ISSN:0975-9646

Intrusion Detection System Using Clustering

Madhulika Deshmukh,

Department of Information Technology,
Vidyalankar Institute of Technology
Wadala(e),Mumbai, India
Mumbai University

Prof. S.K. Shinde

Department of Computer engineering
Lokmanya Tilak college of engineering
Koparkhairne,Navi Mumbai,India
Mumbai University

***Abstract*— Intrusion detection is the process of monitoring and analysing the events occurring in a computer system in order to detect signs of security problems. 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. IDS come in a variety of "flavours" and approach the goal of detecting suspicious traffic in different ways. There are network based (NIDS) and host based (HIDS) intrusion detection systems. 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 are primarily focused on identifying possible incidents, logging information about them, and reporting attempts. In addition, organizations use IDS for other purposes, such as identifying problems with security policies, documenting existing threats and deterring individuals from violating security policies. Over the past ten years, intrusion detection and other security technologies such as cryptography, authentication, and firewalls have increasingly gained in importance. However, intrusion detection is not yet a perfect technology. This has given data mining the opportunity to make several important contributions to the field of intrusion detection.**

***Keywords*— Intrusion detection, monitoring data, logging, data mining,**

I. INTRODUCTION

An intrusion is defined as any set of actions that attempt to compromise the integrity, confidentiality or availability of a resource. Intrusion detection is classified into two types: misuse intrusion detection and anomaly intrusion detection. Misuse detection is based on known attack actions. In this method features are extracted from known intrusions and rules are pre-defined. The important disadvantage of this method is the novel or unknown attacks that cannot be detected.

Anomaly detection is based on the normal behaviour of a subject; any action that significantly deviates from the normal behaviour is considered intrusion. Sometimes the training audit data does not include intrusion data. One problem with anomaly detection is that it is likely to raise many false alarms.

II. PROBLEM DEFINITION

As stated above to detect novel attacks activities of the user, they are logged and identified for differed activities which are suspected to be against normal behaviour. But this method raises the count of number of false alarms. False negatives are associated with signature based IDS. Signature based IDS require the use of signatures incorporated into its database to match the signatures of packets of data entering into the network. Signatures of known viruses and other malicious codes are placed in the database for signature matching. As a result, any attack for which it has the signature can be accurately identified and detected. Unfortunately, newly created malicious code or known viruses with modified signatures are allowed to go undetected within the system and are classed as a false negative. Such a drawback is owed to the inability of signature based NIDS to detect new attacks as stated by McHugh et al. [4].

Apriori is the best-known algorithm to mine association rules. This algorithm was developed by Agarwal and Srikant in 1994. Association rules find frequent item sets whose occurrences exceed a predefined minimum support threshold and deriving association rules from those frequent item sets. These two sub problems are solved iteratively until no more new rules emerge. Minimum support threshold must be defined by user and initial transactional database. This algorithm uses knowledge from previous iteration phase to produce frequent itemsets.

This algorithm uses breadth-first search and a hash tree structure to make candidate itemsets efficient, and then the frequency occurrence for each candidate itemsets will be counted. Those candidate itemsets that have higher frequency than minimum support threshold are qualified to be frequent itemsets.

ALGORITHM

Ck: Candidate itemset of size k
Lk : frequent itemset of size k
L1 = {frequent items}; for
(k = 1;
Lk != □; k++) do begin
Ck+1 = candidates generated from
Lk; for each transaction t in database do
increment the count of all candidates in
Ck+1 that are contained in t
Lk+1 = candidates in
Ck+1 with min_support end
Return ∪k Lk;

Example:

- Let I = {A, B, C,D,E} a set of items
- Let D be a set of DB transactions

- Let T be a particular transaction
- An association rule is of the form A => B where A, B included in I and (A ∩ B = ∅)
- Support: The support of a rule, A => B, is the percentage of transactions in D, the DB, containing both A and B.
- Confidence: The percentage of transactions in D containing A that also contain B.
- Strong Rules: Rules that satisfy both a minimum support and a minimum confidence are said to be strong
- Itemset: Simply a set of items
- k-Itemset: a set of items with k items in it
- Apriori Property: All non-empty subset of a frequent itemset must also be frequent
- Frequent Itemset: An itemset is said to be frequent if it satisfies the minimum support threshold.
- A two-step process
    - *The join step:* Find L_k, the set of candidate of k- itemset; join L_{k-1} with itself.
    - Rules for joining:
        - Order the items first so you can compare item by item
        - The join of L_{k-1} is possible only if its first (k-2) items are in common
    - *The Prune step*:
    - The "join" step will produce all k-itemsets, but not all of them are frequent.
    - Scan DB to see which itemsets are indeed frequent and discard the others.

Stop when "join" step produces and empty set.

III. CLUSTERING

As seen in previous section the intrusion detection system which uses association rules for identifying intrusions requires more processing time for discovering the frequent pattern item sets.To reduce this processing time we are introducing the concept of clustering.

Clustering is the process of grouping of data, where the grouping is established by finding similarities between data based on their characteristics. Such groups are termed as Clusters. Cluster is a collection of data objects similar to one another within the same cluster and dissimilar to the objects in other clusters. Cluster analysis is grouping a set of data objects into clusters Clustering is unsupervised classification of no predefined classes. There are many clustering algorithms available with their own strength and weakness understanding the systems need we are using Density-based clustering algorithm.

Density-based approaches, apply a local cluster criterion, are very popular for database mining. Clusters are regions in the data space where the objects are dense, and separated by regions of low object density (noise). These regions may have an arbitrary shape. A density-based clustering method is presented in [8]. The basic idea of the algorithm DBSCAN is that, for each point of a cluster, the neighborhood of a given radius (e), has to contain at least a minimum number of points (MinPts), where e and MinPts are input parameters.

Density based Clustering

1. Compute the ε-neighborhood for all objects in the data space.
2. Select a core object CO.
3. For all objects co Ɛ CO, add those objects y to CO which are density connected with co. Proceed until no further y are encountered.
4. Repeat steps 2 and 3 until all core objects have been processed.

Algorithm:

Arbitrarily select a point ***p***

Retrieve all points density-reachable from ***p*** wrt ***Eps*** and ***MinPts***.

If ***p*** is a core point, a cluster is formed.

If ***p*** is a border point, no points are density-reachable from ***p*** and DBSCAN visits the next point of the database.

Continue the process until all of the points have been processed.

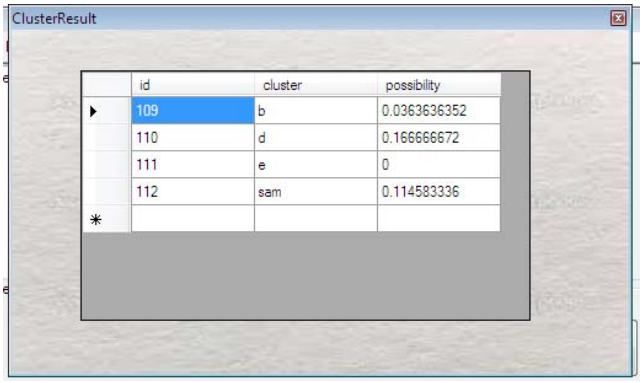

Figure a: Clusters formed after processing audit

IV. AAA PROTOCOL

The goal 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 also include the abuse of privileges or the use of attacks to exploit software or protocol vulnerabilities. Sometimes insider may also misuse their rights which should be detected by the intrusion detection system. For implementing this we use AAA prototcol. Authentication, authorization, and accounting is a term for a framework for intelligently controlling access to computer resources, enforcing policies, auditing usage, and providing the information necessary to bill for services. These combined processes are considered important for effective network management and security.

As the first process, authentication provides a way of identifying a user, typically by having the user enter a valid user name and valid password before access is granted. The process of authentication is based on each user having a unique set of criteria for gaining access. The AAA server compares a user's authentication credentials with other user

credentials stored in a database. If the credentials match, the user is granted access to the network. If the credentials are at variance, authentication fails and network access is denied.

Following authentication, a user must gain authorization for doing certain tasks. After logging into a system, for instance, the user may try to issue commands. The authorization process determines whether the user has the authority to issue such commands. Simply put, authorization is the process of enforcing policies: determining what types or qualities of activities, resources, or services a user is permitted. Usually, authorization occurs within the context of authentication. Once you have authenticated a user, they may be authorized for different types of access or activity.

The final plank in the AAA framework is accounting, which measures the resources a user consumes during access. This can include the amount of system time or the amount of data a user has sent and/or received during a session. Accounting is carried out by logging of session statistics and usage information.

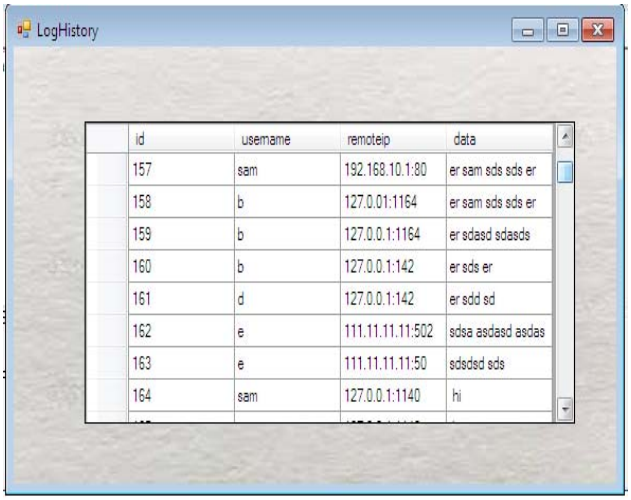

Figure b: Audit log

V. CONCLUSIONS

This paper explains that how using a clustering technique we can reduce the processing time and improve the performance of the system by calculating the probability of intrusion and help detecting the IP and or user to be blocked.

A key feature of this model is that all access is through roles. Controlling all access through roles simplifies the management and review of access controls.

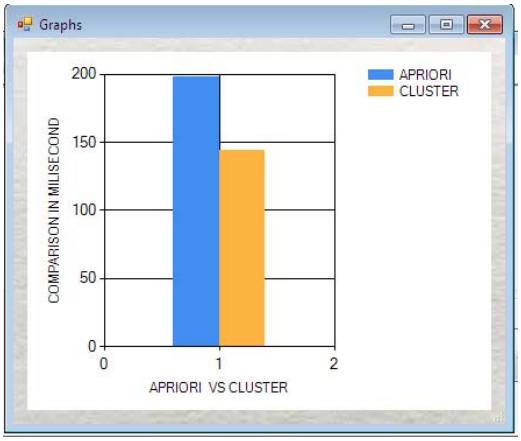

Figure c : Graph showing difference in processing time using Apriori Algorithm and clustering

REFERENCES

[1] Ehsan Saboori, Shafigh Parsazad, *"Automatic Firewall rules generator for Anomaly detection System with Apriori algorithm"*, 2010 3rd International Conference on Advanced Computer Theory and Engineering.

[2] Bace, R. (2000). *Intrusion Detection.* Mcmillan Technical Publishing.

[3] Lippmann, R. P., Fried, D. J., Graf, I., Haines, J. W., Kendall, K. R., Mc-Clung, D., Weber, D., Webster, S. E., Wyschogrod, D., Cunningham. K., and Zissman, M. A. (2000). *"Evaluating Intrusion Detection Systems: The 1998 DARPA Off-Line Intrusion Detection Evaluation"*. In Proceedings of the 2000 DARPA Information Survivability Conference and Exposition, pages 12-26.

[4] McHugh, J., Christie, A., & Allen, J. (2000). *"Defending yourself: The role of intrusion detection systems"*. IEEE Software 17(5), 42-51. Retrieved October 2, 2006, from IEEE Computer Society Digital Library database.

[5] Allen, J., Christie, A., Fithen, W., McHugh, J., Pickel, J., and Stoner. (2000). *State of the Practice of Intrusion Detection Technologies.* Technical report, Carnegie Mellon University. http://www.cert.org/archive/pdf/99tr028.pdf

[6] Klaus Julisch, *Data mining for intrusion detection, A Critical Review*, IBM Research, Zurich Research Laboratory, kju@zurich.ibm.com.

[7] *Association rule apriori algorithm* (www.aplysit.com | www.ivan.siregar.biz) APLYSIT – IT Solution Center Jl. Ir. H. Djuanda 109Ivan Michael Siregar Data Mining 2010.

[8] G. Mohammed Nazer, A. Arul Lawrence Selvakumar, *"Intelligent Data Mining Techniques for Intrusion Detection Models on Network"*, European Journal of Scientific Research ISSN 1450-216X Vol.71 No.1 (2012), pp. 36-45

[9] Ahmed Youssef and Ahmed Emam, *"Network Intrusion Detection Using Data Mining and Network Behavior Analysis"*, International Journal of Computer Science & Information Technology, III (6), pp. 87-98, 2011.

[10] Wenke Lee and Salvatore J. Stolfo and Kui W. Mok, *"Mining Audit Data to Build Intrusion Detection Models"*, In Proceedings of KDD-98 (AAAI), 1998.