CLASSIFICATION PROCEDURES FOR INTRUSION DETECTION BASED ON KDD CUP 99 DATA SET

Shaker El-Sappagh, Ahmed Saad Mohammed, Tarek Ahmed AlSheshtawy

Faculty of Computers & Informatics, Benha University, Egypt.

ABSTRACT

In network security framework, intrusion detection is one of a benchmark part and is a fundamental way to protect PC from many threads. The huge issue in intrusion detection is presented as a huge number of false alerts; this issue motivates several experts to discover the solution for minifying false alerts according to data mining that is a consideration as analysis procedure utilized in a large data e.g. KDD CUP 99. This paper presented various data mining classification for handling false alerts in intrusion detection as reviewed. According to the result of testing many procedure of data mining on KDD CUP 99 that is no individual procedure can reveal all attack class, with high accuracy and without false alerts. The best accuracy in Multilayer Perceptron is 92%; however, the best Training Time in Rule based model is 4 seconds . It is concluded that ,various procedures should be utilized to handle several of network attacks.

KEYWORDS

Intrusion Detection, Data Mining, KDD CUP 99, False Alarms

1. INTRODUCTION

Communication system plays an inevitable role in common people's daily life. Computer networks are effectively used for business data processing, education and learning, collaboration, widespread data acquisition, and entertainment [1]. With the enormous growth of computer networks usage and internet accessibility, more organizations are becoming susceptible to a wide variety of attacks and threats [2]. One of the main challenges in the security management of large-scale high-speed networks is the detection of suspicious anomalies in network traffic patterns due to distributed denial of service (DDoS) attacks or worm propagation [3].

Generally, the major focus of the network attacks is to increase the threat against the commercial business and our daily life, so it becomes a serious problem for the researchers to find a suitable solution for these types of attacks [4]. Network security is becoming an absolute necessity to protect information contained in the computer systems worldwide. With the rapid expansion of computer networks during the past decade, the network grows in size and complexity, and computer services expansion, vulnerabilities within the local area and wide area network become a huge problem [5]. Nowadays, network security is a world hot topic in computer security and defense. Intrusions, attacks, or anomalies in network infrastructure lead mostly in great financial losses and massive sensitive data leaks. Therefore, they decrease the efficiency and quality of productivity of organizations [6]. Reliance on Internet and world wide connectivity has increased the potential damage that can be inflicted by attacks launched over Internet against remote systems. Successful attacks inevitably occur despite the best security precautions [7].

Intrusion detection system (IDS) is a program that tries to find indications that the computer has been compromised [8]. It attempts to detect an intruder breaking into computer system or

DOI: 10.5121/ijnsa.2019.11302

legitimate user misuses system resources. Intrusion detection is an important issue and has captured the attention of network administrators and security professionals. It is the art of detecting unauthorized, inappropriate, or anomalous activity on computer systems. IDSs are classified as network based, host based, or application based depending on their mode of deployment and data used for analysis [9]. In addition, IDSs can also be classified as signature based or anomaly based depending upon the attack detection method. The signature-based systems are trained by extracting specific patterns (or signatures) from previously known attacks while the anomaly-based systems learn from the normal data collected when there is no anomalous activity [10, 11]. The main purpose of IDS is to detect as many attacks as possible with the minimum number of false alarms. In other words, the system must be accurate in detecting attacks. However, accurate systems that cannot handle large amount of network traffic and is slow in decision making will not fulfill the purpose of an intrusion detection system [12].

The huge issue in IDs is presence of huge number of false alerts; this issue is being motivated by several experts to discover the solution for minifying false alerts. . This research will be presented as various data mining classification for handling this issue in IDSs as reviewed.

2. RELATED WORK

This section briefly discusses many techniques of classification used for classifying intrusion detection datasets including decision trees, Bayesian classification, artificial neural network, support vector machines, associative classification, and k-nearest neighbor. Classifiers have been suggested and developed to reduce false alarm of intrusion detection in the area of network security based on different ideas.

Warrender et al. [13] have proposed several intrusion detection methods based on system call trace data. They tested a method that utilizes sliding windows to determine a database of normal sequences to form a database for testing against test instances and classify instances according to those in the normal sequence database. This requires the maintenance of a large database of normal system call trace sequences. Wenke et al. [14] proposed the Mining Audit Data for Automated Models for Intrusion Detection project. It is one of the best known data mining projects in intrusion detection. It is an off-line IDS to produce anomaly and misuse intrusion detection models for network and host systems. Association rules and frequent episodes are applied to replace hand-coded intrusion patterns and profiles with the learned rules. Agarwal et al. [15] proposed a two-stage general-to-specific framework for learning a rule-based model (PNrule) to learn classifier models on KDD 99 data set. Barbara et al. [16] proposed (Audit Data Analysis and Mining), which is an intrusion detector built to detect intrusions using data mining techniques. It first absorbs training data known to be free of attacks. Next, it uses an algorithm to group attacks, unknown behavior, and false alarms.

Abraham [17] proposed (Intrusion Detection using Data Mining Technique), which is a real-time NIDS for misuse and anomaly detection. It applies association rules, Meta rules, and characteristic rules. It employs data mining to produce a description of network data and uses this information for deviation analysis. Zhang et al. [18] proposed a statistical neural network classifier for anomaly detection, which can identify UDP flood attacks. Comparing different neural network classifiers, the back propagation neural network (BPN) has shown to be more efficient in developing IDS. . Xu et al. [19] presented a framework for adaptive intrusion detection based on machine learning. Multi-class Support Vector Machines (SVMs) is applied to classifier construction in IDSs. Li et al. [20] though realized the deficiencies of KDD dataset, developed a supervised network intrusion detection method based on Transductive Confidence Machines for K-Nearest Neighbors (TCM-KNN) machine learning algorithm and active learning based training data selection method. Panda et al. [21] study performance of three well known data mining classifier algorithms namely, ID3, J48, and Naïve Bayes are evaluated on the KDD CUP 99 data set. Mohammed et al. [22] proposed a comprehensive analysis classification techniques are used to predict the severity of attacks over the network. They compared zero R classifier, Decision table classifier and Random Forest classifier with KDD CUP 99 databases from MIT Lincoln laboratory. Sathyabama et al. [23] used clustering techniques to group user's behavior together depending on their similarity and to detect different behaviors and specified as outliers.

Chihab et al. [24] presented five data mining algorithms like (ID3, Naive Bayes, Random forest, C4.5, and multilayer perceptron) to make the comparison between them which applied on network intrusions and get the best proposal of a hybrid classifier based on naïve Bayes and random forest algorithms. The results shows that the hybrid system improved the prediction with reduced, consuming time. Keerthika et al,[25] provided proposal, which focuses on the naïve feature reduction, in addition to feature selection methods such as gain ratio and information gain for reducing the redundant and irrelevant of features. This proposal used naïve Bayes classifier to design intrusion detection system. Tesfahun et al. [26] suggested an effective hybrid layered intrusion detection system by combining misuse and anomaly IDS for detecting both previously known and unknown attacks. The first layer consisted of misuse detector, which is based on random forest classifier for detect and stop known attacks; the second layer involved anomaly detector was built using bagging technique with a staff of one class (SVM) classifiers. The results showed that system can detect previously unknown attacks with a detection rate improvement of (18.73%) by using NSL-KDD dataset. Aggarwal et al. [27] assessed several classification algorithms like Random Forest, Naïve Bayes, C4.5, and Decision Table. They compared these classification algorithms in WEKA with KDD99 dataset. These classifiers were resolved according to metrics like accuracy, precision, and F-score. Random Tree displays the best outcomes aggregate in contrast to the algorithms, which have high detection and low false alarm rate were C4.5 and Random Forest. Mukund et al. [28] proposed the existing algorithms for intrusion detection system to introduce an improved way of using the HDFS (Hadoop Distributed File System). So to reduce the false alarm rate, they used decision tree technique and enhanced it in the process with the multi-system capabilities of the HDFS. Therefore this approach reduced the time taken by the DFS and improved the accuracy of the IDS.

Gupta et al. [29] IDSs monitors the network or malicious activities and forbidden access to devices. IDSs used to protect the data's features and integrity. The proposal was used NSL-KDD dataset to learn the manner of the attacks depending on the methods of data mining such as logistic regression and K-means clustering. Hence, it generates rules for classifying network activities. The results show that linear regression was very effective accuracy in detecting attacks was (80%) while the K-means clustering was showed kind results with (67%) accuracy. Akashdeep Sharma et al. [30] work proposes an intelligent system which first performs feature ranking on the basis of information gain and correlation. Feature reduction is then done by combining ranks obtained from both information gain and correlation using a novel approach to identify useful and useless features. These reduced features are then fed to a feed forward neural network for training and testing on KDD99 dataset. Kabir et al. [31]. Proposes a novel approach for intrusion detection system based on sampling with Least Square Support Vector Machine (LS-SVM). Decision making is performed in two stages. In the first stage, the whole dataset is divided into some predetermined arbitrary subgroups. The proposed algorithm selects representative samples from these subgroups such that the samples reflect the entire dataset. In the second stage, least square support vector machine (LS-SVM) is applied to the extracted samples to detect intrusions on KDD 99 database which is considered a de facto benchmark for evaluating the performance of intrusions detection algorithm.

Liyu Duan and Youan Xiao [32] large volume of the data and unbalanced data, intrusion data were inevitable obstacles. So, solve those issues utilizing the fuzzy c-means procedure to reconstruct feature vectors according to central points. Nathan et al [33] This paper shows deep learning procedure for intrusion detection, which implemented in GPU-enabled TensorFlow and evaluated utilizing KDD 99 dataset. Osamah et al [34] In this paper, introduced learning procedure for intrusion detection according to tree calculation on the KDD-99.

3. INTRUSION DETECTION SYSTEM

IDS can be defined as a combination of software and/or hardware components which monitors computer systems and makes an alarm when an intrusion occurs [35]. The basic architecture of IDS is shown in Figure 1.

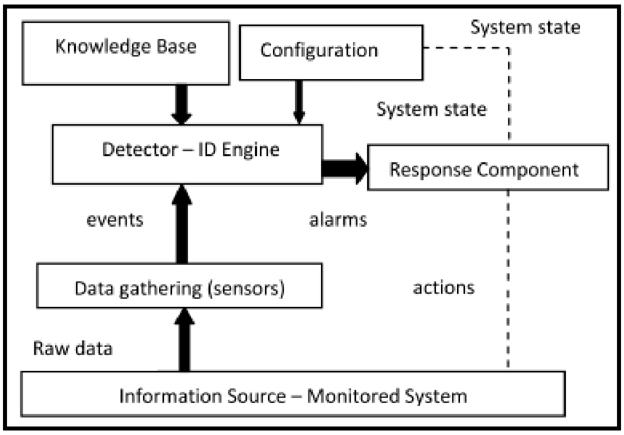


Figure 1: Basic Architecture of IDS [36].

The components in this architectural framework are as follows [36].

- *Data Gathering Device*: responsible for collecting the data from the monitored system.
- *Detector–ID Engine*: processes the data collected from sensors to identify intrusive behavior and send an alarm signal to response component if there is an intrusion.
- *Knowledge Base*: contains pre-processed information provided by network experts and collected by sensors.
- *Configuration Device*: provides information about the current state of IDS.
- *Response Component*: initiates the response (active or inactive) when intrusion is detected.

4. EVOLUTION OF INTRUSION DETECTION

There are many metrics for evaluating the IDS performance. The following is a description of some of these metrics [37]:

- *Predictive accuracy:* The two measures used for evaluating the predictive performance of IDS are: (i) detection rate and (ii) false alarm rate. Detection Rate (DR) also known as True Positive Rate (TPR) is defined as the ratio of number of attacks correctly detected to the total number of attacks, while the False Alarm (false positive) Rate (FAR) is the ratio of the number of normal connections that are incorrectly classified as attacks to the total number of normal connections.

- *True Positive (TP):* IDS producing an alarm when a legitimate attack occurs. False Positive (FP): IDS producing an alarm when no attack occurs. False Negative (FN): IDS producing no alarm when the actual attack occurs. True Negative (TN): IDS producing no alarm when no attack occurs

- *Receiver Operating Characteristics (ROC):* Evaluation of IDS can also be performed using Receiver Operating Characteristics (ROC). ROC graphs depicts trade-offs between detection rate and false alarm rate. In ROC , the point that corresponds to 0% false alarm rate and 100% detection rate represents the perfect IDS (Foster Provost, Tom Fawcett)

- *performance Time:* The performance time of IDS is the total time taken by IDS to detect the intrusion

5. Intrusion Detection Dataset

In this section, brief description of KDD Cup 1999 dataset which was derived from the 1998 DARPA Intrusion detection Evaluation program is provided. It is the most widespread dataset collected over a period of nine weeks for a LAN simulating a typical U.S. Air Force LAN. The dataset contains a collection of simulated raw TCP dump data, where, multiple intrusions attacks was introduced and widely used in the research community from seven weeks of network traffic. The dataset contains 4,898,430 labeled and 311,029 unlabeled connection records. The labeled connection records consist of 41 attributes. In network data of KDD99 dataset, each instance represents feature values of a class, where each class is categorized either normal or attack. The classes in the dataset are characterized into one normal class and four main intrusion classes [38] :

- *Normal:* connections are generated by simulating user behavior.
- *DoS attacks:* use of resources or services is denied to authorized users.
- *Probe attacks:* information about the system is exposed to unauthorized entities.
- *User to Remote attacks:* access to account types of administrator is gained by unauthorized entities.
- Remote to Local attacks: access to hosts is gained by unauthorized entities.

6. Data Mining And Intrusion Detection

Data mining is the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories [39]. Classification is a data mining technique, which arranges data into predefined groups. The goal of predictive classification is to predict the target class accurately for each record in a set of new data, that is, data that is not in the historical data [40]. Intrusion detection can be defined as a classification problem where each audit record can be classified into one of a discrete set of possible categories (i.e. normal or a particular kind of intrusion). Intrusion detection using data mining have attracted more and more interests in recent years by utilizing procedures programs applied to audit data to compute misuse and anomaly detection models [41].

7. Implemented Data Mining Methods Based On Kdd Cup 99

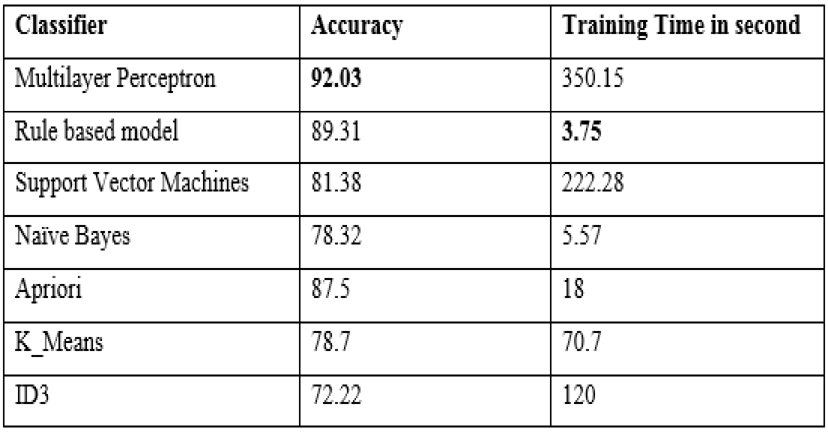
In IDSs, there are important surveys of implemented data mining methods on KDD Cup 99 by various experts.

- *Multilayer perceptron for classification of KDD dataset*: it is consider as the most usually neural network procedure according to one layer for input, hidden and output [41] [42].
- *Rule based model:* simple procedure usually with good rules for depicting the framework in data [43].
- *Support vector machines*: it is a procedure for converting the training data to a feature scope hence getting the best splitting hyperplane [44].
- *Naïve Bayes:* it is a simple procedure according to probabilistic relined underlay an individual structure [44].
- *Apriori Association Rule Mining Algorithm:* It is the procedure to giving frequent item sets according to the dataset and making scan process to identify, most frequent items [45].
- *K Means clustering*: it is a procedure for tasking of dataset points to clusters according to the distance between dataset points and cluster centroid [46].
- *ID3, C4.5, and C5.0 decision tree algorithms:* its procedure for building a decision tree for classification dataset according to training data [47].

8. RESULTS AND DISCUSSION

The fundamental aim of this research is to decide the excellent procedure of data mining procedures to classify KDD99 so as to has a high accuracy and low time in knowing attacks. Furthermore, smoothing the mission of select for expert's in the future on KDD dataset, Good implementing cases of all the seven procedures aforementioned over were assessed. Results are given in the Table 1 to compare the classifiers; for IDs, utilizing accuracy and Training Time for knowing the best procedure for the classifier. Simply as predictably that no sole procedure can reveal all attack class, with high accuracy and without false alarm ratio. The best accuracy in Multilayer Perceptron is 92% however the best Training Time in Rule based model is 4 seconds.

Table 1: Comparison of Seven Procedures



Classifier	Accuracy	Training Time in second
Multilayer Perceptron	**92.03**	350.15
Rule based model	89.31	**3.75**
Support Vector Machines	81.38	222.28
Naïve Bayes	78.32	5.57
Apriori	87.5	18
K_Means	78.7	70.7
ID3	72.22	120

9. CONCLUSION

Many data mining procedures have been focused by the experts in IDS scope and they purpose to minify the great load of analyzing massive quantity of data. KDD Cup'99 data set is suffering from a variance inter the classes which impact miserable detection and is a main issue to data mining procedures. An important issue in designing IDS is minified false alarm ratio and attaining high detection ratio. Utilizing various classification procedures, potential to minify false alarm enhances the detection accuracy and many classification procedures. In this research, utilized by the experts in the performance of IDS structure are discussed and reviewed. From the tentative survey applied this paper specified that various experts suggest various procedures for IDS scope in various group , but still, needed to search.

REFERENCES

[1] Aleksandar Lazarevic, Levent Ertoz, Vipin Kumar, Aysel Ozgur, Jaideep Srivastava, “A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection”

[2] Denning D. E, An intrusion-detection model, IEEE Transactions on Software Engineering, vol. SE-13, no. 2, pp.222-232.

[3] Zesheng C., L. Gao and K. Kwiat, “Modeling the Spread of Active Worms”, Twenty Second Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), vol. 3, pp:1890-1900, 2003.

[4] W. Lee and S. J. Stolfo, “Data Mining Approaches for Intrusion Detection”, the 7th USENIX Security Symposium, San Antonio, TX, January 1998.

[5] Moradi M., Zulkernine M., 2003, “A Neural Network Based System for Intrusion Detection and Classification of Attack”, Natural Science and Engineering Research Council Canada (NSERC).

[6] Symantec Enterprise.: Internet Security Threat Report 2016. https://www.symantec.com/content/dam/symantec/docs/reports/istr-21-2016-en.pdf. [accessed 18.03.17].

[7] Arman Tajbakhsh, Mohammad Rahmati, and Abdolreza Mirzaei, "Intrusion detection using fuzzy association rules", Applied Soft Computing ASOC509, Elsevier B.V, 2008.

[8] J.Hu, Host-Based Anomaly IDS. Springer Handbook of Information and Communication Security, Springer Verlag, 2010, ISBN978-3-642-04116-7 (Print), 978-3-642-04117-4 (Online)

[9] H Wang, J Cao, and Y Zhang, ”A flexible payment scheme and its role-based access control”, IEEE Transactions on knowledge and Data Engineering, vo. 17, no. 3, 425–436, 2005.

[10] Y. Zhang, Y. Shen, H. Wang, Y. Zhang, X. Jiang, “On Secure Wireless Communications for Service Oriented Computing,” IEEE Transactions on Services Computing, no.1, pp. 1.

[11] D. Wang, Z. Zhang, P. Wang, J. Yan, and X. Huang, ”Targeted Online Password Guessing: An Underestimated Threat,” ACM Conference on Computer and Communications Security, pp. 1242-1254, 2016

[12] K.K. Gupta, B. Nath and R. Kotagiri, “Layered Approach Using Conditional Random Fields for Intrusion Detection,” IEEE Transactions on Dependable and Secure Computing, vol. 7, no. 1, pp. 35–49, 2010.

[13] Warrender C., Forrest S. and Pearl M.,“Detecting Intrusions Using System Calls: Alternative Data Models”, in IEEE symposium on security and privacy, pp:133-145, 1999.

[14] Wenke L. and S. J.Stolfo, "A Framework for Constructing Features and Models for Intrusion Detection Systems”, ACM transactions on Information and system security (TISSEC), vol.3, Issue 4, Nov 2000.

[15] Agarwal R., Joshi M.V., “PNrule: A New Framework for Learning Classifier Models in Data Mining”, Tech. Report, Dept. of Computer Science, University of Minnesota, 2000.

[16] Daniel B., J.Couto, S.Jajodia, and N.Wu, "ADAM: A Test Bed for Exploring the Use of Data Mining in Intrusion Detection”, SIGMOD, vol30, no.4, pp: 15-24, 2001.

[17] Abraham T. , "IDDM: Intrusion Detection Using Data Mining Techniques”, Technical report DSTO electronics and surveillance research laboratory, Salisbury, Australia, May 2001.

[18] Zheng Z., J. Li, C.N. Manikapoulos, J.Jorgenson, J.ucles, "HIDE: A Hierarchical Network Intrusion Detection System Using Statistical Pre-Processing and Neural Network Classification”, IEEE workshop proceedings on Information assurance and security, pp:85-90, 2001.

[19] Xu X., “ Adaptive Intrusion Detection Based on Machine Learning: Feature Extraction, Classifier Construction and equential Pattern Prediction”, International Journal of Web Services Practices 2(1-2), pp:49–58, 2006.

[20] Li Y., Guo L., “An Active Learning Based TCM-KNN Algorithm for Supervised Network Intrusion Detection”, In: 26 th Computers and Security, pp: 459–467, October 2007.

[21] Mrutyunjaya P. and M. Ranjan Patra, ” Evaluating Machine Learning Algorithms for Detecting Network Intrusions”, International Journal of Recent Trends in Engineering, vol. 1, no.1, May 2009.

[22] Mohammed M Mazid, M. Shawkat Ali, Kevin S. Tickle,“A Comparison Between Rule Based and Association Rule Mining Algorithms “, Third International Conference on Network and System Security, 2009.

[23] Sathyabama S., Irfan Ahmed M., Saravanan A,”Network Intrusion Detection Using Clustering: A Data Mining Approach”, International Journal of Computer Application (0975-8887), vol. 30, no. 4, Sep. 2011.c

[24] Chihab Y. , Ouhman A., Erritali m. and Ouahidi B.,2013," Detection & Classification of Internet Intrusion Based on the Combination of Random Forest and Naïve Bayes ", International Journal of Engineering and Technology (IJET), 2013

[25] Keerthika G. and Priya D. S.," Feature Subset Evaluation and Classification using Naive Bayes Classifier ", Journal of Network Communications and Emerging Technologies (JNCET) Volume 1, Issue 1, March (2015) 2015

[26] Tesfahun A. and D.Bhaskari L., ,"Effective Hybrid Intrusion Detection System: A Layered Approach", IJCNIS, vol.7, no.3, pp.35-41, 2015

[27] Aggarwal P. and Sharma S.K.," An Empirical Comparison of Classifiers to Analyze Intrusion Detection", Proc. of Fifth International Conference an Advanced Computing and Communication Technologies, 2015.

[28] Mukund Y. and Nayak S., ‘Improving false alarm rate in intrusion detection systems using Hadoop’, 21-24 Sept, International Conference. Vol.3 , 2016

[29] Gupta D., Singhal S, Malik S. and Singha., Network intrusion detection system using various data mining techniques, IEEE publication, 2016.

[30] Akashdeep Sharma ,Ishfaq Manzoor, Neeraj Kumar, A Feature Reduced Intrusion Detection System Using ANN Classifier, Expert Systems With Applications (2017)

[31] E. Kabir, J. Hu, H. Wang, G. Zhuo, A novel statistical technique forintrusion detection systems, Future Generation Computer Systems (2017)

[32] L. Duan and Y. Xiao, "An Intrusion Detection Model Based on Fuzzy C-means Algorithm,", 8th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, pp. 120-123. (2018)

[33] Wang, Zheng. "Deep learning-based intrusion detection with adversaries." IEEE Access 6 , 38367-38384.(2018):

[34] Raheem Esraa and Saleh Alomari ,"An Adaptive Intrusion Detection System by using Decision Tree Osamah Adil", Journal of AL-Qadisiyah for computer science and mathematics Vol.10 No.2,(2018).

[35] Chen M.S., Han J and Yu Philip S., Data Mining: An Overview from a Database Perspective, IEEE Transactions on Knowledge and Data Engineering, vol.8,No.6,1996,pp.866-883.

[36] Christine Dartigue, Hyun IK Jang, Wenjun Zeng, A New data-mining based approach for network Intrusion detection, Proc. of Seventh Annual Communication Networks and Services Research Conference, 2009, pp.372-377.

[37] Foster Provost, Tom Fawcett, Robust Classification for Imprecise Environment, 2000, pp.1-38, Kluwer Academic Publishers.

[38] Chawla N.V, Bowyer K.W, Hall L.O, Kegelmeyer W.P, Smote: Synthetic minority oversampling technique, Journal of Artificial Intelligence Research, vol.16, 2002, pp.321–357.

[39] Dewan Md. Farid, Nouria Harbi, Mohammad Zahidur Rahman , Combining Naive Bayes and Decision Tree for Adaptive Intrusion Detection, Proc. of Intl. Journal of Network Security & Its Applications (IJNSA), Volume 2, Number 2, 2010, pp.12-25.

[40] Domingos P. and Pazzani M., Beyond Independence: Conditions for the optimality of the simple Bayesian Classifier, In proceedings of the 13 th. Conference on Machine Learning, 1996, pp.105-110.

[41] Yeung D. Y. and Chow C., "Prazen-window Network Intrusion Detectors", In: 16 th International Conference on Pattern Recognition, Quebec, Canada, pp:11–15, August 2002.

[42] Yeung D. Y. and Chow C., "Prazen-window Network Intrusion Detectors", In: 16 th International Conference on Pattern Recognition, Quebec, Canada, pp:11–15, August 2002

[43] Witten I. H. and Frank E., " Data Mining: Practical Machine Learning Tools and Techniques", 2 nd edn. Morgan Kaufmann, San Francisco, 2005.

[44] Huy A. N., D. Choi ," Application of Data Mining to Network Intrusion Detection: Classifier Selection Model ", pp:1, 2008.

[45] Mohammed M Mazid, M. Shawkat Ali, Kevin S. Tickle,"A Comparison Between Rule Based and Association Rule Mining Algorithms ", Third International Conference on Network and System Security, 2009.

[46] Kusum K. Bharti, S. Shukla and S. Jain , "Intrusion detection using clustering", vol.1, issue 2, 3, 4, pp.6, 2010.

[47] Amanpreet C., G. Mishra, G. Kumar, "Survey on Data Mining Techniques in Intrusion Detection" , vol: 2, issue.7, pp:2, 2011.

AUTHORS

Shaker El-Sappagh received the bachelor's degree in computer science from the Information Systems Department, Faculty of Computers and Information, Cairo University, Egypt, in 1997,the master's degree from Cairo University, in 2007,and the Ph.D. degree in computer science from the Information Systems Department, Faculty of Computers and Information, Mansura University,Mansura, Egypt, in 2015. In 2003, he joined the Department of Information Systems, Faculty of Computers and Information, Minia University, Egypt, as a Teaching Assistant. Since 2016, he has been an Assistant Professor with the Department of Information Systems, Faculty of Computers and Information, Benha University. He is currently a Postdoctoral Fellow with the UWB Wire-less Communications Research Center, Department of Information and Communication Engineering, Inha University, South Korea. He has publications in clinical decision support systems and semantic intelligence. His current research interests include machine learning, medical informatics, (fuzzy) ontology engineering, distributed and hybrid clinical decision support systems, semantic data modeling, fuzzy expert systems, and cloud computing. He is very interested in the diseases diagnosis and treatment researches. He is a Reviewer for many journals

Ahmed saad Mohammed received the bachelor's degree from the Software engineering Department, Baghdad College of Economic Sciences University, Iraq, Baghdad in 2005. His current research interests include machine learning, data mining and artificial intelligence

Dr. Tarek El-Shishtawy is a professor of Information System. His current work is vice Dean of postgraduates and researches at faculty of computers and informatics. The scientific interests in clude Information Retrieval, Data Mining, and researches related to information systems in developing countries. Dr. Tarek published and refereed many articles in NLP.