Data Mining Techniques for Intrusion Detection and Prevention System

Ashok Chalak [#1],
ME (CompEngg)II Year student
Computer Engineering,
MGM's College of Engineering
University of Mumbai, India

Naresh D Harale[#2]
Head of Department,
Computer Engineering,
MGM's College of Engineering
University of Mumbai, India

Rohini Bhosale [#3],
ME (CompEngg)II Year student
Computer Engineering, MGM's College of Engineering, University of Mumbai, India.

Abstract:
The main purpose of Intrusion Detection Systems(IDS) and Intrusion protection Systems(IPS) for data mining is to discover patterns of program and user activity, and determine what set of events indicate an attack. In the last years, the networking revolution has finally come of age. More than ever before, we see that the Internet is changing computing as we know it. The possibilities and opportunities are limitless; unfortunately, so too are the risks and chances of malicious intrusions. In Network Security, intrusion detection and prevention system is the act of detecting activity or action that attempt to compromise the confidentiality, integrity or availability of a resource. Intrusion prevention techniques, such as user authentication avoiding programming errors, and information protection (e.g., encryption) have been used to protect computer systems is act as first line of defense. We focus on issues related to deploying a data mining-based IDS in a real time of networking environment. To improve accuracy and security, data mining programs are used to analyze audit data and extract features that can distinguish normal activities from intrusions. In this paper present an architecture consisting of sensors, detectors, a data warehouse, and model generation components and we can identify attack and which type of attack on database take place.

Keywords
Data mining, IDS, IPS, Network security.

I. INTRODUCTION

Security for network system is becoming increasingly important ,as more sensitive information is being stored and send though Internet .Anderson while introducing the concept of intrusion detection in 1980[1],define an intrusion as, "An attempt or a threat to be potential possibility of a deliberate ,

- Unauthorised access information.
- Manipulate information.

Most IDSs are based on signatures that are developed by manual encoding of expert programmer. In intrusion detection (IDS) and prevention system (IPS) match activity on the system being monitored to known signatures of attacks.In data mining based approaches to building detection models for IDSs. Data mining-based systems have the advantage of potentially being able to detect new attacks and prevents the attack on network, our goal is to develop a data mining-based IDS that is capable of outperforming signature-based systems at the tolerated false positive rate. Network intrusion detection deals with information passing on the wire between hosts. Typically referred to as "packet-sniffers," network intrusion detection devices intercept packets traveling along various communication mediums and protocols, usually TCP/IP. Once captured, the packets are analyzed in a number of different ways. Some NID devices will simply compare the packet to a signature database consisting of known attacks and malicious packet "fingerprints", while others will look for anomalous packet activity that might indicate malicious behavior. antivirus software typically detects and protects against malware- and there are IDS that detect based on comparing traffic patterns against a baseline and looking for anomalies[2].

This paper introduces Intrusion detection system(IDS) and intrusion prevention system (IPS) that detect attack based on specific signatures of known threats similar to the way the Network Intrusion Detection System (NIDS), which uses a suite of data mining techniques to automatically detect the attacks against computer networks and systems. While the long-term objective of NIDS is to address all aspects of intrusion detection, in this paper we present two specific contributions:

(i) An unsupervised anomaly detection technique that assigns a score to each network connection that reflects how anomalous the connection is, and

(ii) An association pattern analysis based module that summarizes those network connections that are ranked highly anomalous by the anomaly detection module [3].

In intrusion system used the following method:

- Distributed Denial of Service.
- Viruses and Worms.
- P Spoofing.
- Network/Port Scans.

Manuscript received August 5, 2011
Manuscript revised August 20, 2011

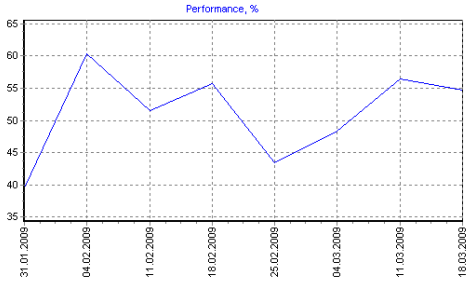

Fig 1.Intrusion Detection and Prevention Scorecard [4].

All IDS need an information source in which to monitor for intrusive behavior. The information source can include: network traffic (packets), host resource (CPU, I/O operations, and log files), user activity and file activity, etc [4].

There are multiple "penetration points" for intrusions to take place in a network system. For example, at the network level, carefully created "malicious" IP packets can destroy a victim host; at the network host level, vulnerabilities in system software can be exploited to yield an illegal root shell. Since activities at different penetration points are normally recorded in different audit data sources, an IDS often needs to be extended to incorporate (additional) modules that specialize on certain components (e.g., hosts, subnets, etc.) of the network systems. The large traffic volume in security related mailing lists and Web sites suggest that new system security holes and intrusion methods are continuously being discovered [6].

The first threat for a computer network system was realised in 1988 when 23-year old Robert Morris launched the first worm, which overid over 6000 PCs of the ARPANET network. On February 7th, 2000 the first DoS attacks of great volume where launched, targeting the computer systems of large companies like Yahoo!, eBay, Amazon, CNN, ZDnet and Dadet. More details on these attacks can be found at [8].

If the network is small and signatures are kept up to date, the human analyst for intrusion detection works well. If some organizations have a large, complex network then intrusion detection become done by the number of alarms and all generate alarm need to review. The sensors on the MITRE network, for example, currently generate over one million alarms per day. And that number is increasing. This situation arises from ever increasing attacks on the network, as well as a tendency for sensor patterns to be insufficiently selective (i.e., raise too many false alarms). Commercial tools typically do not provide an enterprise level view of alarms generated by multiple sensor vendors [9].

Intrusion Detection before Data Mining.

When we start the intrusion detection on our organizations network, that time we didn't focus on data mining, but rather on more issues:

How alarm generated? How much data would we get? How would we show the data? And what type of data we want to monitor or see?

We began to suspect that our system was inadequate for detecting the most dangerous attacks—those performed by adversaries using attacks that are new, stealthy, or both. So we considered data mining with two questions in mind: [9]

- Can we develop a way to minimize what the analysts need to look at daily?
- Can data mining help us find attacks that the sensors and analysts did not find?

Data Mining: What is it?

Data mining is the process of extracting patterns from large datasetbycombiningmethodsfrom statistician artificia l intelligence with database management.

In intrusion detection(IDS)and intrusion prevention system(IPS) we consider some things that are used in data mining for intrusion detection(IDS) and intrusion prevention system(IPS).

- Remove activity from alarm data.
- Identify false alarm generators and attack sensor signatures.
- Identify long, ongoing IP packets.
- Find bad activity.

II. LITERATURE SURVEY

Intrusion detection (IDS) and intrusion prevention system (IPS) a war that must be fought day and night, without rest, on thousands of company and its network. Winning strategies will be as varied as the organizations designing them, but none will succeed without a comprehensive solution for securing the data. Data security depends on a complete but flexible toolset capable of managing, maintaining and securing. The goal of intrusion detection is to monitor network assets to detect anomalous behavior and misuse. This concept has been around for nearly twenty years but only recently has it seen a dramatic rise in popularity and incorporation into the overall information security infrastructure. In current architecture for intrusion detection is shown in Figure 2 network traffic is analyzed by a variety of available sensors. This sensor data is pulled periodically to a central server for conditioning and input to a relational database. HOMER filters events from the sensor data before they are passed on to the classifier and clustering analyses. Data mining tools filter false alarms and identify anomalous behavior in the large amounts of remaining data. A web server is available as a front end to the database if needed, and

analysts can launch a number of predefined queries as well as free form SQL queries from this interface [9].

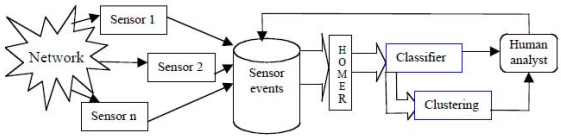

Fig 3 How sensors feed into overall intrusion detection system

In Intrusion detection system (IDS) and intrusion prevention system (IPS) contain some Terminologies:

- Alert/Alarm: A signal show that system has been or is being attacked.
- True Positive: A legitimate attack which triggers an IDS to produce an alarm and show attack has take place
- False Positive: An event generated signal an IDS to produce an alarm when no attack has taken place.
- Alarm filtering: The process of categorizing attack alerts produced from an IDS in order to distinguish false positives from actual attacks.

Feature selection from the available data is vital to the effectiveness of the methods employed. Researchers apply various analysis procedures to the accumulated data, in order to select the set of features that they think maximizes the effectiveness of their data mining techniques. Table I contains some examples of the features selected. Each of these features offers a valuable piece of information to the System. Extracted features can be ranked with respect to their contribution and utilized accordingly [10]

% of same service to same host	# different services accessed
% on same host to same service	# establishment errors
average duration / all services	# FIN flags
average duration /current host	# ICMP packets
average duration / current service	# keys with outside hosts
bytes transfered / all services	# new keys
bytes transfered / current host	# other errors
bytes transfered / current service	# packets to all services
Destination bytes	# RST flags
Destination IP	# SYN flags
Destination port	# to certain services
Duplicate ACK rate	# to privileged services
Duration	# to the same host
Hole rate	# to the same service
Land packet	# to unprivileged services
Protocol	# total connections
Resent rate	# unique keys
Source bytes	# urgent
Source IP	% control packets
Source port	% data packets
TCP Flags	wrong data packet size rate
Timestamp	variance of packet count to keys

Table1 Table of feature that have been applied for (IDS) and (IPS) for data miming.[10].

Data Mining based IDS generally not give good result in a simulated environment and then deployed in a real environment. They generate a lot of false alarms and positive alarms.

Current weaknesses in intrusion detection and intrusion prevention system include new attacks are not detected until someone has generated a rule or signature for that specific attack. Also, most attacks need a alteration in existing malicious code in order to bypass existing signatures. Hence, new signatures are generally created manually.

III. METHODOLOGY

In our paper we will propose the following methods for intrusion detection and intrusion prevention system for data miming. Data Mining may be thought of as the most interesting one in accomplishment of intrusion detection and intrusion prevention system. In IDS and IPS, Data Mining used for to discover consistent and useful patterns of system features that describe user behavior. In intrusion detection and intrusion prevention system can be two types.

- Misuse-based system.
- Anomaly-based system

Misuse-based	Anomaly-based
The attacks uncovered under this are assumed to be true positives.	The normal packets separated under this are assumed to be true negatives.
It risks high porosity towards new and undefined attacks.	It risks the chances of normal but undefined packets to be tagged as abnormal data.
It has a chance of failure to capture many attacks.	It has a tendency to show greater number of false positives.

Table 2 comparison among Misuse-based and Anomaly-based.

Thus we can introduce INIDS (Integrated NIDS). Not only will INIDS be an integrated system which uses both misuse-based and anomaly based approaches, but it also implements a classification rules again on the data.

Data Mining-based intrusion detection systems have demonstrated high accuracy, good generalization to novel types of intrusion, and robust behavior in a changing environment, In Figure 4 we depicted (Peietal.: Data Mining Techniques for Intrusion Detection and Computer Security)[11].

The intrusion detection and intrusion prevention system is an integrated system which uses both misuse-based and anomaly based approaches. Data mining techniques that are used for intrusion detection and intrusion prevention system are as following,

The classification rules used to discover attacks in a TCPdump. These classification rules used to accurately capture the behavior of intrusions and normal activities for data mining system. The classification rule that we use is the decision tree.

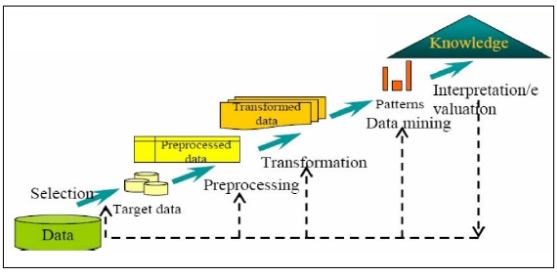

Fig 4. Data Mining Phases.

a) Classification rules

Decision Tree: Decision tree induction is the learning of decision trees from class-labeled training tuples. A decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node holds a class label.

The topmost node in the tree is the root node. To decide which attributes will decide how the tree should form we need an attribute selection measure. The method that we use is called information gain. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends.

For example, a classification model can be built to categorize bank loan applications as either safe or risky. In other word, classification maps a data item into one of several pre-defined categories. These algorithms normally output "classifiers". A prediction model can be built to predict the expenditures of potential customers on computer equipment given their income and occupation. An ideal application in intrusion detection would be to gather sufficient "normal" and "abnormal" audit data for a user or a program, then apply a classification algorithm to learn a classifier that can label or predict new unseen audit data as belonging to the normal class other abnormal class;

b) Knowledge Discovery in Databases (KDD)

KDD can be defined as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Data mining is a particular step in this process in which specific algorithms are applied to extract patterns from data.

The KDD process involves a number of steps and is often interactive, iterative and user-driven [12].

- Getting to know the application domain: trying to understand the data and the discovery task.
- Data Mining: includes first deciding what model, for example, summarization, classification, or clustering is to be derived from the data.
- Using the discovered knowledge: includes incorporating the knowledge into a production system, or simply reporting it to interested parties.

IV. CONCLUSIONS

In this paper we first identify the type of attack take place on network or data base. The classification rules can be used for intrusion detection (IDS) and intrusion prevention system (IPS) to classify the attack and signature. Many possibilities have been considered, even the incorporation of artificial intelligence. We have shown the ways in which data mining has been known to aid the process of Intrusion Detection and the ways in which the various techniques have been applied for intrusion detection (IDS) and intrusion prevention system (IPS).

REFERENCES

[1] J.P Anderson "Computer Security Threat Monitoring and Surveillance. Technical report"

[2] Bradley, T. (n.d.). ,"Introduction to Intrusion Detection Systems(IDS") [Online] Available on http://netsecurity.about.com/cs/hackertools/a/aa030504.html.

[3] "Network Intrusion Detection System (NIDS) Using Data Mining Techniques" [Online] Available on http://etrx.spit.ac.in/ieee_colloquium/Information_Security/spit-265.pdf.

[4] Intrusion Detection and Prevention Scorecard [Online] Available on http://www.strategy2act.com/solutions/scorecard-reports/bsc_intrusion_prevention.html.

[5] What is an IDS? [Online] Available on http://www.idstutorial.com/what-is-ids.php.

[6] "A Data Mining Framework for Building Intrusion Detection Models1".[Online] Available on, http://citeseerx.ist.psu.edu/viewdoc/download.

[7] "Network intrusion detection (IDS) and intrusion protection system (IPS)" [Online] Available on http://www.cdacbangalore.in/design/corporate_site/override/pdf-doc/projects/GYN.pdf.

[8] Available on http://www.securityfocus.com/news/2445.

[9] Eric Bloedorn, Alan D. Christiansen, William Hill,Clement Skorupka, Lisa M. Talbot, Jonathan TivelThe MITRE Corporation."Data Mining for Network Intrusion Detection: How to Get Started"

[10] Data Mining Techniques for (Network) IntrusionDetection Systems Theodoros Lappas and Konstantinos Pelechrinis [Online] Available on http://www.mendeley.com/research/data-warehousing-and-data-mining-techniques-for-intrusion-detection-systems/

[11] Mohammad hossein haratian " An architectural design for a hybrid intrusion detection system for data base ",. [Online] Available on www.citeseerx.ist.psu.edu/viewdoc/download.

[12] "A Data Mining Framework for Constructing Features and Models for Intrusion DetectionSystems".,Wenke Lee,[Online] Available on. http://portal.acm.org/citation.cfm?id=929987&dl=GUIDE&coll=GUIDE