Implementation of Network Forensics Mechanism for Web Attack Detection

[1]Sudhakar Parate, [2]Smita M. Nirkhi, [3]Dr. R.V.Dharaskar
[1,2]Dept. of CSE, G.H. Raisoni College of Engineering Nagpur, India
[3]Director, M.P.G.I Nanded, India

Abstract
Network forensics is the most significant technology to investigate different types of networking attack. Network forensics will help to capture, copy, transfer, analysis and investigation purpose. The most of the web application can easily attack by the hackers even when antivirus, firewall are exist in the system. This system used to identify different types of web attacks by using Kddcup 99 and NSL KDD dataset as a evidence. These digital evidence help in the course of the investigation phase to prepare the next steps. The evidences are just like a log files, that log files take as input and pre-process before training the neural network. Backpropogation algorithm is used for the training neural network and attack detection with the help of different dataset. Finally system generates forensic report, which will help for aiding an investigation.

Keywords
Investigation, Network Forensics, Attack Detection, Evidence, Log Files, Forensic

I. Introduction

The rapid development of application and local network systems that have changed the computing world in the last decade. The highly connected computing world has also equipped the intruders and hackers with advance facilities to fulfil their purpose. The hackers can exploit the vulnerabilities of website to take advantage or gain access to private information. To reduce the probability of successful attacks it is necessary to introduce adequate security precautions. In the propose application feedforward back propagation neural network architecture is very popular because it can be applied to many different tasks. Its architecture help to examine how it is trained and how it processes the pattern behavior. The first term indicate "feed foreword" describes how this neural network processes the pattern and recalls patterns. Each layer of the neural network connections to the next layer.

Neural network is capable to capture and represent complex input/ output relationships. Back propagation algorithm is used for the training neural network. It is a form of supervised training. While using a supervised training method the network must be provided with sample inputs and anticipated outputs. These anticipated outputs will be compared against the anticipated output of the neural network. The weights are adjusts in the various layers backwards from the output layer all the way back to the input layer. The propose system identify attack by using post event log files of Kddcup99 standard dataset. That log files used for training as well as testing purpose.

II. Related Work

In the year 1990, the network forensics issue was discuss by security expert Marcus Ranum and he defined, capture, recording and analysis of evidences [4-5]. The first digital forensics research workshop was held in 2001 related to uncovering facts related to the planned intent, or measured success of unauthorized activities and recovering activities. That workshop framework activity involves the identification, preservation, collection, examination, analysis, presentation and decision and this framework is on the basis for all the proposed models. McGrath and Nelson (2006) interpret network forensics and enable collection of the evidentiary data using non intrusive network traffic record system. In early days network forensics used for the troubleshooting connection issue and it also help to solve various network security problem. For example fail router, a leaky firewall or insecure database [5].

The four types of evidences are there in computer hacking forensics investigator, which are authentication logs, application logs, operation system logs and network device logs. Hackers may delete the log for making attacking trajectory. If the vulnerability is detected then website will be hacked by hackers and they capture all the valuable information [3, 8]. The model proposed by Freiling and Schwittay in 2007, both are for the incident response and computer forensic processes which allowed a management oriented approach in the digital investigations, while retaining the possibility of a rigorous and forensic investigation [14]. The goal of the recovery and detection is stated as to recognize the digital objects which may contain information about the incident and document them. There are many ways to prevent these attacks and some systems available to detect these attacks. Network Forensics occurs for finding the root cause of the attack and investigate in depth, once an attack occurs. Network forensic helps to analyse traffic data logged through firewalls or intrusion detection system or at network devices like routers and switches. A forensics investigation requires the use of disciplined investigative techniques to discover and analyse traces of evidence left behind after a crime has been take place [3, 6].

As we have seen that the network threats are increasing gradually with the passing of time therefore securing the network resources is a big problem. The intrusion detection systems can be classified into following three categories as host based, network based and vulnerability assessment based[12][13]. A host based intrusion detection system evaluates information which is found on a single or multiple host systems, which including contents of operating systems, and application files. The vulnerability assessment based Intrusion detection system is also used to detects vulnerabilities on internal networks and firewall.

The Mark I was the first machine which is used to "learn" to identify optical patterns. that were setup manually because the multilayered perceptron neuron did not have the ability to learn. It was very limited when compared with the infinitely more flexible. In 1969 Minsky and Papert wrote a book in which they described the limitations of single layer Perceptrons. Al-Rashdan [4] has proposed an intelligent model using Hybrid Artificial Neural Networks, supervised and unsupervised learning capabilities to classify and / or detect network using KDDCup'99 dataset. The system operation are divided into three categories : Input Data Collection and Pre-processing, Training, and Detection stage. [2, 14]. Klopf (A. Henry Klopf) in 1972 developed a basis for learning in artificial neurons based on a biological principle for neuronal learning called heterostasis. Werbos (Paul Werbos 1974) developed and used the back-propagation learning method,

however several years passed before this approach was popularised. Back propagation network are probably the most well known and widely applied of the neural networks [4,12].

III. Proposed System

Artificial Neural Networks (ANN) are the most commonly used approaches for Detection Systems. Neural networks are a uniquely powerful technique in multiple class classification, mainly when used in applications where formal analysis would be very difficult or even impossible. neural networks are able to work with imprecise and incomplete data. It means that they can recognize also patterns not presented during a learning phase. That is why the neural networks could be a good solution for detection of a well- known attack.

KDDcup99 is a standard dataset used as a evidence. Then pre-process the data before it use as a input to the system. The system should be trained with the backpropogation algorithm and it will help to detect web attack using neural network.

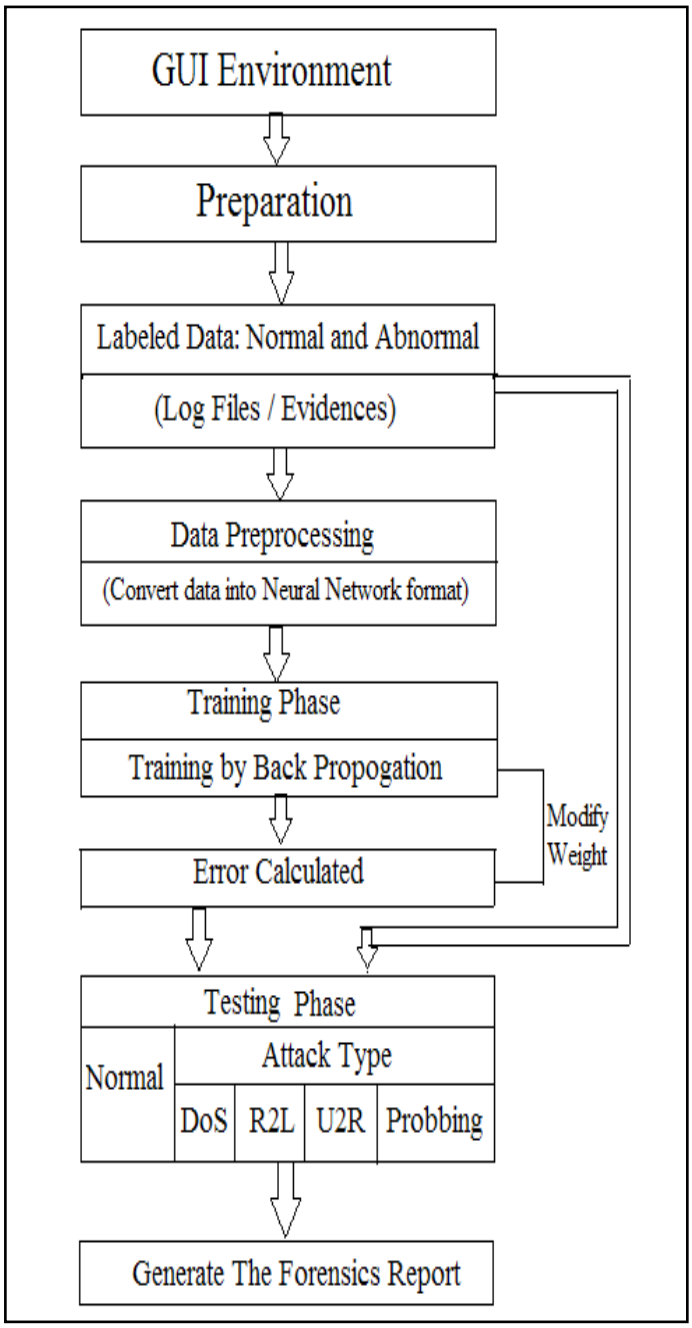

Fig. 1: Propose System for Attack Detection

A. KDD Dataset and Pre-Processing Dataset

The KDDCup99 dataset is used for the experiment. This section describes how the data set is used for our experiment. The 80% from the samples are used for training the neural network while others are used for testing the neural network . The data set is pre-processed so that it may be able to give it as an input to our developed system. This dataset consists of numeric and symbolic features and we converted that text data into numeric form so that it can be given as inputs to our neural network. Then this modified dataset is ready to be used as training and testing of the neural network.

B. A Variety of Attacks are Incorporated in the Dataset Fall Into Four Major Categories

1. Denial of Service Attacks

A denial of service attack is an attack where the attacker constructs some computing or memory resource fully occupied or unavailable to manage legitimate requirements, or reject legitimate users right to use a machine.

2. User to Root Attacks

User to Root exploits are a category of exploits where the attacker initiate by accessing a normal user account on the system (possibly achieved by tracking down the passwords, a dictionary attack, or social engineering) and take advantage of some susceptibility to achieve root access to the system.

3. Remote to User Attacks

A Remote to User attack takes place when an attacker who has the capability to send packets to a machine over a network but does not have an account on that machine, makes use of some vulnerability to achieve local access as a user of that machine.

4. Probes

Probing is a category of attacks where an attacker examines a network to collect information or discover well-known vulnerabilities.

Following table illustrates a number of attacks falling into four major categories:

Table 1: Different Types of Attacks Described in Four Major Categories

Denial of Service Attacks	Back, land, Neptune, pod, smurf, teardrop
User to Root Attacks	Buffer_overflow, loadmodule, perl, rootkit
Remote to Local Attacks	Ftp_write, guess_passwd, imap, multihop, phf, spy, warezclient,warezmaster
Probes	Satan, ipsweep, nmap, portsweep

C. The Structure of a Feed Forward Neural Network

A feed forward neural network differs from the neural networks previously examined. Figure shows a typical feed forward neural network with a single hidden layer. There are numbers of input with different weight for reducing the error. The input vector is interconnected with hidden layer and then it produces the output.

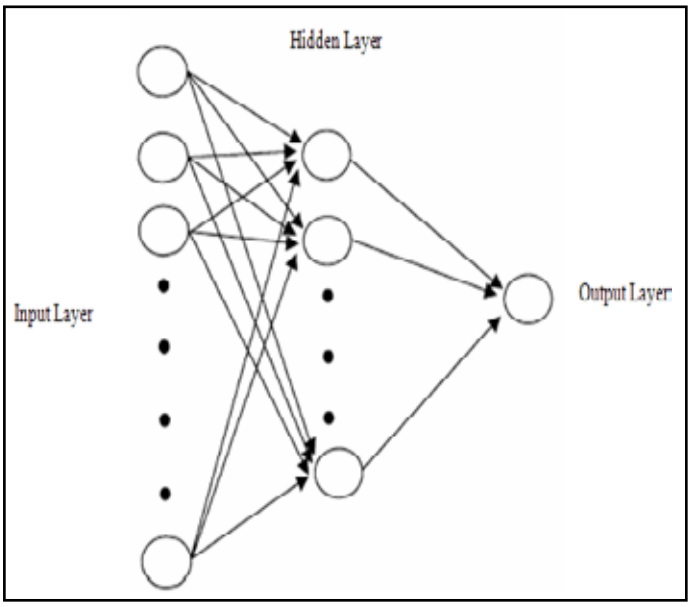

Fig. 2: Feed Forward Neural Network

IV. Experiments and Results

This system is performed to detect all five different classes of from the Kddcup99 dataset including Dos, U2R, Probe, U2L and normal. The distribution of an attack and normal records are 70% and 30% based on the experiment association of any feature with attack class is analyzed. The detection algorithm maps incoming events to attacks and normal activity. The resulting classification can be used to determine the effectiveness of an system. Effectiveness is the ability of an system to maximize the detection rate while minimizing the false alarm rate (false positive rate). In other words, good system reports intrusions when they occur, and does not report intrusions when they do not occur.

A. Pre-processing and Training Neural Network

Convert the information into the format required by the statistical processor. In the pre-processing process all the event are pre-processor before the training action take place because it accept only the numerical value.

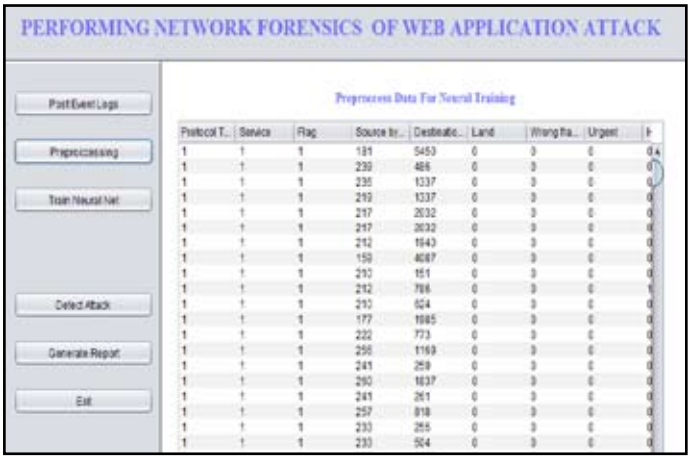

Fig. 3: Preprocessing Input Dataset (KDD 99 and NSL KDD)

Training is the process for which these connection weights are assigned. Most of the training algorithms begin by assigning random numbers to the weight matrix. Next the weights are adjusted based on how valid the neural network performed. Neural network training methods generally classified into the categories of supervised, unsupervised and various hybrid approaches [9].

Supervised training is accomplished by giving the neural network with a set of sample data it may be KDDCUP 99 along with the anticipatd outputs from each of these samples. Supervised training is the most common form of neural network training. As supervised training proceeds the neural network is taken through several iterations, or epochs, until the actual output of the neural network matches the anticipated output, with a reasonably small error[4].

Each iteration is one pass through the training samples. Unsupervised training is similar to supervised training except that are no anticipated outputs are provided. Unsupervised training usually occurs when the neural network is to classify the inputs into several groups. The training progresses through many epochs, it just like as in supervised training. It is very important to understand how to properly train a neural network. There are several methods of neural network training, including back propagation, simulated annealing, and genetic algorithms. Once the neural network is trained, it must be validated.

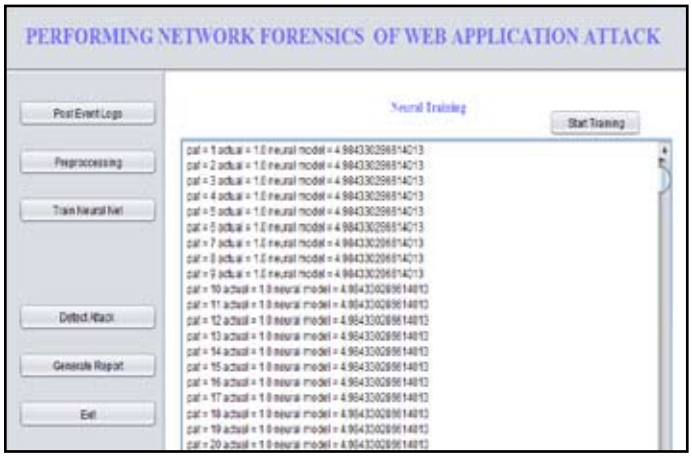

Fig. 5: Training Neural Network

Calculate error is an important part of any neural network. When the neural network is supervised or unsupervised, an error rate must be calculated. The goal of virtually all training algorithms is to minimize the error. It will examine how the error is calculated for a supervised neural network. In Supervised training, there are two components to the error that must be considered. First, we must calculate the error for each of the training sets as they are processed. Secondly we must take the average across each sample for the training set [14-15].

B. Root Mean Square (RMS) Error

The Root Mean Square error allows the neural network to know when enough training has taken place. The RMS error can be calculated at any time after the "calcErrors" method has been called. This process is necessary because the RMS is an average to take the average error across all training set elements, for that you must know the size of the training set. The RMS error is then calculated by dividing the global error by which the product of the training set length and the number of output neurons. The square root of this ratio produces the RMS error. At last after the RMS error has been calculated the global Error is set back to zero. This process is required so that it can help to begin accumulating for a new error13].

C. Testing Phase

Once the neural network has been trained it must be evaluated to see that is it ready for actual use. This final step is important

so that it can be determined if additional training is required. In order to correctly validate a neural network validation data must be set aside that is completely separate from the training data [8]. When 20,000 sample elements provided to neural network for this sample data the group of each element should be classified into is known. Once the network has been properly trained then the second group of 5,000 elements would be used to test the neural network. It is very important that a separate group always be maintained for testing. By using this same set we can predict the anticipated error of the neural network [10, 12].

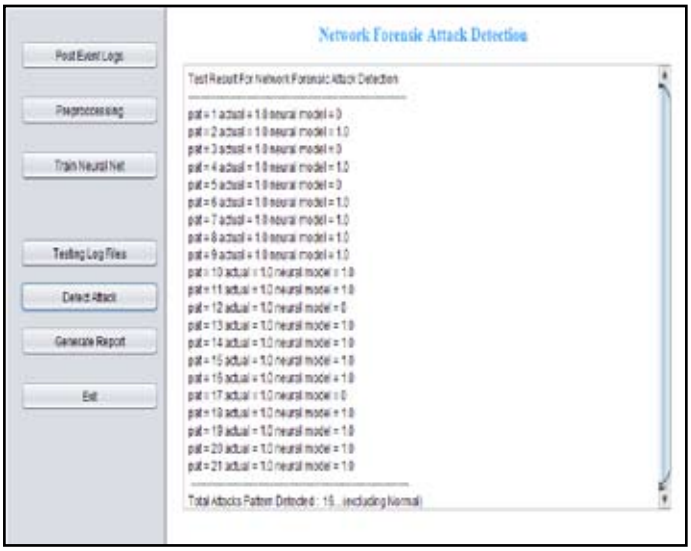

Fig. 6: Web Attack Detection

D. Forensics Report Analysis

Forensics report gives the details of information about the type of attack and attacking scenario. It provide information about forensic evidences for identifying attacking source were found from Web application logs, Web security logs, Web system logs. It shows the information of different forensics tool like Wireshark, MD5, TCPDump and it shows total number of attack of each category.

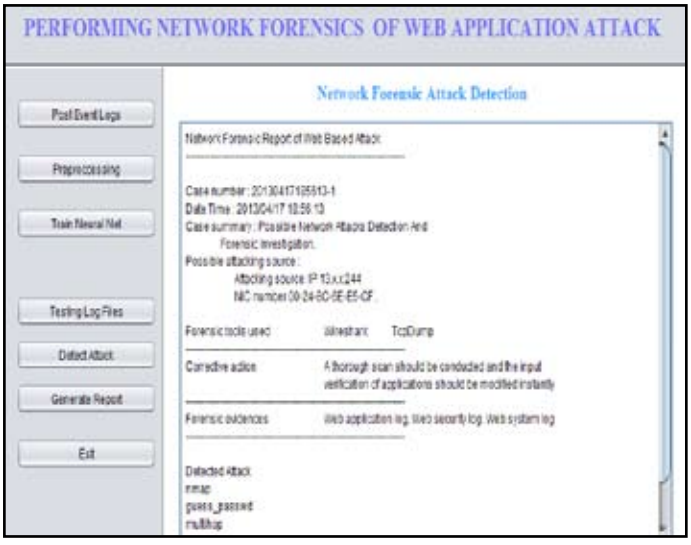

Fig. 7: Forensics Report on Web Based Attack

V. Conclusion

In this paper, the system is implemented by using back propagation algorithm for training the system and detection of various web based attack. This system uses network forensics process, as it provides the investigative capabilities and analysis of evidences. This system uses the standard KDD99 and NSL KDD dataset for measuring the performance. This implementation improves the detection rate with high accuracy. Finally system generate the network forensic report, which having information about different attack that will help for adding an investigation.

References

[1] Nuno Antunes, Marco Vieira,"Defending Against Web Application Vulnerabilities", IEEE transaction on computer, pp 66-72, 2012.

[2] G. Mohay,"Technical challenges and directions for digital forensics", In SADFE '05: Proceedings of the First International Workshop on Systematic Approaches to Digital Forensic Engineering. Washington, DC, USA: IEEE Computer Society, 2005, pp. 155.

[3] Meixing Le, Angelos Stavrou, Brent Byunghoon Kang,"Doubleguard: Detecting Intrusions In Multitier Web Applications", IEEE transactions on dependable AND secure computing, Vol. 9, No. 4, pp. 512-525, July/August 2012.

[4] Jatinder Kaur, Gurpal Singh, Manpreet Singh,"Design & Implementation of Linux Based Network Forensic System Using Honeynet", International Journal Of Advanced Research In Computer Engineering & Technology Vol. 1, Issue 4, pp. 231-238, June 2012.

[5] Reyadh Shaker Naoum, Namh Abdula Abid, Zainab Namh Al-Sultani", An Enhanced Resilient Backpropagation Artificial Neural Network for Intrusion Detection System", IJCSNS International Journal of Computer Science and Network Security, Vol. 12, No. 3, March 2012

[6] Igino Corona, Davide Ariu, Giorgio Giacinto,"HMM-Web: A Framework For The Detection of Attacks Against Web Applications", IEEE International conference on communication, pp. 1-6, 2009.

[7] Stephen D. Wolthusen,"Overcast: Forensic Discovery In Cloud Environments", Fifth International Conference on IT Security Incident Management And IT Forensics, pp. 3-9, 2009.

[8] Ryuya Uda,"Proposal of Method for Digital Forensics in Physical Distribution", Second International Conference on Computer Engineering and Applications, pp. 211-216, 2010.

[9] Amelia Phillips,"Computer Forensics Investigators or Private Investigators: Who is Investigating the Drive?", Fifth International Workshop on Systematic Approaches to Digital Forensic Engineering, pp. 550-557, 2010.

[10] Sohaib Ikram, Hafiz Malik,"Digital Audio Forensics Using Background Noise", International conference on multimedia and expo, pp. 106-110, 2010.

[11] CP Grobler, CP Louwrens,"Digital Evidence Management Plan", ISSA, pp. 1-6, 2010.

[12] Chung-Huang Yang, Pei-Hua Yen,"Fast Deployment Of Computer Forensics With Usbs", International Conference On Broadband, Wireless Computing, Communication and Applications, pp. 413-416, 2010.

[13] D. Reilly, C Wren, T. Berry,"Cloud Computing: Forensic Challenges for Law Enforcement", International conference on internet technology and secured transaction, pp. 1-7, 2010.

[14] Ying Xuan, Incheol Shin, My T. Thai,"Detecting Application Denial-Of-Service Attacks: A Group-Testing-Based Approach", IEEE transactions on parallel and distributed systems, Vol. 21, No. 8, pp. 2103-1216, august 2010.

[15] Sindhu. K. K , Dr. B. B. Meshram, "A Digital Forensic Tool For Cyber Crime Data Mining", IRACST – Engineering Science And Technology: An International Journal (ESTIJ), ISSN: 2250-3498, Vol.2, No.1, 2012.
[16] Varish Mulwad, Wenjia Li, Anupam Joshi, Tim Finin Krishnamurthy Viswanathan,"Extracting Information About Security Vulnerabilities From Web Text", IEEE 2011.
[17] Abraham Yaar, Adrian Perrig, Dawn Song,"Pi: A Path Identification Mechanism to Defend Against Ddos Attacks", IEEE 2003.
[18] Liang Xie, Sencun Zhu,"Message Dropping Attacks In Overlay Networks:Attack Detection And Attacker Identification", IEEE 2006.
[19] Veena H Bhat, Prasanth G Rao, Abhilash R V, P Deepa Shenoy, Venugopal K R, L M Patnaik,"A Data Mining Approach For Data Generation And Analysis For Digital Forensic Application", IACSIT International Journal Of Engineering and Technology, Vol. 2, No. 3, June 2010.
[20] Mohd Taufik Abdullah, Ramlan Mahmod, Abdul A. A. Ghani, Mohd A Zain, Abu Bakar Md S,"Advances In Computer Forensics", International Journal of Computer Science And Network Security, Vol. 8, No. 2, February 2008.
[21] Smita.Nirkhi,"Potential Use of Artificial Neural Network In Data Mining,"International Conference On Computer and Automation Engineering (ICCAE), pp. 339-343, 2010.

Mr. Sudhakar Parate recived the B.E. in Computer Engineering from Bapurao Deshmukh College of Engineering, Seagram, Wardha in 2007 and Pursuing M.Tech in computer science and Engineering from G.H.Raisoni college of Engineering, Nagpur. His main research interests include Digital Forensics, Artificial Neural Network.

Asst. Prof. Ms. S. M. Nirkhi has completed M.Tech in Computer Science &Engineering & currently Pursuing Ph.D in computer science. She has attended 6 STTP workshops along with other training programs. She has Published 15 papers in international conferences & 5 papers in international journals. She had presented paper at International Conference at Singapore. She has 12 years of professional experience. Currently working as Assistant professor in Department of Computer Science & Engineering at GHRCE. Her area of interest include Soft computing, Data mining, web mining, pattern recognition, MANET, Digital Forensics.

Dr. R.V. Dharskar has received Ph.D. Degree (Computer Science &Engineering) from Amravati University, M. Tech. (Computers) from I.S.M. and P.G. Dip. M.Phil., M.Sc. from Nagpur University. He is having more than 29 years of teaching and 23 years of R&D experience in the field of Computers & IT. He is an author of number books on Programming Languages. He has been actively involved in the research on Mobile Computing, Multimedia, Software Engineering, Web Technology, E-Learning and Networking, Digital Forensics. He has authored more than 227 research papers at various International / National Conferences and Journals. His research work has been accepted at IEEE Computer Society of USA, Bristol (UK), Hong Kong (China) etc. He has been invited as a Keynote Speaker, Invited Speaker, and Session Chair for more than 34 International & National Conferences