International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 1 | ISSN : 2456-3307

Intrusion Detection Techniques : A Review

Sandip Hingane[1], **Dr. Umesh Kumar Lilhore**[2]

[1]M. Tech Scholar, Department of CSE, NIIST Bhopal Madhya Pradesh, India
[2]Head, Department of CSE, NIIST Bhopal Madhya Pradesh, India

ABSTRACT

Now, these days Internet technology is widely used everywhere. Most of the Internet-based applications are publically available for all the users. This public nature of Internet-based applications increases security threats. So security of user data and information is an important issue and need high attention in the research area. An intrusion detection system is designed for detection and classification of various attacks. It classifies attacks into normal and abnormal classes. Intrusion detection systems are based on either host based or network based. Various data mining and machine learning methods are widely used by ID systems. In this paper, we are presenting a review of various intrusion detection methods.

Keywords :- Intrusion Detection, Network-Based, Host-Based, Data Mining, Machine Learning.

I. INTRODUCTION

Intrusion detection is defined as identifying unauthorized use, misuse and abuse of computer systems by both inside and outside intruders. The main task of an intrusion detection system (IDS) is to defend a computer system or computer network by detecting hostile attacks on a network system or host device, monitoring the events occurring in a computer system or network and analyzing them for signs of intrusions [1,3].

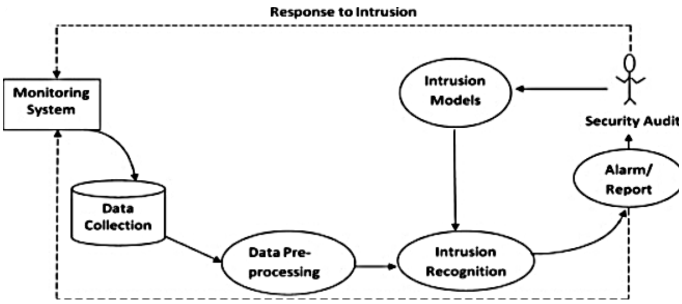

Figure 1. Working of IDS

One of the main challenges in the security management of large-scale high-speed networks is the detection of suspicious anomalies in network traffic patterns. A secure network must provide the following [4]:

• **Data confidentiality:** Data that are being transferred through the network should be accessible only to those that have been properly authorized.

• **Data integrity:** Data should maintain their integrity from the moment they are transmitted to the moment they are actually received. No corruption or data loss is accepted either from random events or malicious activity.

• **Data availability:** The network should be resilient to Denial of Service attacks.

The goal of an ID is to detect malicious traffic. In order to accomplish this, the IDS monitor all incoming and outgoing traffic. There are several approaches to the implementation of IDS. Among those, two are the most popular:

Anomaly detection: This technique is based on the detection of traffic anomalies. The deviation of the monitored traffic from the normal profile is measured. Various different implementations of this technique have been proposed, based on

CSEIT183146 | Received : 03 Jan 2018 | Accepted : 16 Jan 2018 | January-February-2018 [(3) 1 : 129-135]

the metrics used for measuring traffic profile deviation.

Misuse or Signature detection: This technique looks for patterns and signatures of already known attacks in the network traffic. A constantly updated database is usually used to store the signatures of known attacks. The way this technique deals with intrusion detection resembles the way that anti-virus software operates.

II. IDS AND POSSIBLE ATTACKS

2.1 IDS-An intrusion detection system (IDS) inspects all outbound and inbound network action and find out the doubtful patterns that may point to a network or system intrusion or attack from someone trying to crack into or conciliation a system. An ID system gathers and observed information from different areas inside a network of systems to find out probable safety breaches, which contain together called intrusions (attacks exterior from the association) and misuse (attacks from inside the association). IDS use susceptibility assessment; it is an expertise which is design and developed to appraise the security of a network [12].

Data mining techniques can be used to detect intrusions. Applications of data mining have presented a collection of research efforts on the use of data mining in computer security. In the context of security of the data, we are looking for the information whether an information security breach has been experienced [8]. Intrusion detection system is the area where data mining concentrates heavily. There are two fold reasons for this first an IDS is very common and very popular and extremely critical activity. Second, a large volume of the data on the network is dealing so this is an ideal condition for the data mining to use it. The data mining technology has the enormous benefits in the data extracting attributes and the rule, so it is significant to use data mining methods in the intrusion detection

2.2 ATTACK & TYPES-Attack is an unwanted information or unauthorized access to our network which cause damage to our records. Following are the four major categories of networking attacks:

1) **Denial of Service (DoS):** In DoS attack, legitimate networking requests are not served because attacker makes the resources either too busy or full to serve the request. Hence the legitimate user cannot access the services of a machine or network resources. Example: Apache, mail bomb, back etc.
2) **Probing (Probe):** In probing, attacker scans a machine or a network device for gathering the information about weaknesses or vulnerabilities that can be exploited later to compromise the target system. Example: saint, mscan, nmap etc.
3) **User to Root (U2R):** In U2R attacks, an authorized user attempt to abuse the vulnerabilities of the system in order to gain the privilege of root user for which they are not authorized. Example: perl, xterm, Fd-format etc.
4) **Remote to Local (R2L):** In this type of attacks, a remote user tries to gain access as a local user to a local machine by sending packets to a machine over the internet. An external intruder exploits vulnerabilities of the system to access the privileges of a local user. Example: xlock, phf, guest etc.

III. RELATED WORK

Following existing methods are related to IDS detection-

2.1 Machine Learning: Machine Learning [1,2] is the study of computer algorithms that improve automatically through experience. Applications range from data mining programs that discover general rules in large data sets, to information filtering systems that automatically learn users' interests. In contrast to statistical techniques, machine learning techniques are well suited to learning patterns with no a priori knowledge of what those patterns may be [3]. Clustering and Classification are probably the two most popular machine learning problems. Techniques that address both of these problems have been applied to IDSs.

1) Classification Techniques: In a classification task in machine learning, the task is to take each instance of a dataset and assign it to a particular class. A classification based IDS attempts to classify all traffic as either normal or malicious. The challenge in this is to minimize the number of false positives (classification of normal traffic as malicious) and false negatives (classification of malicious traffic as normal). Five general categories of techniques have been tried to perform classification for intrusion detection purposes:

a) Inductive Rule Generation: The RIPPER System is probably the most popular representative of this classification mechanism. RIPPER [7] is a rule learning program. RIPPER is fast and is known to generate concise rule sets. It is very stable and has shown to be consistently one of the best algorithms in past experiments [11]. The system is a set of association rules and frequent patterns that can be applied to the network traffic to classify it properly. One of the attractive features of this approach is that the generated rule set is easy to understand; hence a security analyst can verify it.

b) Genetic Algorithms: Genetic algorithms were originally introduced in the field of computational biology. Since then, they have been applied in various fields with promising results. Fairly recently, researchers have tried to integrate these algorithms with IDSs [19].

c) Fuzzy Logic: Fuzzy logic is derived from fuzzy set theory dealing with the reasoning that is approximate rather than precisely deduced from classical predicate logic. It can be thought of as,"the application side of fuzzy set theory dealing with well thought out real world expert values for a complex problem".

d) Neural Networks: The application of neural networks for IDSs has been investigated by a number of researchers. Neural networks provide a solution to the problem of modeling the user's behavior in anomaly detection because they do not require any explicit user model. Neural networks for intrusion detection were first introduced as an alternative to statistical techniques in the IDES intrusion detection expert system to model [18].

e) Support Vector Machine: Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classifiers. SVMs attempt to separate data into multiple classes (two in the basic case) through the use of a hyperplane [20].

2.2 Feature Selection- "Feature selection, also known as subset selection or variable selection, is a process commonly used in machine learning, wherein a subset of the features available from the data is selected for application of a learning algorithm. Feature selection is necessary either because it is computationally infeasible, to use all available features, or because of problems of estimation when limited data samples (but a large number of features) are present." Feature selection from the available data is vital to the effectiveness of the methods employed. Researchers apply various analysis procedures to the accumulated data, in order to select the set of features that they think maximizes the effectiveness of their data mining techniques.

2.3 Clustering Techniques: Data clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics. Clustering is the classification of similar objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset (ideally) share some common trait often proximity according to some defined distance measure [16]. Machine learning typically regards data clustering as a form of unsupervised learning. Clustering is useful in intrusion detection as a malicious activity should cluster together, separating itself from the nonmalicious activity. Clustering provides some significant advantages over the classification techniques already discussed, in that it does not require the use of a labeled data set for training [22].

2.4 Statistical Techniques-It also known as "top-down" learning is employed when we have some idea as to the relationship were looking for and can employ mathematics to aid our search. Three basic classes of statistical techniques are linear, nonlinear (such as a regression-curve), and decision trees [5]. Statistics also include more complicated techniques, such as Markov models and Bayes estimators. Statistical patterns can be calculated with respect to different time windows, such as day of the week, the day of the month, the month of the year, etc [21].

2.5 Data Mining Techniques-Data mining techniques can be differentiated by their different model functions and representation, preference criterion, and algorithms [10].The main function of the model that we are interested in is classification, as normal, or malicious, or as a particular type of attack [1,2]. We are also interested in link and sequence analysis [13].Additionally; data mining systems provide the means to easily perform data summarization and visualization, aiding the security analyst in identifying areas of concern [6].The models must be represented in some form. Common representations for data mining techniques include rules, decision trees, linear and non-linear functions (including neural nets), instance-based examples and probability models. Here are a few specific things that data mining might contribute to an intrusion detection project [17]:

- Remove normal activity from alarm data to allow analysts to focus on real attacks.
- Identify false alarm generators "bad" sensor signatures.
- Find an anomalous activity that uncovers a real attack.
- Identify long, ongoing patterns (different IP address, same activity).

To accomplish these tasks, data miners employ one or more of the following techniques [11]:

- Data summarization method with statistics, including finding outliers.
- Visualization: It presents a graphical summary of the data.
- Clustering of the data into natural categories
  - **Association rule discovery:** defining normal activity and enabling the discovery of anomalies.
  - **Classification:** predicting the category to which a particular record belongs.

2.6 AODE Classifiers-Averaged one-dependence estimator (AODE) is a probabilistic classification learning technique. It was developed to address the attribute-independence problem of the popular naive Bayes classifier. It frequently develops substantially more accurate classifiers than naive Bayes at the cost of a modest increase in the amount of computation.

Features of the AODE classifier-Like naive Bayes, AODE does not perform model selection and does not use tunable parameters. As a result, it has low variance. It supports incremental learning whereby the classifier can be updated efficiently with information from new examples as they become available. It predicts class probabilities rather than simply predicting a single class, allowing the user to determine the confidence with which each classification can be made. Its probabilistic model can directly handle situations where some data are missing.

AODE has computational complexity $O(ln^2)$ at training time and $O(kn^2)$ at classification time, where n is the number of features, l is the number of training examples and k is the number of classes. This makes it infeasible for application to high-dimensional data. However, within that limitation, it is linear with respect to the number of training examples and hence can efficiently process large numbers of training examples.

IV. CHALLENGES OF IDS

Intrusion Detection Systems (IDS) have become a standard component in security infrastructures as they allow network administrators to detect policy violations [3, 5]. These policy violations range from external attackers trying to gain unauthorized access to insiders abusing their access. Current IDS have a number of significant drawbacks:

- **Service level network attacks:** Current IDS are usually tuned to detect known service level network attacks. This leaves them vulnerable to original and novel malicious attacks.
- **Data overload:** Another aspect which does not relate directly to misuse detection but is extremely important is how much data an analyst can efficiently analyze. That amount of data he needs to look at seems to be growing rapidly. Depending on the intrusion detection tools employed by a company and its size there is the possibility for logs to reach millions of records per day.
- **False positives:** A common complaint is the number of false positives an IDS will generate. A false positive occurs when a normal attack is mistakenly classified as malicious and treated accordingly.
- **False negatives:** This is the case where an IDS does not generate an alert when an intrusion is actually taking place.

V. DATABASE USED FOR IDS

5.1 KDD CUP 99: Software to detect network intrusions protects a computer network from unauthorized users, including perhaps insiders. The intrusion detector learning task is to build a predictive model (i.e. a classifier) capable of distinguishing between ``bad'' connections, called intrusions or attacks, and ``good'' normal connections. The 1998 DARPA Intrusion Detection Evaluation Program was prepared and managed by MIT Lincoln Labs. The objective was to survey and evaluate research in intrusion detection. A standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment, was provided. The 1999 KDD intrusion detection contest uses a version of this dataset [16].

5.2 NSL-KDD: It is a dataset suggested to solve some of the inherent problems of the KDD-99 data set which are mentioned in [5]. Although, this new version of the KDD dataset still suffers from some of the problems discussed by McHugh and may not be a perfect representative of existing real networks, because of the lack of public datasets for network-based IDSs, we believe it still can be applied as an effective benchmark data set to help researchers compare different intrusion detection methods [10, 18].

5.3 Kyoto DATASET: KDD Cup 1999 dataset [16. 22] is the most popular dataset that is used for evaluating the anomaly type intrusion. A KDD Cup 1999 dataset having 4, 90, 000 single vector connection. This dataset contains 41 features which were labeled as normal or as abnormal. In this dataset, attack lays into four categories they are-

- **Denial of Service Attack (DOS)**
- **User to Root Attack (U2R)**
- **Remote to Local Attack (R2L)**
- **Probing Attack**

Table 5.3 Comparisons of Datasets

Data Set	Number of attributes	Classes Identified
KDD Cup-99	42	5
NSL-KDD	42	2
Kyoto-2006	24	3

The above table 5.3 shows a comparison of various data sets used in IDS.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we have presented a review of intrusion detection methods. Also covers working on IDS system, types of attacks. Intrusion detection system mainly used database KDD Cup-99, Kyoto and NSL Kdd.

Still, existing IDS have different challenges which need more attention in future. In future work, we will propose an efficient and improved IDS detection and compared this proposed system with existing IDS system to check its performance.

VII. REFERENCES

[1]. Amreen Sultana, M.A.Jabbar, "Intelligent Network Intrusion Detection System using Data Mining Techniques", IEEE 2nd International Conference on Applied and Theoretical Computing and Communication Technology (ICAC), July 2016, pp 329-334.

[2]. M A Jabbar a, Rajanikanth Aluvalub, Sai Satyanarayana Reddy S," RFAODE: A Novel Ensemble Intrusion Detection System", ELSEVIER 7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India, pp 226-234.

[3]. Levent Koc and Alan D. Carswell," Application of an AODE Based Classifier to Detect DOS Attacks", IJCSNS International Journal of Computer Science and Network Security, VOL.15 No.2, February 2015, pp 24-29.

[4]. Adel Ammar," Adel AmmarA Decision Tree Classifier for Intrusion Detection Priority Tagging", Journal of Computer and Communications, 3, March 2015, pp 52-58.

[5]. Sean T Miller, Curtis Busby-Earle, "Multi-Perspective Machine Learning a Classifier Ensemble Method for Intrusion Detection", ICMLSC '17 Proceedings of the 2017 International Conference on Machine Learning and Soft Computing, Jan 2017, pp 7-12.

[6]. Sharmila Kishor Wagh, Vinod K. Pachghare, Satish R. Kolhe, "Survey on Intrusion Detection System using Machine Learning Techniques", International Journal of Computer Applications (0975- 8887) Volume 78, No. 16, September 2013, pp 30-38.

[7]. Roshani Gaidhane, Prof. C. Vaidya, Dr. M. Raghuwanshi," A Survey: Learning Techniques for Intrusion Detection System (IDS)", International Journal of Advance Foundation and Research in Computer (IJAFRC) Volume 1, Issue 2, Feb 2014, pp 21-29.

[8]. Sejal K. Patel, Umang H. Mehta, Urmi M. Patel, Dhruv H. Bhagat,"A Technical Review on Intrusion Detection System", International Journal of Computer Science & Engineering Technology (IJCSET), Vol. 6 No. 01 Jan 2015, pp 17-25.

[9]. AbdullaAminAburomman,"Surveyoflearningm ethodsinintrusiondetection system", 2016 International Conference on Advances in Electrical, Electronic and System Engineering, 14-16 Nov 2016, pp 362-366.

[10]. Vipin Das, Vijaya Pathak, Sattvik Sharma, Sreevathsan, MVVNS.Srikanth, Gireesh Kumar T", Network intrusion detection system based on machine learning algorithms", International Journal of Computer Science & Information Technology (IJCSIT), Vol. 2, No 6, December 2010, pp 138-152.

[11]. Nutan Farah Haq, Musharrat Rafni," Application of Machine Learning Approaches in Intrusion Detection System: A Survey", (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, 2015, pp 9-19.

[12]. Trupti Phutane, Apashabi Pathan," A Survey of Intrusion Detection System Using Different Data Mining Techniques", International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2014, pp 6801-6808.

[13]. Chirag Modi, Dhiren Patel a, Bhavesh Borisaniya a, Hiren Patel b," A survey of intrusion detection techniques in Cloud", ELSEVIER Journal of Network and Computer Applications, June 2012, pp 42-57.

[14]. Yanjie Zhao," Network Intrusion Detection System Model Based on Data Mining", IEEE SNPD 2016, May 30-June 1, 2016, Shanghai, China, pp 206-212.

[15]. M.A. Jabbar, B.L. Deekshatulu, Priti Chandra, "Computational intelligence techniques for early diagnosis of heart disease", ICETECH, IEEE 2015, pp 127-133.

[16]. C.Ellcan, "Results of the KDD CUP 99 classifier learning "ACM SIG-KDD, Explorations newsletter, Feb-2000, 119-128.

[17]. Theodoros Lappas and Konstantinos Pelechrinis," Data Mining Techniques for (Network) Intrusion Detection Systems", International Journal of Computer and Telecommunications Networking, Vol. 64, Issue 7, June 2015, pp: 206-219.

[18]. Krishna Kant Tiwari, Susheel Tiwari, Sriram Yadav, Intrusion Detection Using Data Mining Techniques, IJACT, Nov 2016, pp21-27.

[19]. Kapil Wankhade, Sadia Patka," An Efficient Approach for Intrusion Detection Using Data Mining Methods", IEEE 2013 International Conference on Advances in Computing, Communications, and Informatics (ICACCI), June 2013, pp 1615-1618.

[20]. Arif Jamal Malik, Waseem Shahzad, Farrukh Aslam Khan, "Network ID using hybrid binary PSO and RF algorithm", Security and Communication Network, March 2012, pp 158-165.

[21]. P. Natesan, P. Balasubramanie, "Multi-Stage Filter Using Enhanced Ad boost for Network Intrusion Detection", International Journal of Network Security & Its Applications (IJNSA), Vol. No. 3, May 2012, pp 226-239.

[22]. Preeti Agarwal, Sudhir Kumar Sharma, "Analysis of KDD Dataset Attributes Class wise For Intrusion Detection". 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015), pp 117-124.