Efficient Method for Intrusion Detection in Multitenant Data Center

Swapnil M. Jawahire [1], Prof. H. A. HIngoliwala[2]

[1] Department of Computer Engineering, JSCOE , Hadpsar Pune, Maharashtra
[2] Professor, Department of Computer Engineering, JSCOE , Hadpsar Pune, Maharashtra

---***---

Abstract - ***Today, cloud computing systems are providing a wide variety of services and interfaces to which different service providers are now renting out spaces on their physical machines at an hourly rate. With the rapid expansion of computer usage and computer network the security of the computer system has become very important. Every day new kind of attacks are being faced by industries. Cloud computing is rapidly changing the scenario of the Internet service infrastructure that provided to users , enabling even small organizations to quickly build Web and mobile applications for millions of users by taking advantage of the scale and flexibility of shared physical infrastructures provided by cloud computing. With wide computing and growth of information in such organizations, the data center is the most demanded infrastructure in industry. In this scenario, multiple tenants save their data and applications in shared data centers, blurring the network boundaries between each tenant in the cloud. In addition, different tenants have different security requirements, while different security policies are necessary for different tenants. Of greater concern for most IT managers in a multitenant environment is securely isolating each tenant's data flows from other tenants operating on the same server. Firewalls, Intrusion Detection System (IDS), Anti-Virus Gateway etc. are now widely deployed in edge-network to protect end-systems from the attacks. However, nowadays sophisticated attacks are distributed in the overall Internet, have fewer characteristics and transform quickly. To address this we propose intrusion detection using artificial intelligence used in a multi-tenant data center. This proposed system will use Self Organizing Map algorithm which learn to characterize normal behavior, to prepare itself to detect any aberrant network activity. Neural networks method is a promising technique which has been used in many classification problems. The neural network component will implement the neural approach, which is based on the assumption that each user is unique and leaves a unique footprint on a computer system when using it. If a user's footprint does not match his/her reference footprint based on normal system activities, the system administrator or security officer can be alerted to a possible security breach. At the end of the paper we will figure out the advantages and disadvantages of Self Organizing Maps and explain how it is useful for building an Intrusion Detection System.***

Key Words: *Artificial Intelligence, Neural Network, Self Organizing Maps, Multitenant Data Center*

1. INTRODUCTION

A cloud data center is an infrastructure that supports Internet services. A cloud data center may be defined from a variety of perspectives, and the most popular ones are categorized by IaaS, PaaS, and SaaS proposed by the NIST [2] and include public cloud, private cloud, hybrid cloud, and other different categories. Other categories include computing, networking, storage from a system's perspective or using (in use), archiving (at rest), and transmission (in motion) from a data perspective.

Despite the tremendous business and technical advantages of the cloud, the security and privacy concern has been one of the major hurdles preventing its widespread adoption. Firewalls, Intrusion Detection System (IDS), Anti-Virus Gateway etc. are now widely deployed in edge-network to protect end-systems from the attacks. When the malicious attacks have fixed patterns, they can be clocked by recording and matching these patterns. This method works well in the past tens of years. However, nowadays sophisticated attacks are distributed in the overall Internet, have fewer characteristics and transform quickly. Confronted with such attacks, the traditional security appliances always have poor performance, so better mechanism is necessary to prevent these attacks.

Traditional security devices such as Firewalls, IDS, WAF, and other devices are deployed with the Middleboxes model inside and outside networks. With the development of cloud computing technology, the deployment of Middleboxes is facing new challenges in the large-scale data center network environment [3].

In this scenario, multiple tenants save their data and applications in shared data centers, blurring the network boundaries between each tenant in the cloud. In addition, different tenants have different security requirements, while different security policies are necessary for different tenants. Network virtualization is used to meet a diverse set of tenant-specific requirements with the underlying physical network, enabling multi-tenant datacenters to automatically address a large and diverse set of tenant's requirements. In this project, we propose the collaborative network security system used in a multi-tenant data center. This system works in three tier mode, with a centralized collaborative scheme and deep packet inspection with an open source UTM [4] system. A security level based protection policy is proposed for simplifying the security rule management. Different security levels have different packet inspection schemes and are enforced with different security plug-in. Along with the additional accuracy of inspection is included in the system for intrusion detection and network policy violations.

2. LITERATURE SURVEY

Intrusion Detection System (IDS) [5] is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of intrusion. It is useful not only in detecting successful intrusions, but also in monitoring attempts to break security, which provides important information for timely counter-measures. Basically, IDS can be classified into two types: Misuse Intrusion Detection and Anomaly Intrusion Detection.

The assemblage and hybridization of various Artificial Intelligence techniques also indicate a bright future in the analysis of IDS and the prediction of its various properties for effective real-time network security. Among which ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.

An artificial Neural Network consists of a collection of treatments to transform a set of inputs to a set of searched outputs, through a set of simple processing units, or nodes and connections between them. Subsets of the units are input nodes, output nodes, and nodes between input and output form hidden layers; the connection between two units has some weight, used to determine how much one unit will affect the other. Two types of architecture of Neural Networks can be distinguished [6] [7]:

- **Supervised Training Algorithms**, where in the learning phase, the network learns the desired output for a given input or pattern. The well-known architecture of supervised neural network is the Multi-Level Perceptron (MLP); the MLP is employed for Pattern Recognition problems.
- **Unsupervised Training Algorithms**, where in the learning phase, the network learns without specifying desired output. Self-Organizing Maps (SOM) are popular unsupervised training algorithms; a SOM tries to find a topological mapping from the input space to clusters. SOM are employed for classification problems.

Matti Manninen [8] focuses on finding out how to make an IDS environment learn the preferences and work practices of a security officer, and how to make it more usable by showing the most often viewed anomalies first. Author seems that neural networks are the most popular selection for this kind of AI with a good reason. The reason is they are able to efficiently use incomplete or distorted data and to figure out relations between events, which help in detecting attacks from multiple sources. They are very fast in classifying the events. They are able to learn and identify new threats that haven't been expressly taught to them. The neural networks return a probability instead of a Boolean value, which makes it possible for them to predict probable following events in case there would be an attack going on. In turn, this would make it possible to defend against them in advance, in case that the system is an online IDS. Wide variety of choices for a neural network type makes it possible to select a type that works in a given application.

N. A. Alrajeh et al [9] presents a critical study on genetic algorithm, artificial immune, and artificial neural network (ANN) based IDSs techniques. Two types of ANN architectures, that is, feedforward ANN and feedback ANN. In feedforward ANN, the signals move in only one direction from input to output. In feedback ANN, the signals move in both directions. ANN concepts are helpful in many areas such as pattern recognition and intrusion detection. ANN based intrusion detection can be helpful to eliminate the shortcomings of rule based IDSs. However, ANN based IDSs can be more effective if properly trained with both normal and abnormal data sets. This paper provide following table of comparison.

In [10] present an approach of user behavior modeling that takes advantage of the properties of neural algorithm. It seems that the multilayer perception with the error back propagation learning algorithm has several disadvantages. It is necessary o fix N, and it will remain constant during all the life of the network. It is a dimensioning parameter for the inputs, and thus a change would require a complete retraining. Which causes the performance of the model will be dramatically reduced.

In [6] [11] considered the two learning algorithms namely supervised and unsupervised and investigated its properties in the classification of post graduate students according to their performance during the admission period. Supervised learning uses the MultiLayer Perceptron (MLP) models and Unsupervised Learning uses the Self-Organizing Maps (SOM) for comparison. The observation on the two results favors unsupervised learning algorithms for classification problems since the correctness percentage is high compared to the supervised algorithm.

System Architecture

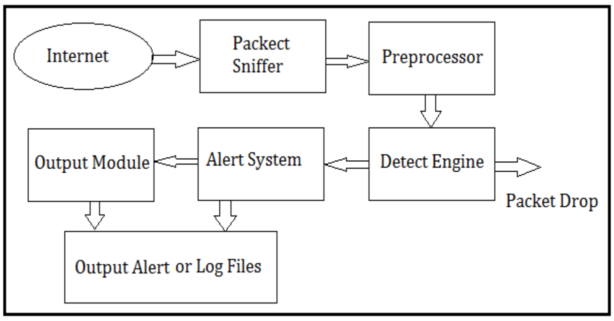

Fig.1 System Architecture

3. TYPES OF NETWORKING ATTACKS

There are four major categories of networking attacks. Every attack on a network can be placed into one of these groupings.

- **Denial of Service (DoS)**: A DoS attacks is a type of attack in which the hacker makes a memory resources too busy to serve legitimate networking requests and hence denying users access to a machine e.g. apache, smurf, Neptune, ping of death, back, mail bomb, UDP storm, etc.
- **Remote to User attacks (R2L)**: A remote to user attack is an attack in which a user sends packets to a machine over the internet, and the user does not have access to in order to expose the machines vulnerabilities and exploit privileges which a local user would have on the computer, e.g. xlock, guest, xnsnoop, phf, sendmail dictionary etc.
- **User to Root Attacks (U2R):** These attacks are exploitations in which the hacker starts off on the system with a normal user account and attempts to abuse vulnerabilities in the system in order to gain super user privileges, e.g. perl, xterm.
- **Probing**: Probing is an attack in which the hacker scans a machine or a networking device in order to determine weaknesses or vulnerabilities that may later be exploited so as to compromise the system. This technique is commonly used in data mining, e.g. satan, saint, portsweep, mscan, nmap etc.

4. INTRUSION DETECTION USING ARTIFICIAL INTELLIGENCE

The application of the capabilities of Artificial Intelligence techniques has been widely appreciated in Computer and Communication Networks in particular, as well as in other fields. AI naturally transformed into Computational Intelligence (CI) with the introduction of the concept of Machine Learning. This is a scientific aspect of AI that is concerned with the design and development of algorithms that allow computers to learn based on data, such as a network intrusion log acquired over a considerable period of time. A major focus of machine learning research is to automatically learn to recognize complex attributes and to make intelligent decisions based on the correlations among the data variables. The machine learning concept can be categorized into three common algorithms viz. supervised, unsupervised and hybrid learning. Supervised learning is the type of machine learning technique in which the algorithm generates a function that maps inputs to the desired outputs with the least possible error. Unsupervised learning is the machine learning technique in which a set of inputs are analyzed without the target output. This is also called clustering. The hybrid learning combines the supervised and unsupervised techniques to generate an appropriate function and to meet a specific need of solving a problem.

2.1 Self Organizing Maps

The Self-Organizing Map [11][12] is a neural network model for analyzing and visualizing high dimensional data. It belongs to the category of competitive learning network. It is a competitive network where the goal is to transform an input data set of arbitrary dimension to a one- or two-dimensional topological map. The model was first described by the Finnish professor Teuvo Kohonen and is thus sometimes referred to as a Kohonen Map. The SOM aims to discover underlying structure, e.g. feature map, of the input data set by building a topology preserving map which describes neighborhood relations of the points in the data set. The SOM is often used in the fields of data compression and pattern recognition. There are also some commercial intrusion detection products that use SOM to discover anomaly traffic in networks by classifying traffic into categories. The structure of the SOM is a single feed forward network, where each source node of the input layer is connected to all output neurons. The number of the input dimensions is usually higher than the output dimension.

The neurons of the Kohonen layer in the SOM are organized into a grid, see figure 2 and are in a space separate from the input space. The algorithm tries to find clusters such that two neighboring clusters in the grid have codebook vectors close to each other in the input space. Another way to look at this is that related data in the input data set are grouped in clusters in the grid [5]. The training utilizes competitive learning, meaning that neuron with weight vector that is most similar to the input vector is adjusted towards the input vector. The neuron is said to be the 'winning neuron' or the Best Matching Unit (BMU). The weights of the neurons close to the winning neuron are also adjusted but the magnitude of the change depends on the physical distance from the winning neuron and it is also decreased with the time.

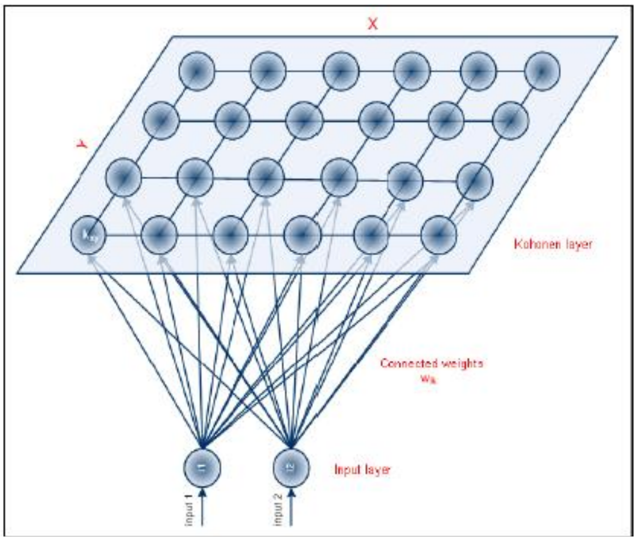

Fig -2: The self-organizing (Kohonen) map

The algorithm must be able to classify such information in a "sensible" way. By sensible we mean that it should preserve three important properties:

1. Preserve as much information as possible about the similarity between packets
2. Separate, as much as possible, packets from different protocols in different groups
3. Most importantly, since our final goal is to detect intrusions, separate packets with anomalous or malformed payload from normal packets

The learning process of the SOM goes as follows:
1. One sample vector x is randomly drawn from the input data set and its similarity (distance) to the codebook vectors is computed by using Euclidean distance measure:

$$\parallel x - m_c \parallel = min\{\parallel x - m_c \parallel\} \text{ .... (1)}$$

2. After the BMU has been found, the codebook vectors are updated. The BMU itself as well as its topological neighbors are moved closer to the input vector in the input space i.e. the input vector attracts them. The magnitude of the attraction is governed by the learning rate. As the learning proceeds and new input vectors are given to the map, the learning rate gradually decreases to zero according to the specified learning rate function type. Along with the learning rate, the neighborhood radius decreases as well. The update rule for the reference vector of unit i is the following:

$$m_i(t+1) = \begin{cases} m_i(t) + a(t)[x(t) - m_i(t)], i \in N_c(t) \\ m_i(t), i\,! \in N_c(t) \end{cases}$$

3. The steps 1 and 2 together constitute a single training step and they are repeated until the training ends. The number of training steps must be fixed prior to training the SOM because the rate of convergence in the neighborhood function and the learning rate are calculated accordingly.

After the training is over, the map should be topologically ordered. This means that *n* topologically close input data vectors map to *n* adjacent map neurons or even to the same single neuron.

5.2 Data Collection

If the network traffic has been examined carefully for different types of events such as downloading, port scanning, surfing etc., it is possible to identify the formal distinctions between them. The idea behind this work is to collect distinct and various kinds of network packets. To collect data we can use any packet sniffer which is available readily. Here in this case we use jpcap packet capture library which is based on WinPcap[13]. Apart from capturing live packets we also a standard DARPA dataset, which we will be using for training purpose [14]. The dataset contain both packets with intrusion and without intrusion.

6. RESULTS

After the data collection, clustering and training of the Self Organizing Maps we pass the packets through the SOM. The result is shown in the fig. The results Fig. 3 show input vectors classification, which represents behavior and its mapping to particular neurons, which form single possible user behavior states. Form states as intrusion – Intrusion, possible intrusion – Intrusion? Normal – Norm. From the test result SOM network represents suitable core for IDS systems.

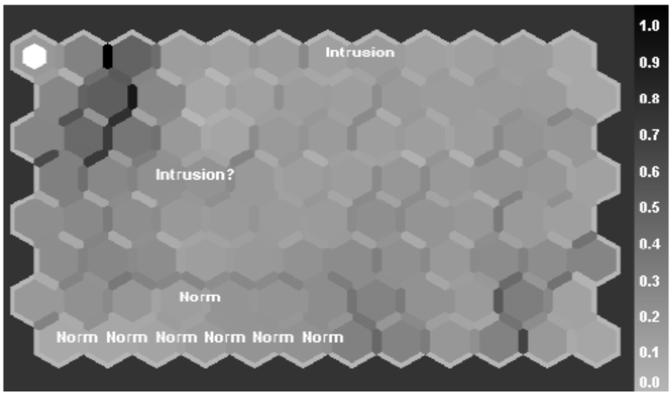

Fig. 3 Results of Experiment

7. CONCLUSIONS

The goal of this article was to design architecture of the system detecting intrusion based on Self Organizing Maps. Classification module of the proposed architecture is self-organizing map. It is observed that neural networks are the most popular selection of AI as it is able to efficiently use incomplete or distorted data and to figure out relations between events, which help in detecting attacks from multiple sources. Among various techniques of neural networks –Self Organizing Maps (SOM) are popular unsupervised training algorithms which overcome the limitations of Multi-Layer Perceptron algorithm (MLP) and other AI techniques.

ACKNOWLEDGEMENT

I wish to thank all the people who have directly or indirectly helped me in completing Paper work successfully. I express my gratitude towards my project guide and Head of Computer Engineering Department Prof. H. A. Hingoliwala for their valuable suggestions and constant guideline during this paper work also acknowledge the research work done by all researchers in this field across the Internet for maintaining valuable document and resource on the Internet.

REFERENCES

[1] Z. Chen, W. Dong, Hang Li, P. Zhang, X. Chen and J. Cao, “Collaborative Network Security in Multi-Tenant Data Center for Cloud Computing”

[2] NIST definition of cloud computing, http:// csrc.nist.gov/ publications/ Pubs NIST IRs.html, 2007.

[3] Qihoo 360 Internet Security Center, Development trend of enterprise security in the internet ages, http://www.gartner.com/technology/mediaproducts /pdfindex.jsp?g=Qihoo issue1, 2013.

[4] K. Y. Zhang, F. Deng, Z. Chen, Y. B. Xue, and C. Lin, UTM-CM: A practical control mechanism solution for UTM system, in Proc. IEEE International Conference on Communications and Mobile Computing, Shenzhen, China, 2010, pp. 86-90.

[5] Fatai Adesina Anifowose, Safiriyu Ibiyemi Eludiora , Application of Artificial Intelligence in Network Intrusion Detection, in Proc. World Applied Programming, Vol (2), No (3), March 2012. 158-166 ISSN: 2222-2510

[6] R. Sathya, Annamma Abraham, “Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification”, in Proc. International Journal of Advanced Research in Artificial Intelligence, Vol. 2, No. 2, 2013

[7] Shelly Xiaonan Wu, Wolfgang Banzhaf, “The use of computational intelligence in intrusion detection systems: A review” in Proc. Applied Soft Computing 10 (2010) 1–35

[8] Matti Manninen, “Using Artificial Intelligence in Intrusion Detection Systems”

[9] Nabil Ali Alrajeh and J. Lloret, Intrusion Detection Systems Based on Artificial Intelligence Techniques in Wireless Sensor Networks, Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2013, Article ID 351047

[10] Herve DEBAR, Monique BECKER, Didier SIBONI, “A Neural Network Component for an Intrusion Detection System”

[11] Vivek A. Patole, Mr. V. K. Pachghare, Dr. Parag Kulkarni “Self Organizing Maps to Build Intrusion Detection System” in Proc. International Journal of Computer Applications (0975 – 8887)Vol. 1 – No. 8, 2010

[12] John A. Bullinaria, “Self Organizing Maps: Algorithms and Applications”

[13] WinPcap, Windows packect Capture Library, http://www.winpcap.org/

[14] Ciza Thomas, Vishwas Sharma, N. Balakrishnan, “Usefulness of DARPA Dataset for Intrusion Detection System Evaluation”.