Network Intrusion Detection Using Data Mining And Network Behaviour Analysis

Deepak Kumar, Diksha Verma, Anjli Tyagi
Department of Information Technology, Dronacharya college of Engineering

***Abstract*- Intrusion detection has become a critical component of network administration due to the vast number of attacks persistently threaten our computers. Traditional intrusion detection systems are limited and do not provide a complete solution for the problem. They search for potential malicious activities on network traffics; they sometimes succeed to find true security attacks and anomalies. However, in many cases, they fail to detect malicious behaviours (false negative) or they fire alarms when nothing wrong in the network (false positive). In addition, they require exhaustive manual processing and human expert interference. Applying Data Mining (DM) techniques on network traffic data is a promising solution that helps develop better intrusion detection systems. Moreover, Network Behaviour Analysis (NBA) is also an effective approach for intrusion detection. In this paper, we discuss DM and NBA approaches for network intrusion detection and suggest that a combination of both approaches has the potential to detect intrusions in networks more effectively.**

I. INTRODUCTION

Nowadays, there exists an extensive growth in using Internet in social networking (e.g., instant messaging, video conferences, etc.), healthcare, e-commerce, bank transactions, and many other services. These Internet applications need a satisfactory level of security and privacy. On the other hand, our computers are under attacks and vulnerable to many threats. There is an increasing availability of tools and tricks for attacking and intruding networks. An intrusion can

be defined as any set of actions that threaten the security requirements (e.g., integrity, confidentiality, availability) of a computer/network resource (e.g., user accounts, file systems, and system kernels) [16,17]. Intruders have promoted themselves and invented innovative tools that support various types of network attacks. Hence, effective methods for intrusion detection (ID) have become an insisting need to protect our computers from intruders. In general, there are two types of Intrusion Detection Systems (IDS); misuse detection systems and anomaly detection systems [14,16,17]. Most commercial IDS employ the misuse strategy in which known intrusions are stored in the systems as signatures. The system searches network traffics for patterns or user behaviours that match the signatures, if a pattern matched a signature; an alarm is raised to a human security analyst who decides what action should be taken based on the type of attack. In such systems, known intrusions (signatures) are provided and hand-coded by human experts based on their extensive experience in identifying intrusions. Current misuse IDS are built based on: expert systems (e.g., IDES, ComputerWatch, NIDX, P-BEST, ISOA) which use a set of rules to describe attacks, signature analysis (e.g., Haystack, NetRanger, RealSecure, MuSig) where features of attacks are captured in audit trail, state-transition analysis (e.g., STAT, USTAT and NetSTAT) which uses state-transition diagrams, coloured petri nets (e.g., IDIOT), or case-based reasoning (e.g., AUTOGUARD) [16]. Anomaly detection [8,12], in contrast to misuse detection, can identify novel intrusions. It builds models for normal network behaviour (called profiles) and uses these profiles to detect new patterns that significantly deviate from them. These suspicious patterns may represent actual intrusions or could simply be new behaviours that need to be added to profiles. Current anomaly detection systems use statistical methods such as multivariate and temporal analysis to identify anomalies; examples of these systems are IDES, NIDES, and EMERALD. Other anomaly detection systems are built based on expert systems such as ComputerWatch, Wisdom, and Sense [16].

Misuse IDS suffer from a number of major drawbacks, first, known intrusions have to be handcoded by experts. Second, signature library needs to be updated whenever a new signature is discovered, network configuration has been changed, or a new software version has been installed. Third, misuse IDS are unable to detect new (previously unknown) intrusions that do not match signatures; they can only identify cases that match signatures.

II. BACKGROUND AND RELATED WORK

Intrusion detection is the process of monitoring and analyzing the data and events occurring in a

computer and/or network system in order to detect attacks, vulnerabilities and other security problems [16]. IDS can be classified according to data sources into: host-based detection and network-based detection. In host-based detection, data files and OS processes of the host are directly monitored to determine exactly which host resources are the targets of a particular attack. In contrast, network-based detection systems monitor network traffic data using a set of sensors attached to the network to capture any malicious activities. Networks security problems can vary widely and can affect different security requirements including authentication, integrity, authorization, and availability. Intruders can cause different types of attacks such as Denial of Services (DoS), scan, compromises, and worms and viruses [17,18]. In this paper, we emphasize on network-based intrusion detection which is discussed in the next sub-section. The primary assumption in intrusion detection is that user and program activities can be monitored and modelled [16,17]. A set of processes represent the framework of intrusion detection, first, data files or network traffic are monitored and analyzed by the system, next, abnormal activities are detected, finally, the system raises an alarm based on the severity of the attack [16]. Figure 1 below shows a traditional framework for ID. In order for IDS to be successful, a system is needed to satisfy a set of requirements. IDS should be able to detect a wide variety of intrusions including known and unknown attacks. This implies that the system needs to adapt to new attacks and malicious behaviours. IDS are also required to detect intrusions in timely fashion, i.e., the system may need to respond to intrusions in real-time. This may represent a challenge since analyzing intrusions is a time consuming process that may delay system response. IDS are required to be accurate in a sense that minimizes both false negative and false positive errors.

Network-based Intrusion Detection

Network-based intrusion detection can be broken down into two categories: packet-based anomaly detection and flow-based anomaly detection. Flow-based anomaly detection tends to rely on existing network elements, such as routers and switches, to make a flow of information available for analysis. On the other hand, packet-based anomaly detection doesn't rely on other network components; it observes network traffic for the detection of anomalies. Flow-based anomaly detection is based on the concept of a network flow and flow records. A flow record is a summarized indicator that a certain network flow took place and that two hosts have communicated with each other previously at some point in time. Typically, the flow record contains both the source and destination IP addresses the source and destination TCP or UDP network ports or ICMP types and codes, the number of packets and number of bytes transmitted in the session, and the timestamps for both the start and end of the network flow. Routers generate these flow records as they observe network traffic. By analyzing flow records and looking for unusual amounts, directions, groupings and characteristics of the network flow, the network behavior analysis software can infer the presence of worms or even DoS attacks in a network. The problem is that these flow records only carry a summary of the information presented for analysis. Basically, this information is the metadata about the network traffic. The actual network packets are not accessible for further analysis [9]. Packet-based anomaly detection software, unlike its flow-based counterpart, does not use third party elements to generate the metadata of the network traffic. Instead, the entire packet-based analysis looks at raw packets as they traverse the network links. Observation of the network traffic can be done using either port mirroring or network taps. Port mirroring, known as SPAN (Switched Port Analyzer), is used on a network switch to send a copy of all network packets seen on one switch port to a network monitoring connection on another switch port. Network taps are used to create permanent access ports for passive monitoring. Test Access Port (TAP) can create a monitoring access port between any two network devices, including switches, routers, and firewalls. A good example to compare the two detection methodologies is that of a large-scale SYN flood denial of service attack. Typically a huge amount of connection request packets are generated by a number of compromised zombie machines. The source addresses are randomly generated.

Data Mining Techniques for Network Intrusion Detection

Many researchers have investigated the deployment of data mining algorithms and techniques for intrusion detection [13,15-23, 32,33]. Examples of these techniques include [16-18]:

Feature selection data analysis: The main idea in feature selection is to remove features with little or no predictive information from the original set of features of the audit data to form a subset of appropriate features [24]. Feature selection significantly reduces computational complexity resulting from using the full original feature set.

Other benefits of feature selection are: improving the prediction of ID models, providing faster and cost-effective ID models and providing better understanding and virtualization of the generated intrusions. Feature selection algorithms are typically classified into two categories: subset selection and feature ranking. Subset selection algorithms use heuristic search such as genetic algorithms, simulated annealing and greedy hill climbing to generate and evaluate a subset of features as a group for suitability. On the other hand, feature ranking uses a metric to rank the features based on their scores on that metric and removes all features that do not achieve an adequate score [34].

Classification analysis: The goal of classification is to assign objects (intrusions) to classes based on the values of the object's features. Classification algorithms can be used for both misuse and anomaly detections [16]. In misuse detection, network traffic data are collected and labelled as "normal" or "intrusion". This labelled dataset is used as a training data to learn classifiers of different types (e.g., SVM,NN,NB, or ID3) which can be used to detect known intrusions. In anomaly detection, the normal behaviour model is learned from the training dataset that are known to be "normal" using learning algorithms. Classification can be applied to detect intrusions in data streams; a predefined collection of historical data with their observed nature helps in determining the nature of newly arriving data stream and hence will be useful in classification of the new data stream and detect the intrusion.

III. NETWORK BEHAVIOUR ANALYSIS

Within the last few years, Network Behavior Analysis (NBA) has been one of these emerging technologies that have been sold as a security management tool to improve the current network security status. The main focus of NBA is to monitor inbound and outbound traffic associated with the network to ensure that nothing is getting into the servers, software, and application systems which helps enhance the overall security of the network at all levels. The author in [1] stated that approximately 25% of large enterprises systems will be using NBA by 2011. The

traditional security model of network is not clear and has too many concerns. First of all, the model have little proactive capability attitude toward preventing any security incidents because the architecture is built with technologies that discover most security events in progress while it misses opportunities to detect and resolve other small threats before it become major problems for the network. Firewalls and intrusion detection systems are typically stationed at a network gateway, which doesn't stop laptops infected with malware or subversive employees from accessing the network. A typical security tactic to overcoming this problem is to deploy firewalls and intrusion detection devices throughout the internal network [4]. This can get extremely expensive and can increase network maintenance and complexity even without addressing many of the security threats.

IV. CONCLUSIONS AND FUTURE WORK

Traditional IDS suffer from different problems that limit their effectiveness and efficiency. In contrast DM and NBA are promising approaches for intrusion detection. In this paper, we discussed DM and NBA approaches for network intrusion detection. We suggested that a combination of both approaches may overcome the limitations in current IDS and leads to high performance ones. NBA can help cover the gap in traditional network systems, which considers a good move for most of industries to integrate NBA with advanced DM to achieve a better performance. NBA can significantly enhance the value of the data generated from IDS that use DM as intrusion detection technique by analyzing and correlating large amount of sequence data. We plan to put the suggested hybrid system model in practice and apply it on real world intrusion detection problems.

REFERENCES

[1] Schwartz, Matthew, "Beyond Firewalls and IPS: Monitoring Network Behavior." February 2006, available on http://esj.com/articles/2006/02/07/beyond-firewalls-and-ips-monitoring-networkbehavior.aspx

[2] K. Scarfone and P. Mell, "Guide to Intrusion Detection and Prevention Systems (IDPS)", NIST Special Publication 800-94, 2007, Available online:http://csrc.nist.gov/publications/nistpubs/800-94/SP800-94.pdf

[3] Conry-Murray, "Anomaly Detection On the Rise", June 2005, available onhttp://business.highbeam.com/787/article-1G1-132920452/anomaly-detection-rise-network-behavioranomaly-detection

[4] Enterprise Strategy Group, "Network Behavior Analysis Systems: The New Foundation of Defensein-Depth", Technical White Paper,

November 2005. http://www.enterprisestrategygroup.com/

[5] Mazu Networks, "What You Can't See Can Hurt You: Ensuring Application Availability through Enterprise-Wide Visibility", November 2006. http://www.developertutorials.com/whitepapers/network-communications/

[6] Liebert, Chris, "Internal Threat Protection with Net-Based Detection, Prevention and Behavioral Systems", October 2006, http://www.mazunetworks.com/resources/analystreports/Internal_Threat_Protection_January_06.pdf.