Retrospect of Machine Learning Techniques for Designing Secure System

Nikita Singh
Amity University
Sector-125, Noida
India

Nidhi Chandra
Amity University
Sector-125, Noida
India

ABSTRACT
Security is the latest research which is of essential concern for many researchers. As the data is generated every second both in host and network so it's necessary to ensure the preservation of their integrity. So, this paper is retrospection of various security measures which are necessary to protect the system from harmful threats/attacks. Machine learning is one of the best technique by which we can apply to get the best results and in turn build an efficient Intrusion Detection System.

General Terms
Intrusion Detection System

Keywords
Host/Network attacks, Machine Learning

1. INTRODUCTION
Research has taken a big leap since 1969 in security and is evolving tremendously in every aspect. Here by every aspect it means security being enhanced by applying various theories from networks, intrusion detection systems, artificial intelligence, mining etc. To strengthen the security more for both hosts as well as network one may apply the machine learning techniques and this thesis work on the same domain to intensify the host/network security. Researchers have given huge emphasis on security with artificial intelligence and have successfully applied the algorithms to increase the intensity of security. The approach discussed here is inspired by human learning style i.e. both supervised as well as unsupervised. By using supervised learning algorithms system/network security performance can be enhanced.

The paper will focus on detailed study of detection and prevention system and to escalate the efficiency of the system it will incorporate the benefits of machine learning algorithms. Since the modern generation is nearly depending on the electronic gadgets for all the purposes, with the increase usage of internet has drawn attention of many intruders who think that these gadgets are the easiest source of gaining personal information. So they emerged with various kinds of attack with simultaneous increase in the number of intrusion detection and prevention techniques. In today's scenario researchers and scholars have plank down efforts by introducing tremendous techniques and tools to monitor the system as well as the network.

The paper is divided into following section: Section II consists of introduction to Computer attacks and its advancement, Section III comprises of the work done so far using machine learning for computer/network security, and Section IV covers introduction to machine learning and various learning algorithms for securing host along with the network.

2. BACKGROUND
Considering the importance of history, I would start with the role played by security in IT industry. The IT security industry has been evolving since past 15 years and is still working to provide the current wave of security innovation. Nowadays security plays a vital role in an industry success and companies are concentrating on innovative ideas to enhance their security measures that ensure full proof security. With current advancement in such technologies have made the developers to give assurance of their product and this has become important as many approaches are followed out there in daily life. To know more about IT security first all security and its type should be covered followed by the attacks for which the same is being used. The history of Security can be analysed under various categories which can be seen in the following Table I:

Table 1. Affected areas by Attacks

Year (from-to)	Affected Areas	Brief Discription
1970-1989	Telephone Networks	Earliest instance of computer security arose out of a need to physically protect hardware. The first example of this occurred after the first computerized switchboard was brought online in 1976. And "by 1982, half of all calls were switched electronically"[1]
	The Modem	As technology matured, the modem became a common tool for remote computing. The FBI began getting involved in remote computer security breaches as it "became increasingly commonplace throughout the 1980s, prompting the passage of Computer Fraud and Abuse Act"[2]
	Viruses	While there may be debate about the first malicious computer programs on the Univac 1108 and IBM S/360 in the 60s or 70s, the most notable concerns persisted after the

		personal computer began making its way into households. The 80s gave rise to the secondary computer security problem of the computer virus. Once IBM and Symantec started researching methods of discovering and deleting viruses from infected computers as well as ways of preventing infection, virus writers began increasing the sophistication of their methods to avoid detection.[2]
1990-2009	The Internet	While most of the security issues in the 80s involved directly communicating with at a computer or network via modem, the Internet created a vastly under protected interconnectivity of networks that made security even more complicated.
	The Firewall	While first created by Cisco in the 80s, firewalls did not become a principle network securing technology for the Internet until there was sufficient need to secure sensitive systems. Firewalls are useful in that they become the central gatekeeper for all incoming and outgoing connections. Centralizing this type of security provides the benefit insuring that certain restricted communications are allowed or disallowed, depending on a predetermined rule.
	Escalated Attacks	As the number of companies and government organizations throughout the world increased their dependency on computers and the Internet, so did malicious attacks.[2] Companies begin steadily increasing their security policies to keep pace with growth. Eventually, the government begins to mandate computer security compliance for major companies.

By observing the historical perspective of Security, we can see that as now technology is emerging at its own rate (which is quite higher) lead our attention towards Host and Network Security whose importance is playing a drastic role in every beings daily life. Day by day all of us are becoming computerized that is we are being dependent on the machines. So we need to secure it at any cost because on Host and Network we are some way revealing our personal information which can be beneficial for an attacker or hacker. There are various security techniques being used these days and these include detection systems, prevention systems, firewalls, anti-viruses etc.

3. RELATED WORK

Evaluating machine learning Algorithms for Detection Network Intrusions

In this paper the author has proposed a machine learning algorithm which has the capability to build n efficient intrusion detection model. Panda et al. have produced experimental results on KDDCup'1999 data sets and demonstrated the effectiveness of choosing a machine learning algorithm for building an efficient intrusion detection model. The machine learning algorithms defined in this paper are Ensemble learning, AdaBoost, Random Forest and Naïve Bayes algorithm. The data set used contains traffic in a simulated military network that consists of hundred of hosts. In learning extremely imbalanced data, the overall classification accuracy is often not an appropriate measure of performance [4]. By using naïve bayes algorithm the paper summarizes the performance measurement parameters of four attack categories namely Probe, DoS, U2R and R2L.

Machine learning based Intrusion Detection Algorithms

Machine learning based intrusion detection approaches have attracted increasing attention in the network intrusion detection research area because of their intrinsic capabilities in discovering novel attacks [5]. Hua Tang et al. have proposed a new approach to detect attacks, which aims to study the method based on machine learning in intrusion detection, including artificial neural network and support vector machine. In this paper, the authors have exhibit the capability of learning technique to detect abnormal behavior via normal behavior. They have proposed their own architecture which shows the process flow of the research.

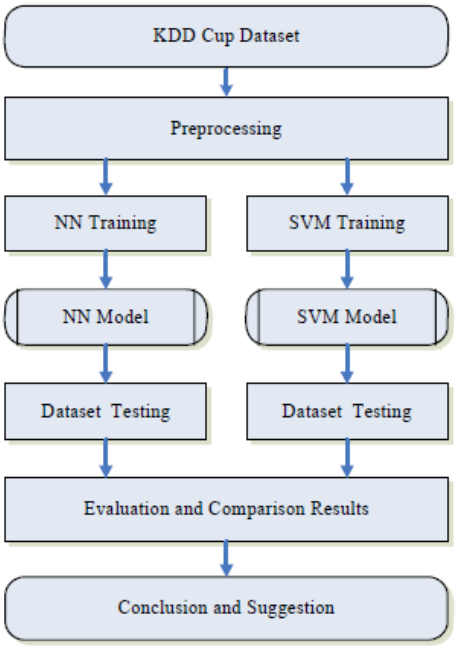

Fig. 1 Architecture of Proposed System

In the figure the first step is to preprocess the data set and in order to compare on different algorithms insert the preprocessed data into NN i.e. Neural Network training and another set into SVM i.e. Support Vector Machines training. Both the processes apply their method in the next step following with the testing of the dataset.

Network Intrusion Detection system based on Machine Learning Algorithms

Network and system security is of paramount importance in the present data communication environment. The main function of Intrusion Detection is to protect the resources from threats [6]. The authors say that machine learning includes a number of advanced statistical methods for handling regression and classification tasks with multiple dependent and independent variables. In this paper the authors discuss about Support Vector machine (SVM), Naïve Bayes and k-nearest Neighbor Algorithm. The following figure 2 gives the proposed architecture of the paper.

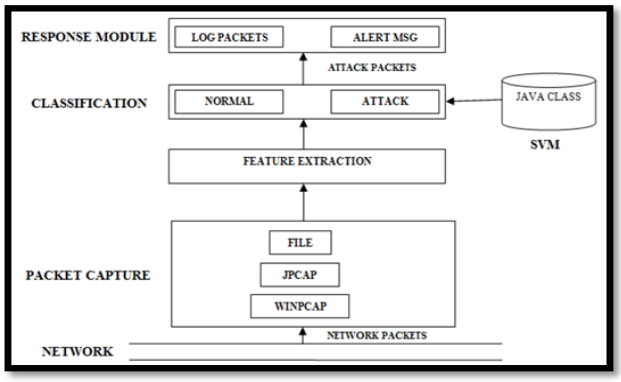

Fig. 2 Proposed Architecture

In the proposed architecture the authors first capture network packets through WINPCAP or JPCAP. After the packet capturing the features are extracted using Rough set theory following with the classification into normal and abnormal content. Finally this classification is analyzed and an alert can be generated to the administrator. The packet capturing is shown in figure 3 given below.

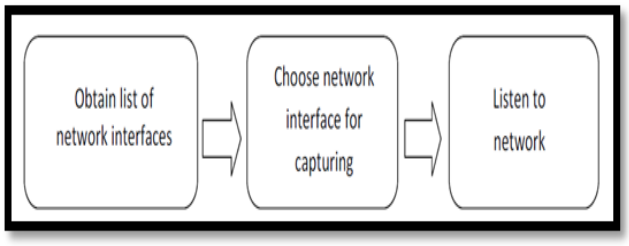

Fig. 3 Packet Capturing

The packet capturing is the most important step and should be analyzed very properly because these packets decide the framework for which the analysis is done using some respective technique.

Classification of Malicious Network Activity

In this paper the author has revealed the benefits of field of adversarial learning that has risen out of need to design learning algorithm that have robustness feature. The project has made use of online-to-batch (O2B) method based on the Perceptron algorithm to train a linear classifier to identify malicious network activity under adversarial conditions, including feature deletion and feature corruption of the data [7]. The authors allow adversary to perform two types of modification: 1) deletion of features entirely by adversary or 2) replace the feature of data by white-Gaussian noise with the same mean and variance as of the original feature. The base of the paper is [8] in which O2B Perceptron algorithm converts an online Perceptron – modified for adversarial robustness to a batch learning using an averaging procedure.
The O2B perceptron are tested on a subset of the KDD Cup 1999 with 250 normal examples and 250 attack samples that represent IP sweeps. In the paper the performance of O2B perceptron model is compared against two binary, linear classification algorithms namely logistic regression and support vector machine (SVM). After the experiment it was observed O2B performed better than the other two algorithms because for the LR and SVM the testing data was corrupted by the adversary while both the training and testing data was corrupted for the O2B perceptron.

4. MACHINE LEARNING AND ITS ALGORITHM

Machine learning is a branch of artificial intelligence. It's the construction and study of the systems that can learn from the system. Arthur Samuel (1959) defined machine learning a field of study that gives computer ability to learn without being explicitly programmed. He started his research by first observing the checker's game where he learned the positions of the checker on which one can win and position on which one can lose. However this is quite traditional definition the new definition is given by Tom Mitchell who defines machine learning as "A computer program is said to learn from Experience E with respect to some Task T and some performance measure P, if its performance on T , as measured by P, improves with experience E". If we consider an E-mail example of spam where we put the e-mails which we don't

want in our inbox then there **Task T** is classifying emails as spam or not spam, **Experience E** is watching you label emails as spam or not spam and **Performance measure P** is the number (or fraction) of emails correctly classified as spam/not spam.

i.)

There are two machine learning algorithms:

- Supervised Learning
- Unsupervised Learning

A. Supervised Learning:

Supervised Learning is a machine learning technique for approximating the input/output behavior of complex systems. To define it more formally, the goal is, given a training set, to learn a function *h:X->Y* so that *h(x)* is a "good" predictor for the corresponding value of y. Here function f is called a **hypothesis**.

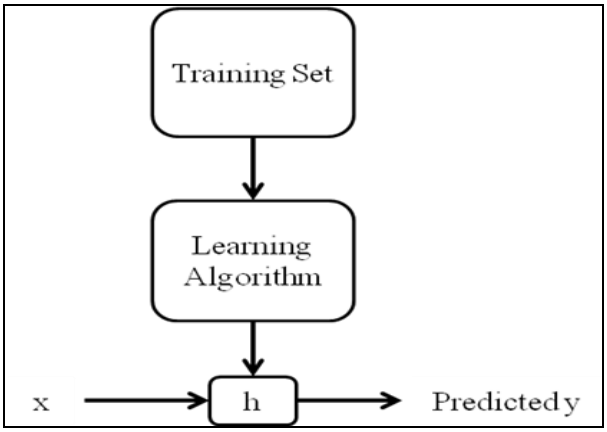

Fig. 4 Process of Supervised Learning

The above example shows the process of supervised learning where learning algorithm is applied on the training set which in turn is creating a function h so that for x input it produces predicted output as y.

B. Unsupervised Learning:

Unsupervised Learning is a machine learning technique in which the computer adapts the behavior on its own and obtains the output by analyzing the input on its own. The following figure shows how unsupervised process is performed:

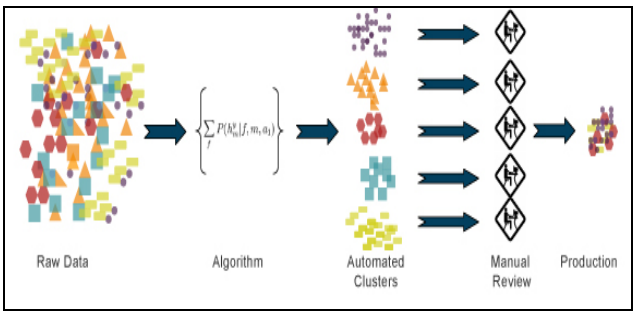

Fig. 5 Process of Unsupervised learning

In the above diagram you can see that in unsupervised learning a data is unknown and needs to be categorized and similar kind of data has to form a separate cluster. By applying the machine learning algorithm one can separate out the clustered data and hence identify the different forms of data. There are many supervised learning techniques which are used today and some have future scope as well. Some of currently used techniques are defined below:

Artificial Neural Network: this technique is highly inspired from central nervous system where it consists of interconnected group of neurons and in most of the cases it has adaptive system which changes its structure during a learning phase. In this technique there are three layers namely input layer, output layer and hidden layer. In the hidden layer the main functioning is defined that consists of activation function which is responsible for the output of a neuron in a neural network. There are three kinds of activation function: threshold function, piecewise-linear function and sigmoid function. The figure shows the pictorial representation of the ANN mathematical model.

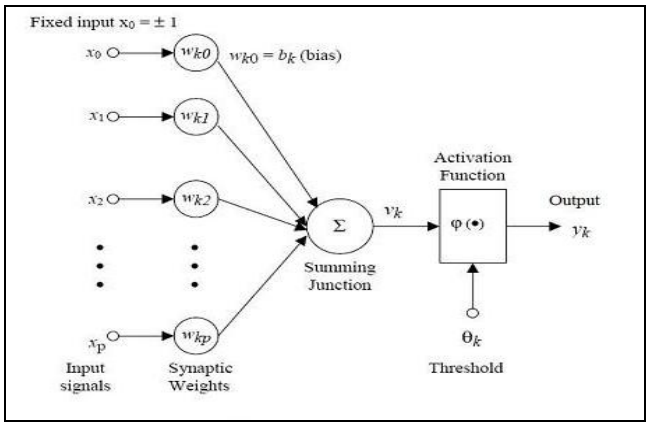

Fig. 6 Mathematical model of artificial neural network [7]

In the diagram you can see first part is the input signals and then comes synaptic weights which are summed together as summing junction. After summation then activation function is applied which decides the output y_k. The artificial neural network can be applied in single layer or multi layer feed forward format.

Support Vector Machine: Support vector machine which is popularly known as SVM is a supervised machine learning technique.

The SVM is a machine learning technique which is used to solve classification problem and is easy to use. In SVM we provide input as a matrix where each row represented as an observation or replicate and each column represented as a feature or variable. The SVM is made learn to classify the required dataset into certain categories which are basically of two types normal and abnormal. The SVM can be classified into linear and non-linear support vector machine. A classic example of SVM can be defined by the following figure 2.6. Suppose there are two classes of data namely class 1 and class 2 and they are in a 2-D plane so in order to classify them separately it requires some classification technique. SVM will classify the two classes by a linear line with the margins which can be small or large according to given data set.

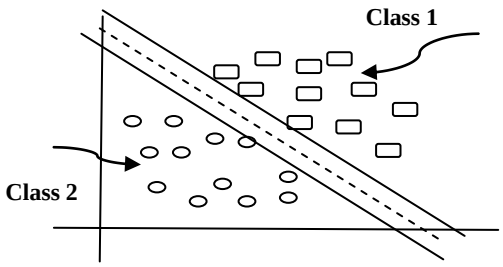

Fig. 7 A simple 2-D example for decision algorithm in an SVM

In the above diagram we can see the linear line separating the two classes and drawing margins across the boundaries. Same way we can separate the packets of the network into categories of normal packet and infected packet (i.e. abnormal or which may harm the network as well as the system).

Perceptron Linear Classifiers: Perception is an essential component of intelligent behavior. In term of neural network Perceptron is the form of feed forward neural network technique which acts as a classifier of two classes which are linearly separable. The following figure 2.7 is showing the perception process. In this process the knowledge gained by past experience is fed into perceptual lens. The current experience is built by using perceptual stream and the manner in which it assemble in perceptual images. Memory and world knowledge is kept as a storage system where the various beliefs and knowledge about the world is present.

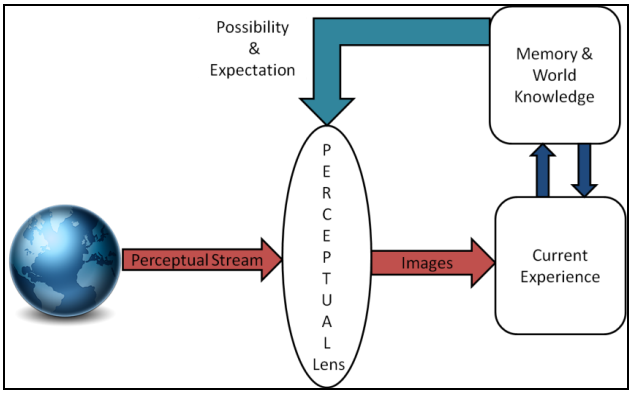

Fig.9 Perception Process

The basic Perceptron architecture consists of three layers: sensor layer, associative layer and output neuron. The number of inputs is represented by X_n in sensor layer which has corresponding weights W_n and an output. The mathematical model is given in figure 2.8.

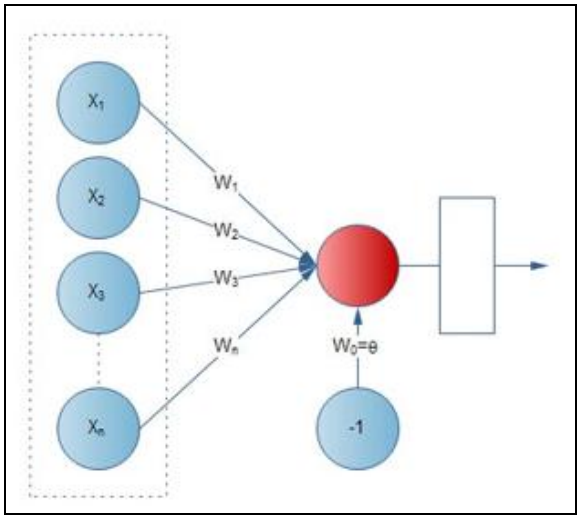

Fig.10 Perceptron Model [8]

In the mathematical model one can observe the inputs X_n having corresponding weights W_n followed by the output. For calculating the output through Perceptron model every input is multiplied by its corresponding weight, then all the multiplied values are summed up which is then evaluated by a function providing the final output.

5. CONCLUSION

The paper is dealing with security aspects of host as well as network. There are various algorithms which are biological inspired and are really beneficial for enhancing the security of the host/network. Our main aim is to analyze all the attacks associated with the system and network. These attacks can be detected using machine learning techniques by which, one can increase the system intelligence in either preventing or detecting (in case if attack is being performed) the threat which can be harmful for the system. The paper covers the latest techniques which are or can be applied to classify the data into secured and non secured so that all the threats are removed successfully from the system.

6. REFERENCES

[1] Krupa, F.,"*The Evolution of the Telephone System*", 1992 From Bell's Electric Toy to the Internet.

[2] Mark Havens, "Brief History of Computer Security", Dallas makerspace, 2010.

[3] William Stallings, "Cryptography and Network Security: principles and practices", Pearson publications, 2008.

[4] Mrutyunjaya Panda, Manas Ranja Patra, "*Evaluating Machine Learning Algorithms for Detecting Network Intrusions*", International Journal of recent trends in Engineering, 2009.

[5] Hua Tang, Zhuolin CAO, "*Machine Learning-based Intrusion Detection Algorithms*", Journal of Computational Information Systems, 2009.

[6] Vipin Das, Vijaya Pathak, Sattvik Sharma, Sreevathan, MVVNS.Srikanth, Gireesh Kumar T, "*Network Intrusion Detection System Based on Machine Learning ALgorithm*", International Journal of Computer Science & Information Technology, 2010.

[7] W. Nicholas Greene, Nathan Newsom, "*Classification of Malicious Network Activity*", 2011.

[8] O. Dekel, L. Xiao, "*Learning to classify with missing and corrupted features*", Machine Learning, 81(2):149-178, 2009.