NETWORK CLASSIFICATION AND MAC FEATURE SELECTION FOR HIGH SPEED NETWORKS

M.Venkat Reddy[1], A.Yashwanth Reddy[2]
[1]Assistant Professor, Sree Dattha Institute of Engineering and Science, Sheriguda, Hyderabad, India.
[2]Assistant Professor, Sree Dattha Institute of Engineering and Science, Sheriguda, Hyderabad, India.

Abstract
Intrusion detection system (IDS) shall be deemed to be an application which screens the system towards spiteful activities and strategy defilements. Some noticed action and violations are characteristically described moreover towards administrators and composed in centralized manner by means of safety information or event management (SIEM) scheme. This scheme pools output at numerous bases thereby employs alarm sieving methods for distinguishing hateful action from untrue alarms. Thus, the proposed IDS system developed the method called PSO-ANN (Particle Swarm Optimization-Artificial Neural Network). It exploits the techniques called feature reduction, optimal feature selection, and ANN classification to classify the intrusions of 802.11 based wireless networks. It significantly improves the classification accuracy and reduces the learning time and testing time of IDS. The feature reduction employs data gains for computing significance of every feature and chooses most important MAC features. Based on the feature reduction, the optimal sets of features get categorized into labeled and unlabeled features.

Keywords: **Feature reduction, Optimal feature selection, PSO; ANN, Classification accuracy, Detection accuracy, Learning time.**

1. INTRODUCTION

The most common classifications are network intrusion detection systems (NIDS) and host-based intrusion detection systems (HIDS). A system that monitors important operating system files is an example of a HIDS, while a system that analyzes incoming network traffic is an example of a NIDS. It is also possible to classify IDS by detection approach: the most well-known variants are signature-based detection (recognizing bad patterns, such as malware) and anomaly-based detection (detecting deviations from a model of "good" traffic, which often relies on machine learning). Some IDS have the ability to respond to detected intrusions. Systems with response capabilities are typically referred to as an intrusion prevention system. Intrusion detection (ID) is a type of security management system for wireless networks. An ID system gathers and analyzes information from various areas within a network to identify possible security breaches, which include both intrusions (attacks from outside the organization) and misuse (attacks from within the organization). Since the ability of IDS is to identify the large variety of intrusions in real time with accuracy and time values are of primary concerns. In this paper, we consider the performance measures of learning machine based IDSs in the critical aspects of classification accuracy, detection accuracy, training time, testing times, and scalability. One of the main problems with IDS is the overhead, which can become prohibitively high. To analyze the system logs, the operating system must keep the information regarding all the actions performed, which invariably results in huge amounts of data, requiring disk space and CPU resources. Next, the logs must be processed and converted into a manageable format and then compared with the set of recognized misuse and attack patterns to identify possible intrusions. Further, the stored

patterns need be continually updated, which would normally involve human expertise. Detecting intrusions in real time is a difficult task. Several artificial intelligence techniques have been utilized to automate the intrusion detection process to reduce human intervention. They existing techniques include neural networks, fuzzy inference systems, evolutionary computation machine learning, etc.

Though they both relate to network security, an IDS differs from a firewall in that a firewall looks outwardly for intrusions in order to stop them from happening. Firewalls limit access between networks to prevent intrusion and do not signal an attack from inside the network. An IDS evaluates a suspected intrusion once it has taken place and signals an alarm. An IDS also watches for attacks that originate from within a system. This is traditionally achieved by examining network communications, identifying heuristics and patterns (often known as signatures) of common computer attacks, and taking action to alert operators. A system that terminates connections is called an intrusion prevention system, and is another form of an application layer firewall [1]. The ANN is based on biological brain neural network functions. ANN is the classifier, it classifies the intruder attack types into various attacks. The ANN consists of three layers. They are input layer, output layer, and hidden layer. One neural node linked with multiple neural nodes. Each node of neural network has multiple links with other nodes for specific computational task. Artificial neural network classifies the intruder types without human intervention. The existing approach uses the approach called Support Vector Machine (SVM). SVM is a discriminative classifier defined by a separating hyperplane. For the given labeled training data, the existing algorithm outputs an optimal heperplane which categorizes the network node features. Network intrusion detection systems (NIDS) are placed at a strategic point or points within the network to monitor traffic to and from all devices on the network. It performs an analysis of passing traffic on the entire subnet, and matches the traffic that is passed on the subnets to the library of known attacks. Once an attack is identified, or abnormal behavior is sensed, the alert can be sent to the administrator [2-4].

An example of an NIDS would be installing it on the subnet where firewalls are located in order to see if someone is trying to break into the firewall. Ideally one would scan all inbound and outbound traffic, however doing so might create a bottleneck that would impair the overall speed of the network. OPNET and NetSim are commonly used tools for simulating network intrusion detection systems. NID Systems are also capable of comparing signatures for similar packets to link and drop harmful detected packets which have a signature matching the records in the NIDS. When we classify the design of the NIDS according to the system interactivity property, there are two types: on-line and off-line NIDS, often referred to as inline and tap mode, respectively. On-line NIDS deals with the network in real time. It analyses the Ethernet packets and applies some rules, to decide if it is an attack or not. Off-line NIDS deals with stored data and passes it through some processes to decide if it is an attack or not. Host intrusion detection systems (HIDS) [5] run on individual hosts or devices on the network. A HIDS monitors the inbound and outbound packets from the device only and will alert the user or administrator if suspicious activity is detected. It takes a snapshot of existing system files and matches it to the previous snapshot. If the critical system files were modified or deleted, an alert is sent to the administrator to investigate. An example of HIDS usage can be seen on mission critical machines, which are not expected to change their configurations.

Some systems may attempt to stop an intrusion attempt but this is neither required nor expected of a monitoring system. Intrusion detection and prevention systems (IDPS) are primarily focused on identifying possible incidents, logging information about them, and reporting attempts. In addition, organizations use IDPS for other purposes, such as identifying problems with security policies, documenting existing threats and deterring individuals from violating security policies. IDPS have become a necessary addition to the security infrastructure of nearly every organization. There are a number of techniques which attackers are using, the following are considered 'simple' measures which can be taken to evade IDS:

Fragmentation: by sending fragmented packets, the attacker will be under the radar and can easily bypass the detection system's ability to detect the attack signature.

Avoiding defaults: The TCP port utilised by a protocol does not always provide an indication

to the protocol which is being transported. For example, an IDS may expect to detect a trojan on port 12345. If an attacker had reconfigured it to use a different port, the IDS may not be able to detect the presence of the trojan. Coordinated, low-bandwidth attacks: coordinating a scan among numerous attackers (or agents) and allocating different ports or hosts to different attackers makes it difficult for the IDS to correlate the captured packets and deduce that a network scan is in progress. Address spoofing/proxying: attackers can increase the difficulty of the ability of Security Administrators to determine the source of the attack by using poorly secured or incorrectly configured proxy servers to bounce an attack. If the source is spoofed and bounced by a server, it makes it very difficult for IDS to detect the origin of the attack.

Pattern change evasion: IDS generally rely on 'pattern matching' to detect an attack. By changing the data used in the attack slightly, it may be possible to evade detection. For example, an Internet Message Access Protocol (IMAP) server may be vulnerable to a buffer overflow, and an IDS is able to detect the attack signature of 10 common attack tools. By modifying the payload sent by the tool, so that it does not resemble the data that the IDS expects, it may be possible to evade detection.

IDPS typically record information related to observed events, notify security administrators of important observed events and produce reports. Many IDPS can also respond to a detected threat by attempting to prevent it from succeeding. They use several response techniques, which involve the IDPS stopping the attack itself, changing the security environment (e.g. reconfiguring a firewall) or changing the attack's content.

2. LIMITATIONS OF EXISTING SYSTEM

Intrusion detection is a problem of great importance to protecting information systems security, especially in view of the worldwide increasing incidents of cyber attacks. Traditional security policies or firewalls have difficulty in preventing such attacks because of the hidden vulnerabilities contained in software applications. Therefore, intrusion detection system is required as an additional wall for protecting systems despite the prevention techniques. Support vector machine is the method that is receiving increasing attention with remarkable results for the design of IDS recently. Unfortunately, the determination of parameters values becomes an optimization problem in the practicability of SVM. IDS is always to deal with huge amount of data causing slow training and testing process and also low detection rate. So feature reduction, optimal feature selection and classification are one of the key topics in IDS to be addressed effectively.

The IEEE 802.11i (WPA2) [6-7] standard is ability to seamlessly integrate their WLANs with their wired LANs, using the Ethernet protocol. This standard provided an optional security feature WEP which was intended to provide the same level of security as a wired LAN. The evolution of security standardization based upon the work of the IEEE has evolved from WEP to WPA which introduced new key management and integrity mechanisms through to WAP2 (IEEE 802.11i) which maintains the management and integrity mechanisms of WPA but introduces AES encryption as well as moving much of the security functionality to the hardware. WEP's simple authentication procedures were easy to break. The encryption key used was shared amongst all clients thus increasing security risks. The case of using weak keys when the initialization vector was added, posed a further vulnerability. The merits of the system are that the encryption key used was shared amongst all clients thus increasing security risks. The demerits are the WEP is ineffective in determining de-authentication and plaintext attacks of WiFi due to the nature of the wireless medium.

Current intrusion detection systems (IDS) examine all data features to detect intrusion or misuse patterns. Some of the features may be redundant or contribute little (if anything) to the detection process. S Chebrolu, A Abraham, P Johnson, Thomas proposes the approach is to identify important input features in building an IDS that is computationally efficient and effective. We investigated the performance of two feature selection algorithms involving Bayesian networks (BN) and Classification and Regression Trees (CART) and an ensemble of BN and CART. Empirical results indicate that significant input feature selection is important to design an IDS that is lightweight, efficient and effective for real world detection systems. Finally, it propose an hybrid architecture for combining different feature selection algorithms

for real world intrusion detection. The proposed feature selection is lightweight, efficient and effective. And also a hybrid architecture involving ensemble and base classifiers for intrusion detection. But it results in high learning times.

A new approach using data mining technique such as SVM and Particle swarm optimization for attaining higher detection rate. PSO is an Optimization method and has a strong global search capability. The SVM-PSO Method is applied to KDD Cup 99 dataset. Free parameters are obtained by standard PSO for support vector machine and the binary PSO is used to obtain the best possible feature subset at building intrusion detection system. The propose technique has major steps: Preprocessing, Feature Reduction using Information Gain, Training using SVM-PSO. Then based on the subsequent training subsets a vector for SVM classification is formed and in the end, classification using PSO is performed to detect Intrusion has happened or not. The proposed approach provides the good detection rate in case of Denial of Service (DoS) attack. But this approach does not apply the intrusion detection for MAC layer features.

3. FEATURES OF PROPOSED SYSTEM

Artificial Neural Networks (ANNs) are system composed of neurons organized in input, output, and hidden layers. The neurons are connected to each other by a set of synaptic weights. An ANN is a powerful tool that has been applied in a broad range of problems such as pattern recognition, forecasting, and regression. During the learning process, the ANN continuously changes their synaptic values until the acquired knowledge is sufficient (until a specific number of iterations is reached or until a goal error value is achieved). When the learning process or the training stage has finished, it is mandatory to evaluate the generalization capabilities of the ANN using samples of the problem, different to those used during the training stage. Finally, it is expected that the ANN can classify with an acceptable accuracy the patterns from a particular problem during the training and testing stage. Several classic algorithms to train an ANN have been proposed and developed in the last years. However, many of them can stay trapped in nondesirable solutions; that is, they will be far from the optimum or the best solution. Moreover, most of these algorithms cannot explore multimodal and noncontinuous surfaces. Therefore, other kinds of techniques, such as bioinspired algorithms (BIAs), are necessary for training an ANN. BIAs have a good acceptance by the Artificial Intelligence community because they are powerful optimization tools and can solve very complex optimization problems. For a given problem, BIAs can explore big multimodal and noncontinuous search spaces and can find the best solution, near the optimum value. BIAs are based on nature's behavior described as swarm intelligence. This concept is defined in as a property of systems composed of unintelligent agents with limited individual capabilities but with an intelligent collective behavior.

The Fig. 1 depicts the block diagram for proposed PSO-ANN IDS system. The Intrusion detection system is process of detecting the intrusion attack types on packets. The IDS system detects the intruder using "signature pattern" and "anomalous misbehavior". The IDS process the MAC Layer 155 network and node features. IDS detect the intrusion packets using entropy and information gain values. The feature selection algorithm selects the best attack features. ANN Classifier classifies the intruder attacks types classification. The proposed IDS system classifies the intruder attacks into three major classifications. They are Injection, flooding, impersonation. Intrusion attack features are classified two common types "Labeled" and "Unlabeled" attacks.

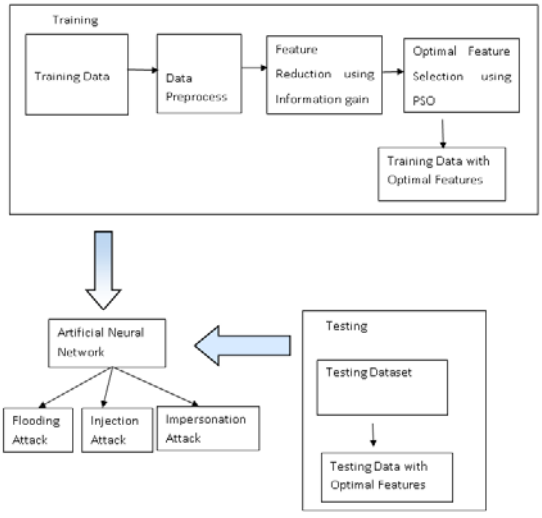

Fig 1. Block Diagram for PSO-ANN IDS system

The proposed IDS system collects the MAC 802.11 dataset with 155 features that contains three general wireless intrusions such as injection, impersonation, and flooding. It divides the dataset into training and testing set. In the training dataset, each feature comprises of different values in packets generated by various attacks. A set of values of a feature is split into distinct values for different classes. The training data set is then separated into many wireless attacks such as Authentication Request, Beacon, CTS, RTS, etc.

The role of feature reduction method is to select important features by discarding irrelevant features in the training data set. The proposed feature selection algorithm calculates information gain and entropy measures on all MAC layer features, making it possible to select important features.

In Feature Reduction algorithm, most of the attackers are labeled than the unlabeled. The proposed IDS adopts PSO (Particle Swarm Optimization) to compute the fitness value of unlabeled attacks in terms of information gain and clusters the attacks of the unlabeled feature under labeled features. This is called as semi-supervised Clustering. In PSO, Particles representing an attacker that moves through an n-optimal features. Each particle maintains a information gain for each features in a vector called pbest. The nbest, is another "best" value that is tracked by the particle swarm optimizer. This is the best value obtained so far by any particle in that particle's neighborhood. When a particle takes the entire feature as its topological neighbors, the best value is a global best and is called gbest. The gbest of features is the labeled features for the unlabeled attack. All attacks under labeled features, the groupings of homogeneous attacks based on optimal features using union set theory.

An Artificial Neural Network has emerged as an important tool for classification. The neural network with appropriate network structure can handle the correlation/dependence between input variables. It works in two phases i.e. Training and Testing. The training phase incorporates that trains the attacker and normal packets. The main purpose of training is to create the dictionary of weights of attacker packets and normal comments. The next phase is to test the packet. The packet will be tested on the basis of the trained weighted dictionary. ANN performs propagation i.e. (BPN), to train the system, by activation of neurons on hidden layer. This step begins training process of BPN by using training data set. The back-propagation algorithm includes a forward pass and a backward pass. The purpose of the forward pass is to obtain the activation value. The backward pass is to adjust weights and biases according to the difference between the desired and actual network outputs. These two passes will go through iteratively until the network converges.

4. CONCULSION

Artificial Neural Network (ANN) design is a complex task because its performance depends on the architecture, the selected transfer function, and the learning algorithm used to train the set of synaptic weights. In this paper we present a methodology that automatically designs an ANN using particle swarm optimization algorithms such as Basic Particle Swarm Optimization (PSO), Second Generation of Particle Swarm Optimization (SGPSO), and a New Model of PSO called NMPSO. An ANN classifier that creates the propagation only the optimal features for classifying the specific 802.11 attacker and normal packets. This paper extends by adding PSO-ANN classifier for only optimal feature set, classify and cluster the attacks based on labeled features. Thus, the effective training and testing set is obtained in proposed approach.

REFERENCES

[1] M. Usha, P. Kavitha "Anomaly based intrusion detection for 802.11 networks with optimal features using SVM classifier" Wireless Networks-Springer Science-Business Media, *DOI 10.1007/s11276-016-1300-5*, May 2016.

[2] Iftihar Ahmad "Feature Selection Using Particle Swarm Optimization in Intrusion Detection" October 2015.

[3] Mathews, Moffat, and Ray Hunt, "Evolution of wireless LAN security architecture to IEEE 802.11 i (WPA2)", *In Proceedings of the fourth IASTED Asian Conference on Communication Systems and Networks*, 2007.

[4] T.M. Khoshgoftaar, S.V. Nath, S. Zhong, and N[1]. Seliya, "Intrusion Detection in Wireless Networks Using Clustering Techniques with Expert Analysis", *Fourth*

International Conference Machine Learning and Applications, 2005.

[5] S. Zhong, T.M. Khoshgoftaar, and S.V. Nath, "A Clustering Approach to Wireless Network Intrusion Detection", 17th IEEE International Conference Tools with Artificial Intelligence (ICTAI), 2005.

[6] AA Bakar, ZA Othman, AR Hamdan, R Yusof, R Ismail, "An Agent Based Rough Classifier for Data Mining", *Eighth International Conference on Intelligent Systems Design and Applications IEEE Computer Societ*, Vol 1, pp. 145-151, 2008.

[7] S Chebrolu, A Abraham, P Johnson, Thomas "Feature deduction and ensemble design of intrusion detection systems", *Computers & Security,* Vol. 24, No. 4, pp. 295-307, 2005.