mainly two types of cryptography techniques are used one is symmetric key cryptography technique and other one is asymmetric key cryptography technique. There are many techniques are available like DES, AES, BLOWFISH, TDES, RSA and others are used for encryption purpose.



Fig. 1. Cryptography Process

Cryptography is the process to provide a secure way of data sharing. In cryptography data is encrypted by the use of a cryptography algorithm. In this process a key is used to encrypt and decrypt data as shown in the above Fig.1, a key is attached with the plaintext that is used by any cryptography algorithm. The key is known only to the sender and receiver. In that case it is too hard to decrypt the text without having knowledge about the key. In this way it prevents data from any unauthorised access and enhances the security of the data

Cryptography algorithms are of two types- symmetric key cryptography algorithm and asymmetric key cryptography based on the keys used. The symmetric key cryptography algorithms use the same key for encrypting and decrypting the data. A key pair known as public key and private key is used in asymmetric key cryptography algorithm. The data is encrypted by the sender using its public key. At the receiver end the private key is used to decrypt the data.

In this paper a dynamic matrix based encryption technique is presented that transform the information into matrix and uses a random key which is hard to guess for an unauthorized user. So it enhances the security of the data and also reduces the time of encryption as compare to the other algorithms. It also provides accurate results as compared to other techniques, so there is no loss of data during the process.

II. RELATED WORK

In this section, the several methodologies and techniques for the encryption used by various researchers are given as follows:-

UdepalSingh [1] et.al proposed an encryption algorithm based on ASCII values of data. The ASCII values of characters are taken and a random key of four characters is generated from the text. Then, text is encrypted by applying various transformations with the help of key. The limitation is that it cannot encrypt sentential form of data and for each plain text a fixed key of four characters is generated. The proposed system uses variable length key whose length varies as the length of the matrix changes. If anyone has to know key first of all it should know the matrix order then key length.

Devendra Prasad, Govind Prasad [2] proposed a random number generator based approach which use simple transposition and substitution to encrypt and decrypt data, in this technique authors overcome the shortcoming of the other techniques in which fixed key length is used which generated by the user, in that case user need to remember that key and key is also fixed so it is easy to predict the key.

Charru [3]et.al proposed a symmetric key encryption algorithm overcoming the drawbacks of existing approaches that generates key of fixed length .In this algorithm ASCII values of data is used to encrypt the data and a key of variable length is generated depending on the size of the input data using some random code. The encrypted data is then operated with XOR operation to increase security level. The key characters should be known to decrypt the data.

Dr.S.Kiran, Dr.T.Bhashakarareddy[4] proposed an using simple logical and substitution operation to encrypt the text in this technique a poly substitution based technique is used and then logical XOR is performed and 2's compliment is taken of the text which converts it into cipher text and in this way it encrypt and decrypt data.

Sonia Dhull, Sonal Beniwa lPreetiKalra[5] in this paper a study on polyalphabetic encryption technique is proposed in which there are three type of polyalphabetic techniques like vignere cipher technique, Vernam Cipher technique , Onetime pad cipher technique, in all these cipher two types of process are used. Substitution method change on alphabet in to another by using secret key and a key is used to convert plane text into cipher text.

Kondwani Magamba [6] et.al proposed an algorithm that uses maximum distance separable key matrix to encrypt the data using variable length hill cipher. The algorithm enhances the security of data, but use of many matrix operations makes the algorithm inefficient.

JGitanjali [7] proposed an encryption technique based on ASCII values and a key is used to encrypt these ASCII values. The palindrome number and alphanumeric id used by the key are also converted into ASCII values. The palindrome number is generated by using the summation of the private key that makes the message difficult to decode for the attacker. The matrix multiplication used increases the security of the data. The proposed system convert the information into the matrix created dynamically and keys are generated randomly. The information is unknown to attacker until it knows the order of the matrix.

Md. PalashUddin [8] et.al devises an algorithm based on ASCII conversions and a cyclic mathematical function to focus mainly on concealing the text to ensure better security. As the function is cyclic the output rotates between 0 and 31. The security can be improved if random keys are used to encrypt the text and Logical XNOR operation is used in order to make the pattern of the ciphertext unpredictable.

III. PROPOSED WORK

In this paper a dynamic matrix based symmetric key encryption technique is presented.

U. Encryption

In this technique first plain text is converted into a matrix and then ASCII of that plain text is taken and after that a random key is generated which has the size equal to the order of the matrix. Then, converting that matrix in its binary equivalent and take a 1's complement of that matrix. The key is XNOR with that Matrix and a XNOR Matrix is generated then a decimal conversion of that matrix is taken. An encoding technique is used to generate an encoding matrix then that encoding matrix is transposed and send it to the receiver as depicted in Fig.2 Flowchart for encryption.



Fig. 2. Flowchart for Encryption process

V. Decryption

In decryption phase inverse process of encryption phase is performed. Firstly, cipher matrix obtained is transposed and opposite encoding technique that used in the encryption phase is performed. A decimal matrix is generated then a binary equivalent of that matrix is taken. Then, XNOR that binary matrix with the random key. Then, perform 1's complement and convert that matrix into its decimal form. After that convert ASCII values into its character form. Thus, we get original plain text matrix and get the plain text as shown in Fig.3 Flowchart for decryption.



Fig. 3. Flowchart for Decryption process

IV. RESULTS

The result analysis over the parameters called Time, Security and Accuracy for the proposed is performed. The comparison with methods called fixed length key encryption and a variable length key based on the text length is performed.

W. Time

The analysis of time for the proposed method and its comparison with the existing methods is shown in the TableI and TableII.

From the graph as shown in Fig.4, it can be analyzed that the execution time taken by the proposed method is less as compared to the existing methods.

Inpu	Size of	Random key	Encryption	Decryption	Elapsed time	
t	matrix	length as per	time(in	time(in	(in	Security
text	created	the size of	microseconds)	microseconds)	microseconds)	Value
size		matrix				
2	2×2	2	356	211	567	High
4	2×2	2	369	225	594	High
6	3×3	3	389	245	634	High
8	3×3	3	453	334	787	High
10	4×4	4	478	344	822	High

TABLE I. TIME ANALYSIS OF THE PROPOSED METHOD

TABLE II. EXECUTION TIME COMPARISON WITH THE EXISTING METHODS

Input	Encryption system with key	Encryption system with key of	Proposed
text	length equal to input(method-1)	variable length(method-2)[3]	algorithm
size	[1]		
2	3220	683	567
4	3679	702	594
6	3861	1154	634
8	4748	1240	787



Fig. 4. Graphical Analysis of Time

X. Security

In this paper, a dynamic matrix based technique is used for encryption that uses three layers of encryption to encrypt data and a small random dynamic key is generated for each instance that reduces the chances to guess the key for an intruder. Thus, in this way the technique enhances the security of the whole process.

Y. Accuracy

In the process of encryption and decryption there is no loss of data occurs. Thus it provides an exact text after decryption. Hence, this method is accurate as compared to others.

V. CONCLUSION

There are various techniques used to encrypt the data like a random number generator based technique using logical operation, ASCII values based technique using cyclic mathematical function. All these techniques are vulnerable to security attacks and also consume lot of time to encrypt data. The proposed method is based on dynamic matrix and uses a random key with three layers of encryption using ASCII values, Matrix, XNOR operation and 1's complement operation and encoding techniques to encrypt data. Thus, this technique imparts better results as compared to other techniques. A comparative analysis for results of execution time, security and accuracy is presented which shows that this method performs better as compared to other algorithms.

In this paper this technique is used only for plain text data. In future it can also be used to encrypt other types of data like multimedia data image etc.

VI.-REFERENCES

- Udepal Singh, UpasnaGarg," An ASCII value based text data encryption System", International Journal of Scientific and Research Publications, Volume 3, Issue 11, November 2013 1 ISSN 2250-3153.
- [2] Devendra Prasad, Govind Prasad Arya, Chirag Chaudhary, Vipin Kumar, "A Text Encryption and Decryption Technique Using Substitution-Transposition and Basic Arithmetic and Logic Operation", International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014.
- [3] Charru, Paramjeet Singh, Shaveta Rani "Efficient Text Data Encryption System to Optimize Execution Time and Data Security"International Journal of Advanced Research in Computer Science and Software Engineering(IJARCSSE), Volume4,Issue7,July2014.
- [4] Dr. S. Kiran, R.Pradeep Kumar Reddy, I.RajaSekhar Reddy," Multiple Text Cryptanalysis By Using Matrix And Logical Operations", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 5, September-October 2014.
- [5] Sonia Dhull, SonalBeniwalPreetiKalra "Polyalphabetic Cipher Techniques used for encryption purpose" IJARCSSE Vol. 3, Issue 2, February 2013.
- [6] KondwaniMagamba, Solomon Kadaleka and AnsleyKasambara, "Variable-length Hill Cipher with MDS KeyMatrix", International Journal of Computer Applications, Volume 57 - Number 13, November 2012.
- [7] J.Gitanjali, Dr.N.Jeyanthi, C.Ranichandra, M.Pounambal, "ASCII Based Cryptography Using Unique ID, Matrix Multiplication and Palindrome Number", The International symposium on Networks, Computers and Communications, IEEE, 2014.
- [8] Md. PalashUddin, Md. Abu Marjan, NahidBinteSadia, Md. Rashedul Islam, "Developing a Cryptographic Algorithm Based on ASCII Conversions and a Cyclic Mathematical Function", Third International conference on informatics, electronics & vision, IEEE, 2014.

Next Generation Mobile Network Technology: 5G

Shadab Pasha Chameli Devi Group of Institutions Indore, India shadabpasha@gmail.com

M.A. Rizvi National Institute of Technical Teachers' Training and Research Bhopal, India Department of Computer Engineering and applications <u>marizvi@nitttrbpl.ac.in</u>

Abstract— In the near future, i.e., beyond 4G, some of the prime objectives or demands that need to be addressed are increased capacity, improved data rate, decreased latency, and better quality of service. To meet these demands, drastic improvements need to be made in cellular network architecture. In this paper an attempt is made to present the challenges features, requirements and architecture of future generation cellular network technologies i.e. fifth generation (5G) that are helpful in improving the efficiency and meeting the future demands of users.

Keywords : 5G, 5G Architecture, 5G Technology, 5G Features, 5G Requirements

I. INTRODUCTION

Mobile phones have really changed the way of communications. Cell phones are the most used communication device today. Today, mobile communications form an indispensable part of the daily lives of millions of people in the world; a situation which is expected to continue and become even more undeniable in the future. The ability to keep in touch with family, business associates, and access to email are only a few of the major reasons for the increasing importance of cell phones. Today's technically advanced cell phones are capable of not only receiving and placing phone calls, but storing data, taking pictures, and can even be used as walkie talkies, to name just a few of the available options. With the advent of 3G technology these mobile phones are converted into SMART PHONES. To facilitate various mobile apps enabled us to do m-commerce (online shopping), m-banking (e-payment facility) and m-learning etc. When Alexander Graham Bell invented telephone he would never thought of the dynamic usage of today's mobile phone. This transformation took place from 1970 's to till now (i.e. Next Generation) is based on various parameters like data rates, technology, service, Switching, core networks, standards, web standard and others

. From 1G to 3G, telecommunication improved rapidly to offer better end user experience than before. Currently 4G and LTE are being rolled out and this is changing the lives of many individuals, encouraging advanced application revolutions. Deployment of Long Term Evolution (LTE) is already started around the world for faster access with lower latency and higher efficiency than its predecessors 3G and 3.5G. LTE-Advanced, which is an evolution of LTE and a "true 4G" mobile broadband, is under development and its initial service rollout is expected by the end of 2015. The network will have no option but to transform. For unlike 4G which is driven by video, 5G will be driven by the Internet of Things. That means the network will have to be intelligent enough to understand how to allocate its resources, capability it does not have at the moment. For example a self-driving car needs a fraction of a millisecond response every time and maybe not the toaster. That means it will have to be content-aware, user-aware and location-aware.

5G Era can be expected to revolutionize the way we communicate by supporting immersive applications that demand exceptionally high-speed wireless connections, a fully-realized IoT, experience lower latency and promote both spectrum and energy efficiency. To realize these benefits, 5G systems will differ fundamentally from their predecessors fueling a series of groundbreaking innovations. This paper presents our view on the future of wireless communication for 2020 and beyond.

The remainder of the paper is organized as follows: In Section II gives the evolution of wireless technologies technologies. Section III trends in 5G mobile services .Section IV Challenges, Requirements and System Architecture of 5G .In Section V major technologies of 5G Network.

II. EVOLUTION OF WIRELESS TECHNOLOGIES

A 1G : First Generation

The first generation of wireless technology, mobile telecommunications which was first introduced in 1980s and completed in early 1990s. It's Speed was upto 2.4kbps.It allows the voice calls in 1 country. 1G network use Analog Signal.

B 2G: Second Generation.

2G technology refers to the 2nd generation which is based on Global System for Mobile Communication (GSM). It was launched in Finland in the year 1991. 2G network use digital signals. It's data speed was upto 64kbps. It also provides services like Short Message Service (SMS) and e-mail. Prominent technologies were Code Division Multiple Access (CDMA), and IS-95..

C 3G: Third Generation.

Networks were upgraded for the most part between 2004 and 2007 to allow for much more data traffic. Your 3G data may be traveling under CDMA, WCDMA, GSM, UMTS, or a number of other terms and frequencies, but all but "vanilla" 3G basically provides data rates at up to or around 2Mbps (that's 2000Kbps).

D 4G: Fourth Generation.

One of the key challenges of 4G technology to further enhance the mobile experience by integrating all the existing mobile technologies (e.g. GSM-Global System for Mobile Communications, GPRS- Global System for mobile Communications Wi-Fi-Wireless Fidelity, Bluetooth).4G is going to offer three main services Ubiquity,Multi-service platform and Low bit cost. 4G is based on Open Wireless Architecture (OWA), supporting multiple wireless air interfaces in an open architecture platform. Applications that are being made to use a 4G network are Multimedia Messaging Service (MMS), Digital Video Broadcasting (DVB) and video chat, High Definition TV Content and Mobile TV.

E 5G: Fifth Generation.

5G is the next step in the evolution of mobile communication. It will be a key component of the Networked Society and will help realize the vision of essentially unlimited access to information and sharing of data anywhere and anytime for anyone and anything. 5G will therefore not only be about mobile connectivity for people. Rather, the aim of 5G is to provide ubiquitous connectivity for any kind of device and any kind of application that may benefit from being connected.IPv6 is a basic protocol for running on both 4G and 5G.

New mobile generations are typically assigned new frequency bands and wider spectral bandwidth per frequency channel (1G up to 30 kHz, 2G up to 200 kHz, 3G up to 5 MHz, and 4G up to 40 MHz), but sceptics argue that there is little room for new frequency bands or larger channel bandwidths. From users point of view, previous mobile generations have implied substantial increase in peak bitrate (i.e. physical layer net bitrates for short-distance communication). However, no source suggests 5G peak download and upload rates of more than the 1 Gbps to be offered by ITU-R's definition of 4G systems. If 5G appears, and reflects these prognoses, the major difference from a user point of view between 4G and 5G techniques must be something else than increased maximum throughput. For example lower battery consumption, lower outage probability (better coverage), high bit rates in larger portions of the coverage area, cheaper or no traffic fees due to low infrastructure deployment costs, or higher aggregate capacity for many simultaneous users (i.e. higher system level spectral efficiency)

III. TRENDS IN 5G MOBILE SERVICES

The current systems that we are using now, like 3GPP LTE or LTE-A, are commonly called 4G mobile communication systems. In future we will use next generation mobile technology i.e.5G.But before going to understand what 5G is all about ,I will try to explain what mega trends in mobile services are these days .

• **Increased number of devices** – As mentioned above due to large number of mobile devices and things that can be communicated to each other through Internet from 12.5 billion in 2015 to 50 billion in 2018^{[2].} Not only the mobile devices but new things (e.g. wearable device, sensor, actuator, etc.) designed to realize this future 5G mobile service and IoT service are also being brought to market as well.
- Increase of Traffic Demands for multimedia and social network services are increasing day by day. Due to that use of mobile and its traffic has been growing tremendously. Therefore it is mandatory for mobile communication technologies to support huge traffic growth. According to the Cisco VNI Global Mobile Data Traffic Forecast, mobile data traffic is expected to surge about 10 times from 1.5EB (Exa Bytes, 1EB = 1,000,000TB) in 2013 to 15.9EB in 2018^[1]
- **Higher dependency on cloud** Due to the various solutions provided by cloud computing, the dependency on cloud is increasing rapidly. As a result, the transition from the current PC era to the new mobile cloud computing era is likely to be accelerated.[3].Cisco the leading company in networking shared the same point that mobile traffic would increase continuously for 70 % % (twice the current 35%) of the total mobile traffic by 2020^{-[4]} In that respect, most 5G mobile services will most likely be provided through mobile cloud computing systems
- Various mobile convergence services As mentioned above due to large number of mobile devices applications and supported convergence services in various field such as virtual reality, ultra high-accuracy location-based service, hologram service, smart healthcare service, etc. Accordingly, development of 5G mobile communication systems that can satisfy all the requirements of these various services should be followed to ensure seamless and reliable supports for the services. It has been observed that these foregoing four mega trends (traffic growth, increased number of devices, higher dependency on cloud computing, and various 5G convergence services) must be considered in the designing of 5G mobile communication systems.

IV. TWO IMPORTANT CHARACTERISTICS OF 5G SYSTEMS

Unlike the current 4G mobile communication systems, 5G systems have some key characteristics that make them different from their precedent. Below I will closely look into the two most important characteristics of the systems: latency and user experienced data rate.

• End to end latency--Latency is the amount of time a message takes to traverse a system. Most of the previous and current mobile communication systems merely focuses on improving data of a user device. But for 5G systems, more attention is towards end to end latency .Due to various high end applications it is the prime duty of mobile communication systems to deliver visual and auditory information within a tolerable latency of about 10 ms and auditory information within a bout 100 ms ^[5]. If the latency is longer, users feel disturbed, unable to enjoy the seamless service. Services like vehicular-to-vehicular (V2V) communications or vehicle-to-infrastructure (V2I) communications require extreme low latency to ensure traffic safety-related services are provided in time. Tactile internet one of the core feature of 5G mobile communication , can support a low latency mobile communication service capable of delivering tactile information in time



Fig .1 Changes of PLT by bandwidth and round trip time changes

In order to make user experience in a mobile communication system as natural, it is necessary to bring end to end latency not more than 10 ms because it is known fact that it takes up to 120m/s for an electronic signal to be delivered through a human's nervous system. This means it takes less than 10 ms for tactile information from a user's hand to reach his brain. Therefore end-to-end latency less than several ms should be guaranteed. In Fig -1 I illustrate the effects of data rate (bandwidth) and round trip time (RTT) on HTTP's page load time (PLT). As seen in the figure, data rate improvement does not have significant effects on reducing HTTP PLT once it exceeds certain level.

From the above observation as shown in Fig.-1, it is clear that if we want to improve data rate without considering RTT would not be effective in further reducing PLT. To address these issues in 5G active R&D on low-latency mobile communication technologies and services in various aspects like physical, medium access control (MAC), network, and transport layers should be followed.

User Experienced data rate -- In addition to its No. 1 goal, achievement of lower end-to-end latency, 5G mobile communication systems have another goal, improvement of user experienced data rates. Then what is the difference between peak data rates and experienced user data rates? ITU- defines them as follows^[6]

- Peak data rate: peak data rate refers to the maximum achievable data rate per user. Future IMT systems should provide very high peak data rate capability that leads to high network capacity enabling new differentiated services and enriching the end user experience.
- User experienced data rate: user experienced data rate is defined as the minimum data rate per user that should be achievable anytime anywhere. Future IMT systems should have the capability to provide anytime, anywhere [gigabit] data rate experience to mobile users. If we want consistent and guaranteed user experienced data rates, 5G mobile communication systems should be able to provide users with high data rates in gigabit anytime and anywhere. It has been observed that performance is degraded at cell edges. So, improving user experienced data rates through upgrading cell edge performance as shown in Table II, has been one of the most important goals that 5G mobile communication systems must achieve.

So far, in this paper I have discussed four trends (traffic growth, increased number of devices, higher dependency on cloud, various mobile convergence services). We also learned 8 key system requirements that ITU-R derived from the trends (user experienced data rates, peak data rates, mobility, latency, connection density, energy efficiency, spectrum efficiency, and traffic volume density). As such, compared to 4G and earlier systems, 5G has additional and different requirements to satisfy, including latency and user experienced data rates, the two most important requirements discussed in detail here. Other important issues to be considered when developing 5G mobile communication systems include: Connection density related to the increase of devices/access points (AP)/base stations and use of more complicated/efficient protocol/algorithm.

V. CHALLENGES OF 5G NETWORK

There are various challenges of 5G mobile communication systems but these three challenges are the most important one. These are as follows

- Secure 5G network that can take care of ultra-high speed and massive connectivity (IoT).
- Build an intelligent SW platform based on real-time analysis of big data and virtualization.
- Provide ICT convergence services by taking advantage of innovative User Interfaces and High Definition multi-media.
- Technology changes rapidly previously in 2G,3G &4G, when new next generation mobile technology introduced users adopt it and replaced the previous one with new one. In 5G network, it co-exists with LTE network through cooperating instead of replacing, with it. In 5G mobile giga speed and massive connectivity are important which is going to be evolved through LTE technologies. This allows 5G revolutionary technologies and various access networks to be integrated based on LTE networks. small cells were introduced in the LTE-A network, which made the network begin to evolve into one with C-RAN based HetNet architecture. As shown in the below table I there is a comparison between SmallCell & HetNet

Parameters	Femtocell 3G	Wi-Fi	DAS	Indoor SmallCell	Outdoor Smallcell	HetNet LTE
Motivation	Indoor coverage improvement and voice traffic offload	Indoor coverage improvement and 3G data traffic offload	Indoor coverage and neutral host deployment style	Coverage and Capacity	Outdoor Capacity	Outdoor Capacity
Demand	Voice 3G	3G mainly & LTE data	All demands (2G,3G,LTE,Wi- Fi) etc.	LTE,3G	LTE,3G	LTE
Frequency	Licensed 2100 Mhz	Outdoor: 2.4 Mhz Outdoor hotspot: 5 Mhz Indoor : 5Mhz	Licensed and Unlicensed	Preference of higher frequency above 1 Ghz	Preference of higher frequency above 1 Ghz	Preference of higher frequency above 1 Ghz
Latency Requirement	Milliseconds	Milliseconds	NA	Milliseconds	Milliseconds	Microseconds
Throughput Requirement	10 Mbps	Depends upon AP Version	NA	3G:N x 20Mbps LTE:N x 150 Mbps	3G:N x 20Mbps N x 150 Mbps	>5 Gbps Uncompressed CPRI for MIMO
Backhaul & Fronthaul	Customer Premises , ADSL	Metro Ethernet : EPON /PTN	Dark Fiber	Backhaul S1 over IP Metro Ethernet/EPON/PTN	NLOS Radio SON Multipoint System Metro Ethernet/EPON/PTN	CPRI/OBSAI over Fiber CPRI over E-band Radio SON Multipoint System

TABLE-I HetNet Architecture Cases and Concerns

VI. 5G REQUIREMENTS AND TARGET VALUES

Any technology which is used in market would have multidimensional requirements. These requirements are categorized in terms of user-who use the technology, performance –experience by user & comparison with previous technologies, How it is going to provide services in a structured way and in the end how this technology is managed so that there would be no problem regarding the smooth usage of technology.

i) in terms of user, for better user experiences, 1 Gbps should be available at cell edges, and data delivery between user and base station should be less than 1 ms. As shown in Table I.

ii) in terms of network, the network should have 1,000 times higher capacity, be more intelligent and SW-based, provide sufficient reliability to allow no hacking, and require low maintenance costs There are eight requirements

Key Performance Indicators	Target Values
User Experienced Data Rate	100Mbps – 1Gbps
Peak Data Rate	10Gbps – 50Gbps
Mobility	Up to 500Km/h
Latency	~ 1ms (radio interface)
Connection Density	$106 - 107 \text{ per Km}^2$
Energy Efficiency	$50 \sim 100$ times more efficient than IMT-A
Spectrum Efficiency	$5 \sim 15$ times more efficient than IMT-A
Traffic Volume Density	1TB - 10 TB/s/Km ²

TABLE-II 8 key performance indicators by ITU-R

VII. 5G SYSTEM ARCHITECTURE

5G system architecture consists of the three layers: Resource, Business enablement layer, and Business application Layer

Infrastructure resource layer –The physical resources of a fixed-mobile converged network, comprising access nodes, cloud nodes (which can be processing or storage resources), 5G devices, networking nodes and associated links. The resources are exposed to the higher layers and to the orchestration entity through virtualization principles.

Business enablement layer – Library of all functions required within a converged network in the form of modular architecture building blocks, including functions realized by software modules that can be retrieved from the repository to the desired location, and a set of configuration parameters for certain parts of the network, e.g., radio access. Those functions and capabilities are called upon request by the orchestration entity, through relevant APIs.As shown in Table II.

Business application layer – Specific applications and services of the operator, enterprise, verticals, or third parties that utilize the 5G network. As shown in Table III different technologies are used for different category.

VIII. 5G MAJOR TECHNOLOGIES

There are many technologies based on various categories

I ABLE-III Category wise major technologies				
S.	Category	Technologies		
No				
1	Transmission of Tactile Multimedia	MPEG Media transport MMT Technology		
2	Cloud based all IT Network	NFV Based virtualized core network, Virtualized RAN		
3	Fast Flexible Transport Network	Packet Optical Transport Network(POTN)		
4	Beyond Cellular Network Architecture	D2D Communication, Contents centric networking (CCN)		
5	Enhanced operation for Multinet / HetNet	Elastic Cell, Multi RAT aggregation (LTE + Wi-Fi)		
6	Massive Input Massive Output(MIMO)	Bean forming		
7	Advanced IoT and New Waveform	Cellular Based Machine Type Communication , NOMA, FBMC		

IX.CONCLUSION

5G will expand the capabilities of the network far beyond what is possible today, exploiting the latest technologies both in the radio and network domains. The many initiatives and discussions on 5G going on around the world by governments, vendors, operators and academia demonstrate the continuing ethos of collaboration and innovation across the industry. 5G systems will need to support on demand composition of network functions and capabilities, thereby creating 5G network slices utilizing the same 5G infrastructure. While the foundational technologies, design principles and basic concepts to achieve such a system in 5G were summarized. In this paper we have discussed various issues of technology. The primary goal of this paper is to understand the features, requirements and technologies that support 5G. It is anticipated by various countries that 5G would commercially available by 2020.

X.REFERENCES

- [1] VNI Global Mobile Data Traffic Forecast 2013-2018, Cisco,
- [2] J. Gubbi et al., Internet of Things (IoT): A vision, architectural elements, and future directions, Elsevier Future Generation Computer Systems, pp.1645-1660, Feb. 2013.
- [3] Forbes, 6 Big Internet Trends to Watch for in 2012, Dec, 2011
- [4] The Mobile Economy, GSMA, 2014
- [5] M. T. G. Pain and A. Hibbs, "Sprint Starts and the Minimum Auditory Reaction Time," J. Sports Sciences, vol. 25, no. 1, pp. 79–86, Jan. 2007
- [6] IMT Vision Framework and overall objectives of the future development of IMT for 2020 and beyond, ITU, Feb. 2014

Analysis of Data Gathering Protocol and Security Issues during Data Aggregation in WSN

Bhat Geetalaxmi Jayram Department of Information Science and Engineering, National Institute of Engineering, Mysore, Karnataka, India. <u>g.nettar@rediffmail.com</u>

D. V. Ashoka

Department of Information Science and Engineering JSS Academy of Technical Education, Bangalore, Karnataka, India. <u>Ashok_d_v@rediffmail.com</u>

Abstract—In wireless sensor networks energy consumes will takes place during data collection. To improve the lifetime of WSN reducing energy consumption is the major challenge. In this paper authors have analysed data gathering protocol which uses multiple mobile sinks with minimum tour length to collect the data from cluster members. This paper also elaborates on main issues involved in implementing security in WSN like size of encrypted message and execution time for encryption at sensor. Authors are going to use the results obtained in this analysis in their research work which concentrates on improving the existing data gathering protocols efficiency to improve lifetime of WSN.

Keywords: Analysis, Consumption, Data Gathering, Efficiency, Privacy Homomorphism

I.INTRODUCTION

WSNs are typically self organizing ad-hoc systems that consist of many small, low cost devices. Each node is equipped with sensor, microprocessors, memory, wireless transceivers, and battery. Once deployed the sensor nodes form a network through short-range wireless communication. They monitor the physical environment and subsequently gather and relay information to one or more sink nodes. The sensed data are relayed towards the sink in hop by hop manner. The energy consumption of the network can be minimized if the amount of data that needs to be transmitted is minimized. WSNs are more and more used for delay sensitive applications such as battle field monitoring. In these applications delay reduction between data generated and data processing becomes mandatory.

II. DATA GATHERING AND ITS NECESSITY IN WSN

Data gathering is recognized as one of the basic distributed data processing procedures in wireless sensor networks for saving energy and reducing medium access layer contention[1]. Data gathering mechanism performs in-network aggregation of data which is needed for energy-efficient information flow[2]. Data gathering protocols can reduce the communication cost, thereby extending the lifetime of sensor networks. The inherent redundancy in raw data collected from the sensors can often be eliminated by in-network data gathering[3]. Data gathering can reduce the number of data packets transmitted and the data conflict thus raises the data accuracy and data collection efficiency through dealing with the redundant data in-network.

III. RELATED WORK

David Jea et al. [4] have showed that as the network scales, using a single mobile element may not be sufficient, and would require multiple of them. The sensor nodes and (or) the mobile elements may not be uniformly placed in practice, necessitating the use of load balancing, so that each mobile element as far as possible, serves the same number of sensor nodes. To overcome this problem, they presented a load-balancing algorithm, and described the mechanism for using multiple mobile nodes for data gathering in wireless sensor networks.

Yanzhong Bi et al. [5] have presented a data-gathering scheme for sensor network with a mobile sink. In this scheme, they distribute three key tasks of a data-gathering period, which are moving the sink, collecting data, and notifying sensors of the sink's positions into three separate phases. The scheme provides design flexibility because of the loose coupling among the three phases. Under the scheme, they have proposed an autonomous moving strategy to take advantage of sink mobility to balance energy consumption among sensor nodes and prolong network lifetime.

Prem Prakash Jayaraman et al. [6] have proposed the use of context aware mobile devices as it carriers data from sensor. The mobile nodes have enough spare capacity to form a distributed shared access network that can be used for sensor data

collection and delivery. This is made possible with available powerful and context aware mobile devices that can work in pervasive environments like VPAN. This paper has proposed the architectural framework of the components on the mobile node that act as the data carrier.

Min Xiang et al [7] proposed energy-efficient intra-cluster data gathering. This technique is designed for preserving the energy as much as possible and increasing life of the network. Here every cluster head works as local control center and will be substituted by the candidate cluster head only when its operation times reach to its optimum value. This will increase the energy consumption ratio of data collected to broadcasted message. But this technique has not considered the buffer overflow of the cluster head.

Deepak Puthal [8] has proposed Mobile Sink Wireless Sensor Network (MSWSN) model which uses a mobile sink to gather the data. In this mobility model the sink moves in the network and covers the whole network area. But this technique has not addressed the visiting schedule of the mobile sink.

IV. ENERGY EFFICIENT DATA GATHERING (EEDG) ADHOC ROUTING PROTOCOL [9]

This protocol uses <u>Vehicle Routing Problem</u> to find Minimum Tour Length (MTL) that Mobile Sinks (MS) can follow to collect data from nodes. Job of MS is to visit certain places and collect data from sensor nodes and send it to centralized Base Station (BS). Different numbers of MS are taken based on different scenarios consisting of different number of nodes. For each MS few number of sensor nodes are attached. In this MS starts collecting data from one point and reach to another point by covering certain nodes visiting only once, so that entire trajectory should be minimum. In this protocol MTL can be achieved by finding nearest neighbor based on earliest deadline. A. *Simulation of EEDG Protocol*

To implement EEDG protocol authors have used NS-2.32 All- in-One tool. Implementation is divided in to two parts:

- 1) Front end design using TCL scripting: Front end includes topology design and connection procedure using Tcl scripting.
- 2) Back end design using C++ : Back end includes implementation of EEDG protocol.

Wireless sensor environment is created using flat grid topology over area 500 X 500. Energy model is included to obtain energy available in each scenario. For Performance Analysis, authors have taken 5 different scenarios i.e. for 20 nodes with 3MS and 1BS, 40 with 4MS and 1BS, 60 with 4MS and 1BS,80 with 5MS and 1BS,100 with 5MS and 1BS. By keeping rate constant as 250kb for each scenario authors have taken the reading for 5 performance metrics like Delay, Packet Drop, Packet Delivery Ratio, Energy available, Control Overhead.

V. RESULTS AND PERFORMANCE ANALYSIS OF EEDG PROTOCOL

Data files for the above performance metrics are collected for 5 different scenarios i.e. 20, 40, 60, 80, 100 nodes. Table 1 shows simulation statistics of EEDG protocol. Based on the analysis it is clear that as number of nodes increases control overhead increases, energy availability increases, packet delivery ratio decreases, delay increases, packet drop increases. As per as energy efficiency of EEDG protocol is concern ,it is good but performance of other performance metric is not up to the mark.

TABLE I. SIMULATION STATISTICS OF EEDG PROTOCOL						
Performance	Different number of nodes topology					
Metrics	20	40	60	80	100	
Overhead	10634	16544	15540	17748	25109	17.115
Energy	11.6534	11.90821	12.45436	12.23052	13.44729	12.33875
Drop	2939	10234	9896	11775	20784	11.1256
Delay	9.175127	12.97177	13.44185	13.6464	19.69865	13.78676
Delivery Ratio	0.726406	0.397086	0.435005	0.395947	0.248945	44.0678



Fig. 1. Performance Metrics Vs Average

Last column of Table 1 represents the average values obtained for the performance metrics. Using these values bar graph is plotted against performance metrics which is shown in Fig.1. This Graph will help authors to compare efficiency of these protocols with any other protocol in their future work.

VI. SECURITY ISSUES IN DATA AGGREGATION [10]

This protocol is energy efficient but not indulged with security of data during aggregation. Prime importance has to be given to security during transmission and aggregation of data while designing sensor network. Availing secure aggregation of data is a challenging task. Main issues involved in implementing security in WSN are size of encrypted message and execution time for encryption at sensor during implementing data encryption at sensor [11].Privacy Homeomorphisms (PH)are encryption functions which allows a set of operations to be performed on encrypted data without the knowledge of decryption function. PH uses positive integer $d \ge$ for computing the secret key. The size of encrypted data increases by a factor of d compared to the original data. Execution times for encryption operation at the sensor increases with d. For instance when d=2, the execution time for encryption of one byte of data in 3481 clock cycles on a MICA2 mote which increases to 4277 clock cycles when d=4 as reported in [11].

The other main aspect of security in sensor networks is the establishment of secrete keys between the sensor and the base station. In the approach proposed in [12] all nodes trust the base station at the network creation time and each node is given a master key which is shared with the base station. To achieve authentication between a sensor and base station, a message authentication code (MAC) is used. The keys for encrypting the data and computing the MAC are derived from the master key using a pseudo random function. All keys derived using these procedures are computationally independent. Hence, if an attacker hacks the key, it would not help in determining the master key or any other key. In scenarios where a key is compromised a new key cam be derived without transmitting confidential information. But issues which are not addressed so far is that, some sensor nodes may be compromised and transmit wrong data values to aggregator that corrupts the aggregation result. Another one is What if the aggregator may be compromised and report malicious aggregate values to base station.

VII. CONCLUSION

Authors have implemented data gathering protocol using mobile sinks which uses vehicle routing problem to find minimum tour length to collect the data from sensor nodes and focused on optimizing important performance metrics Energy Availability. Security is another important aspect in data gathering applications. Even though some research has been pursued in integrating security in data gathering protocols, there is significant scope for future work which can include unaddressed issues like compromise of sensor nodes and aggregator while achieving security in WSN.

XI. ACKNOWLEDGMENTS

This work is sponsored and supported by grant from VGST, Govt. of Karnataka (GRD-128). The authors wish to thank Dr. S. Ananth Raj, Consultant, VGST, and Prof. G.L. Shekar, Principal, NIE., for their encouragement in pursuing this research work.

XII. REFERENCES

- Zhenzhen Ye, Alhussein A. Abouzeid and Jing Ai," Optimal Policies for Distributed Data Aggregation in Wireless Sensor Networks", in Proc.26th, Int. Conf. on Computer Communication, IEEE, Anchorage, AK, May 2007, 1676 – 1684.
- [2] Bhaskar Krishnamachari, Deborah Estrin, Stephen Wicker," The Impact of Data Aggregation in Wireless Sensor Networks", in Proc.22,th Int. Conf on Distributed Computing Systems Workshops, 2002, pp.575 – 578.
- [3] Kai-Wei Fan, Sha Liu, and Prasun Sinha, "Structure-free Data Aggregation in Sensor Networks", IEEE transaction on Mobile Computing, Vol. 6, No. 8, August 2007, pp 929- 941.
- [4] David Jea, Arun Somasundara and Mani Srivastava, "Multiple Controlled Mobile Elements (Data Mules) for Data Collection in Sensor Networks", *in Proc. 1^{st,} Int. Conf* on DCOSS, USA, June 30 – July 1, 2005 pp 244-257.
- [5] Yanzhong Bi, Limin Sun, Jian Ma, Na Li, Imran Ali Khan and Canfeng Chen, "HUMS: An Autonomous Moving Strategy for Mobile Sinks in Data Gathering Sensor Networks", EURASIP *Journal on Wireless Communications and* Networking, Volume 2007, March 2007,pp1-15.
- [6] Prem Prakash Jayaraman, Arkady Zaslavsky and Jerker Delsing, "Sensor Data Collection Using Heterogeneous Mobile Devices", Int. Conf on Pervasive Services, IEEE, pp 161 – 164(2007)
- [7] Min Xiang et all," Energy-Efficient Intra-Cluster Data Gathering of Wireless Sensor Networks", Int. Journal of Hybrid Information Technology Journal of Networks, Vol. 5, Issue 3, pp 383-390. March(2010)
- [8] Deepak Puthal, Bibhudatta Sahoo, Suraj Sharma, "Dynamic Model for Efficient Data Collection in Wireless Sensors Networks with Mobile Sink", IJCST, Vol. 3, issue 1, Jan-March (2012)
- [9] Khaled Almi'ani et al," Energy-Efficient Data Gathering with Tour Length-Constrained Mobile Elements in Wireless Sensor Networks", in Proc. 35,th Int. Conf on Local Computer Networks (LCN), 2010, pp 582 – 589.
- [10] Rajagopalan, Ramesh and Varshney, Pramod K., "Data aggregation techniques in sensor networks: A survey," (2006). *Electrical Engineering and Computer Science*. Paper 22.
- [11] E.J. Duarte-Melo and M. Liu,"Data-Gathering wireless sensor networks : organization and capacity," Computer Networks: The International Journal of Computer and Telecommunications Networking, vol.43, no.4, Nov.2003.
- [12] J. Girao, D. Weshoff and M. Schneider, "CDA: Concealed Data Aggregation for Reverse Multicast Traffic in Wireless Sensor Networks," *IEEE International Conference on Communications*, vol.5, 2005, pp.3044-9.

Adaptive Threshold Memristor Model

Prof. Ibraheem: Department of Electrical Engineering: Faculty of Engineering and Technology: Jamia Millia Islamia New Delhi- 110025 INDIA ibraheem 2k@yahoo.com

Prof. Shahida Khatoon: Department of Electrical Engineering: Faculty of Engineering and Technology: Jamia Millia Islamia New Delhi- 110025 INDIA <u>skhatoon@jmi.ac.in</u>

Md. Isityaque: Department of Electrical Engineering: Faculty of Engineering and Technology: Jamia Millia Islamia New Delhi- 110025 INDIA isitjmi@yahoo.com

Akbal Hanoon: Department of Communication Engineering Amra institute, Southern Technical University, IRAQ: <u>akbalhanoon@yahoo.com</u>

Abstract—This paper discusses various models/window function of memristive devices. These devices are hardly available commercially, but they can be used for various applications ranging from memory and logic to neuromorphic systems. Memristor devices have many advantages like; it is compatible, have no leakage current, good scalability, nonvolatility, and compatible with CMOS technology. In this paper, a new window function/memristive device model ie. Adaptive Threshold Memristor Model (ATM) is presented which can be used with any practical mem device. It enables the device with a fair margin of flexibility and also reduces the size of the device. The present study also incorporates the comparative study of existing models with the proposed ATM model.

Keywords: Memritor, memristive systems, window function.

I. INTRODUCTION (*HEADING 1*)

The "memory resistor" which is popularly known as memristor was termed by circuit theorist Leon Chua in 1971. It is also a fourth class of electrical circuit, joining the resistor, the capacitor, and the inductor. that exhibit the unique properties primarily at the nanoscale. Memristor with variable resistance can store the total charge that pass through the device. Theoretically memristor is a state or series of being linked together. And is a passive circuit element that maintain a relationship between the time integrals of current and voltage across a two terminal element.



Fig1. The four fundamental two-terminal circuit elements: resistor, inductor, capacitor and memristor

$$\frac{d\,\omega}{dt} = f(\omega, i) \qquad \dots (1)$$

$$V(t) = \mathbf{R}(\mathbf{w}, \mathbf{i}) \mathbf{i}(t) \qquad \dots (2)$$

Here i(t) is the memristive device current, w is internal state variable, v(t) is the memristive device voltage and R(w,i) is the memristance of the device. The standard and specific definition of memristor and associated technical terms are described in [9,11]. Hewlett-Packard in 2008 had described the fabrication of the working memristive device in [2]. New device of memristor exhibiting memristive behavior that have appeared in [4], [5]. However the existing devices such as spin transfer torque magneto resistive random access memory (STT-MRAM) are redefined in terms of memristive system [6].

Memristive device can be used in various places for different applications such as physical neural network system [8], memory [7], analog circuit [10], [19], and logic design. Various characteristics are important for the effective use of memristive devices in each of these applications. For being a unique designer a friendly physical model of a memristive device is required. In this paper the characteristics of memristive device are presented in section II. The published memristive device models are described in section III. A new model i.e. Adaptive Threshold Memristive Model which has the merits is desirable in section IV. And In section V a comparison have been made between the various models. They were carried out in the paper is summarized in section VI.

II. CHARACTERISTICS (MEMRISTOR)

The basic difference between a transistor and memristor is that the memristor is the time integral of current which is the total charge that has been pass through the device "widrow 1960". Memristor resistance is not controlled by the

Instantaneous current. Memristor named for memory-resistor that has an electrical resistance and it depends on the history of the current that pass through the device namely, how much current is flow and in what direction. In general the amount of current pass through it more makes it easy to work in future. And when the power of the device is turned off, it remembers its most recent resistance value until power is turned back on. Or in a transistor once the flow of electrons is interrupted by say cutting the power all information is lost. But a memristor can remember the amount of charge that was flowing through it, and much like a memory stick it will retain the data even when the power is turned off.

III. MULTISTABLE STATE MEMORY



Fig2. Memristive device model shows linear ion drift. This device having two regions: doped region and the undoped region. The total resistance is the combination of both the device.

This is the basic unique property of this nonlinear element to store distinguishable Boolean data in a memristive device. It requires a high ratio between the resistances named R_{ON} and R_{OFF} . Several additional characteristics are important for all applications, such as low power consumption, good scalability, and compatibility with conventional CMOS. STT-MRAM exhibits these characteristics except for the high On/Off resistance ratio [11]. Memristive device model and circuit based applications can be design and analyze, these models required distinguishing features. A memristor which is basically two terminal circuit element that will provide hysteresis loops in an i-v plot when an alternating voltage signal applied. An example of an ideal model of a memristor is shown in figure 2. The hysteresis loops are very important when memristive systems are to be identified and the loops normally run through the origin in an i-v plot. So that hysteresis loops will divide here in two half that is positive half and negative half, with mixed positive half and negative half.

PREVIOUSLY PUBLISED MEMRISTIVE MODEL

Requirements From a Memristive Device Model

Several requirements needs to satisfy for a memristive device model: like it must be accurate sufficiently, efficient and problem solving. It is appropriate for this model to be simple, spontaneously derived and closed form. For the general model It is also preferable, so that it can be adjust to suit different types of memristive devices.

Drift Model of Linear ion

Linear ion drift model is suggested for a memristive device model in [3] in this model this device whose physical width D comprise two regions which is shown in figure- 2. The width of first region is W (that is acting like the state variable of the system) having high concentration of dopants (in the first place the oxygen vacancies of TiO_2 , namely TiO_{2-X}). The oxide region of width d-w is the second region (originally TiO_2). Region with the Dopants side have a higher conductance than the oxide region because oxide region has the lower conductance. So distinct and individual assumption are made: Linear ion drift, Ohmic conductance, are in a uniform field, and ion mobility μ_v have equal amount, when takeout the average

Equation (1) and (2) become becomes

$$\frac{d\omega}{dt} = \mu_V \frac{R_{ON}}{D} i(t) \qquad \dots(3)$$

$$V(t) = \left(R_{ON} \frac{W(t)}{D} + R_{OFF} \left(1 - \frac{w(t)}{D}\right)\right) i(t) \qquad \dots(4)$$

IV. Window Function

Here R_{ON} is the low resistance when w(t) = D, and R_{OFF} is the high resistance when w(t) = 0. This state variable w(t) is defined to the physical dimensions of the device, ie, the vale is within the defined interval [0, D]. The window function is a function of the state variable and it is defined as $f(w) = w(1 - w)/L^2$. The boundary condition here is f(0) = 0 and $f(L) = \frac{l-1}{L} \approx 0$. It meets the indispensable boundary condition ($\omega \rightarrow boundraies$) = 0, There is a slight difference when $\omega \rightarrow L$

. The problem of this boundary assumption is when a memristor is driven to the terminal state $\omega \to 0$ and $\omega \to L$, $\frac{d\omega}{dt} \to 0$,

the memristor state will not change even when the external field will change. This is the basic problem of the window function. The second problem of the window function is it persume that the memristors remembers the exact amount of charge that pass through them.



Fig. 7 Non- linear ion drift model I-V curve. (a) Input Sinusoidal voltage for several frequencies and (b) Input voltage of Rectangular waveform.





W(n)=
$$\begin{cases} \frac{1}{2} [1 + \cos 2\pi (\frac{x}{\alpha} - 1)], & 0 \le \frac{x}{\alpha} \le 0.3 \\ 1, & 0.3 \le \frac{x}{\alpha} \le 0.7 \\ \frac{1}{2} [1 + \cos 2\pi (\frac{x}{\alpha})], & 0.7 \le \frac{x}{\alpha} \le 1 \end{cases}$$

In Benderli and Wey, proposed another window function that has been proposed that is little different from Strukov- et al. This window function,

$$f(\omega) = \omega \left(\frac{L - \omega}{L^2}\right) \qquad \dots (6)$$

Approaches zero when $\omega \to o$ and when $\omega \to L$ then $f(\omega) \to 0$. Therefore, this window function meets both the boundary conditions. And it imitates the first function when we consider $x = \frac{\omega}{L}$ instead of ω . There is another problem associated with these two window functions, namely the modeling of approximate linear ionic drift when $0 \ge \omega \ge L$. Both the window functions prodromakis and piecewise window function approximate nonlinear behavior when the memristor is not in its terminal states, $\omega = 0$, or $\omega = L$. This problem is addressed in Joglekar and Wolf where they propose an intresting window function to address the nonlinear and approximately linear ionic drift behavior at the boundaries when $0 \ge \omega \ge L$. The most important problem associated with this model is revealed at the boundaries. On the basis of this model, when the memristor is at the terminal states, no external stimulus can change its state. Biolek- et al tackles this problem with a new window function that depends on x, p, and memristor current, i. Basically x, and p are playing the same role in their model and the only new idea is using current as an extra parameter.

V. COMPARISON BETWEEN THE MODELS

In this section, ATM (Adaptive Threshold Memristive Model) a unique and simple memristive device model have been presented. The essential section of the ATM model is based on an expression for the derivative of the internal state variable that can be fitted to any memristive device type. In this the current and voltage relationship is undefined rather than the other memristive device models and that can be freely chosen from any current-voltage relationship.

$$f_{off}(x) = \exp\left[-\exp\left(\frac{x - \alpha_{off}}{\omega_c}\right)\right] \qquad \dots (7)$$
$$f_{on}(x) = \exp\left[-\exp\left(\frac{x - \alpha_{on}}{\omega_c}\right)\right] \qquad \dots (8)$$

The TEAM model is the most precise and accurate model of a TiO_2 memristive device. TEAM model is still quite complicated. This model is computational and also inefficient. Model with like expressions alternately than the complex equations in TEAM is desired, so that the delicacy of the simple model must be sufficient. This simplified model in equation (7),(8) represents the same and simple physical behavior, but with simpler mathematical functions.

VI. CONCLUSION

The proposed Adaptive Threshold Memristor Model is consisting of different window functions and is more accurate, simple and flexible that can used conveniently. It has almost all salient features of existing different memristive devices. This model proposed is a practical memristive device that is capable to display a threshold and nonlinear dependence on the charge flow. The simulation study carried out demonstrates the superiority of the proposed model over the others. The average is limited between 0.2 - 0.4% which is very small as compared to that offered by other window functions.

VII. REFERENCES

- [1] Leon. O. Chua, "Memristor—The missing circuit element," *IEEE Circuit Theory*, sep. 1971, pp. 507-519, Vol, CT-18, no. 5. K. Elissa, "Title of paper if known," unpublished.
- [2] B. Strukov, G. S, Snider, D. R. Stewar, and R. S. Williams, "The missing memristor found," *Nature*, May 2008, pp. 80-83, vol. 453.
- [3] S. M. Kang, and Leon. O. Chua, "Memristive devices and systems," Proc. IEEE, Feb. 1976, pp. 209–233, vol. 64, no. 2.
 - a. Leblebici, D. Sacchetto, S. Carrara, M. H. Ben-Jamma, and G. DeMicheli, "Memristive devicessfabricated with silicon nanowires schottky barrier transistors," in proc. IEEE International. Symp. Circuits System, May/Jun. 2010, pp. 9-12.
- [4] Campbel, A. Timilsina, A Oblea and, K. A. I, "Compact method for modeling and simulation of memristor devices: Ion conductor chalcogenide-based memristor devices," in proc. IEEE/ACM International. Symposium. Nanoscale Archit, Jun. 2010, pp. 1-4,.
- [5] Dimitrov, Y. Chen, H. Xi, and, X. Wang, "Spintronic memristor through spin-torque-induced magnetization motion," *IEEE Electronics Device Lett.*, Mar. 2009, pp. 294-297, vol. 30, no. 3.
- [6] M. Huang, Y. Ho and P. Li, "Nonvolatile memristor memory: Device characteristics and design implications," in proc. IEEE int. conf. Comput.-Aided Design, pp. 485-490, Nov. 2009.
- [7] Raissi A.Afifi, and A. Ayatollahi, "Implementation of biologically plausible spiking neural network models on the memristor crossbar based CMOS/Nano circuits," in proc. Eur. Circuit Theory Design, pp. 563-566, Aug. 2000.

a. Biolek, V. Biolkova, and D. Biolek, "Pinched hysteresis loops of ideal memristors, mecapacitors, and meminductors must be 'self crossing'," Electron. Lett, Dec. 2011, pp. 1385-1387, vol. 47, no. 25.

- [8] Snider, "Computing with hysteretic resistor crossbars," Applied. Physics. A, Mater. Science. Process., Mar. 2005, pp. 1165-1172, vol. 80, no. 6.
- [9] L. O. Chua, "Resistance switching memories are memristors, "Applied physics. A, Mater. Science. Process., Mar. 2011, pp. 765—783, vol. 102, no. 4.
- [10] Y. Huai, S. wang, Z. Li, Z. Dio, A. Panchula, Y. Ding, E. Chen, and L. C. wang, "Spin-Transfer torque switching in magnetic tunnel junction and spin transfer torqueaccess memory," *J. Phys., Condens. Matter*, Apr. 2007, pp. 1-13, vol. 19, no. 16.
- [11] S. J. Wolf and Y. N. Joglwkar, "The elusive memristor: properties of of basic circuits," Eur. J. Phys., Jul. 2009, pp. 661-675, Vol. 30, no. 4.
- [12]Biolek, Z. Biolek, and V. Biolkova, "SPICE model of memristor with nonlinear dopant drift," *Radio engineering*, Jun. 2009, pp. 201-214, Vol. 18, no. 2, pt. 2.
- [13] M. D. Picket, J.J. Yang, D.A. A. Ohlberg, X. Li, D.R. Stewar, and R. S. Williams, "Memristive switching mechanism for metal/oxide/metal nanodevicess," *Nature Nanotechnology.*, Jul. 2008, pp. 429–433, vol. 3.
- [14] R. S. Williams and D. B. Strukov, "Exponentila icndrift: Fast switching and low volatility of thin –film memristors," *Appl. Physics. A. Mater. Sci. process.*, Mar. 2009, pp. 515—519, vol. 04, no. 3.
- [15] M. Laiho and E. Lehtonen, "CNN using memristors for neighborhood connections," in Proc. International. Workshop Cell. Nanoscale Network. Their Application., February. 2010, pp.1-4.
- [16] R. S. Williamsm, M. D. Pickett, J. L. Borghetti, D. B. Strukov, G. S. Snider, J. J. Yang, and D. R. Stewart, "Switching dynamics in titanium dioxide memristive devices," J. Appl. Phys., pp. 1—6, Oct. 2009, vol. 106, no. 7.
- [17] J. G. Simons, "Genaralized formula for the electric tunnel effects between similar electrodes seperated by a thin insulating film," Jan. 1963, J. Appl Physics., pp.1793—1803, vol. 34, no. 6.
- [18] S.-M. Kang, S. Shin and K.Kim, "Compact models for memristors based on charge-flux constitutive relationships," *IEEE Transaction. Comput Aided Design Integr. Circuits Syst.*, Apr 2010, pp. 590—598, vol. 29, no. 4.

Modeling, Control and Simulation of Integrated Photovoltaic cell-Fuel Cell-Battery Based Distributed Generation System

Avinash Kumar Associate Professor, EEE Dept. CIT, Ranchi <u>bitavinash@gmail.com</u>

Bikash Chandra Saha Asst. Professor, EEE Dept CIT Ranchi <u>bikash.5@gmail.com</u>

Abstract—Utility restructuring, cutting-edge power electronics, public environmental concerns, and expanding power demand are providing the opportunity for emerging generation technologies in the modern world. In order to meet sustained load demands during varying natural conditions, different energy sources need to be integrated with each other for extended usage of alternative energy. The paper presents the modeling and control framework of a photovoltaic cell-battery-fuel cell (PVBFC) integrated stand-alone distributed generation system (DGS) in Matlab/Simulink environment. A new control strategy for charging and discharging controller (CDC) of generic battery (GB), maximum power point tracker (MPPT) for photovoltaic cell (PVC) and fuel flow rate controller (FRC) for fuel cell (FC) are developed. GB works in parallel with PVC to compensate varying natural conditions in day time and FC works independently at night. We have also designed an inverter model for the proposed system by considering the output voltage regulation.

Keywords: PVC, MPPT, PEMFC, GB, DGS, Voltage Regulation

I. INTRODUCTION

The conventional fuel energy sources such as petroleum, natural gas, and coal which meet most of the world's energy demand today are being depleted rapidly. Also, their combustion products are causing global problems such as the greenhouse effect and pollution which are posing great danger for our environment and eventually for the entire life on our planet [1]. The alternative energy sources (PVC, FC, GB etc.) are attracting more attention. Today, new advances in power generation technologies and new environmental regulations encourage a significant increase of distributed energy resources (DERs) around the world. DGS has mainly been used as a Stand-alone power system [2]. A detailed approach to PVC, FC and GB modeling based on a mathematical description of the equivalent electrical circuits are given in [3-4],[5] and [6] respectively. Tracking the maximum power point (MPP) of a photovoltaic (PV) array is usually an essential part of a PV system. Much focus has been on hill climbing [7], and perturb and observe (P&O) methods [8,9]. Hill climbing involves a perturbation in the duty ratio of the power converter, and P&O a perturbation in the operating voltage of the PV array. The system under study in this paper is a stand-alone PVBFC integrated power system, which consists of PVC, GB and a proton exchange membrane fuel cell (PEMFC). A simulation software program known as MATLAB/SIMULINK is used in dealing with modeling, simulation, and control and energy management of the system under study. MPPT for PVC, CDC for GB and FRC for PEMFC are developed to get optimal system performance. The configuration of overall DGS system is a PVBFC integrated voltage source, a voltage source inverter (VSI) with a LC filter and an AC load.

II. MAIN COMPONENTS OF THE PROPOSED DGS

The basic components of the proposed DGS are DERs and their controllers, VSI with LC filter, and an AC load. **A. Modeling of DERs**

The equivalent circuitry of a PV cell shown in Fig.1, the simplest model can be represented by a current source in parallel with a diode, and the non-idealities are represented by the insertion of the resistances Rs and Rp.



Fig.1: Equivalent Model of the Photovoltaic Panel

 $V=I_{ph}-I_d-V_d/R_p$(1)

 $I = V_d - R_s * I_{pv}$(2)

 $I_d=I_{sat}*exp(V_d/V_T-1)$(3)

 $V_T = k^T/q^Q d^N_{cell} N_{ser}$ (4)

Where,

- $I_d = Diode cuurent (A)$
- $V_d = Diode voltage (V)$
- Rs=Series resistance (Ω)
- R_p =Shunt resistance (Ω)
- Isat = diode saturation current (A)
- T = cell temperature (K),
- V_T=temperature voltage (V)
- $k = Boltzman constant = 1.3806e-23 J.K_{-1}$
- $q = electron charge = 1.6022e_{-19} C$
- $Q_d =$ diode quality factor
- Ncell= number of series-connected cells per module
- Nser = number of series-connected modules per string

B. Control Schemes for DERs

Hill climbing involves a perturbation in the duty ratio of the power converter, and P&O a perturbation in the operating voltage of the PV array. In the case of a PV array connected to a power converter, perturbing the duty ratio of power converter perturbs

he PV array current and consequently perturbs the PV array voltage. A flow chart for MPPT algorithm is developed as shown



Fig. 2: Flow chart of MPPT Algorithm.

It can be seen that incrementing (decrementing) the voltage increases (decreases) the power. When operating on the left of the MPP and decreases (increases) the power when on the right of the MPP. Therefore, if there is an increase in power, the subsequent perturbation should be kept the same to reach the MPP and if there is a decrease in power, the perturbation should be reversed. The process is repeated periodically until the MPP is reached. The system then oscillates about the MPP. The oscillation can be minimized by reducing the perturbation step size.



Fig.3: Maximum power point for PV Array.

The Power converter i.e the boost converter boosts DC voltage from VMMP to 500V. This converter uses a MPPT system which automatically varies the duty cycle in order to generate the required voltage to extract maximum power. Basic equation of boost converter is (5)

Where, Vo=Output Voltage Vin=Input voltage

D=duty cycle

The boost converter is designed in such a way that the Vo is again regulated to fixed voltage with the help of controlled voltage source converter of 500V. A flow chart is developed for the CDC of the battery shown in Fig.4. The battery status signal i.e state of charge (SOC) is compared with upper and lower limit of SOC(%) in which limit battery works properly



Fig.4: Flow chart for CDC of Battery

A fuel cell current controlled fuel flow rate regulator is derived from [5]. The fuel flow rate is directly proportional to fuel cell current.

III. SIMULATION MODEL OF PVBFC SYSTEM

The models are integrated in Sim Power Systems (SPS) and used in a simulation of a PVBFC System. As shown in Fig. 9, it consists of three DERs with their controllers, an IGBT voltage source inverter with PWM gate controller and an AC load. The four power (KW) displays show the status of power generated and observed by DERs and AC load. Additional scopes and measurements are displayed inside the subsystem block (light blue color).

A: PWM IGBT inverter

A 3-phase 6-switch DC/AC PWM voltage source inverter is used to convert the power from DC to AC. Fig. 5 shows the main circuit of a 3-phase voltage source inverter, which is actually an IGBT-diode universal bridge.



Fig. 5: IGBT-Diode Universal Bridge

This VSI needs gate pulses which are generated by PWM generator.

B: PWM Generator

The PWM Generator block generates pulses for carrier based PWM converters using two-level topology. The block can be used to fire the forced-commutated devices (FETs, GTOs, or IGBTs) of single-phase, two-phase, three-phase, two-level bridges or a combination of two three-phase bridges.



Fig. 6: Six pulses generated by the PWM Generator.

The pulses are generated by comparing a triangular carrier waveform to a reference modulating signal. The modulating signals can be generated by the PWM generator itself, or they can be a vector of external signals connected at the input of the block. The modulation index (m) of Modulated signal is externally controlled by voltage regulator block.

C: Voltage Regulator

The voltage regulator shown in fig. 7 takes three component of load voltage in per unit (pu) as feedback, applies proper operation to control the modulation index.



Fig.7: Simulation model of voltage regulator

D: LC Filter

The primary function of the AC filter is to filter out the high frequency components caused by the inverter switching operation. The LC-filter aims to reduce the high order harmonics which are the multiples of the carrier frequency of PWM generator. The transfer function of the LC-filter designed by the output voltage to the input voltage is given as follows:

$$G(s) = \frac{1}{1 + LC s^2}$$
(6)



Fig. 8: Simulation model of LC filter.



Fig. 9: Simulation model of PVBFC System

IV. SIMULATION PARAMETERS OF COMPONENTS

Table I shows the parameter details of components in proposed PVBFC system.

Name of	Details about the Component
Components	
PVC	Power=100KW, V ₀ =273.5V at 1000
	W/M ₂ Irradiance
PEMFC	Power=65KW, Vo=500V
GB	Rated Capacity=15Ah,Vo=500V
MPPT	Tracks Maximum Power Point and Varies
	duty cycle of Boost Converter
Boost	Boost the V _{mpp} to fixed 500V
Converter	
FRC	Regulates the fuel flow rate w.r.t fuel cell
	current
CDC	Control SOC in between 40% to 90%.
PWM IGBT	Convert 500Volt(DC) to 280V(rms)
VSI	(3-Phase AC)
PWM	Generator 3 arm,6 pulse, carrier
	Frequency=2000KHz
Voltage	Regulates modulation Index to m=0.875
Regulator	for PWM Generator
AC	Load Vrms=260V,f=50Hz,Power=64kW

Table I:

V. SIMULATION INPUT AND RESULTS

The PVBFC is simulated over 24s in which the first 12s, battery works in parallel with photovoltaic cell to compensate varying natural conditions in day time and fuel cell works independently at night for the next 12s.



Fig.10: Irradiance versus Time signal with varying natural condition

A solar irradiance signal is build with the help of signal builder in Matlab considering different natural conditions as shown in Fig.10.

The signal shows following points explained below:

1. Gradual increment in irradiance from t=0s to t=3s considers morning time.

- 2. Rapid variation in irradiance due to disturbances such as rain, shadow, etc. up to t=5s.
- 3. Gradual increment and decrement of irradiance from t=5s to t=13s considers afternoon and evening time.
- 4. Constant irradiance from t=13s to t=24s due to temperature and other light sources at night.

The following lines explain what is going on during the simulation:

- At t=0s, GB and PVC start working in parallel and, the PEMFC is disabled. Initially SOC is assumed to be 80%. Fig. 11 shows the power (KW) graph of DERs and loads. Since the GB and PVC works in parallel, thereby their power graphs are complementary of each other. Negative power of battery shows that it is being charged by PVC.
- 2. Natural disturbances as explained in Fig.10 is displayed from t=3s to t=5s.
- 3. At t=12s, PEMFC is enabled, it begins to provide power alone. And the some power generated by PVC is directly used to charge the GB.
- 4. Fig.11-18 shows the simulation results.
- 5. Fig. 17 shows the status of battery when it is fully charged at time =8s.



Fig.13: PVC status.



Fig.16: Modulation index of PWM generator and Vdc of VSI.





VI. CONCLUSION

The modeling and control fraimework of the PVBFC system is developed and simulated in MATLAB/ SIMULINK using SIMPOWER Systems library. The simulation results give encouraging output on the performance of PVBFC integrated system and thus validate the effectiveness of the system. It is useful for further research work such as, maximum power point tracking, integration of PV energy with other energy sources.

VII. REFERENCES

- [1] T. Veziroglu: Hydrogen Energy System: A Permanent Solution to Global Problems, Clean Energy Research Institute, USA, Web Site, visit 2004.
- [2] A. Chowdhury, S. K. Agarwal, and D. O. Koval, "Reliability modeling of distributed generation in conventional distribution systems planning and analysis," IEEE Transactions on Industry Applications, vl.39, pp. 1493- 1498, sept.oct., 2003.
- [3] KJAER, S. B.; et al. A Review of Single-Phase Grid- Connected Inverters for Photovoltaic Modules. IEEE Transactions on Industry Applications, v. 41, n. 5, p.p 1292-1306, 2005.
- [4] Moacyr A. G. de Brito, Leonardo P. Sampaio, Luigi G. Junior, Carlos A. Canesin: Evaluation of MPPT Techniques for Photovoltaic Applications, IEEE, Brazil,pp.1039-1044, 2011.
- [5] Souleman Njoya M., Olivier Tremblay and Louis-A, IEEE Members, A Generic Fuel Cell Model for the Simulation of Fuel Cell Vehicles, pp.1722-1729,2009

- [6] Olivier Tremblay, Louis-A. Dessaint, and Abdel-Illah Dekkiche, A Generic Battery Model for the Dynamic Simulation of Hybrid Electric Vehicles. IEEE, p.p 284-289,2007.
- [7] Trishan Esram, and Patrick L. Chapman, Comparison of Photovoltaic Array Maximum Power Point Tracking Techniques, IEEE TRANSACTIONS ON ENERGY CONVERSION, VOL.22, NO.2, pp. 439-449, june, 2007
- [8] Wang NianCHun, ,Wu MeiYue and SHi GuoSHeng, Study on characteristics of photovoltaic cells based on MATLAB simulation, IEEE, 978-1-4244-6255-1,2011
- [9] Femia, Giovanni Petrone, Giovanni Spagnuolo, and Massimo Vitelli, Optimization of Perturb and Observe Maximum Power Point Tracking Method Nicola, Member, IEEE TRANSACTIONS ON POWER ELECTRONICS, VOL.20, NO.4, pp. 963-973, JULY 2005.

Social, Ethical & Other Issues of Networked World

Farhat Hashmi Red Rose School of Institutions Bhopal (M.P), India Farhathashmi20@yahoo.co.in

M.A.Rizvi Dept. of Computer Engineering and Applications National Institute of Technical Teachers' Training and Research Bhopal (M.P), India <u>marizvi@nitttrbpl.ac.in</u>

Abstract—Internet is becoming widely popular among all age group and extensively used in day to day life. People from far distance can now easily communicate through various available online social networking sites. As these social networking sites are gaining more and more interest of people, the risk of hacking the personal information available on these sites is also increasing in same proportion. Security of sensitive data is prime area to explore more in depth. As people are becoming more cyber friendly, cyber crimes are also tremendously growing. So security of sensitive and personal data is one of the most important issues to have in focus. In this paper various privacy issues are discussed and laws against such issues are described in detail.

Keywords : Social Networking Sites; Internet Security; Social Issues ; Ethical Issues; Legal Issues

I. INTRODUCTION

Internet is necessity in these modern days, it is really hard to imagine our life without internet. Each and every field requires internet for its progress. It is a very versatile facility which can help you in completing many tasks easily and conveniently with few clicks. Today we depend on the internet for our various routine needs. We can access information through internet from a remote village globally.

Nowadays, the people know about internet and interested in learning computer and internet courses. It is really easy to learn the internet and small kid can also access the online sites easily without more efforts. If we started to list out the benefits of the internet, really the process don't have any end.

Now jobless people are doing online jobs and make a good income from it based on their work, so internet solved the jobless problems. In short, internet is a virtual world, where we can live an imaginary life using it. The Internet technology provides a new way of conducting business in all electronic sectors. The Internet technology has become an important source of information according to which various business strategies of subjects in tourism industry are formed. You can pay your bills online and purchase various items by going through various websites and choosing among a variety of options. One can get information on any particular thing around the world using internet facility.

However, there are chances for a person to set the privacy settings in such a way that it is not possible for you to search for them over there. A basic overview of internet research considerations advocates a process approach to ethics, which emphasizes the importance of addressing and resolving ethical issues. Rather than prescribing a set of approved practices, we have deliberately chosen to suggest a characteristic range of questions that should be asked by internet researchers as well as those responsible for oversight of such research. Guidelines such as these should not be considered complete or final. More than two decades of research practice provide multiple examples of ethical complications, unique solutions, and best practices. To continue to build, rather than reinvent the wheel, we believe it is vital for researchers to maintain a strong awareness of previous scholarship. To help in this endeavor, we provide a diverse bibliography in Appendix 3. We also recognize that as internet researchers encounter new venues, contexts, inter/actions, and tools, additional issues, questions, and responses will inevitably arise, either as variations of those mentioned in this document or as distinctively new. We encourage researchers to contribute synopses of their ethical challenges or specific case studies to our web-based compendium of cases. [website link here]. Our intent is that this document will remain useful in those new contexts in so far as it suggests starting points and possible analogies for raising new questions and developing new responses. At the same time, we anticipate that as this document comes into use, it may help researchers further identify its limits as they encounter new contexts and ethical issues that emerge in the ongoing evolution of both the internet and our multifaceted efforts to research the communicative engagements these technologies make possible. We intend this document to serve both as a first way of addressing

contemporary ethical challenges and as an occasion for inviting further critical evaluation and ongoing dialogical development of internet research ethics. In sum, we hope this document will prove helpful to researchers, ethicists, and others concerned with the important ethical challenges of internet research, and the Association of Internet Researchers remains grateful for further critical comment and examples that help to carry our dialogues onward from this point.

II. PRIVACY ISSUES OF SOCIAL NETWORKS

Z. Personal identification

Personal identification means associating an identity with an individual to resolve the identity issue of a person, it can be categorized into –Verification and Recognition. Verification (authentication) refers to the problem of confirming or denying a person's identity (Am I who I claim I am?). Identification (Who Am I?) refers to the problem of establishing an identity-either from a set of already known identities (close identification problem) or otherwise (open identification Problem). The term positive personal identification typically refers (in both verification as well as identification context) to identification of a person with high certainty.

AA. Social Networking Sites

For example, Facebook that gathers critical information of users which includes Date of birth, Address, Telephone numbers and more intimate details such as interests, hobbies, relationship status and sexual preference. Go and Google yourself you can find at least two pages of details about you.Users upload video content to a site for sharing either privately or public at Photo sharing Instagram, Flickr, Photobucket, Picasa and video sharing Youtube, etc.The capacity and speed of information technology are accelerating rapidly which can be viewed as every minute of the day:

- 100,000 tweets are sent
- 684,478 pieces of content are shared on Facebook
- 2 million search queries are made on Google
- 48 hours of video are uploaded to YouTube
- 47,000 apps are downloaded from the App Store
- 3,600 photos are shared on Instagram
- 571 websites are created
- \$272,000 is spent by consumers online (source: All Twitter)

(Source: thesocialskinny.com)

BB. Location Based Networks

Sharing personal location with social connections via smart phones is another intrusion in privacy. Due to high penetration of smartphones with photo and video creation and sharing opportunities, the amount of personal content available online has been increasing rapidly in the last years. Once content is shared online, it might be very difficult to take it offline again and it will remain there for public. On Internet once the statement is typed, it can be copied, saved and forwarded. You have given up complete control of your personal information to third party. There has been a number of cases when youngsters have been harassed by pedophiles online and these cases have also led to suicide. Peers intruding into each other's privacy and anonymity with video and multi-media files in ways that harm the subjects of the digital files.' here are no rules to protect individuals from accidentally having an embarrassing photo or video taken of them and then posted on the web.

CC. Privacy concerns regarding Social networking services

People will give up personal information to join social networks on the Internet. Adults are concerned about invasion of privacy, while teens freely give up personal information. This occurs because often teens are not aware of the vulnerable nature of the Internet.

III. ETHICAL ISSUES

DD. Email

Google scans your account and this scanning and cannot be fully turned off, making this an ethical issue in information privacy on the Internet.

EE. Online Transactions

Online activities are threat to privacy of confidential details, can be used unethically or for fraudulent purposes.

FF. Government and Other Agencies

The government holds a lot of important data about people and organizations. It is threat on national security. The degree of accessibility to the government for personal data of citizens raises ethical issues in information privacy. Maintaining its confidentiality should be a prime importance of the government agencies.

GG. Social Networking Websites and Privacy

Through status updates on social networking websites, you let everyone know of your whereabouts.

HH. Internet Cookies

Cookies are another entity which causes privacy concern as they can be used by hackers. Blocking or deleting cookies frequently can protect data from threat. It is users responsibility to keep ourselves abreast of the laws pertaining to information privacy on web.

IV. LEGAL ISSUES

- *II.* Law: While one of the most controversial and un-clear areas of evolving laws is the determination of what forum has subject matter jurisdiction over activity (economic and other) conducted on the internet. It is certainly clear that substantial portions of internet activity are subject to traditional enforcement of laws and regulations.
- *JJ*. Architecture: Internet filtering, encryption programs, basic architecture of TCP/IP protocols and user interfaces falls under the category of private regulation. It is arguable that all other modes of internet regulation either rely on, or are significantly affected by, West Coast Code.
- *KK.* Norms: Conduct is also regulated by social norms and conventions. While certain activities may not be specifically prohibited by the code architecture of the Internet or by governmental law. Certain actions will be censored or self-regulated by the norms of communities one is associated with.

LL. Markets: The Internet creates a virtual marketplace for information. The increase in popularity of the Internet as new means of transacting all kinds of commercial activities has brought the laws of supply and demand to cyberspace.

1) Free speech on the Internet: Article 19 of the Universal Declaration of Human Rights calls for the protection of free expression on all media.

2) Internet censorship: In many countries, speech through cyberspace has proven to be another means of communication that has been regulated by the government - "Open Net Initiative".

3) Section 66: In February 2001, Delhi police arrested two men running a web-hosting company that had shut down a website over non-payment of dues. The owner claimed that he had already paid. The Delhi police had charged the men for hacking under Section 66 of the IT Act and breach of trust under Section 408 of the IPC.

4) Section 66A Removed: In September 2010, a freelance cartoonist was arrested under Section 66A of the IT Act, Section 2 of Prevention of Insults to National Honour Act, 1971 and for sedition under the Section 124 of the IPC. His cartoons depicting widespread corruption in India were considered offensive.

MM. Cyber Crime Control Cells in India

A live example of such an enforcement agency is Cyber Crime Police Station, Bangalore, India's first exclusive Cyber Crime enforcement agency.

Other examples of such enforcement agencies include:

• Cyber Crime Investigation Cell of India's Mumbai Police.

- Cyber Crime Police Station of the State Government of Andhra Pradesh, India.
- Crime Branch of Criminal Investigation Department, in Tamil Nadu, India, has a Cyber Crime Cell at Chennai.
- Cyber Crime Cells have been set up by the Kolkata Police as well as the Criminal Investigation Department, West Bengal.
- NN. Copyright law: It provides one of the most important forms of intellectual property protection on the Internet for at least two reasons. First, much of the material that moves in commerce on the Internet is the work of authorship, such as musical works, multimedia works, audiovisual works, movies, software, database information, etc. which are within the usual subject matter of copyright. Second, because the very nature of an electronic online medium requires that data be "copied"; copyright rights are obviously at issue.

V. TECHNICAL ISSUES

OO. Trademark Concerns

A trademark is a word, image, slogan, or a device to identify the goods or services of any particular party. Trademark violation occurs when one party use the mark of another in such a way so as to create a likelihood of confusion, mistake or deception. Website developers should be aware of the possible trademark issues. If the goods or services sold under the mark will be sold internationally, trademark registrations in other countries should also be considered.

PP. Linking And Framing Concerns

When the image from another web site is incorporated into one's own page by means of an unauthorized IMG link, there is no direct copying by the creator of the link. Therefore, one should not include links to images found on another party's web site without permission.

One can also utilize a link to pass off another's work as one's own. It should be noted that a link to someone's page or image could be defamatory and hence subject to legal liability.

QQ. Children and the Internet

The Internet is becoming lifeline for children and youth. It is therefore vital for elders and peers to educate children and young people regarding risks and responsibilities while using the Internet. Teaching children about the importance of 'netiquettes' and instilling to them the notion of "think before you click" should be our primary goal.

A 25 country survey suggests that top activities for children and youth using the Internet are: Schoolwork (92%), Playing games (83%), Watching video clips (75%) and Social networking (71%) Filtering does not remove the objectionable or illegal content from the Internet this only make it harder to access.

RR. Name System Domain Security Extensions DNSSEC

Name System Domain Security Extensions - or DNSSEC – Provides more confidence to the users in the online activities. It provides additional security so that your browser can verify that DNS information has not been modified.

Attacker can intercept DNS query and returns false information. Internet security can be improved by deploying DNSSEC.

SS. Human Rights

The challenge while using Internet is double-edged:

-it can either promote and expand our rights, or

-be used in a way that damages them.

The use of Internet is- fundamental rights, freedom of expression, access to education, privacy and freedom of association.

TT. Networks & Trust

The issues at Internet Society Trust & Identity are also to redesign or improve existing elements to meet emerging requirements.

The Internet Society must provide channels for secure, reliable and private communication between entities, which can be clearly authenticated in a mutually understood manner.

UU. Open Internet Standards

The Internet has open and non-proprietary standards, to facilitate the smooth functioning of Internet. It is for devices, services and applications to work together on Internet

The core groups behind the development of the standards are:

- The Internet Engineering Task Force (IETF);
- The Internet Research Task Force (IRTF) and;
- The Internet Architecture Board (IAB).

VV. Privacy & Identity

The Internet Society works to support the development of tools and policies that make people powerful to manage their online identities. Privacy of individual means:

- Protecting personal Information
- Understanding Online Identity
- Hiding Identity while Browsing
- Protection against Web Tracking and User Privacy
- Maintaining Data privacy while keeping pace with an evolving environment
- Solutions for enabling data flow

WW. Intellectual Property

Intellectual property (IP) is a term referring to creations of the intellect for which a monopoly is assigned to designated owners by law. Some common types of intellectual property rights (IPR) are copyright, patents, and industrial design rights.

XX. Internet Regulation and Governance

It is needed to include new ideas and perspectives for the global internet users. Internet regulation consists of: Censorship of data and Controlling domain registration, IP address control, etc. The Internet Society works on three key aspects:

- permission less innovation
- open access and
- collaboration

YY. Internet Flexibility and Stability

The Internet is stable, secure and flexible. The open and global nature of the Internet, built on fundamental principles of open standards, voluntary collaboration, reusable building blocks, integrity, permission-free innovation and global reach. It is important that these issues be addressed in the way that they do not damage the global architecture of the Internet. At the same time, using the Internet is always a threat to security. Internet technology focuses on resiliency and security of its naming and routing systems.

ZZ. IP Addressing

Internet is currently addressed with IP version 4 (IPv4) addresses which are unique numeric identifiers. IPv6 is the next generation Internet protocol, offering a much larger address pool in comparison to IPv4. IPv6 will provide enough addresses (2¹²⁸ to be exact). Even if there are no more IPv4 addresses available, the existing IPv4 networks will continue to function. In fact, IPv6 and IPv4 are expected to co-exist for many years to come.

IP addresses are managed by the Internet Assigned Numbers Authority (IANA), and by the Regional Internet Registries (RIRs).

AAA. Routing Security

To prevent future errors and malicious activity, it is important to increase the security of the Internet's routing protocol through secure routing mechanism. There is also opportunity for malicious activities that could damage the routing infrastructure in the future therefore routing security is prime importance.

BBB. Putting a stop to spam

Spam is a global issue that affects every Internet user, but at the moment it feels as though the world is fighting a losing battle to put a stop to it. At the Internet Society they are creating an online library, the Spam Toolkit, to fight against spam.

VI. CONCLUSION

After going into all possible details of privacy issues related to use of social networking sites on internet it has been concluded that security of information is very critical when doing online task. No single way is enough for preventing security of online task. Various social networking sites are using their own way to mitigate risk of being a cyber victim. As cyber crimes are increasing enormously, new cyber laws are also enforcing and are amended as per the requirement. So the only way to be in safe side while doing any online work is to have awareness of current cyber crimes and take every possible precaution against these cyber crimes.

VII. REFERENCES

- [1] Fink, Arlene. Conducting research literature reviews: from the Internet to paper. Sage Publications, 2013.
- [2] Roman, Rodrigo, and Javier Lopez. "Integrating wireless sensor networks and the internet: a security analysis." Internet Research 19.2 (2009): 246-259.
- [3] Chakrabarti, Anirban, and G. Manimaran. "Internet infrastructure security: A taxonomy." Network, IEEE 16.6 (2002): 13-21.
- [4] Cohen, Fred. "Simulating cyber attacks, defences, and consequences." Computers & Security 18.6 (1999): 479-518.
- [5] Cavusoglu, Huseyin, Birendra Mishra, and Srinivasan Raghunathan. "The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers." International Journal of Electronic Commerce 9.1 (2004): 70-104.
- [6] Lichtenstein, Sharman. "Developing Internet security policy for organizations." System Sciences, 1997, Proceedings of the Thirtieth Hawaii International Conference on. Vol. 4. IEEE, 1997.
- [7] Shetty, Nikhil, et al. "Competitive cyber-insurance and internet security." Economics of information security and privacy. Springer US, 2010. 229-247.
- [8] Armando, Alessandro, et al. "The AVISPA tool for the automated validation of internet security protocols and applications." Computer Aided Verification. Springer Berlin Heidelberg, 2005.

A Survey on: PSO-GA based Back Propagation Learning-MLP (PSO-GA-BP-MLP)

Chanda Prasad

Computer Science & Engineering Ramgovind Institute of Technology At-Mahuvam, P.O-Karma, Koderma Jharkhand 825409 chanda.prasad49@gmail.com

Nitish Kumar

Computer Science & Engineering

Ramgovind Institute of Technology At-Mahuvam,P.O-Karma,Koderma Jharkhand 825409 er.nitish169@gmail.com

> Janmenjoy Nayak Department of Computer Science & Engineering Veer Surendra Sai University of Technology, Burla - 768018, Odisha, India <u>mailforjnayak@gmail.com</u>

Abstract

In last few decades, Evolutionary computation and Swarm intelligence are two hot favorites for almost all types of researchers. Moreover, many contributions have been made in two directions: Genetic Algorithm (GA) and Particle Swarm optimization (PSO). But, some limitations in both the algorithms (complicated operator like crossover and mutation in GA and early convergence in PSO), are the major restricted boundaries for solving complex problems. In this paper, a survey of hybridization of Particle swarm optimization and Genetic algorithm and back propagation learning based Multilayer perceptron neural network. The hybridization approches is more accuracy value.

Keywords: EA, PSO, GA, MLP, BP, GA -PSO

I. INTRODUCTION

Evolutionary computation and Swarm intelligence are diversified area of research which make them more favorites than other optimization algorithms. Evolutionary algorithms (like GA) (Holland,1970,1992) [1], Differential Evolution (DE) (Storn, 1999)[2] and Swarm Intelligence (like PSO) (Kennedy and Eberhart) [3]; Group Search Optimization (GSO), (He, Wu, & Saunders, 2009[4]); Ant Colony algorithm(ACO) (Alatas & Akin, 2005[5]; Bee Colony algorithm(BCO) (Alatas, 2010[6]; Karaboga & Basturk,2007[7], Cat Swarm Optimization(CSO) (Chu, Tsai, & Pan, 2006[8]; Pradhan & Panda, 2012[9]), Glowworm Swarm Optimization(GSO) (Krishnanand & Ghose 2006)[10], are widely used in various engineering applications and these methodologies have been combined with different neural network to perform various data mining tasks like classification, prediction and forecasting.

Shi et. al. (2005)[11] have introduced an improved GA algorithm and a new hybrid algorithm using PSO & GA. They performed the simulations for optimization problems and found that the hybrid algorithm performs better than other techniques. A novel Chaotic Hybrid Algorithm (CHA) by using the hybridized PSO-GA method for circle detection has been developed by Dong et. al.(2012)[12]. Liu et. al. (2013) [13] have made an investigation on two Wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization. The comparison performance between the two networks helps to prove that the Wavelet network is more statistically significant than MLP. Ludermir and Oliveira (2013) [14] have used the PSO technique for the identification of factors related to Common Mental Disorders. A hybrid algorithm combining Regrouping Particle Swarm Optimization (RegPSO) with wavelet radial basis function neural network is presented by Nasir et. al. (2013) [15] which is used to detect, identify and characterize the acoustic signals due to surface discharge activity. Sahoo et. al.(2014)[16] have developed a hybrid method by considering both PSO and GA for solving mixed integer nonlinear reliability optimization problems in series, series–parallel and bridge systems.

Most of the conventional optimization techniques are iterative methods in which the selection of initial solution is based on the nature of the problem[17] and are revised using deterministic update rules which usually depend upon the problem structure. However, the improvement of heuristic techniques, like genetic algorithm, particle swarm optimization, attracts the researcher's attention towards these methods, due to the efficiency of solving a complex iterative optimization problem within a rational time complexity. But, for the improvement in computational efficiency, hybridization between two or more algorithms are required. With this intention, in this paper a hybridized PSO-GA based multilayer perceptron has been proposed to perform classification task. The experiments for this purpose will comprise of the following performance comparisons: (1) GA-MLP,(2) PSO-MLP,(3) Hybrid PSO/GA-MLP. Remaining of this paper is organized as follows: Preliminaries, Proposed Method, Result and Analysis, Conclusion and REFERENCES.

1.1 Evolutionary algorithms:

First Evolutionary computing is based on the Darwinian principle of the natural selection and EA is the first discovered by Holland. Evolutionary algorithm is very important for the survival of fitness and gives the approximate solution. In a GA the population is not xed every time the population has been changed and the value will be also changed. If we are choosing the best one then there is more advantages that is the life line is more i.e the age will be increased.

So, we can say that maximum lifetime is increased by their fitness value.

1.2 Genetic Algorithm:

In Evolutionary calculation, Genetic Algorithms have greater role[1]. It is an interesting area for computer scientist, and it was introduced by Holland in 1975[1]. They are interesting because these are very robust in nature and capable of optimize complex results, including a great number of interacting components. GAs includes techniques enhanced by natural evolution to compute and change species to their environment. Generally speaking, a GA modifies a number of solutions to the problem at present, similarly as nature alters species. A GA develops a population of arbitrarily generated individuals by using iterations and performing genetic operations such as crossover, mutation and some other methods [3]. Charles Darwin developed a method of natural selection, which is implemented in Genetic algorithms [1] use selection of fittest individuals.

Genetic algorithm (GA) [1] is a computational model of machine learning inspired by evolution. The development of GA has now reached a stage of maturity, due to the report made in the last decade by academics and engineers all over the world. They are less vulnerable to getting 'stuck' at local optima than gradient search methods. The pioneering work is contributed by J.H. Holland for various scientific and engineering applications. GA is inspired by the mechanism of natural selection, a biological process in which stronger individuals are likely be the winners in a competing environment. Fatnesses (goodness) of the chromosomes are used for solving the problem and in each cycle of genetic operation (known as evolving process) a successive generation is created from the chromosomes with respect to the current population. To facilitate the GA evolution cycle, an appropriate selection procedure and two major fundamental operators called crossover and mutation are required to create a population of solutions in order to and the optimal solution (chromosome).

1.3. Particle Swarm Optimization

Particle swarm optimization (PSO) [3, 18] is a widely used stochastic based search algorithm and it is able to search global optimized solution. Like other population based optimization methods the particle swarm optimization starts with randomly initialized population for individuals and it works on the social behavior of particle to find out the global best solution by adjusting each individual's positions with respect to global best position of particle of the entire population (Society). Each individual is adjusting by altering the velocity according to its own experience and by observing the experience of the other particles in search space by use of equation - 1 and equation - 2. Equation-1 is responsible for social and cognition behavior of particles respectively where c1 and c2 are the constants in between 0 to 2 and rand(1) is random function which generates random number in between 0 to 1.

 $V_i(t+1) = V_i(t+1) + c_1 * rand(1) * (lbest_i - X_i) + c_2 * rand(1) * (gbest_i - X_i)$ (1)(2)

 $X_i(t+1) = X_i(t) + V_i(t+1)$

The basic steps of problem solving strategy of PSO are:

Initialize the position of particles $V_i(t)$ (population of particles) and velocity of each particle $X_i(t)$. Do

Compute fitness of each particle in the population.

Generate local best particles (LBest) by comparing fitness of particles in previous population with new population.

Choose particle with higher fitness from local best population as global best particle (GBest).

Compute new velocity $V_i(t + 1)$ by using equation-1.

Generate new position $X_i(t + 1)$ of the particles by using equation-2.

While (iteration <= maximum iteration OR velocity exceeds predefined velocity range);

Each particle keeps track of its coordinates in hyperspace which are associated with the best solution (fitness) it has achieved so far. (The value of that fitness is stored.) This value is called pbest. Another best value is tracked. The global version of the particle swarm optimizer keeps track of the overall best value, and its location, obtained thus far by any particle in the population; this is called gbest. The particle swarm optimization concept consists of, at each time step, changing the velocity (accelerating) each particle toward its pbest and gbest (global version). Acceleration is weighted by a random term, with separate random numbers being generated for acceleration toward pbest and gbest.[17]

The problem of finding a set of weights to minimize residuals in a feed forward neural network is not a trivial one. It is nonlinear and dynamic in that any change of one weight requires adjustment of many others. Gradient descent techniques, e.g., back propagation of error, are usually used to and a matrix of weights that meets error criteria, though there is not widespread satisfaction with the e

activeness of these methods. A number of researchers have attempted to use genetic algorithms (GAS) to nd sets of weights, but the problem is not well suiied to crossover. Because a large number of possible solutions exist, two chromosomes with high fitness evaluations are likely to be very different from one another; thus recombination may not result in improvement.[16]

1.4. Multi Layer Perception

MLP (Fig.1) is the simplest neural network model which is consists of neurons called perceptions (Rosenblatt, 1958). From multiple real valued inputs, the pecptron compute a single output according to its weights and non-lenear activation functions. Basically MLP network is consists of input layer, one or more hidden layer and output layer of computation perceptions.



Figure 1 MLP with input layer, single hidden layer and output layer, Where f(.) is non-linear activation function.

MLP is a model for supervised learning which uses back propagation algorithm. This consists of two phases. In the 1st phase, error(eq.4) based on the predicted outputs (eq.3) corresponding to the given input is computed (forward phase) and in the 2nd phase, the resultant error is propagated back to the network based on that weight of the network are adjusted to minimize the error (Back Propagation phase). (3)

$$y = f(\sum_{i=1}^{n} w_i x_i + b)$$

Where w is the weight vector, x is the input vector, b is the bias and f(.) is the non-linear activation function.

$$\delta_k = (t_k - y_k) f(y_{in_k}) \tag{4}$$

Where t_k and y_k is the given target value and predicted output value of input kth pattern and δ_k is the error term for kth input pattern.

The popularity of MLP increases among the neural network research community due to its properties like nonlinearity, robustness, adaptability and ease of use. Also it has been applied successfully in many applications [19-30].

PSO-GA II.

They performed the simulations for optimization problems and found that the hybrid algorithm performs better than other techniques. A novel Chaotic Hybrid Algorithm (CHA) by using the hybridized PSO-GA method for circle detection has been developed by Dong et. al.(2012)[12]. Liu et. al. (2013) have made an investigation on two Wavelet of air MLP hybrid frameworks for wind speed predicted using PSO and GA optimization. The comparison performance between the two networks helps to prove that the Wavelet network is more statistically signicant than MLP. Ludermir and Oliveira (2013) have used the PSO technique for the identification of factors related to Common Mental Disorders. Hybrid algorithm combined Regrouping Particle Swarm Optimization (RegPSO) with wavelet arranged on the basis of function neural network is presented by Nasir et. al. (2013)[15] which is used to detect, identify and characterize by the property signals due to surface discharge activity. Sahoo et. al.(2014)[16] developed a hybrid method by considering both PSO and GA for solving mixed integer nonlinear reliability optimization problems in series, series parallel and bridge systems. The conventional optimization techniques are repeated methods in which the selection of initial solution is based on the nature of the problem and are changed using deterministic update rules which usually depend upon the problem structure. However, the improvement of heuristic techniques, like genetic algorithm, particle swarm optimization, attracts the researcher's attention towards these methods, due to the efficiency of solving a complex iterative optimization problem within a rational time complexity. But, for the improvement in computational efficiency, hybridization between two or more algorithms are required. With this intention, a hybridized PSO-GA based multilayer perception has been proposed to perform classification task. The experiments for this purpose will comprise of the following performance comparisons:

(1) GA-MLP,

(2) PSO-MLP.

(3) Hybrid PSO/GA-MLP.

2.1. An efcient GAPSO approach for solving mixed integer nonlinear programming problem in reliability optimization The basic objective of reliability redundancy allocation problem is to nd the number of redundant components and the reliability levels of each component that either maximize the system reliability or minimize the system cost under several constraints. Reliability-Redundancy allocation problem (RRAP) is basically a nonlinear mixed integer programming problem.[16] For further research, one can improve the proposed hybrid approach using advanced crossover, mutation operators for GA and dierent variants of advanced PSO, viz. CLPSO (Liang et al.), DMSPSO (Liang and Suganthan), FIFS (Mendes and Kennedy), QPSO (Sun et al.) etc. Again, the proposed approach can be applied to solve other mixed integer nonlinear programming problems.[16]

- 1. Reliability-redundancy optimization problem
- 2. Series system
- 3. Seriesparallel system
- 4. Complex (bridge) system
- 5. Constraint handling technique for

constrained mixed integer nonlinear problems.

2.2. An investigation of two Wavelet-MLP hybrid frameworks for wind speed prediction using PSO and GA technique

On the basis of the wavelet theory, time series analysis,GA,PSO and ANN,2 hybrid frameworks the wavelet GA-PLP and WAVELET-PSO-MLP are proposed to pridiction the non-stationary speed of wind. As we know that wind is the renewable energies So,in both hybrid approaches the PSO and GA are improving the MLP.[13]

There are three methods are:-

- 1. Physical methods
- 2. Statistical methods
- 3. Intelligent methods

Among the upper three methods, These methods are some advantages and some disadvantages. Physical methods are for long term prediction and statistical and intelligent methods are better performed in short-term prediction. There are two classifications for wavelet are:-

- (1) Wavelet-GA-MLP vs. Wavelet-MLP vs. GA-MLP vs. MLP
- (2) Wavelet-PSO-MLP vs. Wavelet-MLP vs. PSO-MLP vs. MLP.



2.3. Optimization with a new nature inspired method using a fuzzy combination of Genetic Algorithm and Particle swarm optimization:

Optimization is very important technique from which we can minimize the cost and the best solution of a given problem. There is new hybrid method is proposed with the fuzzy combination to the PSO and GA.Here Particle swarm and Genetic algorithm are combine with the fuzzy ,the FIS variable is given to the PSO and GA and evaluated the architecture.[31][32]

2.4. The Pseudo code of Genetic Algorithm :

1. First we generate the initial population of individuals aleatorily P(0).

2. While(numberggenerations <= maximumnnumbersggenerations)

Do:

Evaluation;

Selection; Reproduction; Generation ++; 3. Show results 4. End

III. CONCLUSION:

In this paper we finally survey the PSO+GA hybrid algorithm using BP in MLP network. For any type of problem optimization is very important so, by using such hybrid approach, it is easier to optimize the weights of MLP network. We focused on the accuracy value in hybrized model. The GA-PSO-MLP-BP algorithm is the best algorithm for the classification problem in terms of accuracy. There is two phase in MLP network first is forward phase and second is backward phase, so here weight is easily adjustable according to there output.

IV. REFERENCES

- [1] J. H. Holland, Adaption in Natural and Artificial Systems. Cambridge, MA: MIT Press, 1975.
- [2] Storn, R. (1999). System design by constraint adaptation and differential evolution. IEEE Transactions on Evolutionary Computation, 3(1), 22–34.
- [3] J. Kennedy, R. Eberhart, "Particle swarm optimization", in: Proceedings of the 1995 IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948, 1995.
- [4] He, S., Wu, Q. H., & Saunders, J. R. (2009). Group search optimizer: An optimization algorithm inspired by animal searching behavior. IEEE Transactions on Evolutionary Computation, 13(5), 973–990.
- [5] Alatas, B., & Akın, E. (2005). FCACO: Fuzzy classification rules mining algorithm with
- ant colony optimization, ICNC 2005. Lecture notes in computer science (Vol. 3612). Springer-Verlag, pp. 787–797.
- [6] Alatas, B. (2010). Chaotic bee colony algorithms for global numerical optimization. Expert Systems with Applications, 37(8), 5682–5687.
- [7] Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. Journal of Global Optimization, 39(3), 459–471.
- [8] Chu, S., Tsai, P., & Pan, J. (2006). Cat swarm optimization. Lecture Notes in Computer Science Springer-Verlag, 4099, 854–858.
- [9] Pradhan, P. M., & Panda, G. (2012). Solving multiobjective problems using cat swarm optimization. Expert Systems with Applications, 39(3), 2956–2964.
- [10] Krishnanand, K. N., & Ghose, D. (2006). Glowworm swarm based optimization algorithm for multimodal functions with collective robotics applications. Multiagent and Grid Systems, 2(3), 209–222.
- [11]X.H. Shi, Y.C. Liang, H.P. Lee, C. Lu, L.M. Wanga, "An improved GA and a novel PSO-GA-based hybrid algorithm", Information Processing Letters 93 (2005) 255–261.Na Dong, Chun-Ho Wu, Wai-Hung Ip, Zeng-Qiang Chen, Ching-Yuen Chan,Kai-Leung Yung, "An opposition-based chaotic GA/PSO hybrid algorithm and its application in circle detection", Computers and Mathematics with Applications 64 (2012) 1886–1902.
- [12] Hui Liu, Hong-qi Tian, Chao Chen, Yan-fei Li, "An experimental investigation of two Wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization", Electrical Power and Energy Systems 52 (2013) 161–173.
- [13] Teresa B. Ludermir, & Wilson R. de Oliveira, "Particle Swarm Optimization of MLP for the identification of factors related to Common Mental Disorders", Expert Systems with Applications 40 (2013) 4648–4652.
- [14] Nasir A. Al-geelania, M. Afendi M. Piah, Zuraimy Adzis, Munir A. Algeelani, "Hybrid regrouping PSO based wavelet neural networks forcharacterization of acoustic signals due to surface dischargeson H.V. glass insulators", Applied Soft Computing 13 (2013) 4622–4632.
- [15]L.Sahoo, et al., "An efficient GA-PSO approach for solving mixed-integer nonlinear programming problem in reliability optimization", Swarm and Evolutionary Computation (2014), http://dx.doi.org/10.1016/j.swevo.2014.07.002i.
- [16] Manoj Thakur, "A new genetic algorithm for global optimization of multimodal continuous functions", Journal of Computational Science 5 (2014) 298–311.
- [17] J. Kennedy, R. Eberhart, "Swarm Intelligence Morgan Kaufmann," 3rd edition. Academic Press, (2001). New Delhi, India.

- [18] Hamedi, Mahyar, et al. "Comparison of Multilayer Perceptron and Radial Basis Function Neural Networks for EMG-Based Facial Gesture Recognition." *The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications.* Springer Singapore, 2014.
- [19] Ndiaye, A., et al. "Development of a multilayer perceptron (MLP) based neural network controller for grid connected photovoltaic system." *International Journal of Physical Sciences* 9.3 (2014): 41-47.
- [20] Roy, Moumita, et al. "Ensemble of multilayer perceptrons for change detection in remotely sensed images." *Geoscience and Remote Sensing Letters, IEEE*11.1 (2014): 49-53.
- [21] Hassanien, Aboul Ella, et al. "MRI breast cancer diagnosis hybrid approach using adaptive ant-based segmentation and multilayer perceptron neural networks classifier." *Applied Soft Computing* 14 (2014): 62-71.
- [22] Aydin, Kamil, and Ozgur Kisi. "Damage detection in Timoshenko beam structures by multilayer perceptron and radial basis function networks." *Neural Computing and Applications* 24.3-4 (2014): 583-597.
- [23] Velo, Ramón, Paz López, and Francisco Maseda. "Wind speed estimation using multilayer perceptron." Energy Conversion and Management 81 (2014): 1-9.
- [24] Lee, Sangjae, and Joon Yeon Choeh. "Predicting the helpfulness of online reviews using multilayer perceptron neural networks." *Expert Systems with Applications* 41.6 (2014): 3041-3046.
- [25] Azim, Shambeel, and Swati Aggarwal. "Hybrid model for data imputation: Using fuzzy c means and multi layer perceptron." *Advance Computing Conference (IACC), 2014 IEEE International*. IEEE, 2014.
- [26] Chaudhuri, Sutapa, Sayantika Goswami, and Anirban Middey. "Medium-range forecast of cyclogenesis over North Indian Ocean with multilayer perceptron model using satellite data." *Natural Hazards* 70.1 (2014): 173-193.
Evaluation and Analysis of eLearning Models

Ashish Chatkawar Dept. of Computer Engineering and Applications National Institute of Technical Teachers' and Training and Research, Bhopal, India <u>achatkawar@nitttrbpl.ac.in</u>

Parivesh Kasturia Dept. of Computer Engineering and Applications National Institute of Technical Teachers' and Training and Research, Bhopal, India <u>pkasturia@nitttrbpl.ac.in</u>

M A Rizvi Dept. of Computer Engineering and Applications National Institute of Technical Teachers' and Training and Research, Bhopal, India <u>marizvi@nitttrbpl.ac.in</u>

Abstract—This paper presents the Model of Strategic e-Learning to explain and evaluate e-learning models. This study presents aspects related to e-learning solutions including the educational context of elearning and pedagogical principles which inform e-learning approaches, examples of e-learning approaches and technologies, and the rapid changes being experienced in educational systems. It then provides a summary of what is known about the impacts of e-learning on education, the economy and society. This study concludes by presenting several future research directions.

Keywords: eLearning, A-VIEW, VidyoDesktop, Features of elearning Models, Moodle, Google hangout and People Link

I. INTRODUCTION

There are many different definitions for the term 'e-learning', which can confuse those at the early stages of exploring elearning. However, Jisc uses a definition whose meaning is deemed to be most accurate: "e-Learning can be defined as 'learning facilitated and supported through the use of information and communications technology'. It can cover a spectrum of activities from the use of technology to support learning as part of a 'blended' approach (a combination of traditional and e-learning approaches), to learning that is delivered entirely online. Whatever the technology, however, learning is the vital element."The term 'e-learning' therefore essentially covers the use of computers and technology as a vehicle for knowledge exchange within teaching and learning. What does the 'e' in 'e-learning' represent? We are often asked this and the answer is that the 'e' used to represent 'electronic' but nowadays it merely signifies the use of technology. In some circles within the education sector, some refer to the 'e' as 'enhanced'.

Benefits of e-learning

An immediate potential benefit of considering implementing e-learning is that it can be seen as an additional avenue with which to support teaching and learning practice. E-learning covers such a wide sphere that it is difficult to point out any benefit as a given, so any benefits should initially be termed 'potential' benefits. However, many cite the following as broad benefits that e-learning supports:

- The ability to provide distance learning (learning not on campus)
- A blended learning/teaching approach (using face-to-face and technology)
- The use of technology to support a wide range of educational activity

II. COMMON E-LEARNING MODELS

There are countless tools, techniques, people, acronyms and resources associated with e-learning. Below we cover some of the ones we think you are likely to come across.

S.No	Tools/Terms	Description
1.	Virtual learning	An online space provided by the institution to support e-learning. All forms of
	environment (VLE)	digital media can be delivered using its various tools. There is a wide range of
		VLEs on the market.
2.	Personal learning	A concept of understanding that individuals utilize a range of networks -
	environment (PLE)	combining both institutional and personal networks and devices to learn.
3,	Moodle	An Open Source VLE that is commonly used in the UK.
4.	Podcast, either video or	A method of delivering multimedia content. The video podcasts are sometimes
	audio	called vodcasts/vidcasts.
5	Face-to-face teaching	A method of delivering teaching and learning that is normally used to distinguish
		between the classroom teaching environments and online.
6.	Blended learning	A method of delivering teaching and learning that involves both face-to-face
		teaching and the use of technology together at the same time. For example the
		internet may be used to support a session that includes interactive tasks for the
		learner.
7.	Distance Learning	Delivering teaching and learning remotely, typically using technology and the
		internet Multimedia resource are often incorporated to provide context to text-
-		based resources.
8.	WiKi	An editable tool for working with others that has a trackable history of changes
		(wikipedia is the most popular example). Much like a blog, its strength is that can
		be used to share multimedia resource.
9.	Mobile learning (use of	Using mobile devices including mobile phones to facilitate teaching and learning.
	mobile phones and other	
10	handheld devices)	
10.	Open-source software	Software that is provided under a license that permits the user to have access to
		the source code. Open source software can be used to create, consume and
		delivery multimedia. An example is the audio editing tool 'audacity' which is
11		very popular for creating and editing audio podcasts.
11.	Social media	Social media tools are used to communicate between people on the web and can
		be used to support teaching and tearning. For example it is often desirable to use
		social media tools to facilitate online community opportunities including learner
		collaboration.

III. MODEL- AMRITA VIRTUAL INTERACTIVE E-LEARNING WORLD (A-VIEW)

Amrita Virtual Interactive e-learning World A-VIEW is an advanced multi-modal, multi-platform, collaborative e-learning solution which allows an instructor to interact with a large number of learners transcending geographies on a real-time basis through live audio video streaming and synchronized content sharing. A-VIEW allows the instructor to perform live evaluation of the learners and to get real-time feedback from attendees on the go. A-VIEW can also act as an online interacting tool that can support online social collaboration and interactions with multiple users from various locations simultaneously.

A-VIEW[7] is a versatile e-learning platform which enables to setup Online Classrooms, Meetings and Trainings with ease.

- **A.** Online Virtual Classrooms:- Online virtual classrooms can be easily set up with A-VIEW leveraging its multi-screen display capability. The multi screen display capability of A-VIEW provides a near life classroom experience where the instructor and the content being shared can be seen simultaneously, legibly on large screens.
- **B.** A-VIEW facilitates online workshops:- A-VIEW can be used as an effective tool to conduct online workshops by connecting experts and audience from around the globe through its live video and audio streaming capabilities. This clubbed with its multi-modal and multi-display capabilities can deliver an unforgettable experience in conducting online workshops.
 - IIT Mumbai trains more than 25000 teachers through online workshops using A-VIEW.
 - Dr. Sam Pitroda addressed three thousand students at multiple locations in India.
 - NITTTR Bhopal[5] used A-VIEW to train more than 4000 teachers/ITI instructor organized by Ministry of Labor & Employment[8], DGET, Govt. of India.

C. Meeting Tool:A-VIEW can be used to interact with up to 8 simultaneous users at a time. This capability of A-VIEW lets it pitch itself as a competitive online meeting and social collaboration tool.

Benefits of opting A-VIEW:- for setting up online classes and meetings are

- A-VIEW is Free for Higher Education in India
- Users can contribute to the content
- Designed to act as Educational and Meeting tool simultaneously
- Highly Adaptable
 - Variable Number of Displays
 - o Low Bandwidth Usage (Rural Areas)
- Highly Customizable
 - For Educational System
 - o For Multiple Languages
 - For Existing Infrastructure

Feature of A-VEIW:

The high level features of A-VIEW are

- 1. Seamless live video and audio streaming (Meeting tool): The video conference feature of A-VIEW allows the user to have a seamless live video and audio streaming experience even in situations where the bandwidth is not very high. With its multi-bit-rate streaming options the software adapts itself to the best settings suitable for the bandwidth available. This means that instructors can reach out virtually to hundreds of classrooms and thousands of students spread across geographies effortlessly. A-VIEW supports interactions with multiple users simultaneously which enables its users to use A-VIEW as an effective meeting tool to organize meetings spread across the globe.
- 2. Content collaboration (Documents, Presentations, Video, 2D & 3D models, Whiteboards) :A-VIEW enables collaboration on content through multiple options. The users can collaborate on documents and presentations that are currently being discussed in classroom and can highlight, underline and add other annotations for the entire class to see, if permitted by the instructor.

A-VIEW also allows collaboration on interactive 2D animations and 3D models, unseen anywhere else so far, to provide a near life virtual experience to the learner in understanding complicated concepts with ease.

- **3.** Evaluation and Feedback: The quiz and poll options of A-VIEW helps the instructor to conduct live evaluation of the class that is being taught.
- 4. Video record, playback and edit functions:-A-VIEW records the session conducted through it and classifies and stores it in the library for later reference. The software also lets the user edit the videos recorded through a simple video editing interface.

VI. MODEL- VIDYODESKTOP

The VidyoDesktop[™] app brings high-quality video conferencing to Windows, Mac, and Linux computers so that everyone can participate at the office, from home, or on the road. VidyoDesktop VE (virtual edition) extends high quality conferencing and collaboration to virtual desktop and thin client environments. Connect, unify and empower: Vidyo puts users in the picture and in control. Vidyo works the way you do. It runs on the devices you are using now from smart phones to tablets, desktops to video room systems, bringing HD-quality video and content to every participant.

Features

A. VidyoDesktop:-Connecting Diverse Devices, Everywhere VidyoDesktop and VidyoDesktop VE (virtual edition) extends high-quality video conferencing to Windows, Mac and Linux computers, allowing users on these systems to participate at their office, from home, or on the road. With support encoding at resolutions up to full HD and dual-screen multipoint video, VidyoDesktop delivers a first-class conferencing experience to any workspace. VidyoDesktop VE extends high quality conferencing and collaboration to virtual desktop and thin client environments.

KEY FEATURES

- Ultra HD 4k support to display rich content and multiple full HD participants
- · Multiple user-selectable layouts for continuous presence, active speaker, and shared content
- · Supported in Windows, Mac, Linux and select virtual desktop environments
- · In-conference public and private text chat

- Share any app or desktop into the conference, and switch between multiple streams of shared content
- Far-end camera control of Vidyo and third-party group solutions
- B. VidyoWeb: The Easy Way for Guests to Attend The VidyoWeb browser extension lets participants join conferences from within a web browser on desktop and laptop computers. Unlike VidyoDesktop, which allows credentialed users to control their own virtual conference rooms and send invitations, VidyoWeb is designed for guest participants who simply want an easy way to join a conference.

KEY FEATURES

- Easy access: Guests can join with a single click of the conference link
- HD conferencing in current versions of popular browsers, including Chrome, Internet Explorer, Firefox, and Safari.
- Unsurpassed quality: Dynamic adaption technology optimizes the experience for guests joining from remote locations and wireless networks
- C. VidyoMobile: Quality Meetings on Mobile Devices The VidyoMobile[™] app brings high-quality video conferencing to popular Apple and Android tablets and smartphones. Host a person-to-person or multi-party video conference from your office, home, or in transit on both wireless broadband and WiFi connections. As a full-featured endpoint in your VidyoConferencing[™] solution, VidyoMobile delivers transcode-free video conferencing for natural communication at the pace of conversation, without the broken pictures associated with traditional solutions.

KEY FEATURES

- · Join and host meetings from your Smartphone or tablet
- Make any participant or content share bigger with pinch-to-zoom and real-time panning
- Far-end camera control (FECC) of remote VidyoRoom systems
- · Customizable layout with up to 4 on-screen participants
- Multipoint Content Sharing: toggle to switch between shares from multiple participants
- D. VidyoSlate:-View, share, and annotate content VidyoSlate[™] lets you share content from your tablet into the conference, view content shared by other conference participants, and annotate shared content with pen and highlighter tools. You can also draw and save virtual whiteboards for a graphical collaborative experience. With integrated two-way voice and video support, VidyoSlate turns your iPad into an all-in-one mobile collaboration experience.

KEY FEATURES

- View content on a second screen: when participating from a VidyoRoom or from any Vidyo app, the VidyoSlate app displays shared content from any conference participant
- Create notes and annotations: mark up shared content with your finger or stylus, and save your annotations to your tablet
- Share content from your tablet: share documents, presentations, and spreadsheets from email attachments, browser downloads, and cloud storage including Box, Dropbox, and Google DriveTM
- Draw on whiteboards: use your tablet as an electronic whiteboard to share notes and drawings into the conference, and email saved whiteboards from your tablet

V. MODEL- GOOGLE HANGOUT FOR DISTANCE LEARNING

If seeing, hearing, or sharing with your online students could help the learning process, then Google+ Hangouts could be a great, engaging venue for your students. Considering that many students use Social Media as one of the primary ways to communicate, this might be a way to better engage younger generations.

Feature/Functions

Here are six ways you might use Google+ Hangouts for online education:

1. Screen sharing :- To share your screen, all you have to do is click "Share screen" at the top. Then you can share a window or your entire screen. This is a great feature for teaching software so your students can see exactly how you are doing something.

- 2. Notes & Google Document collaboration:-You can add a note to the chat or you can open up a shared Google document. This is a very powerful tool for meetings, collaboration, and training. Let's say you want to talk to an online student about her term paper. You could open up the paper and go line by line while you are seeing and talking to each other.
- 3. Sketch Pad:-The Sketch Pad feature allows you to draw within the chat window in case you want to get creative or need to whiteboard ideas.
- 4. **Circles could be classes:**-You can invite people through e-mail or circles. If you grouped all of your students into classes, you could invite an entire class in seconds. (If you invite more than nine people besides yourself, when ten people are in the Hangout, the latecomers will receive a message that the Hangout is full.)
- 5. Training that requires personal one-on-one:-Hangouts would be great for language, arts, music, and many other subjects that require personal interaction with an instructor.
- 6. Hangout API:-The Hangout API allows developers to integrate Hangouts into their own Web apps. How useful would it be if your LMS integrated Hangouts into your course room?
- 7. What Hangouts does not do:-Hangouts does not have a way, by default, to record and play back the meetings. As far as I can tell you cannot schedule a meeting through Hangouts.

NITTTR, Chandigarh Technology Enabled Learning (NCTEL), NCTEL is another step in this growth to integrate ICT in T-L system. The total No. of Video Lectures available as on 2015/12/01: 1027 [6] through the google hangout.

MODEL- PEOPLE LINK

Today, when you shop for communication solutions, you can choose a variety of vendors available in the market who deals with various communication solutions. So you might be thinking why should you choose PeopleLink[4]? Companies come to peoplelink because an investment in our products delivers fast, proven ROI (Return on Investment), which consistently delivers a superior end-user experience, and impressive productivity gains, make you choose the best business solutions. With our video conference solutions, you don't need to go to the conference room to have a conference. PeopleLink easily extends your room system conferencing application to remote and desktop users for voice, video and data communications

Video Conferencing Solutions for Education, E-Learning

CCC. E-learning : PeopleLink offers seamless integration with 3rd party hardware products installed in the classrooms like e-podium, projectors, audio and video devices to extend an unparalleled feature of its extensive e-Learning suite accessible by the masses on any device / platform.

Technology plays a crucial role in the transformation of society. Communities are integrating these tools of technologies for their own development processes. Education sector being one of the cornerstones of society's development needs to use these tools to enhance the experience of Teaching / Learning. PeopleLink offers a leading eLearning suite, which can connect and extend the expert's availability to the remotest corner of the globe to connect the students who are unable to access it due to limited availability of necessary Infrastructure.

PeopleLink offers a unique solution for this vertical capable of delivering best in Class Video experience, flexible scalability, and a comprehensive set of data collaboration tools which make learning extremely effective & efficient. PeopleLink elearning suite are backed with a strong integrated approach, focusing on, securing the existing Investments of the user by connecting seamlessly with the available 3rd party hardware for providing an unparalleled Video collaboration experience.

Advantages for Education Vertical

- Gain a competitive edge with PeopleLink eLearning Suite for extending effective content & lecture delivered to the students.
- Leverage the utility of inbuilt video mixing by connecting multiple camera Inputs (Like, the camera focusing on the Lecturer, Integrated Smart board, Virtualizer etc.) and share all the videos simultaneously with the remote students for delivering a real classroom experience.
- Unlimited Inbuilt recording in a nonproprietary format assists in creating a digital Library for students to access the recordings later for revisions.
- Public / Private Chat option extends the feature of conducting group discussions & One to One feedback between the trainer and the students.
- Video Consultation helps educational Institutes in conducting Interviews / Counseling, giving career consultations to its students, organizing Parent Teacher meetings over video etc.
- Virtualize your classrooms to optimize the Operating costs & expand the reach & connectivity even to the remote location.
- Real Time sharing of Media with the students

- Evaluate Training effectiveness, take feedbacks & conduct basic questionnaire based evaluations with Inbuilt Poll option
- Flexible scalability and mobility gives the freedom of joining from any device.
- Inbuilt Whiteboard shares with the ability to offer a real time two way annotation makes learning more comprehensive & fun.
- Enhance Group Learning among students by creating virtual sub groups for conducting focus discussions and conducting team tasks & activities.
- *DDD.* PeopleLink Cloud : Collaborate Efficiently Meet, share and collaborate efficiently with colleagues, customers and suppliers, using customized cloud video conferencing services of Peoplelink. Our cloud based video conferencing is a complete one stop solution that includes infrastructure customized as per your requirement, implementation and management and video-ready managed networks. We offer secure, seamless, unhindered and high quality interaction to the geographically dispersed workforce to communicate effectively and productively.

With our unique SLA and direct services, we ensure you that you are not having sleepless nights

Share Anything on your Desktop/Computer/Laptop/I Pad/Smartphone

PeopleLink with its rich API's customized as per the market demand makes sure that you can do much more than simple connecting to various devices.

Features

- Telepresence class solution that enables you to communicate and collaborate without the need of face to face meeting at 1080P.
- Helps to bring the benefits of personal interaction into virtual meetings.
- Specialized in integrating various video and voice platforms to create TelePresence solutions, e-Education solutions and web based Video solutions.
- Special Customized API's like: Digital White Board, Document Sharing, Desktop/Application Sharing, Recording, File Transfer, Video and Desktop Sync, Lock Conference, E-mail Scheduling, Multiple Room Creation, Public and Private Chat, Support for Android/ IOS, Remote Camera Control, SMS integration, Dual Display, Video Mixer, Voting, Customized Video Layouts, Roll Call, Media File Sharing, Banner / Title, Conference Grouping, Video Rotation, Co-Browsing, Voice Activation etc have been the key differentiator over the years.

VI. CONCLUSIONS

This study proposed the Model of Strategic e-Learning commonly models. In this study we analysed various on-line learning models like A-VIEW, Google Hangout, People Link and VidyoDesktop on various parameters. Potential elements and corresponding features in each model are also illustrated and elaborated in this work.

The actual impact of past e-learning programs on student achievement, and related national economic development and societal changes in developing countries is at this point difficult to estimate due to the paucity of quantified research. Some research has indicated that certain types of e-learning approaches, especially those that blend more traditional teacher centered teaching with technology, and technology that encourages student interaction with the content, does improve student learning. Cross-country comparison research finds significant impacts of education on economic development, and indicates that improved education does lead to a better workforce and higher incomes of graduates. What is needed is research that follows e-learning graduates and others over time to identify how strong the e-learning impact is, and what educational factors are particularly effective. Similarly, the impact of e-learning on societal factors can only be indirectly inferred. So It can be summarized that one can adapt any e-learning models as per one's nee.

VII. REFERENCES

- [1] Xiao Jun1,*, Wang Lamei1, Zhu Xiaoxiao1, : Shanghai Engineering Research Centre of Open Distance Education, Shanghai Open University, Shanghai, P.R.China.
- [2] Dziuban, C., & Moskal, P., "A course is a course is a course: Factor invariance in student evaluation of online, blended and face-to-face learning environments", Internet and Higher Education, 2011, pp. 1-6
- [3] Hiltz, S. R. "Evaluating the Virtual Classroom." In Online Education: Perspectives on a New Environment, Harasim, L. M. (ed.). Praeger, New York, 1998, pp. 133-183.
- [4] Information retrived from url http://www.peoplelink.in.

- [5] Information retrived from National Institute of Technical Teachers' Training and Research, Bhopal url : http://www.nitttrchd.ac.in.
- [6] Information retrived from National Institute of Technical Teachers' Training and Research, Chandigarh url http:// www.nitttrbpl.ac.in.
- [7] Information retrived from Amrita University url http:// www.aview.in
- [8] Information retrived from Advanced Training Institute, Mumbai under D.G.E&T Ministry of Labour and Employment, Government of India
- [9] Coomey, M. & Stephenson, J. "Online Learning: It is all About Dialogue, Involvement, Support and Control According to the Research." In Teaching & Learning Online: Pedagogies for New Technologies, Stephenson, J. (ed.), Kogan Page Limited, London, 2001, pp. 37-52.

Thermohydraulic Performance of Solar Water Heater with Twisted Tapes and Boosted Mirrors

Sunil Kumar Department of Mechanical BIT Sindri, Dhanbad, Jharkhand, skumarbit@gmail.com

S.K Singh Assistant Professor Ex-Director Department of Mechanical BIT Sindri, Dhanbad, Jharkhand, India

Abstract—Heat Transfer in solar water heater could be enhanced by means of Twisted Tapes inside the fluid flow tube which induces swirl flow and turbulence promoter. Experimental investigation for solar water heater with twisted tapes inserted having twist pitch tube diameter ratio ranging from3 to 12 have been carried out for varying mass flow rates. Results concludes that such collectors would be preferable for higher grade energy collection at higher rate. Solar water heater having twisted tape inserted inside the flow tube having with boosted mirrors performed better than the plane one. It has also been observed that heat losses are reduced consequently increasing the thermal perform even better at higher values of intensity of solar radiation by boosted mirrors.

Keywords Heat transfer, solar water heater, turbulence, twisted tapes, boosted mirrors, thermal performance.

I. INTRODUCTION

Energy is the crucial input in the process of economic, social and industrial development. It plays a vital role in our present day life. The degree of development and civilization of a country is measured by the utilization of energy by human beings for their needs. The energy demands increases day to day because of population increasing, industrialization increases and transportation increases. The rate of energy consumption, increasing supply is depleting resulting in inflation and energy shortage. This is called "ENERGY CRISIS". So alternative, non conventional, renewable resources are to be developed for future energy requirement.

1.1 CLASIFICATION OF ENERGY RESOURCES:

Energy classification is based on its nature, quality, availability and storing capacity such as:

- (a) Commercial & non commercial energy resources
- (b) Renewable& non renewable energy resources
- (c) Conventional & non conventional energy resources.
- 1.2 Classification of non conventional energy resources:
- 1.3 Energy plays important role in the development of nation. The energy requirement is increasing with the population of the world. There is co-relation between the development and amount of energy used in the country. After the crisis of natural resources the other resources known as NON CONVENTIONAL ENERGY RESOURCES are used which are as follows:

(I) Solar energy

- (ii) Biogas or bio mass energy
- (iii) Ocean energy & Tidal energy
- (iv) Wind energy
- (v) Geo-Thermal energy
- (vi) Magneto hydrodynamic generation.
- Recent Developed technologies
- (i)Fly-wheel energy storage
- (ii)Hydrogen &Fuel cells.

(1.3) **SOLAR ENERGY**: The earth receives the energy from the sun in the form of electro-magnetic radiation. It is cheap and free from pollution. In India most part receives 4-7KWh of solar radiation per ²meter per day. Though the energy density is low and the availability is not continuous it has become possible to harness this abundantly available energy rapidly.

For many purposes by converting to heat or through direct generation of electricity solar energy can be used in following ways

(i) Conversion of solar energy into electricity

(ii) Conversion of solar energy into thermal energy

(iii) Photosynthesis.

(1.4)The main application of solar energy is:

(i)Heating of building

(ii)Cooling of building

(iii)Photovoltaic conversion

(iv)Solar water heater

(v)Solar air heater

(vi)Solar pond

(vii)Solar engine for water pumping.

II. INTRODUCTION OF SOLAR WATER HEATER:

In thermosyphon solar water heater natural circulation takes place through changes of density of water caused by heat absorbed from solar radiation. The solar heater water rises into the insulated storage tank and the other colder water from the tank flows by natural convection into the lower part of the solar collector. For this type of heater the height of the storage tank above the top of the solar collector is an important design parameter.

TYPES OF SOLAR WATER HEATER:

Following are some simple solar water heater (domestic) which are in general use:

- (i) Black rubber hose as solar water heater
- (ii) Shallow trough of water
- (iii) Japanese style plastic water bag
- (iv) Japanese pipe water heater
- (v) Drain back water heater.

SOME TYPICAL AND COMMERCIAL DESIGN OF WATER HEATER:

- (I) Natural circulation solar water heater(non pressurized)
- (II) Natural circulation solar water heater(pressurized)
- (III) Forced circulation solar water heater.

INSTRUMENTS USED FOR PERFORMING EXPERIMENTS

(I)Small powered centrifugal pump (.25HP)

- (II) Digital microvoltmeter
- (IV) Pyranometer
- (V) Inclined tube manometer.
- (VI) Temperature sensors
- (VII) Thermometer

OTHER APPARATUS USED ARE

(i) Solar collector (a) Flat plate collector (b) Focussing collector

Based on heat transfer fluid

(a)Liquid collector which uses water as fluid

(b) Air collector which uses air or gases.

(ii) Glass cover: The cover plate through which the solar radiation must be transmitted as much solar radiation as possible to the absorber plate.

Also to minimize heat loss from absorber plate to environment.

To shield the absorber plate from direct exposure to weather.

To receive as much energy as possible for the longest period of time each day.

(iii)Absorber plate: Below the cover plate an absorber plate is required to receive the solar radiation which has black surface. The plate is generally metal sheet. The absorber has

(a)High thermal conductivity

(b) Adequate tensile and compressive strength

(c) Good corrosion resistance

(iv) G.I.Tube: In case of liquid flat plate collector metal tubes are attached with the absorber plate. The fluid passes through the tubes which are generally made of cu,al,or G.I.Tubes. The diameter is generally 1 to 1.5 cm but in present set 2.5 cm of G.I.Tubes has been used.





Fig 1



Fig 2

SPECIFICATION OF EXPERIMENTAL SET UP:

No of solar water collector	02
No of pass of the tube	04
Total no of tubes	08
External diameter of tube	2.2 cm
Internal diameter of tube	1.8 cm
Total length of tube	8.8 cm
Length of collector	1.08 m
Width of collector	1.00 m
Thickness of back insulation	8.4 cm
Thermometer	03
Pyranometer	01
Milivolt digital calibrator	01
Material of insulation	Thermocol, glasswool
Material of tube	G.I Tube
Material of absorber plate	Al sheet (blackened)
Back of the collector	plywood
Side of the collector	wooden insulation

Plastic pipes of suitable diameter for flow of water from pump to collector, from collector to storage tank and from water sump to centrifugal pump.

METHODOLOGY: The flat plate solar collector for water heating and air heating have been used for long time but not to be extent it should be due to limitation of heat transfer rate through them. The fluid flow tubes have been tried for better heat transfer by providing twisted tapes and also with boosted mirrors for constant wall temperature.

To analyze and investigate an experimental set was developed consisting one having smooth or conventional tubes and other having twisted tapes of varying geometry.

Simultaneously on both type of solar water heater experiments were performed at outdoor condition.

Effects of boosted mirrors were observed on both type of heater.

Data of intensity of solar radiation, water temperature, tube temperature, flow rates, pressures drops were collected.

Effects of various parameters on the performance of the collector were observed.

Table 2: Observation for 0.023kg/s mass flow

Date 23/03/09

						Panel-I				Panel-II						
SI N	Tim	Pyron Q meter	I _T (W/m ²)	Body Temperat ure (°C)		Tube Temperature (°C)			Body Temperatur e (°C)				Temperature (°C)			
0.	of Obs.	Readi ng (mV)		T _{b1} , T _{b2} , T _{b3}	Tg	Tø	T ₁₄	Tt	T _{st}	T ₅₄ , T ₅₅ , T ₅₆	Ttl	Tg	To	Tzi	Td	Tem p. t _i ,t _o (°C)
1	10.3 0A M	5.27	680	56.6 63.7 63.8	70	70	67	66	65	70.4 70.8 70.8	52	53	61	55	64	26 28
2	11.0 0A M	4.39	567	54.5 61.0 62.2	59	58	54	54	53	64.7 72.4 66.0	50	56	55	52	54	26 29
3	11.3 0A M	5.67	732	53.0 61.0 62.8	58	57	53	53	54	59.7 69.4 63.4	46	54	53	50	52	26 30
4	12.0 0PM	5.51	711	53.4 63.0 65.8	61	60	54	55	55	62.0 72.4 66.5	49	57	56	52	56	26 34
5	12.3 0PM	5.53	714	54.8 67.6 70.1	66	67	60	60	59	66.5 75.2 70.0	50	60	60	54	60	26 37
6	1.00 PM	3.53	436	45.7 52.3 56.4	51	51	44	43	44	51.5 63.4 57.2	51	51	44	43	44	26 29

Table - 3: Observation for 0.23 Kg/s mass flow

Date: 25/03/2009

		Руго-				Panel-I				Panel-II						Initial And
	of Obs.	meter Readin	[W/m ²]	Body Temperatur e (°C)		Tube Temperature (°C)		Body Temperature (°C) Tube Temperature (°C)			ature		Final Temp.			
		(mV)		T _{b1} , T _{b2} , T _{b3}	Tc	T _t	T ₁₆	Trd	Tes	$T_{bb},T_{b\bar{b}},T_{b\bar{d}}$	Tti	T ₁₂	Τđ	Tst	Τđ	(°C)
	10.30 AM	4.72	610	42.2 51.1 55.0	52	54	49	44	48	50.1 43.1 59.2	43	50	50	43	51	27 34
	11 :00 AM	5.54	715	51.1 61.2 61.4	56	57	52	50	51	53.4 47.8 63.4	47	53	53	50	53	27 37
Ī	11:30 AM	5.88	759	53.0 63.2 65.8	57	59	58	53	57	54.4 57.0 68.8	51	59	58	53	57	27 49
	12:00 Noon	5.80	749	52.9 65.1 68.0	62	61	53	55	57	56.0 51.2 70.5	53	61	60	55	59	27 43
	12:30 PM	4.84	62.5	52.7 66.2 68.6	62	61	53	55	57	56.3 50.6 71.4	50	58	56	52	57	27 42
	01:00 PM	434	560	47.8 60.1 61.3	56	58	50	51	52	50.5 46.4 62.1	47	52	51	50	51	27 40

Table-4: Observation for 0.046kg/s mass flow

Date 27/03/09

	Pyron Panel-I			Panel-II						Initial					
of Readi		meter Readin	Body Temperature (°C)	Tube Temperature (°C)			Body Temperature (°C) Tube Temperature (°C)					Final Temp.			
	(mV)		T_{b1},T_{b2},T_{b3}	Tc	Τđ	T ₁₄	Td	$T_{t\bar{t}}$	Tb4 Tb5, Tb5	${\tt T}_{\tt G}$	T12	Τđ	T ₁₄	Td	(°C)
10.30 AM	4.72	610	42.2 51.1 55.0	52	54	59	44	48	50.1 43.1 59.2	43	50	50	43	51	27 34
11.00 AM	5.54	715	51.1 61.2 61.4	56	57	52	50	51	53.4 47.8 63.4	47	53	53	50	53	27 37
11.30 AM	5.88	759	53.0 63.2 65.8	57	59	58	53	57	54.4 47.0 68.8	51	59	58	53	57	27 49
12.00 Noon	5.80	749	52.9 65.1 68.0	62	61	53	55	57	56.0 51.2 70.5	53	61	60	55	59	27 53
12.30 PM	4.84	625	52.7 66.2 68.6	62	61	53	55	57	56.3 50.6 71.4	50	58	56	52	57	27 42
01:00 PM	4.34	560	47.8 60.1 61.3	56	58	50	51	52	50.5 46.4 62.1	47	52	51	50	51	27 40



Graph 1





Graph 3

RESULT AND DISCUSSION

As it is clear from the graph that Ln Nussult No. has a linear relationship with Ln Reynolds No. and increases with it and also found in the graph that increase of Nusselt No.is more in the case of twisted tape inserted tube for lower Reynolds No.Heat transfer is increased because of high turbulence, hence Nusselt No is a function of heat transfer (h) but when velocity is more than Reynolds no the twisted tape gives less effect. Hence effect of twisted tape gives more enhancement in low Reynolds no and in low mass flow rate.

In the second graph the efficiency decreases as increases but for twisted tape inserted tube this decrease is less. Hence twisted tape gives better performance.

In the third graph shows the effect of solar radiation on the efficiency of solar collector. The effiency of solar collector increases with increase of solar intensity in the whole day. But the graph shows that at very intensity the enhancement of efficiency is less but in case of twisted tape the efficiency curve tends to be horizontal at higher intensity. The argument is that at higher intensity of solar radiation heat removal factor is increased hence heat transferred to atmosphere is more but in case of twisted tape the turbulence of water is more in the tubes hence heat transfer is more and gives better performance.

CONCLUSION

Twisted tapes could be inserted inside the flow tubes in solar water heater for enhancing heat transfer rate

Decreasing values of twist pitch to tube diameter ratio lead to increasing values of heat transfer rate and the pressure drop.

Twisted tape generates turbulence superimposed with swirlness inside the flow tube and consequently result in enhanced heat transfer.

Minimum value of thermal enhancement factor has been found to be 1.3 for twisted tape inserted in solar water heater.

Twisted tape solar water heater has better thermal performance than the plane solar water heater under the same operating condition.

Increased twisted tape solar water heater performance is remarkable at low and moderate values of the flow Reynolds No. and monotonous at high values of the Reynolds No.

Twisted tape solar water heater might perform even better at high intensity of solar radiation for high grade energy collection.

REFERENCES

- [1] SUKHATAMI S.P.Solar Energy Principles of thermal collection and storage", Tata McGraw hill publishing co. limited, NEW DELHI.
- [2] RAI G.D. Solar Energy Utilization, Khanna Publishers, NEW DELHI.
- [3] KEITH FRANK Principles of heat transfer," New chand & Brothers, Roorkee, INDIA
- [4] Prasad .B.N&Saini J.S Effect of Artificial roughness on heat transfer & friction factor in solar air heater solar energy."SOLAR ENERGY, 41(6),pp-555-560,1988.
- [5] Saeed, S.Hassan & Sharma D.K.Non Conventional Energy Resources."S.K. Kataria& Son's.New Delhi.
- [6] Garg, H.P & H.Rishikeshan D.S.Enhancement of solar energy on flat plate collector by plane mirrors,"40,pp-295-307,1988.
- [7] Kumar A & Prasad B.N, Investigation of twisted tape inserted solar water heater heat transfer, friction factor& thermal performance results."PERGAMON, Renewable Energy.19, pp-379-398, 2000.
- [8] Gambia W.R, Bundy R.D. & Wansbrough R.W, Heat transfer and pressure drop for water in swirl flow through tubes with twisted tapes."Chemical Engineering Prog.Sys.Ser 57(32),pp-127-137,1961

The Software Reliability Assessment Using Neuro-fuzzy Systems with Encoded Input

Bonthu Kotaiah Department of CS & IT, Hyderabad, India Assistant Professor, Maulana Azad National Urdu University, kotaiah_bonthuklce@yahoo.com

Mr. Mohd. Omar Department of CS & IT, Hyderabad, India Assistant Professor, Maulana Azad National Urdu University, <u>omarmanuu@gmail.com</u>

Abstract—We propose software reliability assessment methods by using neuro-fuzzy based systems and their effectiveness in assessing the Software Reliability. Also, we make a comparison between the neural networks based software reliability growth model and the fuzzy logic based software reliability growth models based on a homogeneous Poisson process applied to software reliability assessment of the entire system composed of several software components. Moreover, we analyze software fault count data to show numerical examples of software reliability assessment with the implementation of Neuro-Fuzzy systems based approach. Furthermore, we investigate the performance of an efficient software reliability assessment methods in this context. Also, we had shown the implementation of the approach by using Java programming language with some programs. We used the normalized root mean square error (NRMSE) as evaluation criteria. The experiments show that the non-parametric models are superior when compared to the parametric models in their ability to provide an accurate estimate when historical data is missing. A comparison among the neural network and fuzzy logic models are provided.

Keywords: Neuro-Fuzzy System, Software Reliability, Software Reliability Assessment, Software Fault, software reliability modeling.

I. INTRODUCTION

In recent years, researchers have proposed several software reliability models to estimate the number of faults during the software testing process. These parametric models have a set of unknown parameters. These parameters must be estimated using observed historic failure data. Statistical estimation methods have been heavily used to solve this problem. For example, least square, maximum-likelihood, instrumental variable method were used to solve the parameter estimation problem.

In the past few years, a number of different software reliability models have been introduced [4] to solve the above-described problem. These software reliability models have been developed in response to the urgent need for software engineers, system engineers and managers to quantify the concept of software reliability prediction. These models were useful in cases like:

- Managing reliability.
- Managing program changes.
- Monitoring test progress.
- Making software engineering and tradeoffs with schedules and cost.

In this paper, we provide a detailed comparison between various models that have been provided in literature for predicting faults in the software testing process. They are commonly known as software reliability models.

II. SOFTWARE RELIABILITY MODEL SELECTIONS

Selection of a particular model is a challenging problem for software reliability prediction. There are two reasons for that. They are the selection of the release time and the value of resource allocation decision. These factors can affect the accuracy of the prediction. In the past, several solutions have been proposed to address the solution for the above described problems. They are [6]:

- Use several software reliability models and select the one that gives the highest confidence.
- Use the recalibration method to compensate for the bias of a model.
- Use an adaptive model as an alternate approach

III. EVALUATION CRITERIA

The model performance was measured in terms of the NRMSE.

NMRSE =
$$\frac{1}{n-l} \sqrt{\frac{\sum_{k=l}^{8} (y(k) - y(k))^{2}}{\sum_{k=l}^{8} (y(k))^{2}}}$$

Where y(k) is the actual accumulated faults and Ay(k) is the estimated accumulated faults using various software reliability models.

IV. COMPARISON AMONG THE MODELS

In this section we cover the analysis part of the experiments. We have used three different data sets as a Benchmark, collected from different projects [1]. The data type is stochastic. Data sets range from military application projects, real time and control applications and finally, operating systems applications. This data are intensively used in literature to evaluate software reliability model [5].

A. NEURAL NETWORKS MODEL

From the observation in Table 1, The Neural Networks (NN) seems to perform relatively well. Although the neural networks are the best predictors in most cases, they have projected the lowest *NRMSE* in two of three projects. For example, in the military application projects they are 0.5644, which is the lowest

NRMSE among the entire project under study. In real time and control project # 1, the *NRMSE* is 1.0755, and for the operating system # ss1c, the *NRMSE* is 0.7714, mean while neural networks perform 66% better than fuzzy logic model.

Java Example for demonstrating the Construction of Neural Network

```
import java.awt.*; import
       javax.swing.*; import
       java.awt.event.*; /**
* Example: The Hopfield Neural Network
* This is an example that implements a Hopfield neural
* network. This example network contains four fully
* connected neurons. This file, Hopfield, implements a
* Swing interface into the other two neural network
* classes: Layer and Neuron.
                           * @author Jeff Heaton
* @version 1.0
*/
public class Hopfield extends JFrame implements ActionListener { /**
* The number of neurons in this neural network.
                                     */
   public static final int NETWORK SIZE = 4; /**
* The weight matrix for the four fully connected
* neurons.
                                     */
                          JTextField matrix[][] =
    new JTextField[NETWORK SIZE][NETWORK SIZE]; /**
* The input pattern, used to either train
* or run the neural network. When the network
* is being trained, this is the training
* data. When the neural network is to be ran
* this is the input pattern.
                                     */
   JComboBox input[] = new JComboBox[NETWORK SIZE]; /**
* The output from each of the four neurons.
*/
JTextField output[] = new JTextField[NETWORK SIZE]; /**
* The clear button. Used to clear the weight
* matrix.
                                     */
   JButton btnClear = new JButton("Clear"); /**
* The train button. Used to train the
* neural network.
                                     */
   JButton btnTrain = new JButton("Train"); /**
                  * The run button. Used to run the neural
* network.
*/
JButton btnRun = new JButton("Run"); /**
* Constructor, create all of the components and position
* the JFrame to the center of the screen.
*/
public Hopfield()
{
setTitle("Hopfield Neural Network"); // create connections
panel
```

```
JPanel connections = new JPanel();
                 connections.setLayout(
       new GridLayout(NETWORK SIZE, NETWORK_SIZE) ); for ( int
             row=0;row<NETWORK SIZE;row++ ) { for ( int</pre>
      col=0;col<NETWORK SIZE;col++ ) { matrix[row][col] = new</pre>
           JTextField(3); matrix[row][col].setText("0");
                 connections.add(matrix[row][col]);
   Container content = getContentPane(); GridBagLayout gridbag = new
   GridBaqLayout(); GridBaqConstraints c = new GridBaqConstraints();
                       content.setLayout(gridbag);
   c.fill = GridBagConstraints.NONE; c.weightx = 1.0;
                           // Weight matrix label
     c.gridwidth = GridBagConstraints.REMAINDER; //end row c.anchor =
                       GridBagConstraints.NORTHWEST;
                                 content.add(
   new JLabel ("Weight Matrix for the Hopfield Neural Network:"),c);
                              // Weight matrix
                    c.anchor = GridBagConstraints.CENTER;
           c.gridwidth = GridBagConstraints.REMAINDER; //end row
                        content.add(connections,c);
                               c.gridwidth = 1;
                            // Input pattern label
                  c.anchor = GridBagConstraints.NORTHWEST;
    c.gridwidth = GridBagConstraints.REMAINDER; //end row content.add(
   new JLabel("Click \"Train\" to train the following pattern:"),c);
                              // Input pattern
 String options[] = { "0", "1"}; JPanel inputPanel = new
                       JPanel();
       inputPanel.setLayout(new FlowLayout()); for ( int
 i=0;i<NETWORK SIZE;i++ ) { input[i] = new JComboBox(options);
                   inputPanel.add(input[i]);
     c.gridwidth = GridBagConstraints.REMAINDER; //end row c.anchor =
           GridBagConstraints.CENTER; content.add(inputPanel,c);
                           // Output pattern label
                  c.anchor = GridBagConstraints.NORTHWEST;
    c.gridwidth = GridBagConstraints.REMAINDER; //end row content.add(
 new JLabel("Click \"Run\" to see output pattern:"),c); // Output pattern
  JPanel outputPanel = new JPanel(); outputPanel.setLayout(new
FlowLayout()); for ( int i=0;i<NETWORK SIZE;i++ ) { output[i] =</pre>
        new JTextField(3); output[i].setEditable(false);
                  outputPanel.add(output[i]);
             c.gridwidth = GridBagConstraints.REMAINDER; //end
                 row c.anchor = GridBagConstraints.CENTER;
         content.add(outputPanel,c);
                  // Buttons
 JPanel buttonPanel = new JPanel(); btnClear = new
 JButton("Clear"); btnTrain = new JButton("Train");
            btnRun = new JButton("Run");
         btnClear.addActionListener(this);
         btnTrain.addActionListener(this);
          btnRun.addActionListener(this);
buttonPanel.setLayout(new FlowLayout());
```

```
buttonPanel.add(btnClear); buttonPanel.add(btnTrain);
buttonPanel.add(btnRun); content.add(buttonPanel,c);
// adjust size and position pack();
Toolkit toolkit = Toolkit.getDefaultToolkit(); Dimension d =
toolkit.getScreenSize(); setLocation( (int)(d.width-
this.getSize().getWidth())/2, (int)(d.height-
this.getSize().getHeight())/2 );
setDefaultCloseOperation(WindowConstants.DISPOSE ON CLOSE);
setResizable(false);
}
/**
* Used to dispatch events from the buttons
* to the handler methods.
* @param e The event
*/
public void actionPerformed (ActionEvent e)
if ( e.getSource() == btnRun ) run();
else if ( e.getSource() == btnClear ) clear();
else if ( e.getSource() == btnTrain ) train();
}
/**
* Called when the neural network is to be ran against
* the input.
*/
protected void run()
                                         {
boolean pattern[] = new boolean[NETWORK SIZE]; int
wt[][] = new int[NETWORK SIZE][NETWORK SIZE]; for (
int row=0;row<NETWORK_SIZE;row++ )</pre>
for ( int col=0;col<NETWORK SIZE;col++ )</pre>
wt[row][col]=Integer.parseInt(matrix[row][col].getText());
for ( int row=0;row<NETWORK SIZE;row++ ) {</pre>
int i = input[row].getSelectedIndex();
if ( i==0)
pattern[row] =
false; else
pattern[row] = true;
ł
Layer net = new Layer(wt);
net.activation(pattern);
for ( int row=0;row<NETWORK SIZE;row++ )</pre>
{ if ( net.output[row] )
```

```
output[row].setText("1"); else
      output[row].setText("0");
             if ( net.output[row]==pattern[row] )
    output[row].setBackground(java.awt.Color.green); else
        output[row].setBackground(java.awt.Color.red);
                                              }
                                              }
                                             /**
* Called to clear the weight matrix.
*/
protected void clear()
{
for ( int row=0;row<NETWORK SIZE;row++ ) for ( int</pre>
col=0;col<NETWORK SIZE;col++</pre>
                                                      )
matrix[row][col].setText("0");
}
/**
* Called to train the weight matrix based on the
* current input pattern.
                                             */
                                  protected void train()
                                              {
 int work[][] = new int[NETWORK SIZE][NETWORK SIZE]; int bi[] =
                      new int[NETWORK SIZE];
    for ( int x=0;x<NETWORK SIZE;x++ ) { if (
   input[x].getSelectedIndex()==0 ) bi[x] = -1;
else
bi[x] = 1;
}
for ( int row=0;row<NETWORK SIZE;row++ ) for ( int</pre>
col=0;col<NETWORK SIZE;col++ ) { work[row][col] =</pre>
bi[row] * bi[col];
}
for ( int x=0;x<NETWORK SIZE;x++ ) work[x][x]</pre>
-=1;
for ( int row=0;row<NETWORK SIZE;row++ ) for ( int</pre>
col=0;col<NETWORK SIZE;col++ ) {</pre>
int i = Integer.parseInt(matrix[row][col].getText());
matrix[row][col].setText( "" + (i+work[row][col]));
}
}
/**
* Main program entry point, display the
* frame.
* @param args Command line arguments are not used */
static public void main(String args[])
JFrame f = new Hopfield(); f.show();
}
}
```

B. FUZZY LOGIC MODEL

The Fuzzy logic model performs relatively well throughout all the experiments. Also, fuzzy logic has the best predictive capability in all the projects. For example, in the following, project military application #40,

and operating system # ss1c, the *NRMSE* are 1.1081, 1.2901, but in the real time control project # 1 the *NRMSE* is 0.9358, which is the lowest when compared between the two projects. As we have seen from this observation of the results, neural networks model gives the best or close to the best predictions in most of the cases. Also, the neural networks model is superior to the parametric models, regression model, and fuzzy logic.

Model	Military #	Real-time	Operating		
	40	Control # 1	Sys. SS1C		
Fuzzy Logic	1.1081	0.9358	1.2901		
Neural	0.5644	1.0755	0.7714		
Network					

Table 1: The performance of these models in term of normalized root mean square error NRMSE

V. ONE-WAY ANALYSIS OF VARIANCE (ANOVA)

The purpose of one-way ANOVA is to find out whether data from several groups have a Common mean. That is to determine whether the groups are actually different in the Measured Characteristic. From the observation in Table 1, we found out that the real time and control project NRMSE is the lowest compared to other models. This is why we decided to investigate this model by implementing an ANOVA test with a significance level, a = 0.1

In Table 2, we show the result of one-way ANOVA test, the *F* value is $F_{(2 \ 405)} = 2 \cdot 8 \ 3$ where 0.1. Is the significant level, and 2 and 405 represent the degrees of freedom. The observed significance level is p-value is equal to 0.06. We can assert that there is a significant difference among groups since p < 0.1.

Source	SS	DF	MS	F	Prob>F
Columns	0.0005	2	0.00025	2.83	0.0603
Error	0.03579	405	0.00009		
Total	0.03629	407			

Table 2: The result of the ANOVA test for the project Real-time and control #1

Data sets for both large and small projects from diverse sources have been analyzed. Results presented here indicate that some perform better than others in most cases. This research shows evidence that a non-parametric approach provides the lowest normalized root mean square error and accurate result for the range of values in the experiments in most cases. The presented results show also that a fuzzy model and neural networks can be used to build prediction models for software reliability.

The results of the fuzzy logic and neural networks models were very promising. The error difference between the actual and estimated response was small. This finding gives a good indication of prediction capabilities of the developed fuzzy model and neural networks.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a comparison between various software reliability models. They include the neural network and fuzzy logic models. The experiments show that the non-parametric models are superior to the parametric models in their ability to provide an accurate estimate of reliability when historical data is missing. Currently we are investigating the use of evolutionary computations to solve the software reliability-modeling problem.

VII. REFERENCES

[1] J. Musa, "Data analysis center for software: An information analysis center," Western Michigan University library, Kalamazoo, MI 1980.

- [2] K.M Atsumoto and K. Inoue, "Experimental evaluation of software reliability growth models," in Proceeding of the IEEE of FTCS18, pp.148-153, 1988.
- [3] S. Brocklehurst, P.Y. Chan, B. Littlewood, and J. Snell, "Recalibrating software reliability models" *IEEE Trans. Software Engineering*, vol.16, pp. 458-470, 1990.
- [4] J.D. Musa and K.Okumoto, "A logarithmic Poisson execution time model for software reliability measurement," in proceeding of the 7th inter. Conf. Software Engineering, pp. 475-478, 1984.
- [5] Y. K. Malaya, N. Karunanithi, and P. Verman, "predictability measures for software reliability models" in proceeding of the 14 the IEEE inter. Conf. Computer Software Applications, pp. 7-12, 1993.
- [6] N. Karunanithi, D. Whitely, and M.K., "Prediction of software reliability using connectionist models," *IEEE Trans. On software Engineering*, Vol. 18, no 7, pp. 563-574, 1992.
- [7] R. Sitte, "Comparison of software reliability growth prediction Neural Networks VSParametric recalibration," *IEEE Trans. On Reliability*, vol. 48, no.3, pp. 285-291, 1999.
- [8] Aljahdali, S., Sheta, A., "Modeling of Nonlinear Systems Using Genetic Algorithms", Proceedings of the IASTED International Conference on Artificial Intelligence and Application, 4-7, September 2001, pp.220-224.
- [9] Aljahdali, S., Sheta, A., and Rine, D., "Predicting Accumulated Faults in Software Using Radial Basis Function Network", Proceedings of the ISCA 17th International Conference on Computers and their Application, 4-6, April 2002, pp. 26-29.
- [10] Aljahdali, S., Sheta, A., and Rine, D., "Prediction of Software Reliability: A Comparison between Regression and Neural network Non-parametric Models", Proceeding of the IEEE/ACS Conference, 25-29, June 2012.

Mr.Bonthu Kotaiah obtained his Bachelor's degree in Computer Applications from ANU in 2001 and M.C.A from ANU in 2008. During the period from September, 2001 to 2011, he worked as an engineer(L-Cube Innovative Solutions), a Corporate Trainer (SyncSoft & Datapro(Vijayawada), COSS(Hyd.)). Presently he is working as Assistant Professor in MANUU, Hyderabad. His research interests include software Engineering, Neural networks, Fuzzy Logic.

Designing Display System using Led and Atmega-16 Controller with Low Power Consumption through Solar System

Prashant Dwivedy Electrical and electronics engineering N I T T T R Bhopal, India <u>dwivedyfamily@gmail.com</u>

Sunil Kumar Meena Electrical and electronics engineering N I T T T R Bhopal, India <u>smeena428@gmail.com</u>

M.A. Rizvi Department of Computer Engineering and Applications <u>marzvi@nitttrbpl.ac.in</u>

Abstract—A spinning display system has been designed using solar energy (photovoltaic cells) as source of energy in driving an electric motor with a mounted LED's on a stick. A continuous moving object appears in a fluid motion in limited frame rate of a rotating object to the human eye called POV (persistence of vision). It builds a stable image in the brain and the rotating disc is not visualised. The project describes applicability of solar energy as a power source in designing an efficient and economical LED spinning display system.

Keywords: Spinning LED; ATMEGA-16; Eclipse Software; Modified Level Shifter; LM-317.

I. INTRODUCTION

Light emitting diodes (LED) are playing an important role in the existing and future electronics world and doing number of useful electronic applications and also last as long as a standard transistor. From the term spinning led we define led displaying an image or text and thus saving lots of power. In this fast moving world we must just be concern for saving power for our future ones, for this we must research on our renewable energies and that too having more and more work out of that limited power, for that this spinning led concept will be very useful for future use. Solar Energy is a nature friendly energy source. Using solar energy we can make the circuit to run even when power supply is not available. The solar cell is used in circuit to provide power to the LED driving circuit. Using solar cell we can make the project more power efficient and environment friendly, 5-12 volt solar cell is used in the circuit as a power source to provide the necessary power to the circuit and a level shifter for low power consumption is also used for increasing the voltage obtained from solar energy.

For spinning led concept we just need a Microprocessor chip ATMEGA -16, led's, programming skills on ECLIPSE software for controlling the on-off connection of led's but the most important part i.e. image display can be distinguished only when motor is mountain at the back of the base containing led for a definite continuous and constant rotation. "Spinning LED Display" is a general name for the type of displays because of the Persistence of Vision phenomena i.e. human eye has a 'shutter speed' relatively slow $1\50$ per second this means that something if been rotated in less than 50th part of second we will be hardly able to see it but if this is happening continuously in a definite manner than surely are brain form a respective image out of it so in this way, the revolutionary approach pursued by this type of display can't be noticed remarkably. And also LEDs are fitted more closely so that it delivers higher quality image. Thus we can form a normal display using limited number of LED's. The challenge was programming a large spinning led programmed to display straight lines, text or pictures and can also be reprogrammed easily. In our display approach; we are working on 2 coordinate axes first is r which is going to be implemented physically and the second one is θ which is going to be implemented virtually. By this approach, they create a 360 degree view in contrast to our display which has a flat view and different sized pixels, and a total number of 24*pi*16 virtual pixels can be created by only using 16 pixels.

II.OBJECTIVES

Our main objective was to create a low cost display system with a different approach so as to reach the standards from a different way. The LEDs, being versatile lighting devices are used in this design which we expect to perform equal or better in some

respects compared to conventional displays. The following objectives were set at the start of the project:

A. High Resolution in a Small Area

We used 8 LEDs in the project. Our LEDs have 1.0mm x 0.5mm dimensions and there is 2.1mm spacing between neighboring LEDs. The LEDs are placed so that they create a screen that has a radius of approximately 33mm. It may be considered as a tiny screen but the pixels at the outermost row will have pixels only as big as a led and we will have smaller and smaller pixel sizes as we progress to the interior. In this respect, the overall screen is to be a high resolution one which we expected to see the image transmitted to it clearly.

B. Full 8 bit intensity control

8 bit intensity control objective was set since the beginning of the project. Since that time, we always progressed without changing this. Currently we use 8 LEDs, all of which are driven by 8 bits.

C. Consistent view without flicker

Another specification that is not changed from the start is the speed of the motor, 3600 rpm, which directly affect the refresh rate, hence the flicker. So we are setting our motor to rotate at 3600 rpm, which will give us a refresh rate of 60Hz, which is considerably higher that human flicker detection frequency

D. An interesting and attractive model

We wanted to have a display with an appealing design and make it simple and interesting with the development of programming. This objective is more important for us, so we made significant progress in this and we are tried to create something interesting and attractive which can be easily made through programming.

E. Wide Viewing Area

In everyday life, we usually share one display with many people, like we do in watching TV. In the areas where more than one people sitting close to a display are present, we expect from a display to be viewable from different viewing angles. Another design concern is, as expected viewing area. The LEDs we use have viewing angles as high as 150 degrees, so we will try to achieve wide viewing angles with these led. The moving of LEDs may spoil a perfect experience, but we will try to reach considerable viewing areas.

F. Higher Refresh Rates And No Flicker

One of the problems of CRT displays was the refreshing problem that LCD displays is able to resolve. Our LEDs are always moving, and at every turn, we refresh all the points. Our motor will have a limited speed, so we will try hard to optimize it for a better display. We plan to turn the motor at 3600rpm, which corresponds to 60Hz operating frequency

G. Smoothness vs. Pixel Based Display

We always talk about pixels when we discuss about cameras and displays. Actually, we have an advantage here because we have no distinct pixel sizes at the axis as we are rotating over. We expect that it will give the eye a more continuous view, but the results may be different. Our objective is to use this interesting fact to achieve a smoother view. We can also make the pixel changes as frequent as we want.

H. Using Renewable Energy

In the 21st century we know how it is necessary to use our renewable sources which are widely available. So we have tried to make our circuit work totally on solar energy so that our other sources of energy which are non - renewable get unaffected.

I. Low Power And High Voltage From Limited Source

We have used level shifter so that we can increase our voltage accordingly and can utilize more energy from a limited input in this way we are able to boost up voltage to drive motor at a higher speed of 3600 rpm.

SOLAR PLATE LM - 317 BATTERY LEVEL SOLAR PLATE MOTOR

III.METHODOLOGY

Fig. 1. Block Diagram

- A. Hardware Description
 - Solar Cell: To utilize energy from the sun directly to the battery is one of the easiest methods to store renewable energy as it requires a minimum of equipment. Solar panel with the highest possible power rating is also easy to carry and handle. Solar panels are more rugged and produce more power. Mostly panel rate around 12 volts and are fairly inexpensive. For better performance, a 16 volt panel will produce more current and will charge the battery faster.
 - 2) LM 317: It is an adjustable voltage regulator IC that simply means it provides Line Regulation and Load Regulations so we can adjust the output voltage by varying the resistance across the adjust pin. For this we need to have a fixed voltage across the battery. Directly connecting the solar panel to the battery may even explode due to the varying output from it. The voltage across R1 is maintained to be 1.25 V using an internal circuit. The Vout is kept constant and given by:

$$Vout = VR1^{*}(1 + R2/R1) + Iadj^{*}R2$$
(1)

This Iadj is in the order of microampere so it can be neglected to have a regulated output voltage.



- 3) Modified Level Shifter: Level Converters or Level Shifter are generally used for converting high voltage level to low or vice versa, i.e. they convert one voltage level to another. They play an important role in VLSI systems and are important circuit component in multi voltage systems. The level shifters can be designed to convert low voltage (5V) to high voltage (20V). We propose a low-power single input level shifter as shown in figure 3 given below. To reduce power consumption, there is no diode-connected TFT in the circuit and the pull-up TFTs and the pull-down TFTs are never turned on simultaneously. In addition, only a single input signal is required by this circuit, which simplifies the interface design considerably. This design minimizes the leakage. Delay can be reduced by using HSPICE tool (90nm).
- 4) *Eclipse Software* : The ECLIPSE software has provided us a platform where we can do various programming according to our requirement.
- 5) ATMEGA-16: The ATMEGA memory consists of three parts:
 - a) Data memory of SRAM : used for temporary storage of data value
 - b) Program memory, which is a Flash Memory, that can be rewritten up to 10,000 times
 - c) Finally the EEPROM memory, which is used for permanent storage of data values or initial parameters for the microcontroller.

ATMEGA –16 has 40 pins, it also has 4 PORTS naming as A0 to A7, B0 to B7, C0 to C7 and D0 to D7, the 9th pin is for RESET while 10th and 30th is for Vcc and 11th and 31th is acting as a ground.

B. Mechanical Assembly

Solar energy is a clean renewable energy source. Solar cells are the heart of any solar energy trapping circuit. We have used solar panel as a power supply and fed into a voltage regulator so as to have a constant voltage charging for the battery. Batteries are the easiest way to store energy. Rechargeable batteries like Li-ion, NiMH and Lead Acid are used depending on the user's requirements. The battery is thus got charged at constant voltage and at desired rate depending on the amount of current supplied. Solar energy is the most abundant but really a least used source of energy. In this project we are using level shifter for increasing voltage received from solar panel. The output from the level shifter is then used as the input for motor which is working on 12 volt. For using the VCC Port 10th or 30th of ATMEGA-16 as the voltage input for LEDs we simply connect one wire from the battery which is directly being charged by solar energy to the Vac Port of ATMEGA-16 and after that we get the

display according to our programming.

C. Algorithm

The main section of our work is coding, which determines that what display will be shown. The ECLIPSE software that we have chosen for feeding it in ATMEGA – 16 microcontroller is slightly based on C language and having some header files which are mainly used for initializing avr, for giving delay and for standard input and output. The port of ATMEGA-16 to which LEDs are fixed is initialized as 255 which in digital means that all are in 'ON' state. Let if we are using B port then we will write DDRB=255 which means that B0 to B7 will all be in 'ON' state so all LEDs will light up, so accordingly we can light any of the led in the similar manner. Some initial steps to be followed are:-

Step 1:- Header files are initialised according to our requirement.

Step 2:- While loop is to be run for infinite times so that we can get a constant display.

Step 3:- Under this while loop the main programming is done. For example PORTB=0b01000010 means at Port B 2nd and 7th pin is in 'ON' state. Between each two code we give a delay of some micro second, it is the main concept of the programming as this delay only gives us a elusion which, during a rotation gives us a display. In this manner with simple programming only we can print anything on the screen whatever we desire from project.

IV.RESULT

A. Result for solar cell

TABLE III.	READINGS OF SOLAR CELL
------------	------------------------

Solar panel size	Power rating	Power voltage max.
362*163*8MM	5W	17.2V

B. Result for battery

6V Application

- a) Output Voltage: Set for 7V
- b) Input voltage
 - i. Battery discharged (6V): 8.75V Min @ 1.5A (this is a little high for panel that are characterized for 6V applications)
 - ii. Battery charged (7V): 9V Min @ 10mA (e.g.)

C. Results for Level shifter

Level shifter Configurations	Power Consumption (PW)	Delay (ns)	Input Voltage (volt)	Output Voltage (volt)
Modified Level				
Shifter	402.2264	2.3376	6	20

FABLE IV.	READINGS OF]	LEVEL SHIFTER
ADLL IV.	INEADINGS OF 1	

D. Result for ATmega16

The display plate containing LEDs is being rotated by a motor which needs a power supply of 12 volt which is being provided by the output of the level shifter. While our display plate i.e. our all eight LEDs get power supply from Vcc which is a port being provided by the ATMEGA-16 and the pin number of it is 10th and 30th. Any of the two pin can be used to activate the display. We require 5 volt power supply for Vcc which is directly taken from the battery charged by solar panels. The display which we

see can be anything depends directly on our programming done in ECLIPSE software and display will occur depending on the programme.

V.CONCLUSION

The goals that have hoped in the beginning of the project have been achieved so we consider our research to be an overall success. It is a unique combination of software and hardware, and it satisfies our goal. This project suggests that solar energy is a reliable power source in building an efficient and economical LED spinning display system. We have shown here that a stable display can be achieved by unique combination of level shifter which is able to boost battery power to drive rotating objects with the LEDs. With a relatively small investment it is possible that the renewable power and a suitable programming code are able to produce any type of desired stable display system.

IV.REFERENCES

Y. Leblebici, S.M. Kang," CMOS Digital Digital Integrated Circuits", Singapore: Mc Graw Hill, 2ndedition, 1999.

- A. P. Chandrakasanet, "Minimizing power consumption in digital CMOS circuits", Proceedings of the IEEE, vol.83, no.4, pp.498-523, 1995.
- Wen-Tai Wang, "Level Shifters for high-Speed V to 3.3V Interfaces in a 0.13um Cu-Interconnection/Low-k CMOS Technology", Proc. Of Symp. On VLSI Circuits, pp.307-310, June 2001.

D. Awachalom and M. Flatté, "Challenges of semiconductor spin- tronics", Nat. Phys., vol. 3, p. 153, Mar. 2007.

R. R. King, A. Boca, W. Hong, D. Law, G. Kinsey, C.Fetzer, M. Haddad, K. Edmondson, H. Yoon, P. Pien, N. Karem, "High-Efficiency Multijunction Photovoltaics for Low-Cost Solar Electricity", IEEE Lasers and Electro-Optics Society, Nov.2008.

7. M.A.Green,"Third Generation Photovoltaics: Advanced Solar Energy Conversion", Berlin, Germany: Springer-Verlag, 2004.

INTERNATIONAL SOCIETY OF RESEARCHERS SCIENTISTS ENGINEERS AND DOCTORS

