A Comprehensive Survey of Intrusion Detection Systems

Vivek Nandan Tiwari[1*] Prof. Kailash Patidar[2] Prof .Satyendra Rathore[3] Prof. Manoj Kumar Yadav[3]
1.PG scholar, IT Dept, SSSIST, Sehore, INDIA
2.Head CSE Dept, SSSIST, Sehore, INDIA
3.Asst. Prof.CSE Dept, SSSIST, Sehore, INDIA

Abstract
Alongside with digital signatures and Cryptographic protocols, Intrusion Detection Systems (IDS) are judged to be the final contour of protection to protect a system. But the major difficulty with today's mainly admired IDSs (Intrusion Detection System) is the invention of massive quantity of false positive (FP) alerts alongside with the true positive (TP) alerts, which is an awkward assignment for the operator to examine to arrange the proper responses. So, there is an immense requirement to discover this area of study and to discover a reasonable solution. A main disadvantage of Intrusion Detection Systems (IDSs), despite of their detection method, is the vast number of alerts they produce on a daily basis that can effortlessly exhaust security supervisors. This constraint has guide researchers in the IDS society to not only extend better detection algorithms and signature tuning methods, but to also focus on determining a variety of relations between individual alerts, formally known as alert correlation. There are a variety of approaches of intrusion detection, such as Pattern Matching, Machine Learning, Data Mining, and Measure Based Methods. This paper aims towards the proper survey of IDS so that researchers can make use of it and find the new techniques towards intrusions.

Keywords: Intrusion Detection System, False positive alert, KDD Cup99, Anomaly detection, misuse detection, Machine Learning.

1. Introduction

An intrusion is a sequence of related actions performed by a suspicious adversary, which result in the form of compromise of a target system. These kinds of actions actually violate a certain security policy of the system. Security policy of a system defines which actions are considered to be malicious for the system and should be prevented in order to maintain the security of the system [2]. The process of identifying and responding to suspicious activities of a target system is called Intrusion Detection. It is a complementary approach to security with respect to the mainstream approaches, such as access control and cryptography [2]. Intrusion detection systems are used to monitor computer systems, as well as the network and to raise alarms when some intrusive activities are detected.

But most of the popular IDSs suffer from generating false alarms in a large volume. False alarms could be of two types. One is called false positive which is generated mistakenly by the IDS as an evidence of malicious behavior of the system, but in reality, it is not such a behavior. The other type of false alarms is called false negative. It is generated by the IDS as an evidence of non malicious event, but in reality, it should be an indication of malicious activity in the system [10]. Previous research on this area reports that this value could be as high as several hundred thousand a day but around 99% of them are false alarms while monitoring intrusion in an active operational network [11]. Network security officers need to investigate each IDS alarm manually whether it is a false or a true alarm. So, it is a quite time consuming, error prone and hard task for the network security officer to investigate manually and take proper action accordingly. Thus we have chosen to address the false alarm problem of IDSs in our survey.

Intrusion Detection Systems is of two types based on sources of audit information [3]:

- Host based Intrusion Detection System (HIDS): It refers to intrusion that take place on a single host system. This type of IDS gets it audit data from host audit trails and monitors activities such as file changes, integrity of system, system logs and host based network traffic. When any suspicious activity found by IDS, it alerts the system administrator or alert the central management server. Server or user or both can block the user request, this judgment is based on the mechanism installed in the local host system [12].
- Network based Intrusion Detection System (NIDS): It is used to monitor the network traffic to protect the system from network based threats. It gets its data from monitoring the network traffic by using sensors and keeps the records in its defined format in the system log. It tries to detect malicious activity like Denial-of-Service (DoS) or Distributed Denial-of-Service (DDoS) [12].

1.1. General Architecture of Intrusion Detection System:

A generalArchitecture of IDS is shown in figure 1. Typically, IDS uses the information available in system configuration data, audit storage and previously known attacks (reference data). The IDS can be placed in the system. It can be located in target system or external to it. In former case if target system is compromised the

IDS can also be invaded, in later case it IDS can be safe. IDS may use active information that is running in the system for reducing the detection time. On detecting anomaly IDS send alarm to Site Security Officer (SSO). For detection of anomaly we set the baseline for normal behaviour in IDS. For detection of true intrusion it is crucial to set the baseline of normal behaviour in IDS, because if it not so system may generate false alarms.

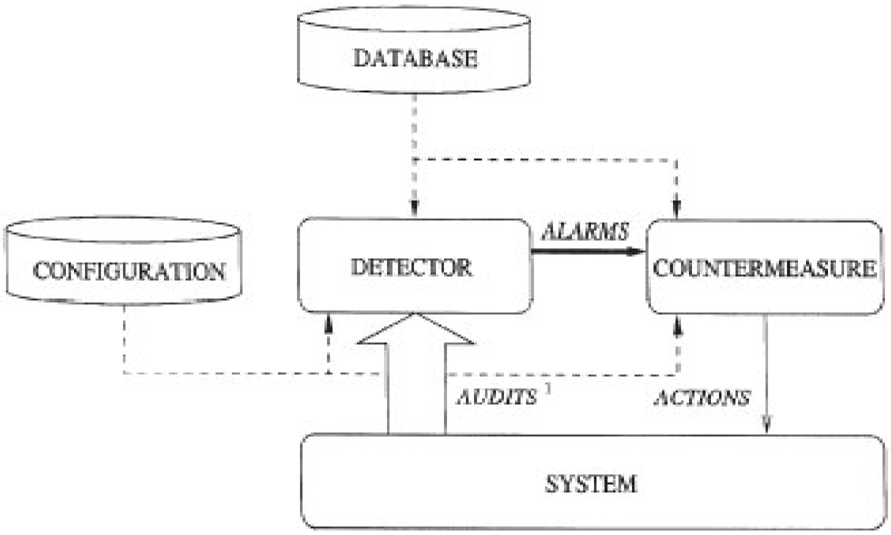

Fig. 1: General Architecture of IDS.

The objective of this paper is to identify the various attacks and defence system against the intrusions. We describe different techniques and approaches of intrusion detection so that researchers can do better comparative studies and find the new approaches of intrusion detection.

Rest of the paper is organized as follow: Section 2 describes traditional IDS briefly, security functions and measures of IDS. Various types of attacks to the network are described in section 3. In section 4 previous work done is analyzed, section 5 states the current problem statement of the ids and finally in section 6 we conclude our paper.

2. Traditional Intrusion Detection Systems

There are two types of intrusion detection system [4].

Anomaly Detection: It refers to the technique which is used to detect the malicious activities based on deviation from normal behaviour. These activities are considered as an attack to the system. It can also detect the unknown intrusions. All that can happen because we can train this type of IDS for unknown abnormal behaviour. For training set we can use the system logs of past activities, database of normal and abnormal behaviour, and systems configuration files. The detection rate of anomaly based IDS is high but it also generate false alarms proportionally.

Three broad categories of anomaly detection techniques exist:

- Unsupervised anomaly detection: These techniques detect anomalies in an unlabeled test data set under the assumption that the majority of the instances in the data set are normal.
- Supervised anomaly detection: These techniques require a data set that has been labeled as "normal" and "abnormal" and involves training a classifier.
- Semi-supervised anomaly detection techniques construct a model representing normal behavior from a given normal training data set, and then testing the likelihood of a test instance to be generated by the learnt model.

Misuse Detection (or Signature-based Detection): Misuse detection or Signature-based detection mainly depends on identifying known signatures. It means in this system we first need to determine the normal behaviour of the user, based on that IDS can define an activity as a normal or a threat to the system. So, this IDS system is used only for detecting known attacks (intrusions). The drawback of this system is that, a slight modification in activity can lead the system to not to generate the alarm, it can or cannot be a malicious activity. The detection rate of these IDS is low but it generates very low false alarms.

IDS provide following security functions:

- **Data Confidentiality:** It checks whether data/information stored in the system is secure or vulnerable to attack. It is the required security function because sometime system uses the sensitive information.
- **Data Availability:** It checks whether the information is available to authorized user or not. Sometimes the valid user cannot access the system information because of DoS attack, so IDS should be tough

against the DoS attacks. Again this is a very required security check.

- **Data Integrity:** It ensures that data is consistent and correct throughout the life cycle of an event. The data should not be changed in between of an event and also a valid/authorized user can have rights to change the data.

Primary criterions of measurements for IDS are as follows [1]:

- **Burglar Alert:** A signal is suggesting that a system has been or is being attacked [7].
- **Detection Rate:** The detection rate is defined as the no. of intrusion instances detected by the system (True Positive) divided by the total no. of intrusion instances present in the test set [8].
- **False Alarm Rate:** Defined as the number of 'normal' patterns classified as attacks (False Positive) divided by the total number of 'normal' patterns [8].

3. Types of Attacks

- **DoS Attack:** Denial-of-Service (DoS) or Distributed Denial-of-Service (DDoS) attack is the type of attack in which computer resources becomes unavailable to authorized users. These attacks slow down the system or deny the services of valid user. Due to this attack a lot of network traffic occurs [3].
- **Probing:** In this type of attack an attacker constantly monitors the network to find its vulnerabilities.
- **Eavesdropping Attack:** It is a network layer attack, in which an attacker captures the packets from the network that are transmitting from a host to others. Attacker can read sensitive and confidential information that is transmitting.
- **User to Root Attack (U2R):** In this attack, attacker starts his activity as a user and takes down the password, next do the dictionary attack and finally attacker gain access as a root user.
- **Remote to User Attack (R2U):** In this attack an attacker sends the packets to a machine over the network but does not have an account on local machine, by using the vulnerabilities of the system attacker gain local access to the system as a user.
- **Man-in-the-Middle Attack:** In this type of attack the attacker situated himself in the middle of two persons in communication, and both persons in communication think that they both communicating to each other but all the conversation is compromised.
- **Smurf Attack:** A smurf attack is an exploitation of the Internet Protocol (IP) broadcast addressing to create a denial of service. The attacker uses a program called Smurf to cause the attacked part of a network to become inoperable [9].

4. Related Work

In related work we explore previous work carried out by various researchers in the field of attack classification of KDD cup dataset in recent years. This section presents brief descriptions of the Data Mining and Machine Leaning approaches used by various researchers

Asak et al. [13] proposed a method for discriminate analysis of Machine learning based Intrusion Detection. In which a feature selection based method is utilized for the classification of individual attack. Author's utilizes system log information as experimental purpose.

Ramani et. al. [14] proposed a Discriminate Analysis based Feature Selection of KDD Intrusion Dataset. In this paper [14], important features of KDD Cup 99 attack dataset are extracted by the use of discriminate analysis method. Author's mentioned that proposed method is suffering by two- class classification or multiclass classification problems.

Kayacik et. al. [15] proposed a work of feature relevance analysis on KDD'99 dataset on the basis of information gain. Feature relevance is expressed in terms of information gain, which gets higher as the feature gets more discriminative. On the basis of result authors sagest that normal, neptune and smurf classes are highly related to certain features that make their classification easier. On the other hand authors told about certain features have no contribution to intrusion detection.

Balakrishnan et. Al[16] proposed a new feature selection algorithm based on InformationGain Ratio. The feature selection decreases the classification time. The author claims that proposed IDS reduce the false positive rates and classification time.

Adetunmbi A.Olusola et. Al [17] proposed the relevance of each feature in KDD '99 intrusion detection dataset to the detection of each class. Rough set degree of dependency and dependency ratio of each class were employed to determine the most discriminating features for each class. Empirical results show that seven features were not relevant in the detection of any class.In this paper, selection of relevance features is carried out on KDD '99 intrusion detection evaluation dataset. Empirical results revealed that some features have no relevance in intrusion detection.

N.S.Chandolikar et. Al [18] in this paper authors evaluate performance to two well known classifiers Bayes Net and J48 algorithms for attack classification. The key ideas are to use data mining techniques efficiently for intrusion attack classification. J48 learning algorithm was found to be performing better than Bayes Net in terms of better accuracy and lower error rate. Experiment performed on KDD cup dataset demonstrates that J48 algorithm is an efficient algorithm for classification. Accuracy demonstrated helps to improve efficiency of intrusion detection system.

Prof. N.S. Chandolikar et. Al [19] in this paper authors present the work on, KDD '99 intrusion detection dataset, which is evaluated to find out most important and relevant features. Proposed work based on selection of appropriate feature for reducing the analysis effort and time. Authors suggest that feature identification helps to improve efficiency of intrusion detection system.

Megha Aggarwal and Amrita [20] present the work on; a comparative analysis which is based on the basis of detection rate, computational time and root mean square error. In this work authors used six feature selection algorithms and their performance is evaluated using Naïve Bayes and C4.5 (J48) classifier. The authors has been observed that Naïve Bayes takes less time to test the dataset but more time in training the set whereas C4.5 does the reverse.

Himadri Chauhan et. Al [21] in this paper, authors presents the comparison of different classification techniques to detect and classify intrusions into normal and abnormal behaviours. J48, Naive Bayes, JRip, and OneR algorithms are used by authors. Authors use the WEKA tool to evaluate these algorithms. The experiments and assessments of these methods are performed with NSL-KDD intrusion detection dataset. The main task of this paper to show the comparison of the different classification algorithms and find out which algorithm will be most suitable for the intrusion detection.

S. Ranjitha Kumari and Dr. P. Krishna kumari [22] in this paper authors have done a survey on four supervised machine learning algorithms: Decision Tree (J48), K-Nearest Neighbour (KNN), Naïve Bayes (NB) and Support Vector Machine (SVM). Authors have shown a comparative analysis of these algorithms based on Accuracy, True Positive Rate (TPR) and False Positive Rate (FPR). Authors have used NSL-KDD dataset for our experiment. On the basis of experimental result, Authors have shown that the performance of Decision Tree (J48) and K-Nearest Neighbor are better than other two algorithms in terms of Accuracy, True Positive Rate (TPR) and False Positive Rat (FPR).

5. Problem Statement

The effectiveness of IDS depends on the capability to detect any abnormal activity in the target system, which is called the sensitivity of IDS. If the IDS are more sensitive, the security of the system would be tighter. To making the IDS more sensitive means to apply tighter signature rules or to be less tolerant to anomalies. As a result, the IDS become more sensitive to its input and generate a lot of alarms each day, even though most of the examined events are not illegal events.

Due to large volumes of IDS false alarms, it is a quite tough task for the security officers to investigate manually which are the real suspicious alarms and thereafter take proper action against them. Even sometimes, some real suspicious alarms are ignored mistakenly by the security officer due to large volumes of false alarms and thereby mistakenly interpret a real alarm to be a false alarm. This is the most dangerous situation when a real instance of an attack is ignored by the security officer and thus the IDS become useless though its functionality remains the same. We have chosen to investigate about this problem in our research and thus our research problem is whether we can reduce the IDS false alarm problem to a reasonable amount, or not.

6. Conclusion

Extensive research is going on in the field of Computer intrusion detection and several IDSs are already developed. But their performance is poor by producing false positives at higher rate. Researchers proposed several intrusion detection approaches and each detection approach is suitable only for detecting a particular type of attack(s). Because of limited attack coverage of each approach, there is an urgent need to arrive of a generic detection approach that handles almost all types of attacks. For that it is required to understand and analyze the techniques that are already investigated by several researchers. Keeping that in view here, we have made an attempt to review the well known intrusion detection approaches. Comparison of various approaches is made to show the strength and weakness of these approaches. We hope this study will be useful for researchers to carry forward research on system security for designs of IDS that not only will have identified strengths but also overcome the drawbacks.

References

[1] Adriana-Christina Enache, Victor Valeriu Patriciu, "Intrusions Detection Based on Support Vector Machine Optimized with Swarm Intelligence", 9th IEEE international symposium on Applied Computational Intelligence and Informatics", P.P. 978-1-4799-4694-5/14, May 2014

[2] Kruegel, Christopher, Fredrik Valeur, and Giovanni Vigna. Intrusion detection and correlation: challenges and solutions. Vol. 14. Springer Science & Business Media, 2005

[3] Subaira.A.S, Anitha.P, "Efficient Classification Mechanism for Network Intrusion Detection System Based on Data Mining Techniques:a Survey" International conference on Intelligent System and Control(ISCO), IEEE, P.P. 978-1-4799-3837, July 2014.

[4] Deepthy K Denatious, Anita John, "Survey on Data Mining Techniques to Enhance Intrusion Detection", International Conference on Computer Communication and Informatics (ICCCI -2012), IEEE, P.P. 978-1-4577-1583, Jan 2012.

[5] A Murali M Rao, "A Survey on Intrusion Detection Approaches", IEEE, P.P. 0-7803-9421-6, 2005

[6] Ming Xue,Changjun Zhu. "Applies Research On Data Mining Algorithm In Network Intrusion Detection", International Joint Conference On Artificial Intelligence, IEEE, 2009

[7] Nitin Mattord, Verma (2008). Principles Of Information Security. Course Technology. Pp. 290–301. Isbn 978-1-4239-0177

[8] Www.Users.Cs.York.Ac.Uk/~Jac/Publishedpapers/Adhocnetsfinal.Pdf

[9] W. Feng, Q. Zhng, G. Hu, J Xiangji Huang, "Mining Network Data For Intrusion Detection Through Combining Svms With Ant Colony Networks" Future Generation Computer Systems,2013

[10] Gula, Ron. "Correlating ids alerts with vulnerability information." (2002).

[11] Pietraszek, Tadeusz. "Using adaptive alert classification to reduce false positives in intrusion detection." In Recent Advances in Intrusion Detection, pp. 102-124. Springer Berlin Heidelberg, 2004.

[12] Mittal, Mitali, Alisha Khan, and Chetan Agrawal. "A Study of Different Intrusion Detection and Prevension System" International Journal of Scientific & Engineering Research 4, no. 8 (2013): 1526-1531.

[13] Asak, Midori, Takefumi Onabura, Tadashi Inoue, and Shigeki Goto. "Remote attack detection method in IDA: MLSI-based intrusion detection using discriminant analysis." In Applications and the Internet, 2002.(SAINT 2002). Proceedings. 2002 Symposium on, pp. 64-73. IEEE, 2002.

[14] Sathya, S. Siva, R. Geetha Ramani, and K. Sivaselvi. "Discriminant analysis based feature selection in kdd intrusion dataset." International Journal of Computer Applications 31, no. 11 (2011): 1-7.

[15] Kayacik, H. Günes, A. Nur Zincir-Heywood, and Malcolm I. Heywood. "Selecting features for intrusion detection: A feature relevance analysis on KDD 99 intrusion detection datasets." In Proceedings of the third annual conference on privacy, security and trust. 2005.

[16] Balakrishnan, Senthilnayaki, K. Venkatalakshmi, and A. Kannan. "Intrusion Detection System Using Feature Selection and Classification Technique."International Journal of Computer Science and Application (2014).

[17] Olusola, Adetunmbi A., Adeola S. Oladele, and Daramola O. Abosede. "Analysis of KDD'99 Intrusion detection dataset for selection of relevance features." In Proceedings of the World Congress on Engineering and Computer Science, vol. 1, pp. 20-22. 2010.

[18] Chandolikar, N. S., and V. D. Nandavadekar. "Comparative Analysis of Two Algorithms for Intrusion Attack Classification Using KDD CUP Dataset,"."International Journal of Computer Science and Engineering (IJCSE) Vol 1 (2012): 81-88.

[19] Chandolikar, N. S., and V. D. Nandavadekar. "Selection of Relevant Feature for Intrusion Attack Classification by Analyzing KDD Cup 99." MIT International Journal of Computer Science & Information Technology 2, no. 2 (2012): 85-90.

[20] Megha Aggarwal, Amrita. "Performance Analysis of Different Feature Selection Methods In Intrusion Detection" International Journal of Scientific & Technology Research Volume 2, Issue 6, June 2013

[21] Chauhan, Himadri, Vipin Kumar, Sumit Pundir, and Emmanuel S. Pilli. "Comparative Analysis and Research Issues in Classification Techniques for Intrusion Detection." In Intelligent Computing, Networking, and Informatics, pp. 675-685. Springer India, 2014.

[22] Kumari, S. Ranjitha. "Intrusion Detection-A Comparative Analysis using Classification Algorithms." Networking and Communication Engineering 5, no. 2 (2013): 85-89.