International Journal of Teknology Informatics and Engineering
e-ISSN :2961-8215;.p-ISSN :2961-9068
Volume 1 No. 2 , Agustus - 2022

Enhancing Performance Using New Hybrid Intrusion Detection System

Candra Supriadi,
Charli Sitinjak,
Fujiama Diapoldo Silalahi,
Nia Dharma Pertiwi,
Sigit Umar Anggono.
Universitas Sains dan Teknologi Komputer

***Abstract*—** Intrusion Detection Systems (IDS) are an efficient defense against network attacks as well as host attacks as they allow network/host administrators to detect any policy violations. However, traditional IDS are vulnerable and unreliable for new malicious and genuine attacks. In other case, it is also inefficient to analyze large amount of data such as possibility logs. Furthermore, for typical OS, there are a lot of false positives and false negatives. There are some techniques to increase the quality and result of IDS where data mining is one of technique that is important to mining the information that useful from a large amount of data which noisy and random. The purpose of this study is to combine three technique of data mining to reduce overhead and to improve efficiency in intrusion detection system (IDS). The combination of clustering (Hierarchical) and two categories (C5, CHAID) is proposed in this study. The designed IDS is evaluated against the KDD'99 standard Data set (Knowledge Discovery and Data Mining), which is used to evaluate the efficacy of intrusion detection systems. The suggested system can detect intrusions and categorize them into four categories: probe, DoS, U2R (User to Root), and R2L (Remote to Local). The good performance of IDS in case of accuracy and efficiency was the result of this study.

Keywords— *data mining, IDS, Clustering data.*

I. INTRODUCTION

The essential component to protect public and private computing infrastructure is Information security technology. The usage of technology applications nowadays makes the security to their resources threatened. The more organization the more threatened their security. Network-intrusion detection is an essential defense mechanism against security threats, which have been increasing in rate lately. It is defined as a special form of cyber threat analysis to identify malicious actions that could affect the integrity, confidentiality, and availability of information resources. Data mining-based intrusion-detection mechanisms are extremely useful in discovering security breaches. An intrusion detection system (IDS) is a component of the computer and information security framework. Its main goal is to differentiate between normal activities of the system and behavior that can be classified as suspicious or intrusive [1].

IDS is required for the number of incidents reported each year, and attack techniques are constantly improving. IDS approaches can be divided into two main categories. Or anomaly detection [2]. The misuse detection method assumes that an intrusion may be detected with the aid of using matching the contemporary pastime with a fixed of intrusive patterns. Anomaly detection structures count on that an intrusion ought to deviate the gadget conduct from its everyday pattern. This approach can be implemented using statistical methods, neural networks, predictive pattern generation and association rules among others techniques. In this study naïve byes classification with clustering data mining techniques is used to extract patterns that represent normal behavior for intrusion detection. This research is describing a variety of modifications that will have made to the data mining algorithms in order to improve accuracy and efficiency. Using sets of naïve byes classification rules that are mined from network audit data as models of “normal behavior.”

To detect anomalous behavior, it will generate naïve byes classification probability with clustering followed from new audit data and compute the similarity with sets mined from “normal” data. If the similarity values are below a threshold value it will show abnormality or normality .

II. RESEARCH DESIGN

This chapter will discuss about the proposed concept of this study. Data mining techniques is used in this study. There are various fields such as marketing, manufacturing, process control, fraud detection and network management that successfully using data mining technique. The research that discusses about implementation of data mining in intrusion detection have been discussed over past five years. This study will apply to data mining for anomaly detection field of intrusion detection. Due to the increasing number of computers connected to publicly accessible networks (such as the Internet), some computer systems cannot ensure network security. There is no ideal solution to prevent event intrusion, so it is very important to detect the event as soon as possible and take the necessary steps to reduce the possibility of damage. One approach to addressing delivery behavior over the network is the attack detection system "IDS". Various techniques are specially applied to detect intrusions, data mining techniques, artificial intelligence techniques, and soft computing techniques. Most data mining techniques such as association rule mining, clustering, and classification have been applied to detect intruders, and classification and pattern mining are important techniques.

Research Design: The research design is using data mining techniques. This data mining and classification technique applies in the field of intruder detection. The anomaly learning approach can detect attacks with high accuracy and high detection rate. However, the false alarm rate is the same as the anomaly approach. To maintain high accuracy and detection rate while reducing false alarm rates, the techniques offered are a combination of three learning techniques. The first step in the proposed approach is to use hierarchical clustering as a pre classification component to group similar data instances based on their behavior. In addition, this C5.0 classifier is used to classify the resulting cluster into attack classes as the final classification task. We have found that data that was misclassified in the previous stage can be correctly classified in the next classification stage. Finally, the CHAID classification was applied.

Proposed Architecture:. The primary concept in the back of that is to use more than one records mining strategies singly to audit the records to calculate an intrusion detection model, consistent with the discovered conduct withinside the records. In the proposed technique, Hierarchical clustering, C5.0 set of rules and CHAID technique are used. It first applies a hierarchical set of rules to the given records set to divide the report records into regular clusters and anomaly clusters. It defines the variety of clusters as 5 to the hierarchy and agencies the facts withinside the records set into regular clusters and anomalous clusters. The cluster anomalies are U2R, R2L, PROBE, and DoS. Records are classified with the cluster index. Then, divide the records set into parts. One element is used for schooling and the alternative element is used for evaluation. In the schooling phase, facts are classified to C5.0 for schooling purposes. Classifier C5.0 facts with classified facts. Then, open the closing facts that aren't classified to C5.0 for classification. Classifier C5.0 will classify unlabeled facts into regular and anomalous clusters. Finally, CHAID which is likewise a classifier that plays the precise cost of every characteristic to be accrued and through putting off the belief of robust independence.

III. RESULT ANALYSIS

Use of java implementation to provide system evaluation. For the time evaluation of the suggested technique, it is necessary to describe a detailed evaluation method. Here we take only one evaluation mode to find the impact of the Intrusion Detection system on the selected technique time. For the experiment using a laptop Pentium® DualCore CPU T00 @2.20Ghz and a 32bit operating system, where data was collected. In the experiment, the laptop executed a data set of fixed records (182679).

Execution time is a measure of the time it takes for a technique to produce results. Throughput is calculated through execution time as evidenced by engineering speed. Memory relates to the amount of storage space required for the entire IDS process. CPU utilization can be calculated that the CPU only for careful calculation method. It calculates CPU load. During the evolution of the results, we have used the KDD99 cup datasets [3 and 4] for training and testing. In 1998, an evaluation of the DARPA detection program was carried out to download a raw TCP/IP dump data [5] for LANs by MIT Lincoln lab to compare the performance of various intrusion detection methods [6 and 7]. In the KDD99 data set, each record consists of a feature set, some of which are discrete or continuous. Qualitative values are non-consecutive labels which can be symbolic or numeric values, e.g. the feature protocol value is one of the symbols. The numeric value of the login feature is 0 or 1 to represent whether the user has successfully logged in or not. For quantitative attributes, the data are characterized by numerical values in a finite interval. An example could be duration. Since feature selection applies only to discrete attributes, not continuous attributes, continuous features are converted to discrete attributes prior to feature selection analysis. To find this method we have used the KDD99 data set [8].

First, apply the K-means clustering hierarchical clustering algorithm on the selected features. After that, we classified the data obtained into Normal or Anomaly clusters by using the Hybrid classifier which is a combination of (Classification C5.0 and CHAID). In this experiment the results of the packet performance, time consuming, memory usage and CPU usage of the algorithm are known on the size of the record set. During collection, the record set comes from the database. For the evaluation mode, there are two parameters: the number of record sets Owned and the size of the initiated record set, where the number of record sets generated is the number of randomly generated record sets and the size of the record set that can be selected from the base data. In this mode, n cycles (that is, the estimated number of record sets) are executed. In each cycle, the respective record sets are executed by the proposed technique. Finally, the outputs of the proposed evaluation system are packet execution, execution time, and execution measured in seconds. Actually, for an algorithm, the execution time that takes not only depends on the algorithm of the algorithm, but also the size of the record set.

The research results are advantages of the proposed technique in terms of notification time, Memory Utilization and CPU Utilization. In general, algorithms that are known to take time usually depend on the size of the record set.

- The proposed Hybrid technique produces good performance then compares the techniques to find a normal performance package.
- The proposed hybrid technique has a lower response time than the comparison technique.
- The proposed hybrid technique uses low memory space during execution compared to the compared technique and is easy to implement and implement.
- The proposed hybrid technique uses a simple structure, the flow is well defined and structure repetition is also minimized. Due to the following facts it takes very little time for execution.

IV. CONCLUSION

The research carried out has increased the speed and accuracy of detection which is the main concern of the proposed work, and presents a cluster rule mining method with a classification method to abnormal detection based on the network. The Presented Approach is a hybrid approach which is a combination of K-mean clustering, K-nearest and the Decision Table Majority rule-based approach. The proposed approach is compared and approximation to the KDD'99 dataset. Considering the relationship between alerts, it proposes an improved cluster with a k-nearest classification; This hybrid approach can find more accurate probabilities of normal and abnormal packets. Compared to other methods, the proposed method can find probabilities from training data and test data with high efficiency. Usually when an attack is carried out, it is very likely that the attack cluster transition will occur. it is based on the sequence of clusters to filter out false alarms generated by the IDS, experimental results prove this method is effective and feasible. Future work should pay more attention to concentration or attention on data mining processes that do not fall into the category of feature selection and anomaly detection. To deal with some of the common challenges in data mining, it may be best to develop a special purpose solution tailored to intrusion detection.

REFERENCES

[1] Om, H. and Kundu, A. “A hybrid system for reducing the false alarm rate of anomaly intrusion detection system” Recent Advances in Information Technology (RAIT), 1st IEEE International Conference on 15-17 March 2012 Page(s):131 - 136 Print ISBN:978-1- 4577-0694-3.

[2] P.R Subramanian and J.W. Robinson “Alert over the attacks of data packet and detect the intruders” Computing, Electronics and Electrical Technologies (ICCEET), IEEE International Conference on 21-22 March 2012 Page(s):1028 - 1031 Print ISBN:978-1-4673-0211-1

[3] V. S. Ananthanarayana and V. Pathak “A novel Multi-Threaded K-Means clustering approach for intrusion detection” Software Engineering and Service Science (ICSESS), IEEE 3rd International Conference on 22-24 June 2012 Page(s): 757 - 760 Print ISBN: 978-1-4673-2007-8

[4] N.S Chandolikar and V.D.Nandavadekar, “Efficient algorithm for intrusion attack classification by analyzing KDD Cup 99” Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on 20-22 Sept. 2012 Page(s):1 - 5 ISSN :2151-7681

[5] Virendra Barot and Durga Toshniwal “A New Data Mining Based Hybrid Network Intrusion Detection Model” IEEE 2012.

[6] Wang Pu and Wang Jun-qing “Intrusion Detection System with the Data Mining Technologies” IEEE 2011.

[7] Z. Muda, W. Yassin, M.N. Sulaiman and N.I. Udzir “Intrusion Detection based on K-Means Clustering and Naïve Bayes Classification” 7th IEEE International Conference on IT in Asia (CITA) 2011.

[8] Dewan M.D. Ferid, Nouria Harbi, “Combining Naïve Bayes and Decision Tree for Adaptive Intrusion detection” International Journal of Network Security and application(IJNSA),vol 2, pp. 189-196, April 2010.