Asian Journal of Engineering and Applied Technology
ISSN 2249-068X Vol. 8 No. 1, 2019, pp. 20-24

Association Rule Mining for Intrusion Detection System: A Survey

D. Selvamani[1] and V. Selvi[2]
[1]Research Scholar, [2]Assistant Professor
[1&2]Department of Computer Science, Mother Teresa Women's University, Kodaikanal, Tamil Nadu, India
E-Mail: selvamani.bhaskar@gmail.com

Abstract **- Many modern intrusion detection systems are based on data mining and database-centric architecture, where a number of data mining techniques have been found. Among the most popular techniques, association rule mining is one of the important topics in data mining research. This approach determines interesting relationships between large sets of data items. This technique was initially applied to the so-called market basket analysis, which aims at finding regularities in shopping behaviour of customers of supermarkets. In contrast to dataset for market basket analysis, which takes usually hundreds of attributes, network audit databases face tens of attributes. So the typical Apriori algorithm of association rule mining, which needs so many database scans, can be improved, dealing with such characteristics of transaction database. In this paper, a literature survey on the Association Rule Mining has carried out.**

Keywords: **Data Mining, Network based Intrusion Detection System, Association Rule Mining, Apriori Algorithm**

I. INTRODUCTION

Almost networks are protected by firewalls. However these firewalls are not always effective against the emerging intrusion attempts. Various methods based on knowledge development and data mining can help to improve Intrusion Detection Systems (IDSs). Data mining is also known as Knowledge Discovery in Database (KDD). The purpose of data mining is to abstract interesting knowledge from the large database. From the analysis of abstracted patterns, decision-making process can be done very easily. Many modern intrusion detection systems are based on data mining and database-centric architecture [4], where a number of data mining techniques have been found. Data mining-based intrusion detection systems can be classified into misuse detection and anomaly detection. Misuse detection attempts to match observed activity to known intrusion patterns. This is typically a classification problem. Anomaly detection attempts to identify behaviour that does not conform to normal behaviour. This approach has a better chance of detecting novel attacks. During recent years, association rule mining, which is one of the important topics in data mining research, is used for anomaly detection. This approach determines interesting relationships between large sets of data items.

II. ASSOCIATION RULE MINING

Association rule-mining is a method to find the relation between the variables in large volumes of data. Strong relationships between the variables are identified to help in creating association rules. While defining the rules, there are various measures used in the databases. The association rule contains two parts. The first antecedent part is used to define, if' part with the help of conjunction and disjunction operator. The second consequent part is used to define the, then' part to find the solution depending on the combination with the antecedent part. The association rules are created by examining the volume of data for frequent if/then patterns and using support and confidence value to classify the majority important relationship between databases. The support is used to indicate how the items frequently appear in the database. The confidence is used to indicate the number of times if/then statement is found in the databases. Normally, the database contains millions of data for processing. The association rule-mining algorithm is a suitable method for discovering relations between variables in large volumes of databases. The method is simple and easy to determine the relation in the large databases.

In the KDD dataset, number of instances is represented by $n = \{N_1, N_2, N_3, .., \infty\}$ and number of features is represented by $m = \{c_1, c_2, c_3, .., c_{40}\}$. An association rule-mining algorithm is proposed to determine relationships among KDD dataset which is represented as X Y $\Rightarrow$. The association between X and Y are whenever X appears Y also be likely to appear. X and Y may be single condition or set of conditions. X is called as the rule's antecedent part and also Y is called as consequent part.

1. Support The rule X $\Rightarrow$ Y holds with support supp if supp % of transactions in KDD Dataset contains X U Y. The rule that has a supp greater than a user-specified support is called as minimum support.

$$Support\ (X) = \frac{Number\ of\ times\ X\ appears}{Total\ Number\ of\ Records}$$

$$Support\ (XY) = \frac{Number\ of\ times\ X\ and\ Y\ appears\ together}{Total\ Number\ of\ Records}$$

2. The rule X $\Rightarrow$ Y holds with confidence conf if conf % of the transactions in KDD Dataset that contain X also contain Y . The rule that has a conf greater than a user-specified confidence is termed as minimum confidence.

$$Confidence\ (X \rightarrow Y) = \frac{Support\ (XY)}{Support\ (X)}$$

The numbers of records are used to find the frequencies of all the features and generate the frequent 1-itemsets based on the minimum support values. Then the candidate 2-itemsets are generated to find the frequency of candidate 2-itemsets. The minimum support value helps to generate the candidate 3-itemsets until the process ends.

III. LITERATURE SURVEY ON ASSOCIATION RULE MINING

Mehrotra, Latika, Prashant Sahai Saxena, and Nitika Vats Doohan [1] a novel model is presented here with decision tree concepts for the data classification. Model that is suggested in this paper is based on the updated ID3 method. It uses a modified gain to select the attribute. This modified gain gives more weightage to most important attribute.

Santra, Palash, *et al.,* [2] This paper proposed expert system architecture for forensic intrusion monitoring, analysis, and evidence generation for cloud logs. Fuzzy data mining technique has been proposed for forensic acquisition. This will reduce the computational effort that would otherwise incur in processing the huge log to identify the attacked area. Further AI techniques are exploited for training and analysis purpose. This helps in identifying various anomalous attacks in cloud environment.

Chen, Hsing-Chung, and Shyi-Shiun Kuo [3] the aim of this paper is to provide an approach based on the association rule mining technique for traffics appeared in the integrated web services, such as HTTP, HTTPS, and FTP traffic, in order to discover the strong attack features or patens of DoS attacks. Association rule mining is employed in this paper to deal with the DoS patens and then find out the strong relations among features of DoS attacks in large well-known dataset, e.g. NSL-KDD. The strong relations which are determined on when the major attack features are discovered from the open dataset would be considered as the strong patterns of DoS attacks.

Muyeed Ahmed, Mir Tahsin Imtiaz, Raiyan Khan and Rashedur M. Rahman [4] Traffic is one of the major problems for any populated city. Currently, there are many traffic alert systems available and almost all of them work with user submitted inputs to give those alerts. We have worked on developing a system that will not depend on any user's manual input. Rather it will be able to retrieve traffic and activity related data from the user's device and vehicle tracking devices automatically to predict traffic and alert users. Our system understands the user's activity using accelerometer sensor data and speed to determine whether the user is sitting at home or going somewhere by a bus or car. Once it is verified that the particular user's location and activity is related to traffic conditions, it takes that user's location related data from his or her device. Using this data from user's devices and the data from vehicle tracking devices, we predict the traffic conditions and let users know about the traffic for particular routes.

Heraguemi, Kamel Eddine, Nadjet Kamel, and Habiba Drias [5] this paper deals with a cooperative multi-swarm bat algorithm for association rule mining. It is based on the bat-inspired algorithm adapted to rule discovering problem (BAT-ARM). This latter suffers from absence of communication between bats in the population which lessen the exploration of search space. However, it has a powerful rule generation process which leads to perfect local search. Therefore, to maintain a good trade-off between diversification and intensification, in our proposed approach, we introduce cooperative strategies between the swarms that already proved their efficiency in multi-swarm optimization algorithm (Ring, Master-slave).

Mehrotra, Latika, and Prashant Sahai Saxena [6] it is found that intrusion detection systems (IDSs) that are signature-based are restricted in their areas of detecting intrusions, because of the fact that the signature-based intrusion detection system is based on matching a signature with the network details. The system using signatures or patterns can detect only known attacks and threats, but they mostly fail when it comes to novel attacks. Thus preventing/detecting the new or special types of attracts whose signature is not specified. Although signature-based IDS does not give false alarms at genuine cases, but still is inept for unknown attacks or masked attacks. Later in the paper, another category of IDS is discussed which is statistical-based intrusion detection system (SBIDS).

Lu, Nannan, *et al.*, [7] in this paper, an evolving mechanism is introduced to extract the rules for intrusion detection. To extract diversified rules as well as control the quantity of rulesets, the extracted rules are examined according to the distance between the rules in the rule set of the same class and the rules in the rule set of different classes. Thereby, it alleviates the problem that the quantity of rules expands unexpectedly with the evolving genetic network programming. The simulations are conducted on a benchmark intrusion dataset, and the results show that the proposed method provides an effective solution to evolve the class association rules and improves the intrusion detection performance.

Gupta, Chetan, Amit Sinhal, and Rachana Kamble [8] in this paper, we suggest a hybrid framework based on association rule mining (ARM) and ant colony optimization (ACO). Combining the properties of association and ant colony may provide better classification in comparison with the previous methodology. In our approach, we consider the dataset of NSL-KDD. It is a dataset that does not include redundant record, and test sets are reasonable which is mentioned. Then, we consider equal proportion of 10,000 dataset from the whole dataset. We first divide it into two parts based on normal establishment and termination. Then, we consider the normal dataset, and for finding the intrusions, we calculate the support value based on the matching factor. Then, we apply ACO technique to check the global optimum value. If the value crosses the limit value, then the node will be added into the final attack category.

Shingo Mabu, Shun Gotoh, Masanao Obayashi, Takashi Kuremoto [9] This paper describes a classification system with random forests, employing weighted majority vote in the classification to enhance its performance. For the performance evaluation, NSL-KDD (Network Security Laboratory-Knowledge Discovery and Data Mining) data set is used and the proposed method is compared with the conventional methods, including other machine-learning techniques (Random forests, SVM, J4.8) in terms of the accuracy and false positive rate.

Khamphakdee, Nattawat, Nunnapus Benjamas, and Saiyan Saiyod [10] proposed a procedure for improving Snort IDS rules, based on the association rules data mining technique for detection of network probe attacks. We employed the MIT-DARPA 1999 data set for the experimental evaluation. Since behavior pattern traffic data are both normal and abnormal, the abnormal behavior data is detected by way of the Snort IDS. The experimental results showed that the proposed Snort IDS rules, based on data mining detection of network probe attacks, proved more efficient than the original Snort IDS rules, as well as icmp.rules and icmp-info.rules of Snort IDS.

Parkinson, Simon, Vassiliki Somaraki, and Rupert Ward [11] a novel method of modelling file system permissions which can be used by association rule mining techniques to identify irregular permissions is presented. This results in the creation of object-centric model as a by-product. This technique is then implemented and tested on Microsoft's New Technology File System permissions (NTFS). Empirical observations are derived by making comparisons with expert knowledge to determine the effectiveness of the proposed technique on five diverse real-world directory structures extracted from different organisations. The results demonstrate that the technique is able to correctly identify irregularities with an average accuracy rate of 91%, minimising the reliance on expert knowledge.

Vinutha, H. P., B. Poornima, and B. M. Sagar [12] Data mining approaches in the field of Intrusion Detection System (IDS) is becoming more popular. The outlier is a current problem faced by many data mining researches. Outliers are the patterns which are not in the range of normal behavior. Outliers in the dataset produce more false positive alarms, and this has to be reduced to increase the efficiency of IDS. We have used Interquartile Range technique to identify the outliers in the NSLKDD'99. In this, the continuous range of input is divided into quartiles and these quartiles are analyzed to target the range of outliers. Then the obtained outliers are removed by a filter called remove with value. The experiment is conducted using Weka data mining tool.

Tiwari, Ravi Raman, Anil Kumar Singh, and Vrijendra Singh [13] In this paper, we intend to propose a SIEM system with the self-learning capability which can produce optimized and efficient correlation directives for analyzing events in a network, system etc. with the least possible human intervention. We propose a SIEM system with classification-based directives, utilising association rule mining to discover relationships between the event logs and generate rules, based on which we construct classifiers which can distinguish between normal and abnormal behaviour.

Dutt, Inadyuti, *et al.,* [14] The proposed research work is based on such a hybrid system which uses misuse detection system for known types of intrusions and anomaly detection system for novel attacks. Frequency episode extraction method is specifically used for misuse-based detection and chi-square test is used for anomaly-based detection. Experiments show that the hybrid intrusion detection system is able to consider the real-time traffic of a network as well as the standard available data set for detecting the efficiency of the system. The proposed system learns and trains itself by detecting known attacks from misuse detection system and novel attacks from anomaly detection system, thereby improving the true positive rates and diminishing false negative rates consequently.

Mabu, Shingo, Wenjing Li, and Kotaro Hirasawa [15] in this paper, probabilistic classification algorithms based on multi-dimensional probability distribution are proposed and combined with conventional class association rule mining of GNP, and applied to network intrusion detection for the performance evaluation. The proposed classification algorithms are based on 1) one-dimensional probability density functions and 2) a two-dimensional joint probability density function. These functions represent the distribution of normal and intrusion accesses and efficiently classify a new access data into normal, known intrusion or even unknown intrusion.

Kaur, Gagandeep, Amit Bansal, and Arushi Agarwal [16] in this paper, we have compared results of detection techniques for SbDS and AbDS for big datasets. Under AbDS, wavelets have been used as a signal processing tool to compute Hurst Index (H), used as a measure for computing degree of self-similarity in network traffic. Deviations beyond threshold were used to detect presence of network anomalies. Under SbDS, two main classification techniques based on J48 and Naïve Bayes have been used to explore the possibilities of having best achievable accuracy with least number of parameters from a big dataset of 41 features. The results of both methodologies have been analyzed for choosing appropriate technique under given constraints.

Herrera-Semenets, Vitali, *et al.,* [17] In this paper, we propose a new data preprocessing strategy which reduces the number of features and instances in the training collection without greatly affecting the achieved accuracy of IDSs. Finally, our proposal is evaluated using four different rule-based classifiers, which are tested on real scan and backscatter data collected by a network telescope.

Jie, Xinchun, *et al.*, [18] Because of the high integration of control, communication, computer and network technology, how to deal with various anomaly behaviors of control systems is a problem that should be solved by researchers. Especially some activities such as data injections, DoS attacks and device failure must be considered. Based on the analysis of dynamic behaviors of industrial process control systems with varying process state variables, a data mining method is proposed on summarizing normal behavior features of the control systems. Depending on association rules, a similarity factor is formulated using a real-time data mining method for describing the likeness between real-time frequent itemsets and normal frequent itemsets. Representative values of change behaviors for process variables and the corresponding generation method are illustrated in detail. On the basis of comparison between several real-time frequent itemsets and the normal frequent itemsets, a reliability parameter is given to describe the abnormal status of a control system within a certain time. Simulation results show that the proposed method can detect anomaly behaviors of a process control system in time, which has practical significance in industrial applications.

Chan, Gaik-Yee, Fang-Fang Chua, and Chien-Sing Lee [19] In this paper, we present our fuzzy association rule-based (FAR) and fuzzy associative pattern-based (FAP) intrusion detection and prevention (IDP) systems in defending against WS attacks at the SaaS level. Our experimental results have validated the capabilities of these two IDP systems in terms of detection of known attacks and prediction of new variant attacks with accuracy close to 100%. For each transaction transacted over the Cloud platform, detection, prevention or prediction is carried out in less than five seconds.

Chandrashekhar, Azad, and Jha Vijay Kumar [20] in this chapter, a novel intrusion detection system has been proposed which is based on the fuzzy min max neural network. The objective of the proposed intrusion detection system is to protect the end user system from the scope of various types of cyber-attacks. The main hurdles in the today's intrusion detections system are the nonlinear separablity, online adaption, preprocessing of the network logs, attribute selection, and the learning of the desired system for the anomalous or the signature detection. The proposed system is tested on the KDD Cup 99 dataset, and the classification accuracy and classification error are used for performance evaluation.

IV. CONCLUSION

Since the substantial advancements in computer era, mining of big data to gain useful knowledge has been a hot topic studied from several aspects. One of the important fields of research is ARM which aims to uncover the subtle relations between the entries of huge bulk data so that some meaningful rules of associations can be generated. The obtained association rules can be exploited to identify which instances correlate in certain dimensions. First, with a view of anomaly cases being relatively rarely occurred in network packet database, we define a rare association rule among infrequent item sets rather than the traditional association rule mining method. In this paper, a literature survey on the Association Rule Mining has carried out for framing the attack rules structure as the future work.

REFERENCES

[1] Mehrotra, Latika, Prashant Sahai Saxena, and Nitika Vats Doohan, "A Data Classification Model: For Effective Classification of Intrusion in an Intrusion Detection System Based on Decision Tree Learning Algorithm", *Information and Communication Technology for Sustainable Development*. Springer, Singapore, pp. 61-66, 2018.

[2] Santra, Palash, *et al.*, "Fuzzy Data Mining-Based Framework for Forensic Analysis and Evidence Generation in Cloud Environment", *Ambient Communications and Computer Systems*. Springer, Singapore, pp. 119-129, 2018.

[3] Chen, Hsing-Chung, and Shyi-Shiun Kuo, "DoS Attack Pattern Mining Based on Association Rule Approach for Web Server", *International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*. Springer, Cham, 2018.

[4] Ahmed, Muyeed, *et al.*, "Clustering and association rule mining-based traffic analysis and prediction of Dhaka", *International Journal of Knowledge Engineering and Data Mining*, Vol. 5 No. 4, pp. 241-276, 2018.

[5] Heraguemi, Kamel Eddine, Nadjet Kamel, and Habiba Drias, "Multi-swarm bat algorithm for association rule mining using multiple cooperative strategies", *Applied Intelligence*, Vol. 45, No. 4, pp.1021-1033, 2016.

[6] Mehrotra, Latika, and Prashant Sahai Saxena, "An Assessment Report on: Statistics-Based and Signature-Based Intrusion Detection Techniques", *Information and Communication Technology*. Springer, Singapore, pp. 321-327, 2018.

[7] Lu, Nannan, *et al.*, "Intrusion Detection System Based on Evolving Rules for Wireless Sensor Networks", *Journal of Sensors*, 2018.

[8] Gupta, Chetan, Amit Sinhal, and Rachana Kamble, "An Enhanced Associative Ant Colony Optimization Technique-based Intrusion Detection System", *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*. Springer, New Delhi, pp. 541-553, 2015.

[9] Mabu, Shingo, *et al.*, "A random-forests-based classifier using class association rules and its application to an intrusion detection system", *Artificial Life and Robotics*, Vol. 21, No. 3, pp. 371-377, 2016.

[10] Khamphakdee, Nattawat, Nunnapus Benjamas, and Saiyan Saiyod, "Improving intrusion detection system based on snort rules for network probe attacks detection with association rules technique of data mining", *Journal of ICT Research and Applications*, Vol. 8, No. 3, pp. 234-250, 2015.

[11] Parkinson, Simon, Vassiliki Somaraki, and Rupert Ward, "Auditing file system permissions using association rule mining", *Expert Systems with Applications*, Vol 55, pp. 274-283, 2016.

[12] Vinutha, H. P., B. Poornima, and B. M. Sagar, "Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset", *Information and Decision Sciences*. Springer, Singapore, pp. 511-518, 2018.

[13] Tiwari, Ravi Raman, Anil Kumar Singh, and Vrijendra Singh, "Self-Learning SIEM System Using Association Rule Mining", *Journal of Advanced Database Management & Systems*, Vol. 2, No. 2, pp. 10-23, 2015.

[14] Dutt, Inadyuti, *et al.*, "Real-Time Hybrid Intrusion Detection System Using Machine Learning Techniques", *Advances in Communication, Devices and Networking*. Springer, Singapore, pp. 885-894, 2018.

[15] Mabu, Shingo, Wenjing Li, and Kotaro Hirasawa, "A Class Association Rule Based Classifier Using Probability Density Functions for Intrusion Detection Systems", *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 19, No. 4, pp. 555-566, 2015.

[16] Kaur, Gagandeep, Amit Bansal, and Arushi Agarwal, "Wavelets Based Anomaly-Based Detection System or J48 and Naïve Bayes Based Signature-Based Detection System: A Comparison", *Ambient*

Communications and Computer Systems. Springer, Singapore, pp. 213-224, 2018.

[17] Herrera-Semenets, Vitali, *et al.*, "A data reduction strategy and its application on scan and backscatter detection using rule-based classifiers", *Expert Systems with Applications*, Vol. 95, pp. 272-279, 2018.

[18] Jie, Xinchun, *et al.*, "Anomaly behavior detection and reliability assessment of control systems based on association rules", *International Journal of Critical Infrastructure Protection*, 2018.

[19] Chan, Gaik-Yee, Fang-Fang Chua, and Chien-Sing Lee, "Intrusion detection and prevention of web service attacks for software as a service: Fuzzy association rules vs fuzzy associative patterns", *Journal of Intelligent & Fuzzy* Systems, Vol. 31, No. 2, pp. 749-764, 2016.

[20] Chandrashekhar, Azad, and Jha Vijay Kumar, "Fuzzy Min-Max Neural Network-Based Intrusion Detection System", *Proceedings of the International Conference on Nano-electronics, Circuits & Communication Systems*. Springer, Singapore, 2017.