MULTI AGENT BASED APPROACH FOR NETWORK INTRUSION DETECTION USING DATA MINING CONCEPT

Ankita Agarwal¹, Sherish Johri², Ankur Agarwal³, Vikas Tyagi⁴, Atul Kumar⁵
¹ankita.890@gmail.com, ²sherish3@gmail.com, ³anur18_mca@yahoo.com, ⁴itengg.vikas@gmail.com, ⁵atul0411@yahoo.com

Abstract: Data mining refers to extracting or “mining” useful data from large amounts of data. Due to the wide availability of huge amount of useful information and knowledge, data mining can be used for market analysis, fraud detection and customer retention. Network intrusion detection system is used to detect any intruder which might have entered into the computer system. In this paper, a multi agent based approach is used for network intrusion detection. An adaptive NIDS will be used. The method proposed is the enhancement of the previous methods. Here more numbers of agents are used which will be continuously monitoring the data to check for any intruder which might have entered in the system. Each agent is trained accordingly so that it can check for any type of intruder entering into the system. Verification through many agents will ensure the safety of the data.

Keywords: Data mining, Network intrusion detection, Data warehousing, multi-agent NIDS.

INTRODUCTION

In a growing number of domains — email spam, counter-terrorism, intrusion detection/computer security, click spam, search engine spam, surveillance, fraud detection, shop bots, file sharing, etc. — data mining systems face adversaries that deliberately manipulate the data to sabotage them (e.g. make them produce false negatives).[6]

The rapid progress of computers and databases has enabled companies to store data about customers and transactions for future use. The sheer amount of data to be analyzed in order to make better decisions require dramatically improved new automated data modeling technologies. A concept of Data Mining is developed. There are two foundation of using data mining techniques: the availability of large amount of data and the data mining modeling techniques.

The data warehouse is typically a combination of detailed demographic data example, on a customer, combined with a historical transactional history, which may include not only the purchases that were made by the customer, but also include contact or interaction data such as what type of promotions were made to each customer, which ones did they respond to, have they called on their own with support related questions, or inquire about a certain product. As networking becomes more widespread, the number of violations to normal operations is increasing. Current firewalls are not sufficient to ensure the security in computer networks, which some intrusions take advantages of vulnerabilities in computer systems or use socially engineered penetration techniques that traditional intrusion prevention techniques are not enough in protection. Network Intrusion Detection System (NIDS) will be another wall for protection [3].

In this paper, a multi agent based approach is used for network intrusion detection. An adaptive NIDS will be used. Here more numbers of agents are used which will be continuously monitoring the data to check for any intruder which might have entered in the system. Each agent is trained accordingly so that it can check for any type of intruder entering into the system.

There are five types of agent based on three data mining techniques, which are clustering, association rules and sequential association rules approaches.[3]

The problem is that current NIDS are tuned specifically to detect known service level network attacks. Attempts to expand beyond this limited realm typically results in an unacceptable level of false positives. At the same time, enough data exists or could be collected to allow network administrators to detect these policy violations. Unfortunately, the data is so volumous, and the analysis process so time consuming, that the administrators don’t have the resources to go through it all and find the relevant knowledge, save for the most exceptional situations, such as after the organization has taken a large
loss and the analysis is done as part of a legal investigation. In other words, network administrators don’t have the resources to proactively analyze the data for policy violations, especially in the presence of a high number of false positives that cause them to waste their limited resources.[5]

RELATED WORK

The problem of alert verification has been addressed using two different kinds of approaches: we have techniques for identifying true positives, and techniques for identifying false positives.

Most of the commercial NIDSs sold in the market are signature-based with a disadvantage in detection of previously known attacks only. Especially, different kinds of attack come every day. The signature-based NIDS will not be functional when new kinds of attack coming. Therefore, many researchers have proposed and implemented different intrusion detection models based on data mining techniques to tackle this problem.[3]

An adaptive NIDS based on data mining techniques is proposed. However, unlike most of the current researches, which only one engine is used for detection of various attacks; the proposed system is constructed by a multi-agent, which are totally different in both training and detection processes. In this stage, three data mining approaches: clustering, association and sequential association, are adopted and different types of agent are built according to the type of attack that might occur in the system.

After training with normal traffic for a network behavior, when new type of attack comes, the proposed system can detect such anomaly by distinguishing it from normal traffic.[3]

The first couple of IDSs of record (CERIAS 2003) that performed data fusion and cross sensor correlation were the Information Security Officer’s Assistant (ISOA) (Winkler and Page 1990; jtruit@dw3f.eis.harris.com 1994) and the Distributed Intrusion Detection System (DIDS) (Snapp et al. 1991; Snapp et al. 1991). ISOA conglomerated the audit information for numerous hosts whereas DIDS conglomerated the audit information from numerous host and network based IDSs.[5]

The data fusion and correlation capabilities of commercial intrusion detection systems spans over a wide range. A few products are specifically designed to do centralized alarm collection and correlation. For example RealSecure SiteProtector, which claims to do “advanced data correlation and analysis” by interoperating with the other products in ISS’s RealSecure line.

Application of Data Mining In Intrusion Detection:

After discussing the various components in intrusion detection system in this section various areas of intrusion detection in which data mining technology are applied are studied. The following are areas in which data mining technology applied or further developed for intrusion.

Data Mining Algorithms for Intrusion Detection:

Data mining algorithms can be used for misuse detection and anomaly detection. In misuse detection, training data are labeled as either “normal” or “intrusion.” A classifier can then be derived to detect anomalies & known intrusions [24, 20]. Research in this area has included the application of classification algorithms, association rule mining, and cost-sensitive modeling. Anomaly detection builds models of normal behavior and automatically detects significant deviations from it [9]. Supervised or unsupervised learning can be used. In a supervised approach, the model is developed based on training data that are known to be “normal.” In an unsupervised approach, no information is given about the training data [22]. Anomaly detection research has included the application of classification algorithms, statistical approaches, clustering, and outlier analysis [23, 19, 20, 21, and 22]. The techniques used must be efficient and scalable, and capable of handling network data of high volume, dimensionality, and heterogeneity [18].

Association and Correlation Analysis Helps to Select And Build Discriminating Attributes:

Association and correlation mining can be applied to find relationships between system attributes describing the network data [12]. Such information can provide insight regarding the selection of useful attributes for intrusion detection. New attributes derived from aggregated data may also be helpful, such as summary counts of traffic matching a particular pattern [18].

Analysis of Stream Data:

Due to the transient and dynamic nature of intrusions and malicious attacks, it is difficult to perform intrusion detection in the data stream environment. However, an event may be normal on its own, but considered malicious if viewed as part of a sequence of events. Thus it is necessary to study what sequences of events are frequently encountered together, find sequential patterns, and identify outliers [15]. Other data mining methods for finding
evolving clusters and building dynamic classification models in data streams are also necessary for real-time intrusion detection.[2]

**PROPOSED WORK**

A feature extractor converts the data from a monitored system into features which will be used in both training and network intrusion detection stages. Figure shows the overall system architecture.

In Detection Engine, a Feature Distributor allocates necessary feature vectors to each agent. In this stage, three main data mining approaches, clustering, association rules and sequential association rules are used. The results among all agents are gathered by the decision makers for making conclusion on the final decision of the system.

For each detection agent, corresponding trainer is built for updating agent in an adaptive manner. Same as the Detection Engine, a Feature Distributor assigns necessary feature vectors to each training node. Each training node is built in a corresponding data mining approach and updated corresponding agent adaptively the agent trainer. An anomaly detection model is based on normal behavior only and deviations from it. In other words, the normal behavior of the network is profiled. This model is possibly high in false alarm rate as previously observed. System behaviors may be recognized as anomalies, but the adaptive ability of this model to the environment is expected to be higher. The Feature Extractor has corresponding functions for each kind of statistics, and it is flexible to use. For example, the statistical figures of the frames of packet will be collected (min/average/max value, unique item, …, etc.). Those statistics form feature vectors for both detection and training processes. Currently, the system supports the following frame feature extraction.

- **Min/Ave/Max value of a field.**
- **Number of unique item of a field.**
- **Time range covered by a frame.**
- **Number of packets in a frame.**
- **Number of packets in a frame after passing filter.**
- **Number of connection attempted to open in a frame.**

The use of multi agents in the path of a single passage will detect any intruder which might have affected the network.

Intrusions can be launched from several different locations and targeted to many different destinations. Distributed data mining methods may be used to analyze network data from several network locations in order to detect these distributed attacks [15,16],[2]

In a centralized firewall policy, it is important for packet filtering process whether its filtering rules are disjoint or not, or even worse in conflict. Thus it is very common to discover that some of these filtering rules are interrelated and thus its ordering may create very different results or anomalies, thus resulting in an incorrect firewall policy.[4]

The proposed system as compared to the previous methodology will ensure that the network is free from any intrusion and the data is fully secured. This might be a more expensive method than the previous ones but it is more accurate than the previous ones.

**CONCLUSION & FUTURE DEVELOPMENT**

The extension of the model with additional attributes can help to unearth further mistakes. The analysis of statistical properties of router configurations appears to be a promising approach to assist operators in detecting mistakes.[1]

Unlike most of the current research, which use only one agent per engine for detection of various attacks, the proposed system is constructed by a number of agents in a single engine. The NIDS can broaden its view on different behaviors of the network traffic by each of the agents with its own strength on capturing a kind of network behavior. Firewall policy rules are one of most important element of network security system. It plays the vital role in management of any organization’s network and its security infrastructure.

Thus the management of policy rule is a significant task for the network security. There have been a number of tools and techniques used to perform anomaly detection and rule editing by utilizing given set of existing policy rules. However, one of the assumption and thus its limitation is that firewall and its rules are set to be static and thus without an ability to reflect the network behavior observed by firewall [4].

For future development, the following directions are proposed: (i) to develop more agents which are strength on other aspects, (ii) to set the thresholds by the system with minimum human interrupt and also (iii) to introduce incremental updating mechanism for the detection agents.[3]

**REFERENCE**

[1]. Franck Le, Sihyung Lee, Student Member, IEEE, Tina Wong, Hyong S. Kim, and Darrell Newcomb.” Detecting Network-Wide and Router-Specific Misconfigurations Through Data Mining”

[2]. S.V.Shirbhate Dr V.M.Thakare Dr S.S.Sherkar .” Data Mining Approaches For Network Intrusion Detection System”


[5]. S TERRY BRUGGER. “Data Mining Methods for Network Intrusion Detection”

[6]. Vuda Sreenivasa Rao, Dr. S. Vidyavathi. “DISTRIBUTED DATA MINING AND MINING MULTI-AGENT DATA”


[8]. IBM’s Aglet Mobile Agent Implementation.


[17]. Yue Jiang, Dr. Bojan Cukic. “Snort - an network intrusion prevention and detection system”


