ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. 2, Special Issue I, January 2016 in association with
KARUR VELALAR COLLEGE OF ARTS AND SCIENCE FOR WOMEN, KARUR
Conference on Emerging Trends and Functional Applications of Computer Technology - 12th January 2016

A MODERN APRIORI ALGORITHM IS USED TO EVADE NETWORK INTRUSION DETECTION SYSTEM

1.Ms.E.M.Roopa Devi , Assistant Professor/IT, Kongu Engineering College roopasen5@gmail.com
2. Dr.R.C.Suganthe, Professor/CSE Kongu Engineering College, rcsuganthe@kongu.ac.in

Abstract

With the development of hacking and exploiting tools and development of new ways of intrusion, intrusion detection and deterrence is becoming the major challenge in network security. The growing network traffic and data on Internet is making this task more challenging. There are various approaches being used in intrusion detections, but unfortunately any of the systems so far is not completely faultless. The false positive rates make it extremely hard to analyze and react to attacks. In this paper, we represent a Mordern Apriori algorithm is proposed which helps in finding the long patterns with low support as well as short patterns with high support. A non decreasing elliptical function is used for calculating length decreasing support. Our technique is used to generate attack rules that will detect the known and unknown attacks .

Keywords
Intrusion Detection System, Association rule mining, KDD dataset

1. Introduction

With the attractiveness of Internet in the world and the development of computer networks have been and people learn, work closely together.In Internet advanced security measures, and are always under innovative and inventive attacks. The goal of intrusion detection systems is to detect attack from network data traffic and generate an alarm.

Two key advantages of using a data mining approaches for IDS are the following.

- It can be used to automatically generate the detection models for IDSs, so that new attacks can be detected automatically.
- It can be used to build IDSs for a wide variety of computing environments

Intrusion detection is of two types: Network based intrusion detection system (NIDS) and Host based intrusion detection system (HIDS). NIDS examines network traffic and monitors multiple hosts. These are placed in the networks to monitor whole network and report the network administrator about malicious activity. In host based intrusion detection system the system detects malicious packets which enter the host system. It does not detect whole network. This type of IDs has an host to identify intrusion by analyzing system calls, application logs and other activities. Two methods are used for intrusion detection: Signature based detection and Anomaly based detection. Signature based systems are used to detect known attacks and require preceding knowledge of attack patterns. Anomaly detection systems assume that an intrusion is an deviation of the system behavior from its normal pattern. Association rule mining is usually used to find the interesting rules from a large database depending upon the user defined support and confidence. A frequent item set is defined as one that occurs more frequently in the given data set than the user given support value. Association rule mining means generating interesting rules from audit data to detect unknown intrusions.

2. Previous Work

Data mining Approaches for intrusion detection (Wenke and Salvatore 1998)The association rules algorithm and the frequent episodes algorithm were used to compute the intra- and inter- audit record patterns of user behavior. Here, the learning agents continuously compute and provide the updated (detection) models to the detection agents. Send mail system call data and the network tcp dump data, were used for the experiments**.** The issues in this work is accuracy of the detection models depends on sufficient training data and the right feature set.AdaBoost-Based Algorithm for Network Intrusion Detection(Weiming et al. 1999) In AdaBoost-based algorithm, decision stumps are used as weak classifiers. The decision rules are provided for both categorical and continuous features. The relations between categorical and continuous features are handled naturally, without any forced conversions between these two types of features. A simple over fitting handling is used to improve the learning results. In the specific case of network intrusion detection, we use adaptable initial weights to make the tradeoff between the detection and false-alarm rates. Intrusion detection using neural networks and support vector machines (Mukkamala S. et al, 2002) The performance of support vector machines and neural networks in intrusion detection, using the DARPA data for intrusion evaluation were discussed. All classifications were performed on the binary (attack / normal) basis. Both SVMs and neural networks deliver highly-accurate (99% and higher) performance, with SVMs showing slightly better results. Further, when a reduction is performed to reduce the 41 features to the 13 most significant, both SVMs and neural networks again were able to train to deliver accurate results. A Novel Rule-based Intrusion Detection System Using Data Mining(LeLi. et al, 2010) Association rule mining is an effective data mining method, which can extract the database features of the relationship between the items. But issues in association rules are that it uses the constant support value irrespective of length

3. Association Rule Mining for Intrusion Detection

Data mining generally refers to the process of mining knowledge from a big amount of data. Association rule mining is the task of discovering correlations and patterns from the large dataset

Association rule problem : Given a set of items I={I1,I2,…,Im} and a database of transactions D={t1,t2, …, tn} where ti={Ii1,Ii2, …, Iik}, the Association Rule Problem is to identify all association rules A =>B which satisfy minimum support and confidence where A is subset of I and B is subset of I[13]. The support of the rule is the percentage of transactions that contains both

A and B in all transactions and is calculated as

Support= Ø(AUB)/Ø(N)

Confidence= Support(AUB)/Support B

It is frequently used in data mining have been developed in machine learning research community. Frequent pattern and association rule mining is one of the few exceptions to this tradition. Apriori algorithm is act as a devise more efficient algorithms of frequent itemset mining.

4. Proposed Work

Association rule mining is an effective data mining method, which can extract the database features of the relationship between the items. The popular data mining approaches is to find frequent item sets from a transaction dataset and derive association rules is Apriori algorithm we propose a network intrusion detection system model that analyses the various item set generated, specifically on attribute relation. In the system, apply association rule mining to generate attack signatures from the large network traffic data. We propose an Intrusion detection system as shown below

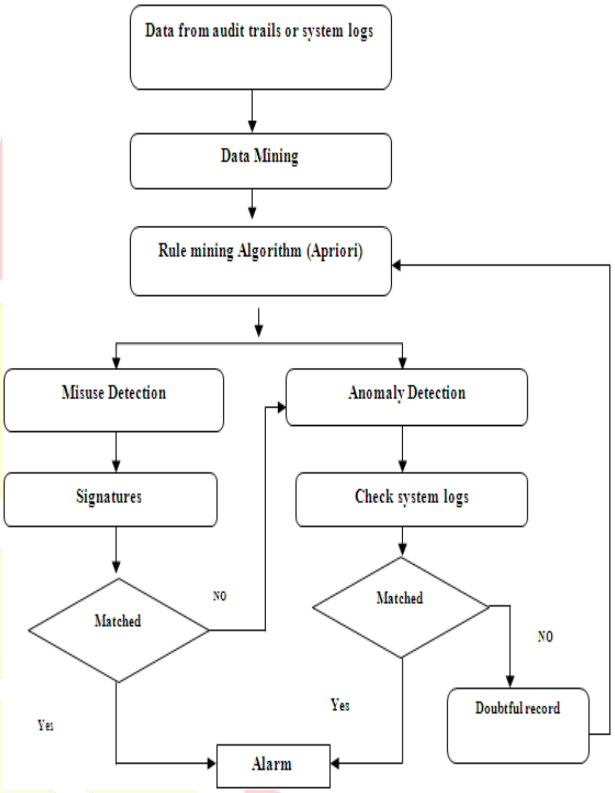

Figure 1: Proposed Intrusion Protection System

The modules of our system are

a. Analysis of KDD Dataset

b.Data preprocessing

c.Breadth first search Apriori Algorithm(Morden Apriori Algorithm)

d.Detection of Attacks

a. Analysis of KDD Dataset

The network intrusion dataset from the KDD archive popularly referred to as the KDD 99 Cup dataset The KDD training dataset consist of 10% of original dataset that is approximately 494,020 single connection vectors each of which contains 41 features and is labeled with exact one specific attack type *i.e.*, either normal or an attack.

b. Data Preprocessing

Preprocessing of NSL KDD dataset is an important step in order to make suitable input for SVM. Dataset preprocessing can be categorized in to following properties: (i) Dataset transformation: Since the training set of NSL KDD dataset 4,900,000 connections with 42 features including attacks such as normal or abnormal. We need to transform nominal features to numeric values to provide suitable input for classification. Also target class (last feature) has to be assigned with numeric value. Here we have assigned zero for normal connection and one for deviation. In this step useless data are filtered and modified. (ii) Dataset Normalisation: This step is important to enhance performance of intrusion detection system when datasets are too large. (iii) Dataset Discretization: Dataset discretization is used in continuous feature selection of intrusion detection and creates homogeneity between values that are of different data types.

c. Breadth first search Apriori Algorithm

A breadth first search can be applied in the frequent itemset generation in Apriori algorithm to detect the attacks.

It is used to create a model to generate rules.

For Creating frequent sets let's define :

Ck as a candidate itemset of size k

L_k as a frequent itemset of size k

Main steps of iteration are:

1. Find frequent set L_{k-1}

2. Join step: Ck is generated by joining L_{k-1} with itself (Cartesian product $L_{k-1} \times L_{k-1}$)

3. Prune step (Apriori property): Any (k–1) Size itemsets that is not frequent cannot be a subset of a frequent k size itemsets, hence should be removed.

4.Frequent set L_k has been achieved

d. Detection of Attacks

Using rules generated by Breadth First Search Apriori Algorithm the Anomaly and Misuse attacks can be detected. The different type of attacks like Denial of Service, Probing, Remote to Local, User to Root and others, it is a challenge for any intrusion detection system to detect a wide variety of attacks.

S.No	Categories	Attack Types
1.	DOS	Back, Land, Neptune,pod, smurf, Teardrop
2.	U2R	Buffer_overflow, loadmodule, perl, rootkit
3.	R2U	ftp_write, guess_passwd, imap, multihop, phf, spy,warezlient, warezmaster
4.	Probe	IPsweep,nmap, satan,portsweep

Table 1: Classes of Attacks

5. System Evaluation and Results

In KDD dataset there are 41 features for each connection record that are divided into discrete sets and continuous sets according to the feature values. It consists of number of total records 494021.There are many measures available for evaluating system performance. For evaluating intrusion detection results following measure are generally used.In order to know how to read the data from the audit data,it need to analyze how the audit data is being recorded.

1. True positive (TP) means number connections that were correctly classified as intrusion.
2. True Negative (TN) means number of connections that were incorrectly classified as intrusion.
3. False positive (FP) means number of intrusion connections that were incorrectly classified as normal.
4. False negative FN) means number of normal connections that were incorrectly classified as intrusion

Detection rate can be calculated as

$$\textbf{Accuracy}=\boxed{\frac{\mathbf{TP+TN}}{\mathbf{TP+TN+FP+FN}}}$$

Dataset	Normal Patterns	Abnormal Patterns	Detection rate
DS1	5263	1658	88%
DS2	4798	1897	86%
DS3	4879	1358	87%
DS4	5364	2257	89%

Table 2: Distribution of Records and Accuracy obtained in 10% KDD dataset

Rule mining for symbolic feature were the Minimum support is 50% and minimum confidence is 80%.Network-based intrusion detection detects intrusions based on signatures. The rules generated when duration, protocol_type, src_bytes,dst_bytes and label these attributes are selected.Some sample rules are shown in the table.

ID	Antecedent	Consequent	Length	Support	Confidence
1	protocol_type=icmp	label=smurf	2	0.58991	0.9899
2	protocol_type=icmp & src_bytes	label=smurf	3	0.58619	0.9981
3	label=smurf	src_bytes	2	0.59991	1.001
4	label=smurf &src_bytes	protocol_type=icmp	2	0.58992	0.9976
5	label=smurf & protocol_type=icmp	src_bytes	3	0.59983	1.000

Table 3: Rules generated from Association rule mining on selected attributes

When all 41 attributes selected we could generate 593 rules and 192 frequent itemsets For experimentation the support and confidence value are kept constant.

6. Conclusion

In this paper discusses the association rules and the Modern Apriori algorithm which is applied to Intrusion Detection problems. Rules are generated for detecting new attacks based on the information of known attacks. The Modern Apriori algorithm is used to detect the unknown attacks with high accuracy rate and high efficiency. . Exploring other data mining techniques with length decreasing support algorithm for removing duplicates from the generated ruleset.

References

1.Dorothy E. Denning. An Intrusion-Detection Model. IEEE Transactions on Software Engineering,13(2):222–232, 1987. IEEE.

2.T.F. Lunt, “A survey of intrusion detection techniques”, Computers and Security12 (4)(1993) 405–418.

3.C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, W.-Y. Lin, Intrusion detection by machine learning: a review, Expert Systems with Applications 36 (2009) 11994–12000.

4.L. Khan, M. Awad, B. Thuraisingham, A new intrusion detection system using support vector machines and hierarchical clustering, The VLDB Journal 16 (2007) 507–521.

5 .V. Vapnik, The Nature of Statistical Learning Theory, Springer, 1999.

6. C.-H. Tsang, S. Kwong, Ant colony clustering and feature extraction for anomaly intrusion detection, in: Swarm Intelligence in Data Mining, in: Studies in Computational Intelligence, vol. 34, Springer, 2006, pp. 101–123.

7. D. Duan, S. Chen, W. Yang, Intrusion detection system based on support vector machine active learning, Computer Engineering 33 (1) (2007) 153–155.

8. J.C. Platt, Fast training of support vector machines using sequential minimal optimization, 1999, pp. 185–208.

9.V. Jaiganesh, “Intrusion Detection Using Kernelized Support Vector Machine With Levenbergmarquardt Learning”, International Journal of Engineering Science and Technology, 2012.