Volume 1, Issue 2, April 2014 ISSN: 2348-8565 (Online)
International Journal of Modern Engineering and Research Technology
Website: http://www.ijmert.org Email: editor.ijmert@gmail.com

Type of Attack in Computer Network using Intrusion Detection System with Data Mining Techniques – A Survey

Balram Purswani
Ph.D. Scholar
Mewar University
Chittorgarh (Rajashthan)
Email: balpurg@gmail.com

Dr. Samar Upadhyay,
Associate Professor
Head Computer Application,
Government Engg. College,
Jabalpur (M.P.) [INDIA]

Abstract—*An* ***Intrusion Detection System (IDS)*** *is a device or software application that monitors network or system activities for malicious activities or policy violations and produces reports to a management station. Some systems may attempt to stop an intrusion attempt but this is neither required nor expected of a monitoring system. Intrusion detection and prevention systems (IDPS) are primarily focused on identifying possible incidents, logging information about them, and reporting attempts. In addition, organizations use IDPSes for other purposes, such as identifying problems with security policies, documenting existing threats and deterring individuals from violating security policies. IDPSes have become a necessary addition to the security infrastructure of nearly every organization. They search for potential malicious abnormal activities on the network traffics; they sometimes succeed to find true network attacks and anomalies (true positive). However, in many cases, systems fail to detect malicious network behaviors (false negative) or they fire alarms when nothing wrong in the network (false positive). Hence applying Data Mining (DM) techniques on the network traffic data is a potential solution that helps in design and develop a better efficient intrusion detection systems. Data mining methods have been used build automatic intrusion detection systems. The central idea is to utilize auditing programs to extract set of features that describe each network connection or session, and apply data mining programs to learn that capture intrusive and non-intrusive behavior. In addition, Network Performance Analysis (NPA) is also an effective methodology to be applied for intrusion detection. Here, we discuss DM and NPA Techniques for network intrusion detection and propose that an integration of both approaches have the potential to detect intrusions in networks more effectively and increases accuracy[3].*

Keywords:—*Intrusion Detection, Network Intrusion Detection System, Data Mining Techniques, Network Performance Analysis.*

I. TERMINOLOGY

Burglar Alert/Alarm: A signal suggesting that a system has been or is being attacked.

Detection Rate: The detection rate is defined as the number of intrusion instances detected by the system (True Positive) divided by the total number of intrusion instances present in the test set.

False Alarm Rate: defined as the number of 'normal' patterns classified as attacks (False Positive) divided by the total number of 'normal' patterns.

True Positive: A legitimate attack which triggers an IDS to produce an alarm.

False Positive: An event signaling an IDS to produce an alarm when no attack has taken place.

False Negative: A failure of an IDS to detect an actual attack.

True Negative: When no attack has taken place and no alarm is raised.

Noise: Data or interference that can trigger a false positive or obscure a true positive.

Site policy: Guidelines within an organization that control the rules and configurations of an IDS.

Site policy awareness: An IDS's ability to dynamically change its rules and configurations in response to changing environmental activity.

Confidence value: A value an organization places on an IDS based on past performance and analysis to help determine its ability to effectively identify an attack.

Alarm filtering: The process of categorizing attack alerts produced from an IDS in order to distinguish false positives from actual attacks.

Attacker *or* ***Intruder:*** An entity who tries to find a way to gain unauthorized access to information, inflict harm or engage in other malicious activities.

Masquerader: A user who does not have the authority to a system, but tries to access the information as an authorized user. They are generally outside users.

Misfeasor: They are commonly internal users and can be of two types:

1. An authorized user with limited permissions.
2. A user with full permissions and who misuses their powers.

Clandestine user: A user who acts as a supervisor and tries to use his privileges so as to avoid being captured.

II. Problem Statement

The rapid development of computer networks and mostly of the Internet has created many stability and security problems such as intrusions on computer and network systems. Further the dependency of companies and government agencies is increasing on their computer networks and the significance of protecting these systems from attacks is serious because a single intrusion can cause a heavy loss or the consistency of network becomes insecure. During recent years number of intrusions has dramatically increased. Therefore it is very important to prevent such intrusions. The hindrance of such intrusions is entirely dependent on their detection that is a key part of any security tool such as Intrusion Detection System (IDS), Intrusion Prevention System (IPS), Adaptive Security Alliance (ASA), checkpoints and firewalls. Hence accurate detection of network attack is imperative. A variety of intrusion detection approaches are available but the main problem is their performance, which can be enhanced by increasing the detection rates and reducing false positives.

III. Intrusion Detection System

Intrusion detection is defined as the process of intelligently monitoring the events occurring in a computer system or network, analyzing them for signs of violations of security policy. The primary aim of Intrusion Detection System (IDS) is to protect the availability, confidentiality and integrity of critical networked information systems. Intrusion Detection Systems are an important component of defensive measures protecting computer systems and networks from abuse. When an IDS is properly deployed it can provide warnings indicating that a system is under attack. It is critical for intrusion detection in

order for the IDS to achieve maximal performance.[4]

An intrusion detection system can be described at a very macroscopic level as a detector that processes information coming from the system to be protected. This detector can also launch probes to trigger the audit process, such as requesting version numbers for applications. It uses three kinds of information: long-term information related to the technique used to detect intrusions (a knowledge base of attacks for example), configuration information about the current state of the system, and audit information describing the events that are happening to the system. The role of the detector is to eliminate unneeded information from the audit trail. It then presents either a synthetic view of the security-related actions taken during normal usage of the system, or a synthetic view of the current security state of the system. A decision is then taken to evaluate the probability that these actions or this state can be considered as symptoms of an intrusion or vulnerabilities. A countermeasure component can then take corrective action to either prevent the actions from being executed or change the state of the system back to a secure state.

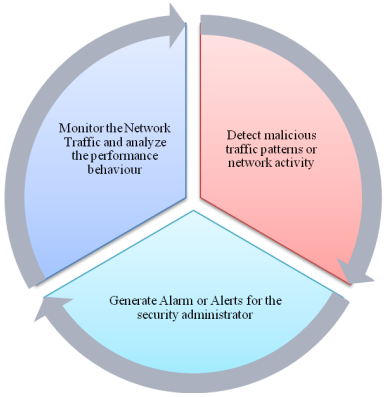

Figure 1. Traditional Network IDS

1. ***Classification of Intrusion Detection:*** Intrusions Detection can be classified into two main categories. They are as follow:

***A. Host Based Intrusion Detection*:** HIDSs evaluate information found on a single or multiple host systems, including contents of operating systems, system and application files.

B. Network Based Intrusion Detection: NIDSs evaluate information captured from network communications, analyzing the stream of packets which travel across the network.

2. Components of Intrusion Detection System: An intrusion detection system normally consists of three functional components. The first component of an intrusion detection system, also known as the event generator, is a ***data source***. Data sources can be categorized into four categories namely Host-based monitors, Network-based monitors, Application-based monitors and Target-based monitors. The second component of an intrusion detection system is known as the ***analysis engine***. This component takes information from the data source and examines the data for symptoms of attacks or other policy violations. The analysis engine can use one or both of the following analysis approaches:

A. Misuse/Signature-Based Detection: This type of detection engine detects intrusions that follow well-known patterns of attacks (or signatures) that exploit known software vulnerabilities. The main limitation of this approach is that it only looks for the known weaknesses and may not care about detecting unknown future intrusions.

B. Anomaly/Statistical Detection: An anomaly based detection engine will search for something rare or unusual. They analyses system event streams, using statistical techniques to find patterns of activity that appear to be abnormal. The primary disadvantages of this system are that they are highly expensive and they can recognize an intrusive behavior as normal behavior because of insufficient data

C. The third component of an intrusion detection system is the ***response manager***. In basic terms, the response manager will only act when inaccuracies (possible intrusion attacks) are found on the system, by informing someone or something in the form of a response.

IV. NETWORKING ATTACKS

This section is an overview of the four major categories of networking attacks. Every attack on a network can comfortably be placed into one of these groupings.[2](Figure. 2)

1. Denial of Service (DoS): A DoS attack is a type of attack in which the hacker makes a computing or memory resources too busy or too full 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. are all DoS attacks.

2. 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, which s/ he 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.

3. 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.

4. 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. saint, portsweep, mscan, nmap etc.

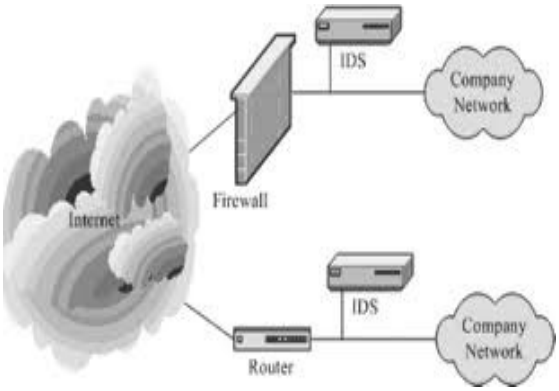

Figure 2. Simple Network

V. DATA MINING TECHNIQUES FOR NETWORK INTRUSION DETECTION

Many researchers have investigated the deployment of data mining algorithms and techniques for intrusion detection. Examples of these techniques include.[1][5]

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. 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 valuate 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.

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. In misuse detection, network traffic data are collected and labeled as "normal" or "intrusion". This labeled 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 behavior 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. Data may be non sequential or sequential in nature. Non-sequential data are those data where order of occurrence is not important, while sequential data are hose data where the order of occurrence with respect to time is important to consider. Using data mining and specially classification techniques can play a very important role on two dimensions; the similarity measures and the classification schema. Temporal data can be classified into discrete temporal sequential data such as logs time or continuous temporal sequential data such as observations.

Clustering Analysis: Clustering assign objects (intrusions) to groups (clusters) on the basis of distance measurements made on the objects. As opposed to classification, clustering is an unsupervised learning process since no information is available on the labels of the training data. In anomaly detection, clustering and outlier analysis can be used to drive the ID model. Distance or similarity measure plays an important role in grouping observations in homogeneous clusters. It is important to formulate a metric to determine whether an event is deemed normal or anomalous using measures.

Association and Correlation Analysis*:* The main objective of association rule analysis is to discover association relationships between specific values of features in large datasets. This helps discover hidden patterns and has a wide variety of applications in business and research.

Association rules can help select discriminating attributes that are useful for intrusion detection. It can be applied to find relationships between system attributes describing network data. New attributes derived from aggregated data may also be helpful, such as summary counts of traffic matching a particular pattern.

Stream Data Analysis*:* Intrusions and malicious attacks are of dynamic nature. Moreover, data streams may help detect intrusions in the sense that 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 perform intrusion detection in data stream, real-time environment. This helps identify sequences of events that are frequently encountered together, find sequential patterns, and identify outliers. Other data mining methods for finding evolving clusters and building dynamic classification models in data streams can be applied for these purposes.

Distributed Data Mining: Intruders can work from several different locations and attack many different destinations. Distributed data mining methods may be utilized to analyze network data from several network locations, this helps detect distributed attacks and prevent attackers in different places from harming our data and resources.

Visualization and Querying Tools: Visualization data mining tools that include features to view classes, associations, clusters, and outliers can be used for viewing any anomalous patterns detected. Graphical user interface associated with these tools allows security analysts to understand intrusion detection results, evaluate IDS performance

and decide on future enhancements for the system.

VI. CONCLUSION

This research is centered on "Developing an algorithm to implement efficient intrusion - detection system". The outcome of the proposed work will yield the expected result and fulfill the following objectives:

1. Algorithm to implement Intrusion detection system based on Data Mining Technique.
2. Rationally Reduce number of false positive alarms.

REFERENCES :

[1] M. M. Pillai, J. H. P. Eloff, H. S. Venter, "An Approach to Implement a Network Intrusion Detection System using Genetic Algorithms", Proceedings of SAICSIT, pp:221-228, 2004.

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

[3] Dartigue, C.; Hyun Ik Jang; Wenjun Zeng;" A New Data-Mining Based Approach for Network Intrusion Detection" Seventh Annual Communication Networks and Services Research Conference (CNSR), 11-13 May 2009

[4] Tanase, Matthew, " One of These Things is not Like the Others: The State of Anomaly Detection", http://en.wikipedia.org/wiki/Intrusion_detection_syste 2010, http://www.symantec.com/connect/articles/one-these-things-not-others-state-anomalydetectio

[5] M. Hossain "Data Mining Approaches for Intrusion Detection: Issues and Research Directions", http://www.cse.msstate.edu/~bridges/papers/iasted.pd

* * * * *