Abstract

It’s observed that all the existing intrusion detection system’s accuracy is low. And lots of alerts produced by them are false alerts which decrease its efficiency. It’s also observed that all existing intrusion detection systems focus on low-level attacks, and generate isolated alerts. In this proposed research different data mining techniques are used to reduce false alarm rate in intrusion detection system and improve its efficiency and they are K-Nearest Neighbor, K-Means and Decision Table based approaches for anomaly detection. Proposed research operates on the KDD-99 Dataset; which is used worldwide for evaluating the performance of different intrusion detection systems. In the proposed system first the clustering can be applied on the KDD’99 dataset. Then it can be classified into four categories: U2R, R2L, Denial of Service and Probe. The main goal of this proposed system is to reduce the false alarm rate of IDS and try to improve the efficiency.

1. Introduction

As we know that Internet plays a vital role in day today’s life and all the organizations are became online so the importance of securing data from the Internet is also increasing. If a system is embraced for even a small time, it could lead to huge losses to the organization. Everyday new techniques and tools are developed to stop these malicious attempts to access or corrupt the data. To stop the intrusion attempts by an attacker mostly firewall has been used. But firewalls are having some fixed configurations they only block attacks based on source, destination ports and IP addresses which is not sufficient to provide security from all the attacks. Therefore, we need such a type of Intrusion Detection system, which could analyze the load of the packet to detect these attacks. [1, 2, 3 & 4]
impulse of the work is to design a system which acts as a mediator between the user and the operations to achieve security goals.

Mostly People use the intrusion detection system in order to recognized attacks in network-based system. The operations include bunch of rules to identify the attacks and to reach and read personal files that is located in personal computer or the owner would like to send somewhere. Computers connected directly to the Internet are subject to relentless probing and attack. While protective measures such as safe configuration, up-to-date patching, and firewalls are all prudent steps they are difficult to maintain and cannot guarantee that all vulnerabilities are shielded. IDS provides defense in depth by detecting and logging hostile activities. An IDS system acts as "eyes" that watch for intrusions when other protective measures fail [7]. Major concern of this paper is to improve Intruder Detection and to examine the potential of how the IDS might support with Proposed IDS to accomplish this. The major object of the proposed effort is to suggest a new effective intrusion detection system which combines different functionality. The proposed Intrusion detection system affects the implementation performance and analysis of security. The concept of security and the word intrusion detection system might be threatening and complex.

2. Literature Survey

An intrusion detection system (IDS) inspects all inbound and outbound network activity and identifies doubtful patterns that may indicate a network or system attack from someone attempting to break into or compromise a system. Basically Intrusion detection (ID) is a type of security management system for computers and networks. An ID system gathers and analyzes information from various areas within a computer or a network to identify possible security breaches, which include both intrusions (attacks from outside the organization) and abuse (attacks from within the organization). IDS uses vulnerability assessment (sometimes referred to as scanning), it is a technology, which is developed to assess the security of a system or network [Onm 1].

An important problem of Intrusion Detection is how to effectively divide the normal behavior and the abnormal behavior from a large number of raw data’s attributes, and how to effectively generate automatic intrusion rules after collected raw network data. To accomplish this, various data mining algorithms must be studied, such as correlation analysis of data mining algorithms, sequence analysis of data mining algorithms, classification of data mining algorithms, and so on [Hand 10].

Intrusion Detection System (IDS) have become an important building block of any sound defense network infrastructure. Malicious attacks have brought more adverse impact on the network than before increasing the need for effective approach to detect and identify such effects more effectively. Naive Bayes is one of the classification models that predicts very fast due to the less complexity functioning of it. Fast prediction is also the reason for a lot work done in recent years using Bayesian approach. In [Hari 1] a new hybrid model has suggested that ensembles Naive Bayes (statistical) and Decision Table Majority (rule based) approaches.

In [Subramanian 2] authors have discussed on network security through Intrusion Detection Systems (IDSs). We have already known that IDS most efficient technique against network attacks since they allow network administrator to detect policy violations. However, traditional IDSs are vulnerable to original and novel malicious attacks. Also, it is very inefficient to analyze from a large amount volume data such as possibility logs. In addition, there are high false positives and false negatives for the common IDS’s. Furthermore in this paper authors have discussed also on data mining technique and how its help full in IDS system. Thus, how to integrate the data mining techniques into the intrusion detection systems has become a hot topic recently. Herr, authors presented the whole techniques of the IDS with data mining approaches in details.
In [Vidit 3] author discussed on Intrusion Detection System (IDS) where IDS is the most important technique to achieve higher security in detecting unknown/malicious/ abnormal activities for a couple of years. Anomaly detection is one of intrusion detection system. Current anomaly detection is often associated with high false alarm with moderate accuracy and detection rates when it’s unable to detect all types of attacks correctly. To overcome this problem, authors have suggested a hybrid learning approach. In this approach they have combine two different technique one is K-Means clustering and second is Naïve Bayes classification. In this authors have used clustering technique of all data into the corresponding group before applying a classifier for classification purpose. Authors have performed experiment using KDD Cup ’99 dataset. Result show that the presented approach performed better in term of accuracy, detection rate with reasonable false alarm rate.

In [Chandolikar 4] an algorithm for adaptive network intrusion detection using naïve Bayesian classifier and decision tree is presented, which performs balance detections and keeps false positives at acceptable level for different types of network attacks, and eliminates redundant attributes as well as contradictory examples from training data that make the detection model complex. The presented algorithm also addresses some difficulties of data mining such as handling continuous attribute, dealing with missing attribute values, and reducing noise in training data. Due to the large volumes of security audit data as well as the complex and dynamic properties of intrusion behaviors, several data mining based intrusion detection techniques have been applied to network-based traffic data and host-based data in the last decades.

In [Virendra 5] I have study that authors evaluated the performance of various rule based classifiers like: JRip, RIDOR, NNe and Decision Table using ensemble approach in order to build an efficient network intrusion detection system. Basically they are using KDDCup’99 intrusion detection benchmark dataset (which is a part of DARPA evaluation program) for experimentation.

In [Muda 7] I have analyzed and describe an adaptive network intrusion detection system that uses two-stage architecture. In the first stage a probabilistic classifier is used to detect potential anomalies in the traffic. In the second stage a HMM based traffic model is used to narrow down the potential attack IP addresses. Various design choices that were made to make this system practical and difficulties faced in integrating with existing models are also described.

In [Dewan 8] presented a hybrid IDS by integrated signature based (Snort) with anomaly based (Naïve Bayes) to enhance system security to detect attacks. This research used Knowledge Discovery Data Mining (KDD) CUP 99 dataset and Waikato Environment for Knowledge Analysis (WEKA) program for testing the proposed hybrid IDS. Accuracy, detection rate, time to build model and false alarm rate were used as parameters to evaluate performance between hybrid Snort with Naïve Bayes, Snort with J48graft and Snort with Bayes Network.

3. Proposed Work

3.1 Proposed Technique

In this section we are presenting the overall idea on a new proposed concept for intrusion detection system, which will increase the efficiency as compared to the existing intrusion detection system. The proposed concept is using different data mining techniques. Data mining has been applied in many different fields such as network management, advertising, scam detection, process control, and business.

From some few years onwards, a number of research techniques have been applied on data mining to solve various problems in intrusion detection. Data mining will be used for inconsistency detection field of intrusion detection. Newly, it is not possible for different systems to hassle security for network intrusions. Certainty is that there is no unspoiled approach to duck or shield intrusions.
from various events, it is very important to detect or identify them at the preliminary step of occurrence and take essential or required actions for plummeting or decreasing the destruction. One of the best ways to handle doubtful behaviors within a network is IDS. For intrusion detection, lots of techniques have been useful some of them are as soft computing technique, artificial intelligence technique and data mining technique. Most of the data mining techniques like, grouping, association rule mining and classification have been functional on intrusion detection, where pattern mining and classification is the significant technique.

### 3.2 Proposed concept

Proposed concept is going to be present here. As shown in figure 1 we are getting the general idea behind this, which will improve efficiency as compare to existing intrusion detection system. The proposed concept is using data mining techniques in hybridized way. In this K-Means data mining technique has applied for abnormality detection field of intrusion detection. Inconsistency learning technique is capable to identify destructions with high accuracy and to get large detection rates. On the other side, false alarm rate using abnormality technique equally elevated. To maintain the elevated detection rate and accuracy even as at the same time to decrease the false alarm rate, the proposed technique is the combination of different learning approaches. For the first step in the proposed technique, data instances based on their similar behaviors are assembled together by utilizing a K-Means clustering as a pre-classification component. In the next step, the consequential clusters are classified into attacks classes as using K-nearest-neighbor classifier technique. In this step whatever data that has been misclassified throughout the previous phase might be appropriately classified in the consequent classification phase. At the last Decision Table Majority rule based approach is applied. Next are the proposed IDS, which divided into subsequent module:

1) Database Creation (Suggested Technique)
   i. Selecting and generating the data source (KDD 99’)
   ii. Data scope alteration and pre-processing

2) Data mining Techniques
   i. K-Means (Cluster Technique)
   ii. K-Nearest (Classification)
   iii. Decision Tree Based Approach

3) Proposed System
   i. K-Means with K-Nearest Neighbor and Decision Tree Based Approach

4) Performance
   i. Time Analysis
   ii. Memory Analysis
   iii. CPU Analysis

![Figure 3.1: Block Diagram of Proposed Concept.](image-url)
3.3 Proposed Algorithm

Input: Dataset KDD’99, a sample K Normal Cluster NC, Abnormal cluster AC
Output: K is abnormal or normal
Algorithms:
A) K- Means
B) K- Nearest Neighbor Classification
C) Decision Tree Approach

4. Conclusion

The prime concern of the proposed research is to improve the detecting speed and accuracy and presents more efficient cluster rules, mining method with classification method to abnormal detecting experiment based on network. I am expecting that the presented approach is an effective approach, which is the combination of different Data mining rule based approach. Considering the dependent relations between alerts, I expect that this approach can find more accurate probability of normal and abnormal packets compared with other method & thus experimental results will prove this method is effective and achievable.

References


Biography

Chordia Anita S. received the Bachelor degree in Computer Engineering from the North Maharashtra University, Jalgaon in 2004 and pursuing M. tech (CSE) from S.V.C.S.E. Alwar, Rajasthan.