International Journal of Advanced Research in Applied Science and Technology ISSN: xxxx: xxxx
Vol.1, No.1, May 2016

ONLINE INTRUSION ALERT AGGREGATION AND PREVENTION USING GENERATIVE DATA STREAM MODELING

$ PriyaVarshini G R *Subhashree R #Parkavi M
Department of Information Technology
Dr. Mahalingam College of Engineering and Technology
Pollachi
$datchu13031994@gmail.com *shreesubha1208@gmail.com #parkavi@gmail.com

ABSTRACT

The performance of a Network can be improved by increasing the throughput, Bandwidth and by reducing the Network load. The Intrusion Detection System (IDS) can be used to detect attack at an instance and to produce alerts. Number of alerts produced at a single instance will cause the Network traffic. Here, the traffic is overcome by producing Meta-alerts for similar attack type. The attacks will be classified, then the similar attacks will be clustered together and Meta-alerts will be produced. The alerts will be clustered using server log file based on clustering algorithm. The malicious attacks will be detected and will be blocked in the Network coupled with Firewalls to prevent further attacks.

Keywords–Intrusion Detection, Alert Aggregation, Meta-alerts

INTRODUCTION

Nowadays with the wide spread use of internet all over the world, Security has an inevitable requirement in providing the network security to the whole network. There are different types of threats that are affecting the network. Firewall is software that acts as a bridge between the network and the system to avoid the attacks that are entering into the network. All the attacks cannot be detected using the firewall; some malicious attacks will enter the network. Intrusion detection (IDS) system is a device or a software application that monitors network or system activities for malicious threats and reports to the system. They can defend various threats and attacks by continuously monitoring the actions of the attackers entering into the host or network. The IDS can detect the attacks by using the TCP/IP connections or server log files. The server log files can be found by using TCP Dump, Wire Shark etc...The single attack instance is the occurrence of attack by an attacker at a certain point in time. The attack instance will results in thousands of alerts. The similar attacks can be grouped together by a technique called Alert Aggregation and meta-alerts must be generated for the clusters. The alerts can be sorted based on the source, destination, and attack type.

INTRUSION DETECTION

It is the act of identifying intruders on the network. This can be at many different levels, at the network (point of entry) firewall, at any point in the network or at a specific host. It is part of knowing what is happening on your network, Intruders can cause

harm to the general health of the network. The obvious reason for doing intrusion detection is to detect suspicious activity on the systems. Intrusion detection can have the side effect of saving bandwidth, especially for known signatures of typical bandwidth wasting software like p2p. Virus and worms have pretty distinct patterns which can usually easily be detected with IDS which can save bandwidth. Network worms and virus can bring a network to a halt and are one of the first problems that can be addressed with IDS. Security is a big issue for all networks in today's enterprise environment. Hackers and intruders have made many successful attempts to bring down high-profile company networks and web services. Many methods have been developed to secure the network infrastructure and communication over the Internet, among them the use of firewalls, encryption, and virtual private networks. Intrusion detection is a relatively new addition to such techniques. Using intrusion detection methods, we can collect and use information from known types of attacks and find out if someone is trying to attack your network or particular hosts. The information collected this way can be used to harden your network security, as well as for legal purposes. To evaluate the efficiency of an intrusion-detection system three parameters required are Accuracy, Performance and Completeness. There are some additional parameters also fault tolerance, timeliness.

FIREWALLS

A **firewall** is an integrated collection of security measures designed to prevent unauthorized electronic access to a networked computer system. A network firewall is similar to firewalls in building construction, because in both cases they are intended to isolate "network" or "compartment" from another. Firewall inspects traffic through the Network. It allows traffic specified in the policy. It can be a Can be hardware or software Ex. Some routers come with firewall functionality- ipfw, ipchains, pf on Unix systems, Windows XP and Mac OS X have built in firewalls. To protect private networks and individual machines from the dangers of the greater Internet, a firewall can be employed to filter incoming or outgoing traffic based on a predefined set of rules called firewall policies.

ATTACK CLASSIFICATION

Networks are prone to Attacks which causes the major impact. Common internet attacks methods are broken down into categories. Some attacks gain system knowledge or personal information, such as eavesdropping and phishing. Attacks can also interfere with the system's intended function, such as viruses, worms and Trojans. The other form of attack is when the system's resources are consumes uselessly, these can be caused by denial of service (DOS) attack. Here we have categorized the five types of attacks that spread across the network. The categories are DOS, U2R, R2L, probe and normal

.They can be classified based on their characters and attributes. The denial-of-service attack prevents normal use of your computer or network by valid users. Some of the DOS attacks are smurf, Neptune, land, pod, port sweep etc. After entering into the network the attacker may cause the abnormal termination of the application that exists, may overload the system or network by continuously flood it with traffic until the shutdown occurs. These attacks can harm the system resources and causes the performance degradation across the network.

PROPOSED SYSTEM

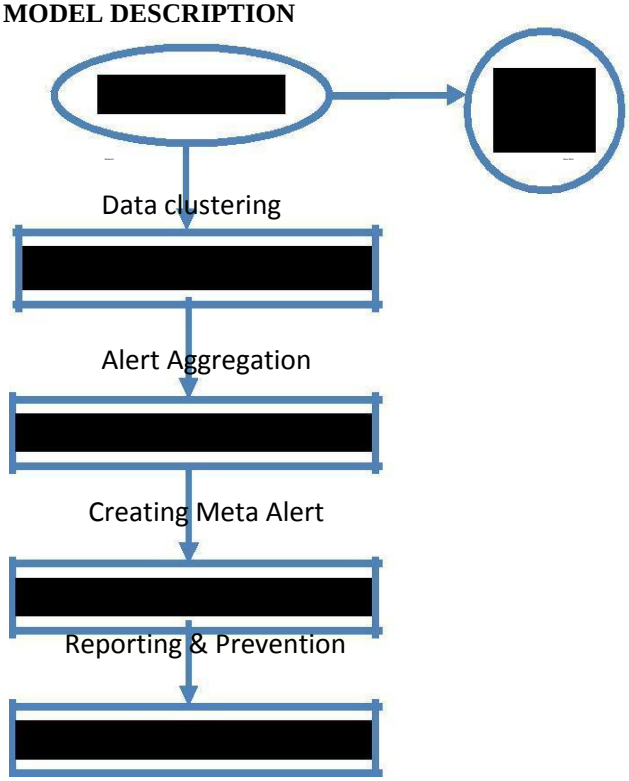
Mostly IDS and monitoring networks enables us to quickly detect and react to unauthorized access. IDSs are used in order to stop attacks, recover from them with the minimum loss or analyze the security problems. They help to defend against the various threats to which networks and hosts are exposed to by detecting the actions of attackers or attack tools in a network or host-based manner. IDS usually focus on detecting attack types and create alerts at each instance of attack.

This results in enormous amount of alerts which results in network traffic. The proposed system has the advantage of reducing alert rate by producing 1meta-alerts for same attack type. The IDS can detect the attacks by using the TCP/IP connections or server log files. The server log files can be found by using TCP Dump, Wire Shark etc...The log file may contain the details of source IP, Destination IP, Port address, and the time of occurrence. The type of attacks can be identified by generating the class label for the similar type of attacks. The type of attacks can be classified and clustered as a group by using the K-means clustering Algorithm. The quite simple sorting of alerts, can be made according to their source, destination, and attack type.

The single attack instance is the occurrence of attack by an attacker at a certain point in time. The attack instance will results in thousands of alerts at a time and can be grouped based on clustered dataset. Here, the unauthorized access can be prevented by blocking the corresponding IP. The working of the proposed system can be described in the Figure.

The advantages of the proposed system are the alert duplication is avoided by clustering the same type of attacks. It is dynamic where the alerts can be produced at a particular instance which has been initiated by the attacker can be grouped based on the type. It reduces the Network load and increases its performance by reducing the number of attacks produced at a time for the same attack type.

MODEL DESCRIPTION



DATASET GATHERING

The real time data can be collected from server log files or by using the wire shark, TCP dump. Here, the data is allocated by using wire shark by continuously monitoring the network flow and transfer of data. The following are the attributes which defines the occurrence of attack and other instances .The attributes are source IP, Destination IP, Source port, destination port and the date and time of occurrence. Based on this the type of attack can be grouped.

DATA PREPROCESSING & CLUSTERING

Data preprocessing is mainly performed to remove the unfilled data columns and irrelevant data, Missing Values can be ignored by deleting the corresponding rows and to identify the classes of attacks with their occurrences.

Clustering is to group data points close or similar to each other. Clustering is unsupervised or undirected process. Search for groups or clusters of data points (records) that are similar to one another will be done during the process. A good clustering

algorithm should produce high quality clusters with high intra-class similarity and low inter-class similarity. The quality of a clustering result depends on the similarity measure used and implementation of the similarity measure.

Here, the similar type of attacks can be clustered by using **K-means clustering Algorithm**

1. It accepts the number of clusters to group data into, and the dataset to cluster as input values.

2. It then creates the first K initial clusters (K= number of clusters needed) from the dataset by choosing K rows of data randomly from the dataset. *For Example*, if there are 10,000 rows of data in the dataset and 3 clusters need to be formed, then the first K=3 initial clusters will be created by selecting 3 records randomly from the dataset as the initial clusters. Each of the 3 initial clusters formed will have just one row of data.

3. K-Means re-assigns each record in the dataset to only one of the new clusters formed. A record or data point is assigned to the nearest cluster using a measure of distance.

4. Stable clusters are formed when new iterations or repetitions of the K-Means clustering algorithm does not create new clusters as the cluster center or Arithmetic Mean of each cluster formed is the same as the old cluster center.

The characteristics of k-means are Top-down approach and partitioning algorithm

RESULTS AND DISCUSSIONS

In the existing system the aggregation of alerts cannot be done where the alert can be produced for each attack instances. To aggregate the similar type of attacks certain attributes such as source IP, destination IP and port address will be considered. The type of attacks can be identified by applying the clustering algorithm. Here, the k means algorithm is used for the data clustering .The k denotes the number of clusters where it can be assigned based on the clusters instances. Here, we have the data obtained after the pre-processing and the type of attacks and their instances are categorized.

REFERENCES

- T.Pietraszek, ”Alert classification to reduce False positives in Intrusion Detection”,2006
- M.Halkidi, Y.Batistakis, and M.Vazirgiannis,”On clustering validation technique “,2001
- F.Autrel and F.Cuppens,’using and Intrusion Detection Alert similarity operator to aggregate and fuse Alerts”,2005
- Bi-Ru Dai, Jen-Wei Huang, Mi-Yen Yeh, and Ming-Syan Chen,” Adaptive Clustering for Multiple Evolving Streams”,2006
- Jo.ao B. D. Cabrera1, Carlos Guti´errez and Raman K. Mehra,” Infrastructures and

Algorithms for Distributed Anomaly-Based Intrusion Detection in Mobile Ad-hoc Networks"

- Vincent S. Tseng, *Member, IEEE*, and Ching-Pin Kao," A Novel Similarity-Based Fuzzy Clustering Algorithm by Integrating PCM and Mountain Method",2007

- Fenye Bao, Ing-Ray Chen, MoonJeong Chang," Trust-Based Intrusion Detection in Wireless Sensor Networks",2011

- S. Axelsson, "Intrusion Detection Systems: A Survey and Taxonomy,"
  Technical Report 99-15, Dept. of Computer Eng.,Chalmers Univ. of Technology, 2000