IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 10, October 2019

A Novel Intrusion Detection System using Decision Support System and Data Mining Techniques

Prajeeth Kumar M.J[1], Boominathan P[2]

School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, India[1, 2]

Abstract: The Internet plays a vital role in day-to-day human life activities. It became essential to protect these human activities from unknown internet threats such as Cyber terrorism, Identity theft and many others from the same category. There exist many approaches, which deliver security to some extent, but the ultimate goal of the efficient intrusion detection system is still a challenging task. A data mining based intrusion detection system is proposed in this paper. The proposed intrusion detection system ensures the use of feature extraction and feature selection for data mining and processing. A Packet sniffer based approach works well for network packet tracking, which is used by the proposed intrusion detection system. Data mining along with proper decision support tool can work effectively for intrusion detection. The proposed system works efficient and accurate when tested with KDDCup'99 data.

Keywords: Decision Support Tool, Intrusion Detection, Feature Eradication, Entropy Function, Feature Election

I. INTRODUCTION

As network-based systems play an increasingly important role in advance and modern society, providing security to the systems which are in-network becoming an important and major issue than ever before. As per the Pentagon's 2005 statistics, 79,500 intrusion attacks were encountered among them 1,800 were successful attacks. Hence in order to deal with rapidly variation in patterns of muggers, computer scientists have strict vigil on its sophistication. Various ID systems that are able to avoid attacks are created by CS communities. Attack detection systems are categorized into two main types where Unknown detection systems help to detect changes in the normal behavior of models which are based on large data sets and the corruption detection system compares the system's behavior with pattern extracted from previously known violations. But both of this models carry their advantages and disadvantages because unknown detection systems are able to detect new attacks but suffers from high wrong assumption where corruption detection systems detect known attacks with very high accuracy with the help of pattern matching mechanism but unable to detect novel one as they don't carry patterns of new attacks. The proposed system deals with corruption detection.

Attackers are in a role in each field of communication and network which is in today's world more connected human beings. As most of the money transactions, sensational data transmission, heavy business deals, defense deals are taking place through the web. Intruders are waiting for the opportunity when he gets the chance to enter in other secure place and steal and harm to them. In today's world, many electronic media are suffering from this problem as they can get easily affected by intruders. So in order to make them secure day-night researcher are in a role to devolve such invention which gives more security to the world with fewer expenses and make their life secure from others.

Earlier, some artificial intelligent systems make the use of decision support tools, effective algorithms and programming, naïve Bayes approaches in order to detect intrusions in systems. But there were some issues while choosing attributes from large data sets as they are not generating proper results and responding late in order to give results. As large data sets are the combination of compatible as well as incompatible features which causes noise and redundancy into design of systems, so In worst case, it is difficult to choose better parts of features which help in improving the performance and efficiency of system, hence, so far decision tree classifiers have not been tried for intrusion detection. This paper proposes a good solution to find intrusion by analyzing the packets for decision supporting tool classification when dealing with the large data sets by taking these four types of attacks in the role: R2L, DOS, Probe, and U2R. The mechanism of eradication of underground data from massive datasets is known as Data Mining. The actual data mining mechanism is the mechanical and semi-mechanical analysis of massive data to extract previously anonymous alluring patterns like a bunch of data sets, uncommon records. Data mining's contribution in the area of network security dependent IDS in sequence allowing analysts to target and determine novel real-time attacks that occur in the network by discarding normal data from highly intense data sets. The distorted alarm signals are used for determining abnormal activities that expose absolute attacks.

DOI 10.17148/IJARCCE.2019.81006

II. RELATED WORK

Weiming Hu.et.al [1], proposed two online Adaboost-based intrusion detection algorithms. The first algorithm uses decision stump as weak classifiers which are totally based on a traditional online Adaboost process. The second algorithm, based on online Gaussian Mixture Models (GMMs) uses weak classifiers. Along with this work, they proposed a distributed intrusion detection framework, in which a local parameterized detection model is constructed in each node using the online Adaboost algorithm.

Gudadhe M [3], proposed the novel approach for detecting intrusion over a network by making the efficient use of the decision tree concept. In this paper, the author tried to work on many novel methods by keeping the agenda of intrusion detection as more as possible in a short period of time.

Yun Wang.et.al [4], has analyzed the problems which Gaussian distributed WSN faces by distinguishing detection probability on the basis of applications that are required and some network parameters for single as well as multiple sensing approaches. In this paper, he tried to examine the effects of detection probability for various network parameters.

Sanjay Kumar Sharma.et.al [5], proposed an anomaly related ID system by using k-means clustering with the help of naïve Bayes classifiers. For the evaluation of performance with other algorithms, authors made the use of KDD'99 data sets. In which he tried to judge his approach in terms of Efficiency, wrong alarm and intrusion finding rate. By making these performance tests he came to conclude that his approach gives better outcomes than NIDS.

Mrutyunjaya Panda.et.al [6], proposed a system for intrusion detection in which he tried to figure out the outcomes of the number of rule-based classifiers. For that, he did experiment with JRip, NNge, and RIDOR with the help of the ensemble concept with the help of KDDCup'99 datasets. For getting the betterment of this method the author did the various comparison with the number of information mining methods where he got success to entitle his approach as better as compared to some other methods.

Christine Dartigue.et.al [7], proposed an intrusion detection technique with the help of binary classifiers for various types of attacks which is totally based on data mining by selecting optimal features and increasing the speed of the attack detection. This technique used to advance the intrusion detection process by catching the attacks which occur infrequently in data. This classifier used to give accuracy by combining the nearest correct values.

Safaa Zaman.et.al [8], proposed the same concept but with more clarity as he discussed a very new method "ESVDF" for feature selection from the large data sets. This method works on two essential concepts which consist of the rank of the feature which is evaluated by the (SVDF) and interaction between the features which are resolved either with the help of (FSR) or (BER). For his experiments for feature selections, he used DARPA datasets.

Mitchell R.et.al [2], proposed the intrusion detection system which captures the intrusion on the basis of the behavior of the packet to secure head-ends, data aggregation points of the modern electrical grid. The proposed system tried to investigate the impact of the attacker by using head-ends and data aggregation points.

Shina Sheen.et.al [9], has defined the benefits of machine learning approaches like decision tree in intrusion detection. For experiment purpose, they have taken data which is a large dimension and filter them for getting optimal features by applying feature selection methods.

Dayu Yang.et.al [10], proposed an approach for detecting intrusion from the network while online in co-operating with feature eradication. The author compared his experiments by taking the online and offline scenario and succeeded to give better results in the online network. In associate with this approach, a novel agreement blend method is engaged to corporate outcomes from various classifiers in order to gain more efficiency. For the experiment, the purpose the author has used KDD Cup 99 data sets.

ZHI-XIN YU.et.al [11], proposed a flexible data mining related ID model for the varieties of data. The authors discussed two algorithms, among them heuristic bundling algorithm for the verity of data to differentiate unknown behavior from known behavior and fuzzy-mining algorithm for creating intrusion similarities database.

Tarek Abbes.et.al [12], proposed the mixture of identical patterns and various ways of protocol study where the first approach of disclosure works on a multi-pattern recognition approach, and another approach takes advantage of an adequate decision tree characteristics.

DOI 10.17148/IJARCCE.2019.81006

III. PROPOSED WORK

The proposed intrusion detection systems working methodology is explained in the below diagram.

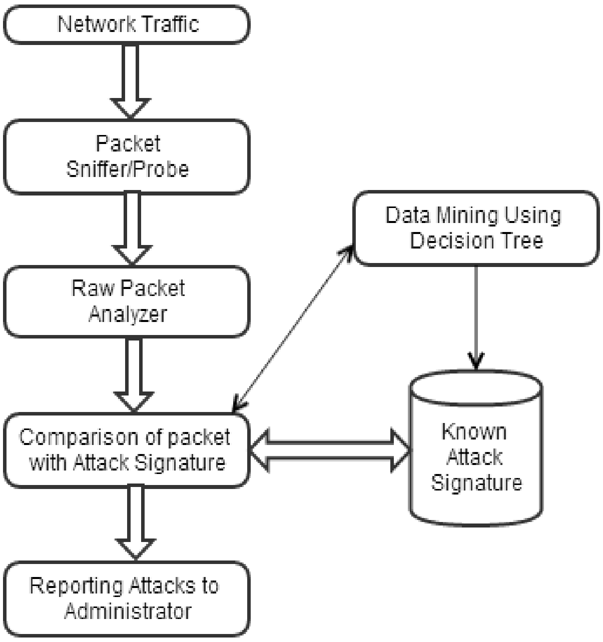

Network Traffic
Packet Sniffer/Probe
Raw Packet Analyzer
Comparison of packet with Attack Signature
Reporting Attacks to Administrator
Data Mining Using Decision Tree
Known Attack Signature

Fig. 1 Flow Diagram of Proposed Model

Step 1) Data present in the network is known as the network traffic. Network traffic contains data that is transferred from one host system to another using a different network path and packet format. Network traffic contains useful packets as well as packets that are a possible threat to the network. The proposed intrusion system considers the above-mentioned traffic as input for processing. Network traffic has different behavior in the network; the rate of traffic burst in-network at a particular time can be very high or very low. The proposed intrusion detection system considers this kind of traffic behavior in the system for the detection of intrusion.

Step 2) Packet Sniffer: The packet sniffer is packet software that enables the capturing of packets that are passing through the network. The packet contains data as well as different types of fields. Using the packet sniffer software, the proposed intrusion detection system captures the various important fields from the packet.

Step 3) Raw packet analyzer: In this step, the captured data from the packet sniffer is analyzed for the classification. The proposed system captures header information and tries to classify the packets based on the already known attack in the system. The separated information of the packet is given as input to the next phase of classification and data mining.

Step 4) Comparison of Packet with attack signature: This step compares captured packet information for the possible attack detection. Packet information is compared for each type of attack present in the system database. If the packet fields matched with the already present attack signature in the database, then that packet is marked as an intrusion. For the remaining packets, entropy is calculated. The entropy calculation is done for the mention of the field in the algorithm of the proposed intrusion detection system.

Step 5) Data mining and Decision tree for attack detection: the proposed system uses the decision tree to detect the possibility of attack for the packet. The decision tree calculates the entropy of the packet for each type of attack using the algorithm of the proposed intrusion detection system. Based on the entropy calculation, the proposed intrusion detection system detects the attack for a particular packet. Data mining is used in this step for retrieving already known attack information. For the entropy calculation, the information about the previously-stored packet is also needed. To retrieve this information from the database the proposed intrusion detection system uses data mining.

Step 6) Reporting attack to the Administrator: When the attack is found out, it is informed to the end-user. If the attack is of the known type, then it is presented to the administrator with that attack information. If a new type of attack behavior is detected by the system, then it is also presented to the admin. Admin can configure the rules for newly detected attack behavior. The rules for the new attack are stored into the attack database for future intrusion detection.

A. Algorithm for the Proposed Intrusion Detection System

The proposed intrusion detection system uses the packet header information as well as the packet data for the attack detection. The mathematical model uses the entropy calculation and parameter comparison to detect the attack. The model of proposed intrusion detection is described below:

DOI 10.17148/IJARCCE.2019.81006

Input: packet (p), Attack signature existing in system (alarm), Attack information (DB)
Output: packet (p) is an attack or normal.

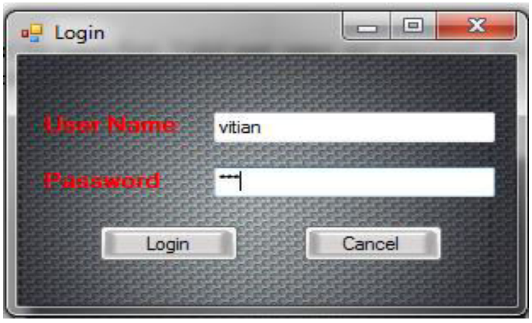
***Algorithm*:**

1. Separate out the header information of each packet that is coming into the system. Header information contain source IP (S), destination IP(d), size of packet(z), port id(p) and data information (d)
2. For each of the attacks stored in the database retrieve the behavior information of the attack like Source IP (Si), destination IP, packet size, port id.
3. For each packet in the system follow the following steps for attack detection:

a. If S > Si, input packet size is greater than the threshold packet size of Attack then detect this packet as Ping of death attack. Set the Alarm as the ping of death.
b. Else if d! =readable, input data from the packet contains garbled values that are not readable by the system; also the header information is corrupted then alarm the packet as a NUCK attack.
c. Else if the summation of P > Threshold, the number of the packet for a particular time is more than the threshold then mark alarm as a SMURF attack.
d. Else if check the other parameters of the packet for the different attack features and is known attack found then set alarm for the known attack.
e. Else consider the packet as a normal packet.

IV. PROPOSED INTRUSION DETECTION SYSTEM WORKFLOW

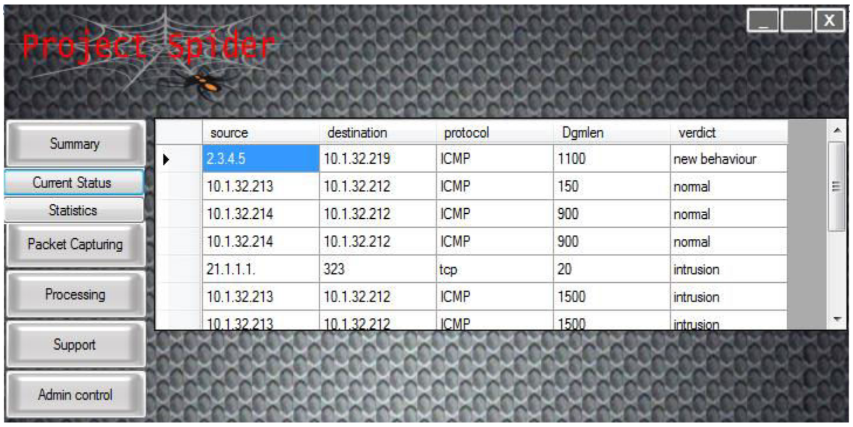
Step 1: Login screen prompted for user login. A valid user name and password allows the user to proceed further into the system. The login screen is as shown in figure 2.



User Name: vitian
Password: ***

Fig. 2 Login Screen

Step 2: The main page shows the statistics of the system as shown in figure 3. Statistics show the real-time tracing of the packet that is coming into the system from the network. Statistics include Source IP, Destination IP, protocol, packet length, and verdict information. Users can check the status of the packet on the summary page. The summary page shows the behavior of the statistics for the fixed interval of time.

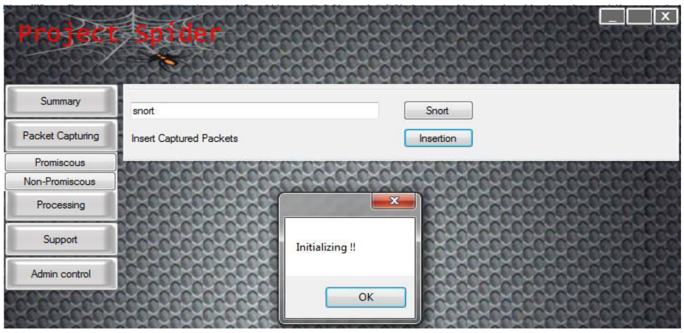


source	destination	protocol	Dgmlen	verdict
2.3.4.5	10.1.32.219	ICMP	1100	new behaviour
10.1.32.213	10.1.32.212	ICMP	150	normal
10.1.32.214	10.1.32.212	ICMP	900	normal
10.1.32.214	10.1.32.212	ICMP	900	normal
21.1.1.1.	323	tcp	20	intrusion
10.1.32.213	10.1.32.212	ICMP	1500	intrusion
10.1.32.213	10.1.32.212	ICMP	1500	intrusion

Fig. 3 Main Summary Page

Step 3: For capturing real-time network data, the user can select the packet capture command on the UI. Packet capture command will start capturing packet, which the user wants to test for possible attack descriptions. Figure 4 shows the UI for the Packet capturing step.

DOI 10.17148/IJARCCE.2019.81006



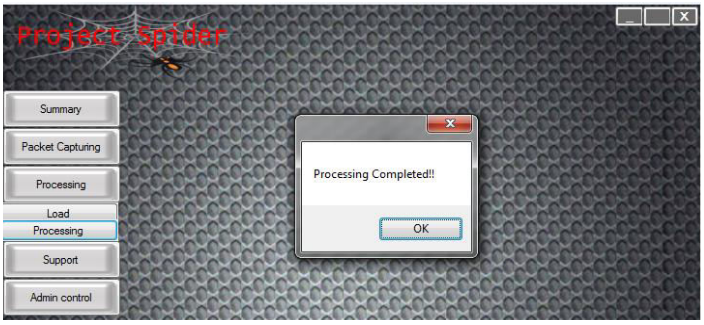
snort

Insert Captured Packets

Initializing !!

Fig. 4 capture packet UI

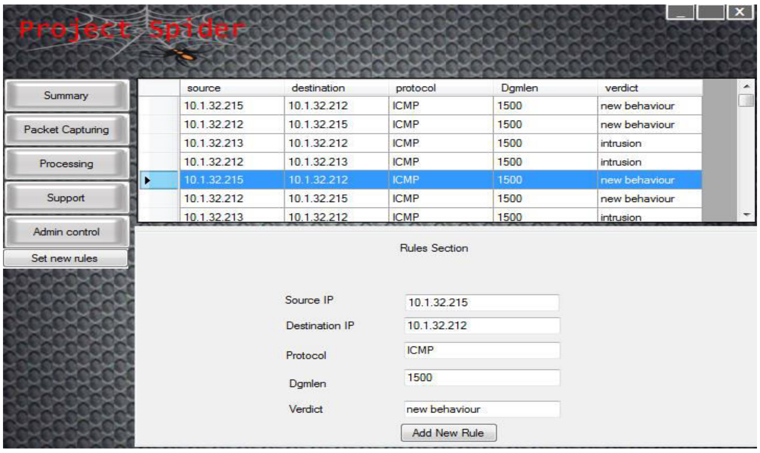
Step 4: Processing, after capturing data for a particular interval of time, the proposed intrusion detection algorithm for attack detection is applied to the captured data. Captured data is processed for the possible attack and the notification is given to the end-user about the possible attack in the packet. Figure 5 shows the processing UI for the system. After this, the proposed algorithm works and processes available network intrusions. Once the processing is completed, the summary detail about the captured packets is shown in the summary form shown in step 2.



Processing Completed!!

Fig. 5 Processing UI for system

Step 5: System allows the user to add the new type of behavior or the new type of attack to the system. When the data is captured and results are generated then depending upon the statistics of the data, the user can manually go and change the behavior of the unknown data to attack or one attack behavior to another behavior. Figure 6 shows the UI for the new rule configuration.



source	destination	protocol	Dgmlen	verdict
10.1.32.215	10.1.32.212	ICMP	1500	new behaviour
10.1.32.212	10.1.32.215	ICMP	1500	new behaviour
10.1.32.213	10.1.32.212	ICMP	1500	intrusion
10.1.32.212	10.1.32.213	ICMP	1500	intrusion
10.1.32.215	10.1.32.212	ICMP	1500	new behaviour
10.1.32.212	10.1.32.215	ICMP	1500	new behaviour
10.1.32.213	10.1.32.212	ICMP	1500	intrusion

Rules Section

Source IP: 10.1.32.215
Destination IP: 10.1.32.212
Protocol: ICMP
Dgmlen: 1500
Verdict: new behaviour

Fig. 6 UI for new rule configuration

V. RESULT ANALYSIS

The proposed intrusion detection algorithm provided better results when experimented with KDDcup'99 data sets. These experiments are carried out with various types of algorithms such as Naïve Bayes, kNN, eClass1, etc. eClass1 and eClass2 are Dynamic Bayesian Networks. In this scenario, the proposed model generates better results than other algorithms when dealing with a large number of packets. This model initially tracks packets, then process and categorize them in intrusions and normal packets.

DOI 10.17148/IJARCCE.2019.81006

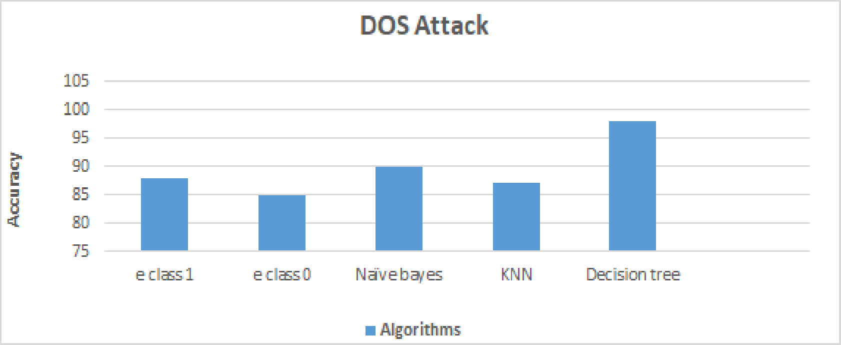


Fig. 7 Comparison of algorithms

The proposed intrusion detection algorithm tested on approximately 2GB of data sets, which takes around 25-30 minutes to filter the packets with respective categories. After getting the result, it has been observed that packets are filtered with unknown attacks, normal behavior, intrusion, etc. where normal behavior packets are the packets that are already in the database, the system assumes that this is known attacks and not harmful. Unknown behaviors are the new one, which is newly arrived, and intrusion is the attacked packed which is more harmful. After analyzing the packets statistics, the administrator can rename unknown behavior as normal or intrusion based on its behavior. After comparing it with the results of other relevant algorithms, we found that the proposed intrusion detection algorithm gives more accuracy over them.

VI. CONCLUSION

The proposed intrusion detection system gives better results when tested with similar algorithms. The combination of decision support tools and data mining techniques enhances the overall working of the algorithm. The proposed intrusion detection system generates better outcomes with a large number of packets. This outcome categorizes the packets in normal behavior, unknown behavior, and intrusion. This categorization is based on the respective occurrence of the matched packet in the database. The proposed intrusion detection system ensures more accurate results with Denial of Service than other relevant algorithms. Efficient computational cost, speed, and accuracy are some key characteristics of the proposed intrusion detection system. The proposed intrusion detection system allows admin to add new rules for newly detected behavior. The proposed intrusion detection system is tested with KDD/DARPA data sets. For better analysis or the results of the proposed intrusion detection system, the number of experiments with different data sets and a variety of data is needed.

REFERENCES

[1]. Weiming Hu, Jun Gao "Online Adaboost-Based Parameterized Methods for Dynamic Distributed Network Intrusion Detection". IEEE Transactions on Cybernetics Volume-44 (1), pp. 66-82, 2014

[2]. Mitchell R, Ing-Ray Chen "Behavior-Rule Based Intrusion Detection Systems for Safety-Critical Smart Grid Applications". IEEE Transactions on Smart Grid, Volume-4 (3), pp. 1254-1263, 2013.

[3]. Li Yun, Liu Xue-cheng, Zhu Feng "Application of data mining in intrusion detection". International Conference on Computer Application and System Modeling (ICCASM), pp. 153-155, 2010.

[4]. Yun Wang "Gaussian versus Uniform Distribution for Intrusion Detection in Wireless Sensor Networks". IEEE Transactions on Parallel and Distributed Systems, Volume-24 (2), pp. 342-355. 2014.

[5]. Sanjay Kumar Sharma, Pankaj Pande, Susheel Kumar Tiwari and Mahendra Singh Sisodia "An Improved Network Intrusion Detection Technique based on k-Means Clustering via NaIve Bayes Classification", IEEE International Conference on ICAESM, pp. 417-422, 2012.

[6]. Mrutyunjaya Panda, Manas Ranjan Patra "Ensembling Rule-Based Classifier for Detecting Network Intrusions", IEEE Conference on ARTCom, pp. 19-22, 2009.

[7]. Christine Dartigue, Hyun Ik Jang, and Wenjun Zeng "A New Data-Mining Based Approach for Network Intrusion Detection", IEEE Conference on Communication Networks and Services Research, pp. 372-377, 2009.

[8]. Safaa Zaman and Fakhri Karray "Features Selection for Intrusion Detection Systems Based on Support Vector Machines", IEEE International Conference on TENCON, pp. 1-8, 2009.

[9]. Shina Sheen, R Rajesh "Network Intrusion Detection using Feature Selection and Decision tree classifier", IEEE International Conference on TENCON, pp. 1-4, 2008.

[10]. Dayu Yang, Hairong Qi "A Network Intrusion Detection Method using Independent Component Analysis", 19th International Conference on Pattern Recognition (ICPR), pp. 1-4, 2008.

[11]. Zhi-Xin Yu, Jing-Ran Chen, Tian-Qing Zhu "A Novel Adaptive Intrusion Detection System Based on Data Mining", Proceedings of IEEE International Conference on Machine Learning and Cybernetics (Volume: 4), pp. 2390-2395, 2005.

[12]. Tarek Abbes, Adel Bouhoula, Michael Rusinowitch "Protocol Analysis in Intrusion Detection Using Decision Tree", Proceedings of IEEE International Conference on Information Technology: Coding and Computing (Volume 1), pp. 404-408, 2004.