© 2017 IJSRST | Volume 3 | Issue 6 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Data Mining Methods for Attacks Recognition & Prevention

Shalini

Asst. Prof., Dept. of Computer Science & Engineering, Jyoti Vidyapeeth Women's University, Jaipur, Rajasthan, India

ABSTRACT

In modern era, due to several benefits digital system and services has gain high attention in each and every field, especially in communication area. Amount of users are increasing day by day. However, this technique facilitates its user in number of ways within effective time period but with the growth of users and chip technology some fears has also comes in front of its users. Security and trustworthy information is one of the most issues with the use of such technique. Different novel approaches made an endeavor to to impacting availability, confidentiality, and integrity of critical data that poses a serious problem for their detection and exploits safety vulnerabilities. On the other hand human labeling of the available network audit data instances is usually tedious, time consuming and expensive. Therefore it is essential for a system administrator that he/she use one or more security tools to protect information from passing before curious eyes or, more importantly, from falling into wrong hands. This paper will examine the intrusion detection systems, one of the relative new technologies in information security. It aims to explore, in high level, the intrusion detection systems available today, as well as new developments in this area by using data mining methodologies. Apart of simple reviewing of accessible technique this study has also focus on current research issues of this field.

Keywords: Intrusion, Intrusion Detection System, Anomaly & Misuse Detection

I. INTRODUCTION

Over past few decades, popularity of computer networks has increases in rocket high direction in the era of information society. As the network system and their related applications gain advancements the security issues have also increases. In current scenario no one detached security system is capable to detect all types of attack with accuracy, day by day naive attackers launch powerful attacks which can bring down an entire network. However, in the last few decades a lot of approaches have been intensively proposed to defend the system, but complete network security still is a critical issue due to the wideness of the network system, the field lacks an integrated approach with high detection rate for minority attacks namely R2L and U2R. These types of attacks are more dangerous than majority attacks like DoS and Probe. Most of the current accessible security methods face several of troubles to detect such types of attack [1, 2].

An attack is defined as any set of actions that compromise the integrity, confidentiality or availability of a resource [3], [4]. Mostly attacks are the violation of information security policy. A security system like Intrusion Detection System (IDS) inspects all inbound and outbound network activity and identifies suspicious patterns that may indicate a network or system attack from someone attempting to break into or compromise a system. The first system that was implemented to provide the security against different categories of attacks was host (System) based that located in servers to examine the internal interfaces [5]-[7]. After that with the advancement of the computer networks the focus gradually shifted toward network-based. Network attack prevention system performs packet logging, real-time traffic analysis of IP network, and tries to discover if an attacker is attempting to break into the system [8]-[10]. The system gathers and analyzes information from various areas within a computer or a network to identify possible security breaches, which include both intrusions and misuse. Today, prevention from different types of attack categories is one of the high priority and challenging tasks for network administrators and security professionals. Typically, an IDS can be categorized into following categories.

- Signature Based IDS. IDS monitor the network and compare actual behavior with known suspicious patterns that are maintained in a database of attack signatures. Matching behavior indicates the existence of attack and generates an alert. The database does not cover any unknown or newly introduced threat whose signature is not available. If any unknown attack occurs, IDS cannot detect it as its signature does notmatch with those in the database. This indicates that success of intrusion detection is limited by the availability of the recent attack signatures in the database. These systems have proved efficient for known attacks.

- Anomaly Based IDS. Signature based IDS effectively detect known attacks but are ineffective for unknown attacks. In order to overcome this limitation, anomaly based IDS compare actual behavior with the baseline that defines the normal state of the system that is, parameters such as protocols, traffic load, and typical packet size [11]. Deviation from the baseline indicates the anomalous behavior and generates an alert. Sometimes normal behavior can be misclassified as attack due to incomplete description of normal behavior.

- Hybrid IDS. Hybrid IDS makes combined use of signature based and anomaly based ones in order to gain advantages of both [12].That is, they try to increase detection rates of known attacks and decrease false positive rates of novel attacks.

II. INTRUSION & INTRUSION DETECTION SYSTEMS

An intrusion is a set of actions that try to compromise the honesty, privacy, or accessibility of a source and falls into one of four categories [13]:

- **Denial of Service (Dos) Attacks:** In a denial-of-service (DoS) attack, an attacker attempts to prevent legitimate users from accessing information or services. The most common DoS attacks will target the computer's network bandwidth or connectivity. Bandwidth attacks flood the network with such a high volume of traffic that all available network resources are consumed and legitimate user requests cannot get through. Connectivity attacks flood a computer with such a high volume of connection requests, that all available operating system resources are consumed and the computer can no longer process legitimate user requests [14].
- **Probe Attacks:** It is an attack in which attacker scans and determines the weaknesses or the vulnerabilities in machine or network device that could be later useful for attacker.
- **Remote to Local (R2L) Attacks:** Attackers does not have an account on the victim machine, hence tries to gain access from a remote machine and exploits this access in order to send packets over the network.
- **User to Root (U2R) Attacks:-** User to root or user to super user exploits are a class of attacks where an attacker starts out with access to a normal user account on the system and then exploits vulnerability to gain root access. These attacks involve the semantic details which are very difficult to capture at an early stage at the network level. The most common exploits in this class of attacks are regular buffer overflows, which are caused by regular programming mistakes and incorrect environment assumptions.

Detection of an attack is the process of monitoring and analyzing the events occurring in a computer system in order to detect signs of security problems. The several dissimilar systems provide unique functions and mechanisms for attack detection with the aim to detect, filter, or prevent system from attack to provide data security and ensure continuity of services provided by a network. Traditional methods for attack detection are based on extensive knowledge of signatures of known attacks. With the rapid advancement in the network technologies including higher bandwidths and ease of connectivity of wireless and mobile devices, the focus of detection system has shifted from simple signature matching approaches to detecting attacks based on analyzing contextual information which may be specific to individual networks and applications. Broadly detection system can be categories in two categories: (i) Host-based Intrusion detection system (HIDS) (ii) Network-based Intrusion detection system (NIDS) [15].

A. Host Based Detection System

A host-based detection system (HIDSs) monitors and analyzes the internals of a computing system rather than the network packets on its external interfaces. It relies on events collected by the hosts. It can be classified based on the type of audit data. Traditionally, the HIDS analyses particular information stored in logs (such as syslogs) and also captures network packets entering/leaving the host in order to check for signs of attacks (such as denial-of-service attacks, backdoors, Trojan horses, unauthorized access attempts, malicious code being run, or buffer overrun attacks).

- **Advantages**
  - The HIDS monitor activities within an individual computer system, thus it can also detects those attacks that cannot be detected by NIDS.
  - The HIDS data sources are normally generated on a plaintext so it can efficiently operate in encrypted environment.
  - Topology does not change HIDS performance.
- **Disadvantages**
  - HIDSs provide poor real-time response and cannot effectively protect against one-time catastrophic events.
  - Are not well suited by detecting network scans or other such surveillance that targets an entire network.

B. Network Based Detection System

A "network intrusion detection system (NIDS)" monitors attack or unauthorized activity on a network. They are also called packet-sniffers. They generally have a signature database against which they compare network packets. These systems have been incapable of operating in switched environments, encrypted networks and high-speed networks. An NIDS needs dedicated hardware, and forms a system which can check packets travelling on one or more network lines, in order to find out if any malicious or abnormal activity has taken place.

- **Advantages**
  1. It does not replace primary security such as firewalls, encryption, and other authentication methods.
  2. By using small number of sensors at hubs, routers etc. easily monitor huge network.
  3. It uses packet sniffing.
- **Disadvantages**
  1. It may have difficult processing, all packets in a large or busy network and therefore, may fail to recognize an attack launched during periods of high traffic.
  2. Modern switch-based networks make it more difficult: Switches subdivide networks into many small segments and provide dedicated links between hosts serviced by the same switch. Most switches do not provide universal monitoring ports.
  3. The encrypted information cannot be analyzed by this system.

III. DATA MINING BASED IDS

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. The various data mining techniques that are used in the context of intrusion detection can be point out as [16].

1. Correlation analysis: Correlation is often used as a preliminary technique to discover relationships between variables. More precisely, the correlation is a measure of the linear relationship between two variables.

2. Feature selection: A subset of features available from the data is selected for the application of a learning algorithm.

3. Machine learning: Machine learning explores the study and construction of algorithms that can learn from and make predictions on data

4. Sequential patterns: It is used to excavate connection between data, time series analysis gains more focus on the relationship of data in times.

5. Classification: It is a technique of taking each instance of a dataset and assigning it to a particular class. Typical classification techniques are: inductive rule generation, genetic algorithms, fuzzy logic, neural networks and immunological based techniques.

6. Clustering: It is a technique for statistical data analysis. It is the classification of similar objects into a series of meaningful subset according to certain rules, so that the data in each subset share some common trait.

7. Deviation analysis: Deviation analysis can reveal surprising facts hidden inside data
8. Forecast: Finding certain laws according to historical data, establishing models and predicting types, characteristics of the future data, etc based on the model.

With the increase in computerization and storage of more and more sensitive data on the data servers, the security of the data servers is a major issue. As the intrusion detection systems are being used for monitoring networked devices where they look for the behavior patterns of various anomalous and malicious behaviors in the audit data. Making comprehensive IDS requires more time and expertise.

IV. RELATED WORK

Since the born of intrusion detection system a number of approaches has been developed in direction to detect the network attacks. The concept of attack detection was firstly introduce [17] in 1980, the author launch a risk arrangement model that build up a security observing surveillance system based on detecting anomalies in user behavior. The lead was then taken by [18] at the SRI International and the first model of detection has been introduces [19]. The system was rule based for monitoring traffic on network and achieve high precision rate. The rule based methods extract features from various audit streams, and detect attacks by comparing the feature values to a set of attack signatures provided by human experts. The signature database has to be manually revised for each new type of attack that is discovered. If the network is small and signatures are kept up to date, the human analyst solution to intrusion detection works well. But when organizations have a large, complex network the human analysts quickly become overwhelmed by the number of alarms they need to review. Therefore, the process of encoding rules is expensive and slow.

To reduce the average control path latency incurred between request and response of the system as well as the increasing the detection rate of network attack groups a approach has been proposed in [20]. The projected approach categorizes the network attacks in to four groups and use independent feature set to achieve the efficiency and accuracy in detecting the network attack groups. In same context an alternative technique, which based on a combination of time series and feature spaces, for using machine learning algorithms to automatically detect anomalies in real time has been proposed in [21]. According to the author the proposed technique can work well for a real network environment, and it is a feasible technique with flexible capabilities to be applied for real-time anomaly detection. To enhance the detection rate of attack in adhoc network a novel approach has been proposed in [22].

The approach has used the Supervised Learning in Quest (SLIQ), a fast scalable classifier for detecting intrusion. In same context to enhance the detection rate of attacks a novel Fuzzy Genetic Algorithm has been proposed in [23]. The proposed approach have use supervised learning algorithm, to classify attacks in the datasets. A hybrid detection system has been proposed in [24] by combining the Naïve Bayesian (NB) and Support Vector Machine (SVM). The Naïve Bayesian (NB) and Support Vector Machine (SVM) has been combined to maximize the accuracy, which is the advantage of NB and diminish the wrong alarm rate which is the advantage of SVM. A genetic algorithm (GA) based approach has been proposed in [25] to solve a problem of network intrusion. The GA resolve on the KDD trophy 99 facts set to construct a imperative set that preserve to recognize attacks scheduled on the system. The [26] presents the detailed investigation on learning techniques for Intrusion detection system (IDS). The survey pointed out some issues of existing intrusion detection system like: low detection rate, detailed classification of attack, large training time required to train the network.

In [27], authors have proposed layered IDS with the aim to enhance prediction accuracy of intrusions. Simulation result shows that proposed model drastically raise the prediction of minority attacks without hurting the prediction performance of the majority class. In [28] authors have Immune Artificial System to develop proactive anomaly detection and prevention system. Negative selection algorithm is tuned to make it evolve and facilitate to detect malicious activities.

In [29], authors have proposed a hybrid of two anomaly component and one signature based detection component. KNN is used as anomaly component in first stage. In the second stage normal traffic is classified using anomaly based identification method while attacks are detected using signature based method.

To improve the detection rate of attacks a novel attack detection model has been proposed [30]. The proposed model performs well for all the classes of attack. In this framework authors use four tiers architecture to enhance the adaptability of the cyber-attack detection. The data collection and preprocessing of the proposed model is included in first tier of proposed model. The Second tier is meant for the feature extraction technique, third tier is dedicated to classification of cyber-attacks and fourth tier is dedicated to user interface for reporting the events. The Second tier is meant for the feature extraction technique, third tier is dedicated to classification of cyber-attacks and fourth tier is dedicated to user interface for reporting the events.

Gaikwad et al [31] introduced a technique based on fuzzy clustering and ANN approach. This method could be applicable to overcome the issues of weak stability detection as well as low precision detection. The restore point in this method was employed for registry keys, system files roll back, project database and installed programs. Fuzzy clustering will generate different subsets for training in order to reduce the amount of subset size and complexity. Then each subset is trained with different type of artificial neural network and finally processed to obtain significant results.

Lin GU et al [32] proposed empirical study for right choice of unstable growing demand in processing big data which entails huge burden of storage, data center communication and computation which brings substantial operational xpenditure for data providing centers. Apart from traditional cloud service, an important characteristic of big data was found to be the tight coupling of computation and data computation tasks were performed only with relevant data. But the means to improve the IDS is not clearly conveyed so far by any of the researchers. Thus, the main aim of this paper is to implement a clear picture of the IDS using distributed big data concept.

V. CURRENT ISSUES WITH IDS

Initially, traditional security systems have many limitations like time consuming statistical analysis, requiring regular updating, non-adaptive, low accuracy, inflexibility and are tuned specifically to detect known major service level network attacks. Attempts to expand beyond this limited realm typically results in an unacceptable level of false positives. At the same time, enough data exists or could be collected to allow network administrators to detect these policy violations. Unfortunately, the data is so voluminous, and the analysis process so time consuming, that the administrators don't have the resources to go through it all and find the relevant knowledge. In other words, network administrators don't have the resources to proactively analyze the data for policy violations, especially in the presence of a high number of false positives that cause them to waste their limited resources. These inadequacies of present security system motivate to improve the detection performance for the minority and novel type of attacks, while maintaining a reasonable overall detection rate.

VI. CONCLUSIONS

This paper has presents some basics of the Intrusion Detection System with the recent trends of security techniques using data mining. Furthermore, some issues with the current IDS are also pointed out to help newcomers in the field of intrusion detection system and are useful for people looking for a quick review of recent development in this field.

VII. REFERENCES

[1] Overview of Attack Trends, 2002. Last accessed: November 30, 2008. http://www. cert.org/archive/pdf/attack_trends.pdf.

[2] Kapil Kumar Gupta, Baikunth Nath, Kotagiri Ramamohanarao, and Ashraf Kazi. Attacking Confidentiality: An Agent Based Approach. In Proceedings of IEEE International Conference on Intelligence and Security Informatics, pages 285–296. Lecture Notes in Computer Science, Springer Verlag, Vol (3975), 2006.

[3] Jian Pei Shambhu J. Upadhyaya Faisal Farooq Venugopal Govindaraju. Proceedings of the 20th International Conference on Data Engineering (ICDE‟04) 1063-6382/04 $ 20.00 © 2004 IEEE

[4] Debar, H., Dacier, M., and Wespi, A., A Revised taxonomyfor intrusion detection systems, Annales des Telecommunications, Vol. 55, No. 7–8, 361–378, 2000.

[5] Jackson, T., Levine, J., Grizzard, J., Owen, and H., “An investigation of a compromised host on a honeynet being used to increase the security of a large enterprise network,” IEEE workshop on Information Assurance and Security, IEEE, 2004.

[6] D. Y. Yeung, and Y. X. Ding, “Host-based intrusion detection using dynamic and static behavioral models,” Pattern Recognition, 36, 2003, pp. 229-243.

[7] X. Xu, and T. Xie, “A reinforcement learning approach for host-based intrusion detection using sequences of system calls,” In Proc. of International Conference on Intelligent Computing, Lecture Notes in Computer Science, LNCS 3644, 2005, pp. 995-1003.

[8] Krasser, S., Grizzard, J., Owen, H., and Levine. J., “The use of honeynets to increase computer network security and user awareness,” Journal of Security Education, vol. 1, 2005, pp. 23-37.

[9] Shon T., Seo J., and Moon J., “SVM approach with a genetic algorithm for network intrusion detection,” in Proc. of 20th International Symposium on Computer and Information Sciences (ISCIS 2005), Berlin: Springer-Verlag, 2005, pp. 224-233.

[10] X. Xu, X.N. Wang, “Adaptive network intrusion detection method based on PCA and support vector machines,” Lecture Notes in Artificial Intelligence (ADMA 2005), LNAI 3584, 2005, pp. 696-703.

[11] S. Rajasegarar, C. Leckie, J. C. Bezdek, and M. Palaniswami,“Centered hyperspherical and hyperellipsoidal one-class support vector machines for anomaly detection in sensor networks,” IEEE Transactions on Information Forensics and Security, vol. 5, no. 3, pp. 518–533, 2010.

[12] A. L. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” IEEE Communications Surveys and Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016.

[13] Mohammad Sazzadul Hoque, Md. Abdul Mukit, and M. A. N. Bikas, "An implementation of intrusion detection system using genetic algorithm," International Journal of Network Security & Its Applications (IJNSA), vol. 4, (2012).

[14] Frank Kargl, Jörn Maier, Stefan Schlott, Michael Weber “Protecting Web Servers from Distributed Denial of Service Attacks” ACM 1-58113-348-0/01/0005. May 1-5, 2001.

[15] Asmaa Shaker Ashoor, Prof. Sharad Gore ―Importance of Intrusion Detection System (IDS)‖ International Journal of Scientific & Engineering Research, Volume 2, Issue 1, January-2011

[16] Ranju Marwaha “Intrusion Detection System Using Data Mining Techniques– A Review” International Journal of Advanced Research in Computer Science and Software Engineering, Volume 7, Issue 5, May 2017

[17] James P. Anderson, “Computer security threat monitoring and surveillance,” Technical Report 98-17, James P. Anderson Co., Fort Washington, Pennsylvania, USA, April 1980.

[18] Dorothy E. Denning, and P.G. Neumann “Requirement and model for IDES- A real-time intrusion detection system,” Computer Science Laboratory, SRI International, Menlo Park, CA 94025-3493, Technical Report # 83F83-01-00, 1985.

[19] H. S. Javitz and A. Valdes. The SRI IDES Statistical Anomaly Detector. In Proceedings of the IEEE Symposium on Security and Privacy, pages 316–326. IEEE, 1991.

[20] Prof.S.S.Manivannan, Dr.E.Sathiyamoorthy “Detection System to detect the Network Attack Groups using the Layer wise Individual Feature Set” International Journal of Engineering and Technology (IJET), Vol 5 No 4 Aug-Sep 2013

[21] Kriangkrai Limthong “Real-Time Computer Network Anomaly Detection Using Machine Learning Techniques” Journal of Advances in Computer Networks, Vol. 1, No. 1, March 2013.

[22] P.Sreenivasul, K.RameshReddy “ A Scalable Classifier for Intrusion Detection in Adhoc Networks” International Journal of Advanced Engineering and Global Technology Vol-2, Issue-4, April 2014

[23] Miss. M. R. Yadav, Prof. P. B. Kumbharkar “ Intrusion Detection System with Supervised Learning Algorithms” International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 4, April 2014

[24] Amit D. Sagale, 2 Swati G. Kale “Combining Naive Bayesian and Support Vector Machine for Intrusion Detection System” IJCAT International Journal of Computing and Technology, Volume 1, Issue 3, April 2014

[25] Sunil Kumar, Surjeet Dalal “Optimizing Intrusion Detection System using Genetic Algorithm” International Journal of Research Aspects of Engineering and Management ISSN: 2348-6627, Vol. 1, Issue 1, FEB 2014, pp. 42-45

[26] Roshani Gaidhane, Student, Prof. C. Vaidya, Dr. M. Raghuwanshi “Survey: Learning Techniques for Intrusion Detection System (IDS)” International Journal of Advance Foundation and Research in Computer (IJAFRC) Volume 1, Issue 2, Feb 2014. ISSN 2348 – 4853.

[27] Dr. Neelam Sharma, Yatendra Mohan Sharma “Exploration of Novel Layered Models for Improving Minority Attack Detection in IDS” International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 1, January 2014

[28] Saurabh, P. and Verma, B. (2016), "An efficient proactive artificial immune system based anomaly detection and prevention system", Expert Systems with Applications, Vol. 60, pp.311-320.

[29] Guo, C., Ping, Y., Liu, N. and Luo, S.S. (2016), "A two-level hybrid approach for intrusion detection.", Neurocomputing, In Press, Corrected Proof, DOI: http://dx.doi.org/10.1016/j.neucom.2016.06.021.

[30] Shailendra Singh, Sanjay Silakari “An Ensemble Approach for Cyber Attack Detection System: A Generic Framework” 14th ACIS, IEEE 2013. Pp 79-85.

[31] Gaikwad, Sonali Jagtap, D.P. Kunal Thakare and Vaishali Budhawant. Anomaly Based Intrusion Detection System Using Artificial Neural Network and fuzzy clustering., International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, November- 2012; 1(9).

[32] Lin Gu, Deze Zeng, Peng Li, and Song Guo. Cost Minimization for Big Data Processing in Geo-Distributed Data Centers,IEEE Transactions on Emerging Topics in Computing;2014.