A Survey on Data Mining Approaches for Network Intrusion Detection System

Anirudha A. Kolpyakwar
Assistant Professor
& Computer Engineering
JCET Yavatmal, India

Mangesh G. Ingle
Assistant Professor
& Computer Engineering
JCET Yavatmal, India

Ritesh V. Deshmukh
Assistant Professor
& Computer Engineering
JCET Yavatmal, India

ABSTRACT
Data mining has been gaining popularity in knowledge discovery field, particularity with the increasing availability of digital documents in various languages from all around the world. Network intrusion detection is the process of monitoring the events occurring in a computing system or network and analysing them for signs of intrusions. In this paper, intrusion detection & several areas of intrusion detection in which data mining technology applied are discussed. Data mining techniques are used to discover consistent and useful patterns of system features that describe program and user behaviour. Data mining can improve variant detection rate, control false alarm rate and reduce false dismissals. By using these set of relevant system features to compute classifiers that recognize anomalies & known intrusion.

Keywords
Intrusion Detection, Data Mining, Misuse Detection, Anomaly Detection.

1. INTRODUCTION
In the digital Network, users are facing the new challenges of electronic attacks. In this context, Intrusion detection is the important technology which gives us remedial solution to this problem. Now days, most of the research is going on in this direction. An intrusion is defined as any set of actions that threat the integrity, confidentiality, or availability of a network resource such as user accounts, file systems, system kernels, and so on. According to Webster's an intrusion as the act of thrusting in, or of entering into a place without invitation, right, or welcome [1] Intrusion is defined as the act of wrongfully entering upon, grasping, or taking possession of the property of another. [2] Intrusion is coming into place without permission. In Intrusion Detection (ID), collects the information and analyzing it for uncommon or unexpected events. Intrusion detection is the process of monitoring and analyzing the events which occurred in a computer system in order to detect signs of security problems. [3] Over the past few years, intrusion detection and other security technologies such as cryptography, authentication, and firewalls have increasingly gained importance in digital data [4] Intrusion detection is data analysis process. The main theme of our approach is to apply data mining techniques to intrusion detection. Data mining is the process of extracting patterns from large amount of stored data. [5] Now days the main reason of applying Data Mining for intrusion detection systems is the enormous volume of existing and newly appearing network data that requires processing. [6] Traditional intrusion detection systems face many limitations. So this has led to an increased interest in data mining for intrusion detection. Data mining can improve variant detection rate, control false alarm rate and reduce false dismissals. This paper has been divided into six sections. Section I defines the overview of data mining approaches for network intrusion detection system. Section II portrays the basic idea of intrusion detection. Section III methods of intrusion detection are discussed. Section IV highlights the various components of intrusion detection system. Section V highlights some areas of intrusion detection in which data mining are applied and section VI finally conclude by discussing the outcome of study.

2. INTRUSION DETECTION
Intrusion detection is the process of monitoring the events occurring in a digital network and analyzing them for signs of possible incidents [7]. The security of our digital network and data is at continual risk. Due to the extensive growth of the Internet and increasing availability of tools and tricks for intruding and attacking networks have prompted that intrusion detection is become a critical component for network administrator. The purpose of intrusion detection is to detect security violations in information systems. Intrusion detection is a passive approach to security as it monitors information systems and raises alarms when security violations are detected. Examples of security violations include the abuse of privileges or the use of attacks to exploit software or protocol vulnerabilities. So there is need of one of the tool which automatically detects the intrusions in digital network. Hence an intrusion detection system is software that automatically detects the intrusions occurred in system.

3. INTRUSION DETECTION METHODLOGIES
Intrusion detection system uses many methodologies to detect incidents. Incidents have many reasons, such as malware e.g., worms, spyware, attackers gaining unauthorized access to systems from the Internet, and authorized users of systems who misuse their privileges or attempt to gain additional privileges for which they are not authorized. Traditionally, intrusion detection techniques are classified into two broad categories: misuse detection and anomaly detection [8].

3.1 Misuse Intrusion Detection
Misuse detection searches for the traces or patterns of well-known attacks which are stored as signatures [9, 10]. These signatures are provided by human expert based on their extensive knowledge of intrusion techniques. In this process if a pattern matched is found, this signals an event for which an alarm raised. After that security analyst evaluate the alarms to decide what action to take for e.g. shutting down part of the system, alerting the relevant internet service provider of suspicious traffic, or simply nothing unusual traffic for future reference. In misuse detection, each instance in data set is labelled as normal or intrusion and a learning algorithm is trained over labelled data [10]. From this discussion we can say that only known attacks that leave characteristic traces can be detected. This is one of the drawbacks of misuse detection.

Also there is need to update the signature whenever new software version arrive or changes in network configuration occur because the systems are dynamic. A key advantage of misuse detection technique is their high degree of accuracy in detecting known attacks and their variations [9, 10]. A typical misuse detection system is as shown in fig 1.

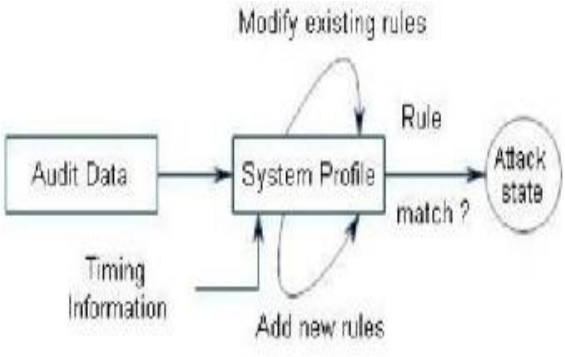

Fig.1. Misuse Detection

3.2 Anomaly Intrusion Detection

Misuse detection system unable to detect new or previously unknown intrusions occurred in computer system or digital network. Novel intrusions can be found by anomaly detection. Anomaly detection uses a model of normal user or system behavior and flags significant deviations from this model as potentially malicious [9, 10]. This model of normal user or system behavior is commonly known as the user or system profile. Strength of anomaly detection is its ability to detect previously unknown attack. A typical anomaly detection system is as shown fig 2.

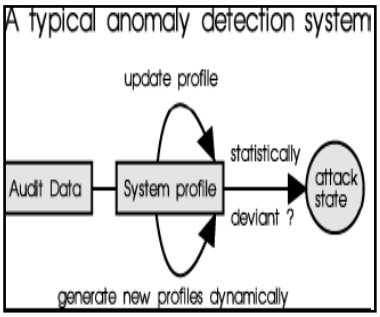

Fig.2. Anomaly detection

The anomaly detection system is effective against novel or unknown attacks. There is no need of prior knowledge about specific intrusions in anomaly detection technique. One of the drawbacks of anomaly detection is the high percentage of false positives [10]. Intrusion detection systems are also categorized according to the kind of input information they analyse. So this is classified into host-based, network-based, wireless and Network Behaviour Analysis (NBA) intrusion detection system [7].

3.3 Host-based intrusion detection

Host-based intrusion detection system analyses host-bound audit sources such as operating system audit trails, system logs, or application logs.

3.4 Network-based intrusion detection

Network-based intrusion detection system analyses network packets that are captured on a network. Network packet is the data source for network intrusion detection system. In the past few years, a growing number of research projects have applied data mining to intrusion detection in network data [10, 14, and 15].

3.5 Wireless Intrusion Detection

Wireless intrusion detection system monitors wireless network traffic and analyses its wireless networking protocols to identify suspicious activity involving the protocols themselves. It cannot identify suspicious activity in the application or higher-layer network protocols such as TCP, UDP that the wireless network traffic is transferring [9].

3.6 Network Behaviour Analysis

Network Behaviour Analysis which examines network traffic to identify threats that generate unusual traffic flows, such as distributed denial of service (DDoS) attacks [6], certain forms of malware such as worms, backdoors, and policy violations e.g., a client system providing network services to other systems. Network behaviour analysis systems are also deployed to monitor flows on an organization's internal Networks, and are also sometimes deployed where they can monitor flows between an organization's Networks and external networks such as the Internet.

4. COMPONENTS OF INTRUSION DETECTION SYSTEM

From the above discussion, intrusion detection is the monitoring and analysing digital data. So typical components used in an intrusion detection system are,

4.1 Sensor or Agent

Sensors and agents monitor and analyse activity. The term sensor is typically used for intrusion detection systems that monitor networks, including network-based, wireless, and network behaviour analysis technologies. The term agent is typically used for host-based intrusion detection system technologies [7, 14].

4.2 Management Server

A management server is a centralized device that receives information from the sensors or agents and manages them. Some of the management servers perform analysis on the event information that the sensors or agents provide and can identify events that the individual sensors or agents cannot. Matching event information from multiple sensors or agents, such as finding events triggered by the same IP address, is known as correlation. Some small intrusion detection system deployments do not use any management servers, but most intrusion detection system deployments management server. In larger intrusion detection system deployments, there are often multiple management servers [7].

4.3 Database Server

A database server is a repository for event information recorded by sensors, agents, or management servers. Many intrusion detection systems have database servers.

4.4 Console

A console is a program that provides an interface for the intrusion detection system's users and administrators. Console software is typically installed onto standard desktop or laptop computers. Some consoles are used for intrusion detection system administration only, such as configuring sensors or agents and applying software updates, while other consoles are used only for monitoring and analysis. Some intrusion detection system consoles provide both administration and monitoring capabilities.

5. 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 detection.

5.1 Data Mining Algorithms for Intrusion Detection

Data mining algorithms can be used for misuse detection and anomaly detection. In misuse detection, training data are labelled as either "normal" or "intrusion." A classifier can then be derived to detect anomalies & known intrusions [10, 12]. Research in this area has included the application of classification algorithms, association rule mining, and cost-sensitive modelling. Anomaly detection builds models of normal behaviour 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 [15]. Anomaly detection research has included the application of classification algorithms, statistical approaches, clustering, and outlier analysis [2, 8, 12, 13, and 15]. The techniques used must be efficient and scalable, and capable of handling network data of high volume, dimensionality, and heterogeneity [11]. Classification algorithm about Data Mining can be used to construct classifier, after the invasion of a large number of data sets being trained [12]. Classifier can be used for intrusion detection. Clustering analysis algorithm can also be used to construct the network model of normal behaviour, or intrusion behaviour model [2, 13]. Association analysis algorithm can be used to describe the invasion of behaviour patterns of association rules, through these rules intrusion detection can come [12].

5.2 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 [11].

5.3 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.

5.4 Distributed Data Mining

Intrusions can be launched from several different locations and targeted to many different destinations. Distributed data mining methods may be used to analyse network data from several network locations in order to detect these distributed attacks [6, 14].

5.5 Visualization and Querying Tools

Visualization tools should be available for viewing any anomalous patterns detected. Such tools may include features for viewing associations, clusters, and outliers. Intrusion detection systems should also have a graphical user interface that allows security analysts to pose queries regarding the network data or intrusion detection results [11]. These are the areas in which data mining technologies are applied and developed for intrusion detection.

6. ACKNOWLEDGMENTS

Intrusion detection system has tremendous demand in this digital era which enables us to detect security violation in information system. Intrusion detection systems based on data mining are generally more precise and require far less manual processing and input from human experts. Different data mining approaches like classification, association rule, clustering, and outlier detection are the few techniques frequently used to analyse network traffic or data to gain knowledge that helps in controlling intrusion. Today the main reason of using Data Mining for intrusion detection systems is the enormous volume of existing and newly appearing network data that will be useful for future pattern generation and recognition in the digital forensics research.

7. REFERENCES

[1] Meng Jianliang, Shang Haikun BianLing"The Application on Intrusion Detection Based on K-means Cluster Algorithm.",International Forum on Information Technology and Applications IEEE, pp150-152, 2009.

[2] Allen, J., Christie, A., Fithen, W., McHugh, J., Pickel, J., and Stoner, E." State of the practice of Intrusion Detection Technologies",Technical report. Camegie Mellon University. http://www.cert.org/archive/pdf/99tr028.pdf, 2000.

[3] Kanwal Garg , Rshma Chawla "Detection Of DDOS Attacks Using Data Mining" International Journal of Computing and Business Research (IJCBR)ISSN (Online) : 229-6166Volume 2 Issue 1, 2011.

[4] U.Fayyad,G.Piatetsky-Shapiro,P.Smyth, "From Data Mining To Knowledge Discovery in Databases", articles in Karen Scarfone and Peter Mell "Guide to Intrusion Detection and Prevention Systems (IDPS) "National Institute of Standards and Technology Special Publication pp 800-94, 2007.

[5] Mannila, H." Data Mining: Machine Learning, Statistics, and Databases." In Proceedings of the 8th International Conference on Sci-entific and Statistical Database Management, pages 1–8,1996.

[6] Foong Heng Wai ,Yin Nwe Aye, Ng Hian James "Intrusion Detection in Wireless Ad-Hoc Networks" CS4274 Introduction to mobile computing, 2004.

[7] Paul Dokas, levent Ertoz, Vipin Kumar, Aleksandar Lazarevic, Jaideep Srivastava, Pang-Ning Tan "Data Mining For Network Intrusion Detection", 2003.

[8] Han Jiawei & Kamber Micheline "Data Mining: Concepts and Techniques"(Second Edition) San Francisco, Morgan Kaufmann Publishers, 2006.

[9] Wenke Lee and Salvatore J. Stolfo "Data Mining Approaches for Intrusion Detection",1998.

[10] Li Bo, Jiang Dong-Dong "The Research of Intrusion Detection Model Based on Clustering Analysis" International Conference on Computer and Communications Security IEEE, 2009.

[11] Imen Brahmi, Sadok Ben Yahia, and Pascal Poncelet"MAD-IDS: Novel Intrusion Detection System using Mobile Agents and Data Mining Approaches.", 2010.