An Assessment on Efficient Intrusion Detection and Classification Technique

Ms Priya Sharma, Mr Anurag Jain

Abstract— Now these day Computers becomes vital part of everyday life and hence use of internet becomes more and more. Due to internet, computers are becomes vulnerable of different kinds of security threats. Therefore it is required that we need to have efficient security method in order to avoid leakage of important data or misuse of data. This security method is called as Intrusion Detection System (IDS). Since from last two decades IDS becomes core area of many researchers and many methods are already presented for efficient intrusion detection and classification. Most of methods are out dated as many new attacks generated by hackers. In this project our main aim is to presented scalable and efficient method for intrusion detection and classifications. For intrusion detection, we are not using traditional methods, rather we are focusing on using distributed approach, which not only improves the scalability but also improves efficiency.

Index Terms— Misuse Detection; Anomaly Detection, IDS

1. INTRODUCTION

With the increase of malicious network activities, considerable attentions have been paid to intrusion detection system. The network intrusion detection system is designed to classify anomalous behaviors by examining the dynamic characteristics of network connection records; its role is becoming more important as a vital part of the network security architecture [1, 2, 3].

In general, intrusion detection approaches are categorized into two methods: misuse detection and anomaly detection. Misuse detection, considered a rule-based approach or a signature recognition technique, uses stored signatures of known intrusion instances to detect an attack. This approach is highly successful in detecting occurrences of previously known attacks. However, it fails to detect new attack types and variants of known attacks whose signatures are not stored [4, 5, 6]. In anomaly detection techniques, usually a profile for normal behavior is initially established. The observed behavior of the subject is then compared with its normal profile, and an intrusion is signaled when the observed behavior of a subject deviates significantly from its normal profile. The primary advantage of anomaly-based detection is the ability to detect novel attacks for which signatures have not been defined. However, minimizing the false-positives is still a significant challenge for the approach [7, 8, 9].

In order to overcome the limitations of these two conventional intrusion detection methods, hybrid intrusion detection methods that combine the misuse detection method and the anomaly detection method have been proposed. Aydin et al. [10] developed a hybrid intrusion detection system by combining packet header anomaly detection and network traffic anomaly detection which are anomaly-based IDSs with the misuse-based IDS Snort. Xiang et al. [11] studied a multiple-level hybrid classifier, which combines the supervised tree classifiers and unsupervised Bayesian clustering to detect intrusions. A multi-level hybrid intrusion detection method that uses a combination of supervised, unsupervised and outlier-based methods was developed by Gogoi et al [12]. Goel et al. [13] investigated a parallel misuse and anomaly detection model, which used C4.5 based binary decision for misuse and CBA (Classification Based Association) based classifier for anomaly detection. Ren et al. [14] proposed a hybrid intrusion detection system based on hierarchical clustering and multiple-level decision trees. Most previous hybrid detection systems independently train a misuse detection model and an anomaly detection model, and then aggregate the results of the detection models. In this case, the detection rate can be improved but the intrusion detection systems still have a high false positive rate. Therefore, this study proposes a novel network intrusion detection system based on hierarchical approach that integrates a misuse detection model and an anomaly detection model for reducing false alarms and time complexity. In the proposed detection system, Random Forest (RF) based misuse detection model is constructed to detect well-known attacks. The RF is an ensemble learner based on randomized decision trees, a method well suited for analyzing high dimensional data and robust against the noise. The other classified set including normal and unknown attack data are considered uncertain data that are analyzed then using an anomaly detection method. In this study, Self-Organizing Map (SOM) based unsupervised anomaly detection model is trained for detecting unknown attacks that deviate significantly from normal patterns with the data classified by the misuse detection model. In this study, the number of attacks used as input for the unsupervised anomaly detection based on SOM can be considerably reduced by excluding the known attacks through the misuse detection first. These can make it possible for the intrusion detection system to construct the normal profiles precisely and achieve higher detection performance for test data. Especially, the proposed detection system employs the feature reduction before classifier training by considering the similarity of feature responses through a clustering analysis based on the feature space reduced by factor analysis in order to remove features that are redundant and contribute little to the detection intrusion detection system, thus helping to improve detection efficiency with low computational cost.

2. Intruders Detection System

Intrusion prevention requires a well-selected combination of "baiting and trapping" aimed at both investigations of attacker's. Diverting the intruder's attention from protected resource is another task of IDS. Data generated by intrusion detection systems is carefully examined (this is the main task of each IDS) for detection of possible attacks (intrusions) [2].

Figure 1 shows the component of IDS .An intrusion detection systems always has its core element- a sensor works as an analysis engine responsible for detection intrusions. This sensor contains decision- making mechanisms regarding intrusion. Sensor receives raw data from three major information sources: own IDS, knowledge base, system log and audit trails. The system log may include configuration of file system, user authorization etc. This information is created on the basis of a further decision-making process [1,3].

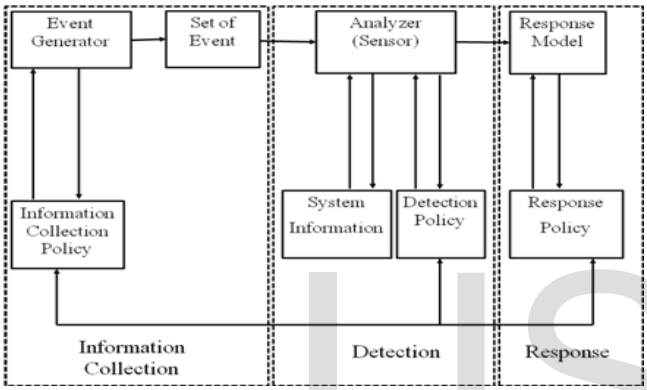

Figure 1 Components of IDS

The role of the sensor is to filter information and discard any irrelevant data obtained from the event set associated with suspicious activities. The sensor is integrated with the component responsible for data collection from an event generator. The collection manner is determined by the event generator policy that defines the filtering mode of event notification information. The event generator produces a policy –consistent set of event that may be a log (or audit) of system events, or network packets. In certain cases, no data storage is employed when event data streams are transferred directly to the analyzer .Once an intrusion has been detected, IDS issues alerts notifying administrators of this fact. The next step is undertaken either by the administrators or the IDS itself, by taking advantage of additional counter measures (specific block functions to terminate sessions, backup systems, routing connections to a system trap, legal infrastructure etc.) [1, 2].

An IDS is an element of the security policy. Among various IDS tasks, intruder identification is one of the fundamental ones. It can be useful in the forensic research of incidents and installing appropriate patches to enable the detection of future attack attempts targeted on specific persons or resources. Intrusion detection may sometimes produce false alarms, for example as a result of malfunctioning network interface or sending attack description or signatures via email [2]. The general activities performed by the general Intrusion detection system shown in Figure 2.

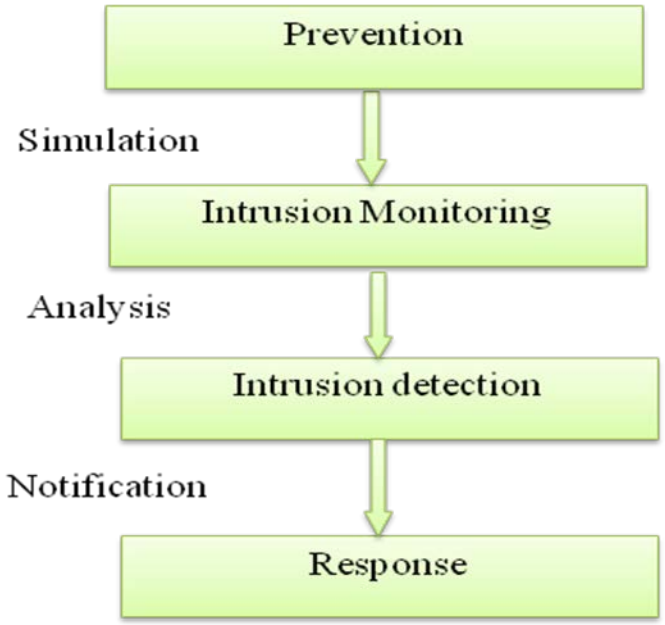

Figure 2: Flow chart for activities performed by IDS

Intrusion detection is the process of monitoring the events occurring in a computer system or network [2]. The purpose of IDS is to analyze the traffic that goes through it and to detect possible intrusions to the system. An IDS is an important part of the policies related to security issues. IDS can freeform various task but identification of intruders is one of the most basic function. It helps in gathering the evidence in computer crime. It also helps in the research of digital forensic in order to understand the activity of attacker.

There are basically two types of intrusion detection system [3]:

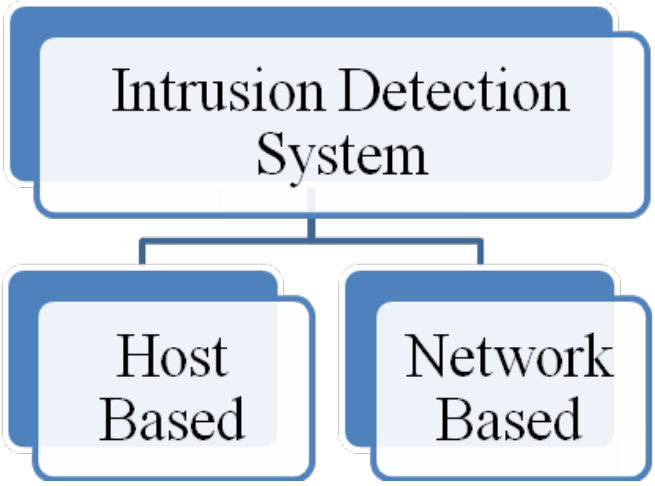

Figure 3: Classifications of IDS

- Host-based intrusion detection system: HIDSs evaluate information found on a single or multiple host systems, including contents of operating systems, system and application files.
- Network based Intrusion Detection: NIDSs evaluate in-

formation captured from network communications, analyzing the stream of packets which travel across the network.

3. Related Work

Varun, C., Arindam, B., Vipin, K [15], this survey attempts to provide a comprehensive and structured overview of the existing research for the problem of detecting anomalies in discrete/symbolic sequences. The aim is to provide a global understanding of the sequence anomaly finding problem and how previous techniques relate to each other. The key grouping of this survey is the categorized of the existing research into three different types, based on the problem formulation that they are trying to solve. These problem formulations are: 1) detecting anomalous orders with respect to a database of normal sequences; 2) detecting an anomalous subsequence between long sequences; and 3) detecting a pattern in a sequence whose frequency of occurrence is anomalous. They display how each of these problem formulations is characteristically different from each other and discuss their related in various application domains. They review techniques from many dissimilar and disconnected application domains that address each of other formulations. Within each problem formulation, our group techniques into typed based on the nature of the dependent algorithm. For each category, they provide a basic anomaly detection technique, and show how the existing techniques are alternative of the basic technique.

Frey, B.J., Dueck, D[16], Clustering data by identifying a subset of representative examples is important for processing sensory signals and detecting patterns in data. Such "exemplars" can be found by randomly choosing an initial subset of data points and then iteratively refining it, but this works well only if that first choice is close to a good solution. They devised a method called "affinity propagation," which takes as input measures of similarity between pairs of data points. Real-valued messages are interchange between data points up to a high-quality set of exemplars and comparative clump gradually emerges. We used affinity propagation to cluster images of faces, detect genes in microarray data, detecting representative sentences in this manuscript, and detect cities that are efficiently accessed by airline travel. Affinity propagation found clusters with much lower error than other methods, and it did so in less than one-hundredth the amount of time.

Davis J J, Clark A J. Data [18], Data pre-processing is widely recognized as an important stage in anomaly detection. This paper reviews the data pre-processing techniques used by anomaly-based network intrusion detection systems (NIDS), boil down on which aspects of the network traffic are analyzed, and what feature creation and selection methods have been used. Inspiration for the paper comes from the large crack data pre-processing has on the accuracy and capacity of anomaly-based NIDS. The review search that many NIDS limit their view of network traffic to the TCP/IP packet headers. Time-based statistics can be divided from these headers to detect network scans, network worm behaviour, and denial of service attacks. A number of other NIDS perform deeper examination of request packets to find attacks against network services and network applications. Today's approaches analyze full service responses to find attacks targeting clients.

R. Goel, A. Sardana, and R. C. Joshi [13]. In order to achieve high capability of arrangement in intrusion detection, a compressed model is proposed in this paper which binds horizontal compression with vertical compression. One R is utilized as horizontal compression for attribute reduction, and AP is used as vertical consolidation to select small illustrative exemplars from large training data. so it can computationally compress the larger volume of training data with scalability, Map Reduce based parallelization approach is then implemented and calculated for each step of the model compression process above mentioned, on which similar but efficient classification methods can be directly used.

W. Ren, L. Hu, K. Zhao[14] The Internet connects hundreds of millions of computers beyond the world running on many hardware and software platforms providing communication and commercial services. However, this inter connectivity among computers also start malicious users to misuse resources and mount Internet attacks. The continuously develop Internet attacks pose severe challenges to develop an adaptable adaptive security oriented methods. Intrusion detection system (IDS) is one of most valuable component is used to find the Internet attacks. In literature, different techniques from different disciplines have been used to develop efficient IDS. Artificial intelligence (AI) dependant techniques plays prominent role in development of IDS and has many benefits over other techniques. However, there is no comprehensive review of AI based techniques to observe and understand the current status of these techniques to solve the intrusion detection problems. In this paper, many AI based techniques have been reviewed focusing on elaborate of IDS. Related studies have been associate by their source of audit data, processing criteria, technique used, dataset, classifier design, feature reduction technique employed and other experimental environment setup Advantages and Disadvantages of AI based techniques have been discussed.

4. Statementof The Problem

In network security various techniques have proposed to detect unauthorized use, misuse and abuse of computer systems by both system insiders and external intruders. Due to the rapidly increasing unauthorized activities, Intrusion Detection System (IDS) as a component of defense-in- depth is very necessary because traditional techniques cannot provide complete protection against intrusion for example:

- The correlation of alarm is not précised,
- The detection and prediction of false positive and false negative rate is high,
- The measurement of abnormal behaviour using the pattern structure has limited scope.

5. Conclusion

As rapid increase in unauthorized activities and abuse of computer system by both system insider and external intruder trends to increase the degree of network security. In order to increase network security various technique has been proposed but having a deficiency over IDS system in some of the situation

i.e. if correlation alarm is not precise, reduction and prevention of false positive and false negative is high , at last having insufficient measurement of pattern recognition. In order to overcome all these deficiency from IDS, system over network, there is need to solve the problem of correlation and SVM theory resolves the problem of unknown and rapidly evolving harmful attacks.

REFERENCES

[1] S. X. Wu and W. Banzhaf, "The Use of Computational Intelligence in Intrusion Detection Systems: A Review," Applied Soft Computing, vol. 10, 2010, pp. 1–35.

[2] Peyman Kabiri and Ali A. Ghorbani, "Research on Intrusion Detection and Response: A Survey", International Journal of Network Security, vol. 1, 2005, pp. 84-102.

[3] C. F. Tsai, Y. F. Hsu, C. Y. Lin, and W. Y. Lin, "Intrusion Detection by Machine Learning: A Review," Expert Systems with Applications, vol. 36, December 2009, pp. 11994– 12000.

[4] K. Shafi and H. A. Abbass, "An Adaptive Genetic-based Signature Learning System for Intrusion Detection," Expert Systems with Applications, vol. 36, 2009, pp. 12036–12043.

[5] A. Ahmed, A. Lisitsa, and C. Dixon, "A Misuse-Based Network Intrusion Detection System Using Temporal Logic and Stream Processing," Proc. International Conference on Network and System Security (NSS 11), 2011, pp. 1 – 8.

[6] S. Petrovic and K. Franke, "A New Two-Stage Search Procedure for Misuse Detection," Proc. International Conference on Future Generation Communication and Networking (FGCN 07), 2007, pp. 418 – 422.

[7] P. G. Teodoro, J. D. Verdejo, G. M. Fernández, and E. Vazquez, "Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges," Computers & Security, vol. 28, 2009, pp. 18–28.

[8] Jonathan J. Davis and Andrew J. Clark, "Data Preprocessing for Anomaly based Network Intrusion Detection: A Review", Computers & Security, vol. 30, 2011, pp. 353–375.

[9] F. Palmieri and U. Fiore, "Network Anomaly Detection through Nonlinear Analysis," Computers & Security, vol. 29, October 2010, pp. 737–755.

[10] M. A. Aydın, A. H. Zaim, and K. G. Ceylan, "A Hybrid Intrusion Detection System Design for Computer Network Security," Computers & Electrical Engineering, vol. 35, May 2009. pp. 517–526.

[11] C. Xiang, P. C. Yong, and L. S. Meng, "Design of Multiple- Level Hybrid Classifier for Intrusion Detection System using Bayesian Clustering and Decision Trees," Pattern Recognition Letters, vol. 29, May 2008, pp. 918–924.

[12] P. Gogoi1, D. K. Bhattacharyya, B. Borah, and J. K. Kalita, "MLH-IDS: A Multi-Level Hybrid Intrusion Detection Method," The Computer Journal, vol. 57, 2013, pp. 602-623.

[13] R. Goel, A. Sardana, and R. C. Joshi, "Parallel Misuse and Anomaly Detection Model," International Journal of Network Security, vol. 14, July 2012, pp. 211-222.

[14] W. Ren, L. Hu, K. Zhao, and J. Chu, "A Multiple-Level Hybrid Intrusion Detection System based on Hierarchical Clustering and Decision Trees," Journal of Computational Information Systems, vol. 9, 2013, pp. 5421–5428.

[15] Varun, C., Arindam, B., Vipin, K.: Anomaly Detection, A Survey. ACM Computing Surveys, 2009,41(3):1-58.

[16] Frey, B.J., Dueck, D. Clustering by Passing Messages between Data Points. Science, 315(5814), 972-976 (2007)

[17] Davis J J, Clark A J. Data pre-processing for anomaly based network intrusion detection. A review [J]. Computers and Security, 2011, 30(6): 353-375.