International Journal of Research in Advent Technology, Vol.4, No.4, April 2016
E-ISSN: 2321-9637
Available online at www.ijrat.org

MINDS: Machine Intelligence Based Network Intrusion Detection System

Purushottam R. Patil
Ph.D.Research Scholar, Computer Sci. & Engg., Faculty of Engg. & Tech, JNU,Jodhpur, RAJ, India.
Email: purupatil7@gmail.com

Dr. Yogesh Sharma
Dean, Faculty of Computer Application, Professor (Maths.), JNU, Jodhpur, RAJ, India
Email: dryogesh121@rediffmail.com
Dr. Manali Kshirsagar
Vice President (Academic), ADCC Infocad, IT Park, Nagpur,M.S, India
Email: manali_kshirsagar@yahoo.com

Abstract*:* Network basedIntrusion detection systems(NIDS) aredesigned to identify and prevent the misuse of computer networks and systems. Effective and intelligent design approaches of NIDS using Machine Learning and Evolutionary computation is major research topic in Network and system security domain. The evolving network security systems need be part of the life system and this is possible only by entrenching knowledge into the network.Intrusion detection does not, in general, include prevention of intrusions. Effectiveness, adaptability and extensibility are the quality measures of NIDS. NIDS said more effective when it has high intrusion detection rate (TP Rate), low false alarm (FP Rate) and area under curve of ROC is 1. In this paper an NIDS hasproposed an intrusion based on adaptive genetic algorithm with clustering. A clustering algorithm will be used for creating intrusion and non-intrusion clusters. Adaptive Genetic Algorithm will be used to generate the perceived traits of normal and abnormal clusters, and Artificial Neural Networks for detecting abnormal packets similar to the ones given during learning sessions and for an artificial intelligence that detects anomalies not presented during learning. The Adaptive mutation operation of the GA will generate the optimal rule for detection of intrusion in the network to ensure the security. TheDataset for implementation Work is KDD Cup'99.

Index Terms—Network based intrusion detection system (NIDS), Clustering,genetic algorithm(GA), artificialneural networks (ANN), detection rate.

I. INTRODUCTION

Today, the number of Internet users is continuously increasing, along with new network services. As the internet grows, network security attack threats have become more serious. Many security vulnerabilities are exposed and exploited by attacks. Recent reports on Internet security breaches indicate that the frequency and the damage costs are continuously rising. The acronym "C.H.E.W."to be known are 1. **C**ybercrime – the notion that unknown person to attack you with the primary motive being financial gain . 2.**H**acktivism – attacks motivated by ideological differences. The primary focus of these attacks is not financial gain but rather persuading or dissuading certain actions or "voices." 3. **E**spionage – direct motive of gaining information on another organization in pursuit of political, financial, capitalistic, market share or some other form of leverage and Finally 4.**W**ar (Cyber) – the notion of a nation-state or transnational threat to an adversary's centers of power via a cyber-attack[2]. Attacks could focus on non-military critical infrastructure or financial services or more traditional targets, such as the military-industrial complex. One recent network attack trend is the use of network traffic. The accurate and rapid detection of network traffic anomaly is crucial to enhance the effective operation of a network. It is often difficult to detect the time when the faults occur in a network. An Intruder is a person who attempts to gain unauthorized access to a system, to damage that system, or to disturb data on that system. This person attempts to violate Security by interfering with system Availability, data Integrity or data Confidentiality. An Intrusion Detection System (IDS) is a network security technology originally built for detecting vulnerability exploits against a target application or computer. IDS come in a variety of "flavors" and approach the goal of detecting suspicious traffic in different ways. There are network based (NIDS) and host based (HIDS) intrusion detection systems. NIDS is a network

security system focusing on the attacks that come from the inside of the network (authorized users). Some systems may attempt to stop an intrusion attempt but this is neither required nor expected of a monitoring system.

In This paper, proposed NIDS is an integration of Machine Learning techniques, Classification (ANN), Clustering (K-means) and Evolutionary algorithm (GA). This paper aims to implement an NIDS to improve detection rate.

The rest of this paper is structured as follows: Section II presents the literature survey. Section III outlines proposed system methodology used in this research. Then the steps for implementation described in Section IV and the results are explained accordingly in Section V. Finally, section VI outlines conclusions and indicates areas for future work

II. LITERATURE SURVEY

In this section Survey of 8 research papers has been performed and views regarding Measurement factors like TP,FP, Detection rate etc. are discussed.

Dewan Md. Farid and Mohammad Zahidur Rahman [3] have proposed an algorithm for generating the minimal rule set for network intrusion detection. The proposed algorithm has detected network intrusions. Their algorithm has analyzed the large volume of network data and has considered the complex properties of attack behaviors to improve the performance of detection speed and detection accuracy. The experimental results have marked that this algorithm minimized false positives, as well as maximize balance detection on the 5 classes of KDD99 data. In their experiments proposed algorithm reduced the number of false positives by up to 90% with acceptable misclassification rates.

Adel Nadjaran Toosi and Mohsen Kahani [4] presented an evolutionary soft computing approach for intrusion detection. It successfully demonstrated its usefulness on the training and testing sub-set of KDD Cup 99 dataset. The ANFIS network was used as a neuro-fuzzy classifier for intrusion detection. A fuzzy decision-making engine was developed to make the system more powerful for attack detection, using the fuzzy inference approach. At last, given paper proposed a method to use genetic algorithms to optimize the fuzzy decision-making engine. Experimentation results showed that the proposed method is effective in detecting various intrusions in computer networks.

K. M. Faraoun and A. Boukelif [5] have studied the possible use of the neural networks learning capabilities to classify and detect network intrusions from a collected dataset of network traffic trace. The experiments showed that the neural networks were more suitable for 2-category classification problem, the discrimination between attacks classes remains a hard task. Since the high computation intensity and the long training cycles were the main obstacles to any neural networks NIDS, they have proposed a new learning schema to reduce the number of used samples using a k-means clustering algorithm. The obtained results demonstrate that the proposed technique performed exceptionally in terms of both accuracy and computation time.

Muna M. Taher Jawhar and Monica Mehrotra [6] have proposed a Network Intrusion Detection System and it was a new kind of defense technology of the network security. In this paper, they had given a new method by using Hamming and MAXNET for anomaly Intrusion Detection System and comparison with MLP network working with the same assumed parameters and testing with the usage of KDD dataset. The results were encouraging. The detection rate of the model was 95.0% and false negative is 4.94% which was relatively high when compared with conventional NIDS and other design with a neural network.

Marjan Bahrololum et al. [7] have applied decision tree model for known attacks identification. Here the processing time was decreased. Experimental results in standard dataset KDD99 have proved that the proposed method was able to achieve accuracy better than SOM. Unknown attacks would be detected by applying the unsupervised NN based on hybrid of Self-Organizing Map (SOM) for clustering attacks into smaller categories and supervised NN based on Back propagation for detailed clustering.

Mr.Vivek A. Patole et al. [8] have presented the Self-Organizing Map. It was an extremely powerful mechanism for automatic mathematical characterization of acceptable system activity. In the given paper they have described how they used Self-Organizing Maps for building an Intrusion Detection System. At the end of the paper they have figured out the advantages and disadvantages of Self-Organizing Maps and have explained how it was useful for building an Intrusion Detection System.

Hai Nguyen, Katrin Franke et al. [9] have proposed a new search method to get the globally optimal subset of relevant features by means of the correlation feature selection (CFS) measure. Classification

accuracies of c4.5 and BayesNet performed on KDD cup'99 data set is 99.41 %, 98.82 % respectively and of constraints and variables on the number of features inthe full set. We used a branch-and-bound algorithm in order to solve that M01LP. Experimental results showed that our approach outperforms the best-first-CFS and genetic algorithm CFS methods by removing much more redundant features and still keeping the classification accuracies or even getting better performances.

Samaneh Rastegari et al. [10] has presented a statistical rule learning technique which utilizes an entropy and volume-based approach for extracting seven simple features of network traffic over a short time window for intrusions detection. The combination of rule-based machine learning and statistical measurement and provided a generic and flexible model for our changing situations. This model is not dependent on the features extracted from packets headers and can be deployed at any point of network infrastructure. Two sources of data were used to evaluate our proposed algorithm (ESR-NID) and to compare against four existing techniques. Based on the results, ESR-NID produced an acceptable trade-off between accuracy and size of the final rule set. The performance of ESR-NID in this study on the combined DARPA/CAIDA dataset for 2 rules TP rate, FP Rate ,Detection Rate is 99.7, 96, 98.4 respectively.

Vinod Mahajan, Bhupendra Verma,[11] resented a distance based semi supervised clustering and probabilistic assignment technique to build the classifier and it has been achieved successfully. It permits both labeled and unlabeled instances to be used in training the classifier. The classifier achieves 94.8% accuracy at K=50.

III.PROPOSED SYSTEM METHODOLOGY

Proposed Machine intelligence Based Network intrusion detection system using network traffic anomaly detection approach an integration of k-means clustering, GA and Artificial Neural Network. The Input to the system is KDD Cup'99 Preprocessed dataset as shown in Table 1.

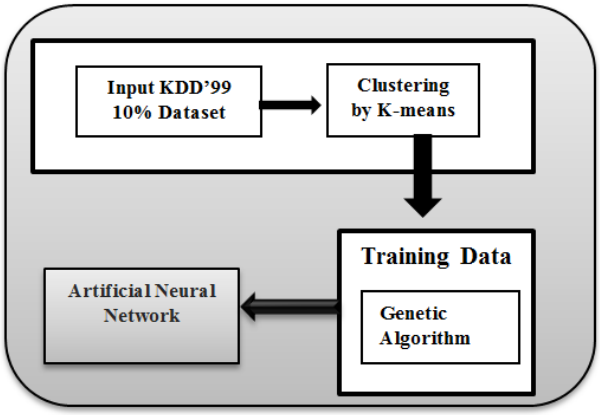

Fig. 1 Proposed Methodology

The methodology for implemented work is as follows:

1. Apply K-means clustering algorithm to Input Dataset (10 % KDD'99) to form clusters of intrusion and normal sets of data
2. Apply Genetic Algorithm to clusters to formulate rule from the perceived traits of normal and abnormal clusters (e.g. HHHLLHL)
3. Train and test the artificial neural network using the knowledge gained through GA
4. Use trained ANN for detecting abnormal packets similar to the ones given during learning sessions and for an artificial intelligence that detects anomalies not presented during learning.
5. Verification and validation of proposed method using network traffic data sets.

IV.STEPS FOR IMPLEMENTATION

Redundant data and insignificant features may confuse the classification algorithm, leading to the discovery of inaccurate or ineffective knowledge. Weka an open source machine learning tool used for preprocessing. WEKA **W**aikato **E**nvironment for **K**nowledge **A**nalysis. With it, a specialist in a particular field is able to use ML to derive useful knowledge from databases that are far too large to be analyzed by hand [12].

A. Data preprocessing in Weka

Firstly, Run Weka software, launch the explorer window and select the Preprocess tab. Then Open the KDD dataset in *.arff* or *.CSV* format, and enter what information do you have about the dataset (e.g. number of instances, attributes and classes)? What type of attributes does this data-set contain (nominal or numeric)? What are the classes in this dataset? Which attribute has the greatest standard deviation? What does this tell you about that attribute? After

Sr. No.	Category	Attack Type	KDD'99 Dataset 10% (Actual Nos.)	No. of Records (Prepro-cessed)
1	Normal	Normal	972788	87832
2	Back	DOS	2203	968
3	Pod		264	206
4	Land		21	19
5	Smurf		280790	641
6	Teardrop		979	918
7	Neptune		107201	51882
8	Nmap	PROBE	231	158
9	Satan		1589	906
10	Portsweep		1247	651
11	Ipsweep		1040	461
12	Phf	R2L	4	4
13	Guess_pwd		53	53
14	FTP_WRITE		8	8
15	Imap		12	12
16	Spy		2	2
17	Multihop		7	7
18	Warezclient		1020	893
19	Warezmaster		20	20
20	Buffer_overflow	U2R	30	30
21	Loadmodule		9	9
22	Rootkit		3	3
23	pearl		10	10
Total			**1369531**	**145693**

entered the dataset under Filter choose the Standardize filter and apply it to all attributes [14]. Results are as shown in Table 1.

B. Clustering (K-means)

The k-means algorithm is an algorithm to cluster n objects based on attributes into k partitions, where k < n. Simply speaking k-means clustering is an algorithm to classify or to group the objects based on attributes/features into K number of group. K is positive integer number. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroids.

C. Genetic Algorithm

A genetic algorithm (or GA) is a search technique used in computing to find true or approximate solutions to optimization and search problems. Genetic algorithms are categorized as global search heuristics. GA are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology's such as inheritance, mutation, selection, and crossover (recombination) [13].

TABLE 1: DATA PREPROCESSING RESULTS

1. Selection Process in Genetic Algorithm

Selection is the first stage in a genetic algorithm. In Selection process the individual genes are chosen for later processing (crossover) from the population. A selection procedure is given as follows:For each individual, the fitness values are calculated. Then the fitness values are normalized.

2. Fitness Function

For the selection process, the fitness function should be calculated.

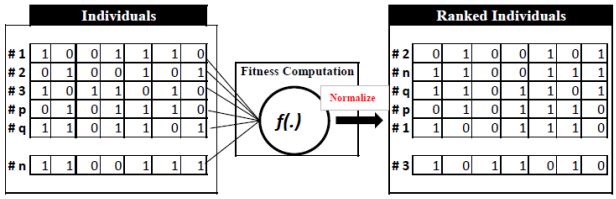

Figure2. Fitness Function

With reference to the fitness function, the chromosomes are selected. In fitness calculation, each individual can be selected more than once. In order to avoid that the below formula is used.

$$P_{(\text{selection})}=\mathbf{F}_{(\mathbf{parents})}/\sum_{i}(\text{parents})$$

3. Crossover in Genetic Algorithm

Cross over is a process of taking more than one parent solutions and producing a child solution from them. Here Random methods for selection of the chromosomes using MATLAB command

R= randi(imax, n)

4. Mutation in Genetic Algorithm

Mutation alters one or more gene values in a chromosome from its initial state. The Adaptive mutation operation of the GA will generate the optimal rule for detection of intrusion in the network to ensure the security. (E.g. Rule HHHHLHHHLL)

D. Artificial Neural Network:

Artificial Neural Networks will be used for detecting abnormal packets similar to the ones given during learning sessions and for an artificial intelligence that detects anomalies not presented during learning. For applying the ANN, training and testing the network needs to be done. For training the neural network, the above process will be performed by mutation operation in genetic algorithm.

E. Metrics for performance evaluation

This section introduces the metrics for NIDS performance evaluation with its merits and demerits for such an evaluation and analyzes them in a unified framework [15].

1. Detection rate and false alarm rate

Let TP be the number of attacks that are correctly detected, FN be the number of attacks that are not detected, TN be the number of normal traffic packet/connections that are correctly classified, and FP be the number of normal traffic packet/connections that are incorrectly detected as attack. The security requirement is determined by the TP rate and the usability requirement is decided by the number of FP. The concept of finding the optimal trade-off of the metrics used to evaluate an NIDS is an instance of the more general problem of multi-criteria optimization. In this setting, we want to maximize or minimize two quantities that are related to a trade-off, which can be done via two approaches. The first is to directly compare the two metrics via a trade-off curve. The second approach is to find a suitable way of integrating these two metrics in a single objective function to optimize.

2. Receiver Operating Characteristic (ROC) Curve

ROC curves are used to evaluate classifier performance over a range of tradeoffs between TP rate and FP rate. ROC curve is a plot which has the x-axis as the false alarm rate and y-axis as the detection rate.

Advantages: 1. Ability to separate error cost consideration from the NIDS performance.

1. The ROC curves remain Invariant under changing class distributions.

Disadvantage: small changes in false alarm rate may cause drastic differences in detection rate when the normal traffic abounds in comparison to the attack traffic in the network traffic.

3. The Area under ROC Curve (AUC)

AUC is a convenient way of comparing NIDSs. AUC is the performance metric for the ROC curve. A random NIDS has an area of 0.5 whereas an ideal one has an area of one [13].

4. Accuracy

The commonly used NIDS evaluation metric on a test Data is the overall accuracy.

Overall Accuracy = TP+TN/TP+FP+TN+FN

5. Precision

Precision (P) is a measure of what fraction of test data detected as an attack is actually from the attack class.
P = TP/TP+FP

6. Epoch

One iteration through the process of providing the network with an input and updating the network's weights typically many epochs are required to train the neural network

7. MSE (Mean Square Error)

Various metrics can be used to grade the
Performance of the neural network based on the results of the testing set
i.e. $M.S.E = (\text{target} - \text{Output})^2$

V. EXPERIMENTAL RESULTS:

In this proposed system supervised training of ANN is performed, both the inputs and the outputs are provided. The network then processes the inputs and compares its resulting outputs against the desired outputs. Errors are then calculated, causing the system to adjust the weights which control the network. This process occurs over and over as the weights are continually tweaked. The attack under study is 21 selected out of 37, Numbers of Clusters 2, Number of iterations 10, Population Size 50 i.e. Epochs 50.

TABLE II. RESULT PARAMETERS

Class	TP Rate	FP Rate	AUC
DOS	0.99	0.01	0.99
PROBE	0.92	0.0005	0.995
R2L	0.65	0.003	0.97
U2R	0.84	0.0015	1
Avg.	**0.85**	**0.003**	**0.988**

TABLE III: AVERAGE DETECTION

TABLE IV: COMPARISION OF DETECTION RATE

Ref. No.	DR (%)
10	98.4
11	94.8
16	92
17	92
18	97.13
Proposed	98.8

Figure3. Comparison of Detection Rate

VI. CONCLUSION

This paper has presented an adaptive NIDS system which has the ability to learn from experience in order to improve their performance and to adapt changes in the environment. The experiments demonstrate that the proposed system, MINDS has an excellent accuracy i.e. 98.8 % and amazing in attack U2R where it is 100 %. Results suggest that integrated system improves accuracy.

REFERENCES

[1] Global application & network security report 2014-2015, Radware Ltd, 2014.

[2] Myung-Sup Kim, Hun-Jeong Kang, Seong-Cheol Hong, Seung-Hwa Chung, and James W. Hong, "A Flow-based Method for Abnormal Network Traffic Detection", 2003.

[3] Dewan Md. Farid, Mohammad Zahidur Rahman, "Anomaly Network Intrusion Detection Based on Improved Self-Adaptive Bayesian Algorithm", Journal Of Computers, Vol. 5, No. 1, January 2010.

[4] Adel Nadjaran Toosi and Mohsen Kahani, "A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers", Elsevier B.V., 2007.

[5] K. M. Faraoun and A. Boukelif, "Neural Networks Learning Improvement using the K-Means Clustering Algo. to Detect Network Intrusions", I.J. of Computational Intelligence Volume 3 Number 2, 2005.

[6] Muna M. Taher Jawhar and Monica Mehrotra, "Anomaly Intrusion Detection System using Hamming Network Approach", Int Journal of Computer Science & Comm, Vol. 1, No. 1, January-June 2010.

[7] Marjan Bahrololum, Elham Salahi, Mahmoud Khaleghi, "An Improved Intrusion Detection Technique based on two Strategies Using Decision Tree and Neural Network"Journal of Convergence Information Technology Volume 4, Number 4, December 2009.

[8] Mr. Vivek A. Patole, Mr. V. K. Pachghare, Dr. Parag Kulkarni, "Self-Organizing Maps to Build Intrusion Detection System", International Journal of Computer Applications, 2010.

[9] Hai Nguyen, Katrin Franke and Slobodan Petrovi´cNISlab, "Improving Effectiveness of Intrusion Detection by Correlation Feature Selection",IEEE Digital Library,2010.

[10] Sampada Chavan, Khusbu Shah, Sanghan Mitra Mukherjee, Ajith Abraham et. al. 2004. "Adaptive Neuro-Fuzzy Intrusion detection Systems",ITCC'04, IEEE.

[11] Ajith Abrahama, Ravi Jain, Johnson Thomas, Sang Yong Hana 2007 " D-SCIDS: Distributed soft computing intrusion detection system" Journal of Network and Computer Applications, pp.81–98.

[12] http://www.cs.waikato.ac.nz/ml/index.html DOA 05 Feb.2016

[13] Muhannad Harrim, "Genetic Algorithm", Wetern Michigan University, USA.

[14] Preeti Aggarwal, Sudhir Kumar Sharma, "Analysis of KDD Dataset Attributes-class wise For Intrusion Detection",Elsevier.2015,

[15] Ciza Thomas," Performance Enhancement of Intrusion Detection Systems using Advances in

Sr.No.	Category	Attack Class	TPR	FPR	AUC
01 .	Back	DOS	0.99	0.01	1
02 .	Pod		1	0.001	1
03 .	Land		1	0.002	1
04 .	Smurf		0.99	0.04	0.98
05 .	Teardrop		1	0.001	0.998
06 .	Neptune		0.98	0.01	1
07 .	Nmap	PROBE	1	0.001	1
08 .	Satan		0.69	0	0.98
09 .	Portsweep		0.99	0	1
010 .	Ipsweep		0.98	0.001	0.999
011 .	Phf	R2L	0.91	0.001	0.97
012 .	Guess_pwd		0.90	0	0.995
013 .	ftp-write		0	0.001	0.92
014 .	Imap		0	0.004	0.91
015 .	Multihop		0	0	0.85
016 .	Warezmaster		0.95	0.01	1
017 .	Buffer_overflow	U2R	0.70	0	1
018 .	Loadmodule		1	.002	1
019 .	Rootkit		0.68	0.001	1
020 .	pearl		1	0.003	1

Sensor Fusion", Supercomputer Education and Research Centre , IISc, Bangalore, 2009.

[16] Samaneh Rastegari, Chiou-Peng Lam, Philip Hingston"A Statistical Rule Learning Approach to Network Intrusion Detection", IEEE 2015.

[17] Vinod Mahajan, Bhupendra Verma," Implementation of Distance BasedSemi Supervised Clustering and Probabilistic Assignment Technique forNetwork Traffic Classification", IJERA, ISSN: 2248-9622, Vol. 2, Issue 2, Mar-Apr 2012.

[18] Marjan Bahrololum, Elham Salahi and Mahmoud Khaleghi, December 2009 "An Improved Intrusion Detection Approach based on two Strategies Using Decision Tree and Neural Network, "Journal of Convergence Information Technology Vol. 4, No. 4.

Purushottam Rohidas Patil obtained his Bachelor's degree in Computer Engineering in 2000 from North Maharashtra University, Jalgaon, M.S, India. Then he obtained his M.E. CSE in 2009.

from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, M.S, India and Currently, he is a Ph.D Research Scholar at Faculty of Engineering and Technology, Jodhpur National University, Jodhpur, RJ, India. His current research interests are Data Mining, Network Security, Human Computer Interaction. He is Members of Professional bodies like LMISTE, CSI, IAENG, CSTA-ACM.

Dr.YogeshSharma received his B.Sc. Degree in 1995, M.Sc.(Mathematics) degree in 1997,and Ph.D. (Mathematics) in 2001 from Jai Narayan Vyas University, Jodhpur, RJ, India. His research interest includes Fractional Calculus, Differential Operators, Matrix Variable,Lie Theory, operation Research, He is currently working as Dean Faculty Of Computer Application and Professor & Head, Applied Science, Faculty of Engineering and Technology, Jodhpur National University, Jodhpur,(RJ), India. He is Members of Professional bodies like Vijnana Parished Anusandhan Patrika, Allahabad, Indian Academy of Mathematics, Indore, Rajasthan Ganita Sandesh, Society of Special function, International Scientific Committee USA, and reviewer 8 Mathematical Society. Dr.Sharma has published 9 books,41 research papers at Reputed International and National Journals and attended and participated 9 national and international Conferences. He is approved research guide at Career Point University, Jodhpur National University, JJTU, Jhunjunu,Rajasthan, India.

Dr. Manali Makarand Kshirsagar, Ph.D.(Computer Science), M.E.(Computer Science & Engineering), B.E. (Computer Technology) and Diploma (Computer Technology), M.B.A.(Fin.& Marketing).

Her Specialization at the PG and Ph.D. level are Data Warehousing, Data Mining, Business Intelligence and Bioinformatics. Other areas of interest are Advanced Operating Systems, Wireless Sensor Networks, Business Analytics

She has Total 23 years of experience in Academics and Industry. Her Job Profile is Education Leader, full Professor at UG and PG level, Ph.D. Supervisor, Head of the Department, Dean (Students Activities).Dr. Manali Kshirsagar currently held the position as Vice President(Academy) ADCC Infocad Ltd. Nagpur, India. She is Members of professional societies and positions held in local chapters, Fellow of Institution of Engineers(India), Life Member of Indian Society for Technical Education(New Delhi), Chair of CSI Nagpur Chapter, Elected Member of IE(I) Nagpur Local Center. She has Above 50 technical and research publications in International Journals, International and National conferences. Guiding 08 Ph.D. students at present.