International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 1, 2017 ISSN: 2348-6600
Reference ID: IJCS-158 PAGE NO: 958-963
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

PERCEIVING INTRUSION USING DATA MINING TECHNIQUES

M.Deepa[1], Dr.P. Sumitra[2]
[1]Ph.D Research Scholar, Department of Computer Science,
Legithasai2010@gmail.com
[2]Professor, Department of Computer Science,
sumithravaratharajan@gmail.com
[1,2]Vivekananda College of Arts and Sciences for Women (Autonomous),Elayampalayam

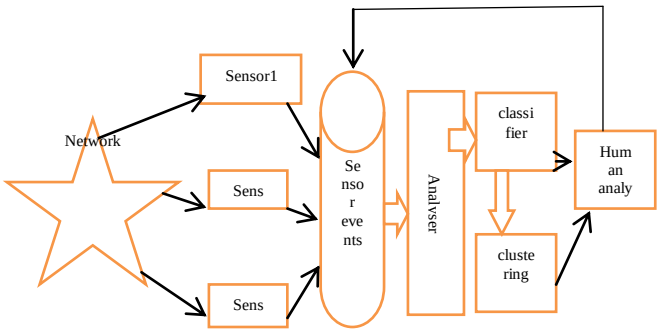
ABSTRACT

In the soviet most of the activities are done through Internet. The increased usage of internet has leads to the unauthorized access of the information. There are several types of attacks present in the current world. Now a days Intrusion detection is the most common challenge of the information security. Regarding perceiving intrusion so many techniques are present but data mining is the tremendous approach for detecting intrusion. In this paper converse the numerous data mining approaches for perceiving intrusion on the network based on accuracy, detection rate and alarm.

Keywords: Data mining, intrusion detection, classification, clustering

I. INTRODUCTION

The Internet or the World Wide Web (WWW) is described by connecting the sequence of computers together to form a network. This architecture allow the people to perform countless communications like Skype, Facebook, WhatsApp and etc.On the other side the trespassers develop many threats for attacking and broken the secure communication. So security is the important aspects of information society.**Firewall** analyzes packet headers and enforces policy based on protocol type, source address, destination address, source port, and/or destination port. Packets that do not match policy are rejected.**Intrusion Detection System** is a device or application that analyzes whole packets, both header and payload, looking for known events. When a known event is detected a log message is generated detailing the event. Fig. 1 demonstrates the overall architecture of Intrusion detection system.



International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 1, 2017 ISSN: 2348-6600
Reference ID: IJCS-158 PAGE NO: 958-963
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

II. Intrusion detection systems

Intrusion detection system (IDS) is a type of security management system which gathers and analyzes the information from various partsof a computer or a network to detect possible security holes, which includes both intrusionsand misuse activities. TheIntrusion detection systems (IDS) can be classified into following ways.there are

- Active Intrusion detection systems (AIDS)
- PassiveIntrusion detection systems(PIDS)
- Network Intrusion detection systems (NIDS)
- Host Intrusion detection systems (HIDS)
- Signature-basedIntrusion Detection system(SIDS)
- Anomaly-based Intrusion Detection system(BIDS)

Active Intrusion detection systems

An active Intrusion Detection Systems (AIDS) is designed toroboticallychunkmistrusted attacks without any intercession required by an operator.

Passive Intrusion detection systems

A passive IDS is a system which configured to only monitor and analyze network traffic activity and alert an operator to inform about the potential vulnerabilities and attacks. A passive IDS is not capable of performing any protective or corrective functions on its own.

Network Intrusion detection systems (NIDS)

Network Intrusion Detection Systems (NIDS) usually have two important components there are network sensor and a Network Interface Card (NIC). The network sensor monitors all network traffics on that segment.

Host Intrusion detection systems (HIDS)

A Host Intrusion Detection Systems (HIDS) is installed on workstations and can only monitor the individual workstation and it cannot monitor the entire network. These systems monitor the operating system activities and write data to log files and trigger alarms if any traffic occurred. Host based IDS systems are used to monitor any intrusion attempts on critical servers.

The drawbacks of Host Intrusion Detection Systems (HIDS) are

• Difficult to analyses the intrusion attempts on multiple computers.

• Host Intrusion Detection Systems (HIDS) can be very difficult to maintain in large networks with different operating systems and configurations

• Host Intrusion Detection Systems (HIDS) can be disabled by attackers after the system is compromised.

Knowledge-based (Signature-based) IDS

Signature-based detection systems have a preconfigured set of signatures that are compared with network traffic to detect an attack or intrusion. Just as virus scanners use signatures to recognize viruses, IDSs use signatures to recognize common attacks or exploits used by hackers.

Behavior-based (Anomaly-based) IDS

In behavior based IDS the security administrator define a normal activities. It monitors all the activities and analyses it with the predefined activities and trigger the alarm if doesn't match with configured profile.

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 1, 2017 ISSN: 2348-6600
Reference ID: IJCS-158 PAGE NO: 958-963
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

III. NETWORK ATTACKS

A. Denial of service(DoS) attacks

In this attack attackers disrupt a host or network servicesthey make the some computing or memory resources toobusy. For example: ping of death, SYN flood etc.

B. Probing Attack

In this attack the attackers scans the network to gather theinformation and then uses it to exploit the system. Forexample: port scanning etc.

C. Remote to Local Attack(R2L)

This occurs when an attacker who does not have anaccount on a remote machine sends packet to that machineover a network and exploit some vulnerability to gain localaccess. For example: password guessing etc.

D. User to Root Attack(U2R)

In this an attacker starts out with access to a normal useraccount on the system and is able to exploit vulnerabilityto gain root access to the system. For example: bufferoverflow attack etc.

IV. DATA MINING

Data mining is a process of extracting hidden information from various data sources. Now a day a datamining is abetter way for designing an Intrusion detection system. Data mining have several algorithms these algorithms are classified by

- *Association based algorithms*
- *Classification based algorithms*
- *Clustering based algorithms*
- *Prediction based algorithms*
- *Sequential patterns*
- *Decision tree based algorithms*

Association based algorithms

Association is one of the best-known data mining technique. In association, a pattern is discovered based on a relationship between data items in the same transaction. The association technique is also known as *relation technique*. The association technique is used in *market basket analysis* to identify a set of products that customers frequently purchase together.

Classification based algorithms

Classificationtechnique is based on machine learning. In classification each item in a set of data classified into one of a predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics

Clustering based algorithms

Clustering is a data mining technique that makes a meaningful or useful cluster of objects which have similar characteristics using the automatic technique. The clustering technique defines the classes and puts objects in each class, while in the classification techniques, objects are assigned into predefined classes.

Prediction

The predictionis one of a data mining techniques that discovers the relationship between independent variables and relationship between dependent and independent variables. For instance, the prediction analysis technique can be used in the sale to predict profit for the future

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 1, 2017 ISSN: 2348-6600
Reference ID: IJCS-158 PAGE NO: 958-963
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

Sequential Patterns analysis

Sequential patterns analysis is one of data mining technique that seeks to discover or identify similar patterns, regular events or trends in transaction data over a business period.In sales, with historical transaction data, businesses can identify a set of items that customers buy together different times in a year. Then businesses can use this information to recommend customers buy it with better deals based on their purchasing frequency in the past.

Decision trees

The A decision tree is one of the most common used data mining techniques because its model is easy to understand for users. The following are top 10 algorithms that frequently used by most of the researches.

- C4.5
- k-means
- Support vector machines
- Apriori
- EM
- PageRank
- AdaBoost
- kNN
- Naive Bayes
- CART

1) C4.5

C4.5 is asupervised learning method, since the training dataset is labeled with classeswhich constructs a classifier in the form of a decision tree. A classifier is a tool in data mining that takes a bunch of data and cracks to predict which class the new data belongs to. Decision tree learning creates something similar to a flowchart to classify new data.

2) k-means

k-means is a semi-supervised method which creates *k* groups from a set of objects so that the members of a group are more similar. It's a popular cluster analysis technique for exploring a dataset. Cluster analysis is a family of algorithms designed to form groups such that the group members are more similar versus non-group members. k-means can be used to pre-cluster a massive dataset followed by a more expensive cluster analysis on the sub-clusters.

3) Support vector machines

Support vector machine (SVM) is a supervised learning method that acquires a hyperplane to classify data into 2 classes. At a high-level, SVM performs a similar task like C4.5 except SVM doesn't use decision trees at all. A hyperplane is a function like the equation for a line, $y = mx + b$. In fact, for a simple classification task with just 2 features, the hyperplane can be a line.

4) Apriori Algorithm

The Apriori algorithm is a unsupervised learning method that follows association rules and is applied to a database containing a large number of transactions.Association rule is a data mining technique for learning correlations and relations among variables in a database.Apriori can also be modified to do classification based on labeled data.The basic Apriori algorithm is a 3 step approach:

- **Join.** Scan the whole database for how frequent 1-itemsets are.

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 1, 2017 ISSN: 2348-6600
Reference ID: IJCS-158 PAGE NO: 958-963
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

- **Prune.** Those itemsets that satisfy the support and confidencemove onto the next round for 2-itemsets.
- **Repeat.** This is repeated for each itemset level until we reach our previously defined size.

5) Expectation Maximization

Expectation-maximization (EM) is an unsupervised data mining technique that used as a clustering algorithm (like k-means) for knowledge discovery.In statistics, the EM algorithm iterates and optimizes the likelihood of seeing observed data while estimating the parameters of a statistical model with unobserved variables.

6) PageRank

PageRank is a link analysis algorithm designed to determine the relative importance of some object linked within a network of objects. Link analysis is a type of network analysis looking to explore the associations (a.k.a. links) among objects.

7) AdaBoost

AdaBoost is a boosting algorithm which constructs a classifier which takes a bunch of data and attempts to predict or classify which class a new data element belongs to.Boosting is an ensemble learning algorithm which takes multiple learning algorithms and combines them. The goal is to take an ensemble or group of weak learners and combine them to create a single strong learner. A weak learner classifies with accuracy barely above chance. A strong learner has much higher accuracy, and an often used example of a strong learner is SVM.

8) K-Nearest Neighbor

k-Nearest Neighbors, is a supervised classification algorithm.K-NN builds no such classification model. Instead, it just stores the labeled training data.When new unlabeled data comes in, K-NN operates in 2 basic steps:

First, it looks at the k closest labeled training data points — in other words, the k-nearest neighbors.

Second, using the neighbors' classes, K-NN gets a better idea of how the new data should be classified.

9) Naive Bayes

Naive Bayes is a supervised learning method. It is a group of classification algorithms that share one common assumption: Every feature of the data being classified is independent of all other features given the class.

10) CART

CART stands for classification and regression trees. It is a decision tree learning technique that outputs either classification or regression trees. Like C4.5, CART is a classifier.Classification trees output classes, regression trees output numbers.

V. CONCLUSION

This paper shows the implementation performed based onvarious data mining algorithms. There are severalintrusion detection tools are available with competing feature to detect four different type of attacks. To increase thedetection rate and to reduce false alarm combination ofalgorithm gives more accuracy.

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 1, 2017 ISSN: 2348-6600
Reference ID: IJCS-158 PAGE NO: 958-963
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

VI. REFERENCES

1. Rowayda A. Sadek, M. Sami Soliman and Hagar S. Elsayed"Effective Anomaly Intrusion Detection System based on NeuralNetwork with Indicator Variable and Rough set Reduction" IJCIinternational Journal of Computer Science Issue, Vol,10, Issue 6,No 2, 2013

2. Ahmed A. Elngar, Dowlat A. El A. Mohamed and Fayed F. M.Ghaleb "A Real-Time Anomaly Network Intrusion DetectionSystem with High Accuracy" 2013 Inf. Sci. Lett. 2, No. 2, 49-56(2013)

3.HeshamAltwaijry ,SaeedAlgarny "Bayesian based intrusiondetection system" 2012 Journal of King Saud University 2012

4.Renuka Devi Thanasekarn "A Robust and Efficient Real TimeNetwork Intrusion Detection System Using Artificial NeuralNetwork In Data Mining" 2011 International Journal of InformationTechnology Convergence and Services(IJITCS) Vol. 1, No. 4, 2011

5. Naveen N C, Dr. R Srinivasan , Dr. S Natarajan "A UnifiedApproach for Real Time Intrusion Detection Using Intelligent DataMining Techniques" 2011, IJCA Special Issue 2011

6. Tao Peng, WanliZuo "Data Mining for Network IntrusionDetection System in Real Time" IJCSNS International Journal ofComputer Science and Network Security, VOL.6 No.2B, February2006

7.https://www.reference.com/technology/internet-communication-a0b802f85dc624db

8. http://www.kdnuggets.com/2015/05/top-10-data-mining-algorithms-explained.html