An Approach for Detecting and Preventing DoS Attacks in LAN

Majed Tabash[1], Tawfiq Barhoom[2].

[1]*Faculty of Information Technology, Islamic University – Gazs , Palestine.*

[2]*Faculty of Information Technology, Islamic University – Gaza , Palestine.*

Abstract **-- Nowadays, Denial of service (DoS) attacks, have become a major security threat to networks and to the Internet, DoS is harmful to networks as it delays legitimate users from accessing the server, In general, some researches were done to detect and prevent DoS from occurring in a wide area network (WAN), but fewer researches were done on Local Area Network (LAN.), yet, detecting and preventing DoS attacks is still a challenging task, especially in LAN.**

In this paper, we propose an approach merging methods from data mining to detect and prevent DoS attacks, by using multi classification techniques to achieve a sufficient level of accuracy and reduce false alert alarm.

And secondly, we will evaluate our approach in comparison with other existing approaches.

Our work is based on EGH Dataset to detect DoS attacks, in addition, our approach is implemented using Rapidminer, the experimental results show that the proposed approach is effective in identifying DoS attacks, our designed approach achieves significant results.

In the best case, our accuracy is up to 99.96%, we used two component of security; Snort tool and PfSense firewall, and compared our approach with other approaches, and we found that our approach achieves best accuracy results in most cases.

Keywords**— Data Mining, DoS attacks, intrusion detection, Misuse Detection, Multi Classification.**

I. INTRODUCTION

Due to the growing rate of communications between computer systems, organizations have become increasingly depending on information being stored and processed on network-based systems, threats to Networks have become more dangerous.

One of these are the attacks of DoS that have become a major threat to the security of Networks.

DoS is a major threat to network security, and as the name implies, it aims to deny the service of network resources to legitimate users [2].

Attackers establish the most sophisticated and recent types of DoS attacks known as amplification attacks, to increase the effect of normal DoS attacks [1].

A. Underlying Organization

In 2002, the European Gaza Hospital (EGH) was the first hospital in Palestine to implement computerized systems, like healthcare, human resources, finance and store system.

The computerized system contributed to facilitate the scheduling, booking and record attendance for patients and issuing computerized reporting, which contributed significantly to the arrival of the results and decision making processes.

B. EGH systems fault

There is a problem that makes networks and servers, not work efficiently, so the information systems which depend on the network does not continue with services, and in this situation make us search for solutions to deal with this situation.

Servers and networks miss functions at EGH, matches the results of DoS attacks.

C. Infrastructure of EGH Network

We mentioned above an overview about EGH system, fault, Benefits also, now we present the current EGH network Infrastructure as shown in Figure 1.

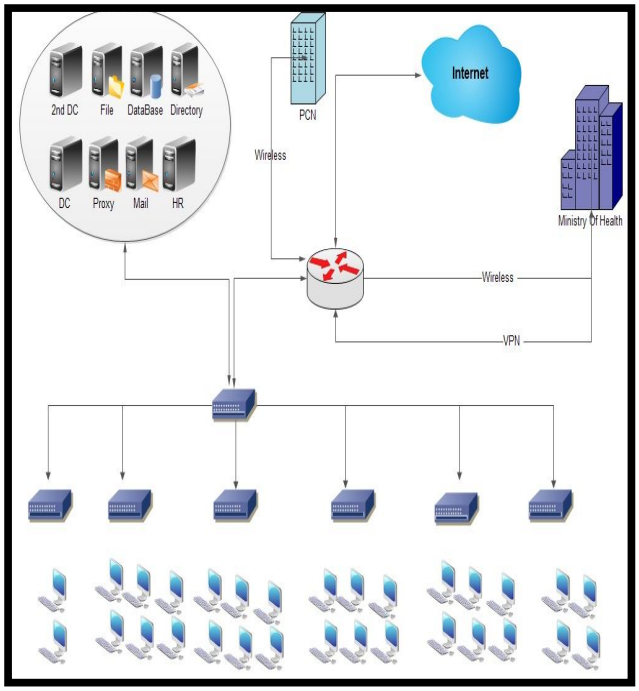

Fig 1 EGH Network Infrastructure.

There is only a router, connected between EGH network, and the external networks.

N.B.: we mentioned above the EGH systems fault, so we decided to protect the network and systems from DoS attack, by our approach (DPDoS) ,details in section III.

D. Intrusion Detection System(IDS)

Intrusion detection system (IDS), is the process of detecting and identifying unauthorized or unusual activity on the system [3].

The types of IDS are divided into two categories: network based (NIDS), and host based (HIDS).

NIDS collects data at the network level, transparent to the other computers, their sensors are located somewhere in the network to monitor network traffic, also capture network traffic and compare the traffic with a set of known attack signatures [4].

HIDS monitors specific files, logs and registry settings on desktop computer, and can alert on any access, modifications, detection and monitoring [4].

In this research, we propose detecting and preventing DoS attacks in LAN.

DPDoS approach is based on multi classifiers and is capable to detect and prevent DoS attacks, this approach uses classification techniques K-Nearest neighbour (KNN), and decision tree (DT.), to classify network traffic into DoS or normal classes.

In addition, the experiment results show that our proposed method is effective.

The remainder of this paper is organized as follows, Section I: Introduction, Section II: Presentations of related works, Section III: Presents the methodology, and finally, Experimental results and conclusions are presented in Sections IV and V respectively.

II. Related Works

In this section, we survey some of related researches in detecting and preventing DoS attacks, and classify them into three categories according to detection and prevention of DoS techniques, that these researches were based on.

Detecting and preventing DoS attacks based on, Intrusion Detection, Data Mining Technique and Router Method.

A. Intrusion Detection

Several researchers used intrusion detection to detect and prevent DoS attacks, Sujatha et. al. in [5], suggested sets of detection algorithms for DoS detection and blocking at the application layer, their method allows legitimate users to use user's signatures that are calibrated, by using occasional CAPTCHAs or AYAH (Are You A Human) page, also, each user has a signature and all signatures are stored in the database, so the AYAH results dynamically in determining signature is an attack, or a legitimate user.

However, their technique uses AYAH or CAPTCHAs in an annoying manner for users, and that is time-wasting.

Fang et. al. in [6], suggested an intrusion prevention system(IPS), named Cumulative-Sum-based Intrusion Prevention System, which detects malicious behaviours, and detects as well attacks and distributed attacks launched to remote clients and local hosts, so when the packets pass through a switch, the switch duplicates the packets, send the original packets to their destinations and delivers the duplicated packets to a IPS, IPS consists of Packet Analyzer, Intrusion Detector, and Response Manager, the detection accuracy of the researcher's method was (98.4%) at its best.

Their results introduces a high security level for the environment, but the drawback of the researcher's proposed technique appears in its failure under huge attacks on the switch, because the packet duplicate according to the switch is sent to (CSIPS), and the switch will be in deny of service during these attacks.

ZENG et. al. in [7], introduced a new defense mechanism against (DDoS) attacks based on the three way handshake process of discarding the aggressive handshake requests.

The average time a (SYN) packet stays in the half-connection queue no more than 1 second.

However, we see the results with normal access service is almost the same result in the case of defense mechanism, in addition, the utilization rate of the memory with defense method is almost equal to the circumstance of no defense measure, and also, some of legitimate connections will be discarded during this method.

B. Data Mining Technique.

Devendra et. al. in [8], introduced a model to improve accuracy rate of intrusion detection using a decision tree algorithm, and Stratified weighted sampling, the goal was to identify attacks with a high detection rate and a low false rate, they used supervised learning with a preprocessing step for intrusion detection, the preprocessing step to the (KDD CUP 99) dataset, which is classified into three phases, data preprocessing stage, fusion decision stage and data callback stage.

Fusion Decision stage, to filter false rates and improves detection rates, and Data Callback stage, to update and test date pool for undetermined samples, and the stratified weighted sampling techniques to generate the samples from the original dataset.

Then used these samples in the decision tree algorithm to classify the records as normal or attacked, however, the drawback of their model appears in the accuracy rate with (94.74 %) and false rate with (2.81%), and that is not sufficient compared with the great danger by DoS attacks negative effects on the network and its resources.

Portony et. al. in [9], Introduced a method for clustering similar data instances together, and used distance metrics on clusters to define an anomaly, the author studied two basic thesis: Firstly, data examples having the same classification should be close to each other in feature space under some reasonable metric, while examples with different classifications should be far apart, Secondly, the number of examples in the training set that represent normal traffic, is overwhelmingly larger than the number of intrusion examples, the clusters were labeled based on cluster size; the biggest cluster (>98%) will be labeled as normal and others as anomalous, the training and testing were done using (KDD CUP 99) data set.

Their solution detects new types of intrusions while maintaining a low false positive rate, also is effective when almost network traffic is normal class and homogenous, however, it depends on cluster 'size', and that may be not accurate in detecting DoS attacks when almost data are anomalous, the big cluster actually anomalous will be considered as normal, so if any assumption doesn't achieve its criteria, consequently the system accuracy will fall and give a high false alert.

H. Nguyen et al. in [10], proposes a (K-NN) classifier method which detects the DoS attack by classifying the network status into normal, pre-attack and attack, so this method has many advantages such as easy implementation, short time computation and accuracy of 91% for detection of DoS attacks, but the drawback of the researcher's method is that it doesn't achieve the sufficient accuracy in detection of DoS attacks, as DoS attacks represents a critical threat to Networks, Systems and resources.

C. Router Method.

Yun Ling et. al. in [11], suggested a defense mechanism to store and detect the validity of outgoing (SYN), and incoming (SYN/ACK) in the edge router, so the mapping table is accomplished in a hash map table by checking pairs of outgoing (SYN) request and the incoming (SYN/ACK) and store them in the database.

The researcher's suggested technique guarantees that each packet sent by the client is valid, mainly divided into two modules, storage module and inspection module.

The storage module, utilizes hash function to store source and destination address information into the database, the inspection module by using the mapping table which is constructed in the storage process to detect whether there are some abnormal cases, or not.

The advantage of the researcher's method shows an accurate detection (SYN) flood or abnormal case attack, based on mapping table which is constructed in the storage module, but the drawback of the researcher's method is the Integration process of storing the packet information such as the source and destination IP addresses is difficult in the case of congestion in the network flow, this leads to incorrect values in the mapping table of bloom filter data structure.

Yi Zhang et. al. in [12], proposed detection and prevention mechanism based on per-IP Traffic Behaviour Analysis, It uses the concept of Cumulative-Sum algorithm to analyse the behaviour of every IP address, the researcher's approach is divided into three layers, application layer, network layer and driver layer, the application layer provides the user with a user-friendly operating platform.

Moreover, Network layer extracting flow features and storing them into the corresponding IP record, and determines whether the traffic behaviour of each IP is abnormal (Detection Module) and updating the data buffer or not, Furthermore, the driver layer consists of two modules of packet capture, that are network card set to the promiscuous mode, and the data packet classification algorithm, data is captured and stored in the data buffer, then the system automatically filters the attacker's traffic and forwards normal user traffic.

However, the researcher's proposed approach for detection of malicious IP addresses is followed by blocking those IP addresses, thus this approach does not work effectively against (DDoS) attacks, as most of the packets are spoofed, it can result in blocking a legitimate client whose IP addresses are spoofed by attacker.

III. The Methodology

The main objective of this paper is to develop an approach to detect and prevent DoS attacks in LAN, that can be valid for the Security domain in an accurate way.

To achieve this objective, we used ensemble method from data mining and defense mechanism.

The methodology of our work starts with understanding the domain, data acquisition, DoS identification labelling, cleaning and preprocessing, processing stage, choosing the data mining task, choosing the data mining algorithm, applying the data mining algorithm, evaluating the data mining algorithm for the dataset, defense mechanism and finally the comparing phase.

A. Data Acquisition

We collected the dataset from EGH network by capture live packet data from a network interface (Wireshark Program), there are many ways to capture traffic, we chose capturing data remotely, because there are situations which prevent access to the server physically or quite simply for security reasons and performance risks, consequently, remote packet capture system, is necessary to execute a server program (rpcapd) on server and authentication, in addition, authorized client lists to connect to the server [13].

Then we started capturing packet and collected datasets over a period of three months (180 hours in working days).

EGH dataset is composed of 15919 profiles, where contained on 7984 normal profiles and 7935 DoS profiles which are used in our research.

Its dataset consists of 9 attributes, Table 1 presents the attributes and their description of the EGH dataset, and we added a new attribute called attack-type, to be used for label attribute in following steps.

Table 1 EGH Dataset Description

Attribute	Description	Selected
No.	Packet Number	
Time	Time of session initiation	√
Source IP	Source IP of the packet	√
Destination IP	Destination IP of the packet	√
Protocol	Protocol Type	√
Length	Packet Length	√
Source Port	Source port of the packet	√
Destination Port	Destination port of the packet	√
Info	Connection Information	
Attack-Type		√

There was no labeled EGH dataset, so to evaluate our method, we need to label each record in dataset by identifying Attack-Type attribute by DoS and normal, we manually labeled an Attack-Type attribute for each record in the EGH dataset, we have done this step after clustering method to EGH dataset.

C. Preprocessing

In order to detect and prevent DoS attacks in LAN by applying data mining method, the data should first be preprocessed to get better input data for data mining techniques ([15],[14]), in the data preprocessing step, we did some preprocessing of the dataset before loading the data set to the data mining software, irrelevant attributes should be removed because they add noise to the data and effects the model accuracy negatively.

The attributes marked as selected and seen in Table 1, are processed via the Rapid Miner environment [16] to apply the data mining methods on them, the attributes such as the No. and info. are not selected to be part of the mining process; this is because it did not provide any knowledge for the dataset processing.

D. Processing Stage

We present our strategy which we followed to achieve our goal, which tries to develop an approach to detect and prevent DoS attacks, that can be valid for information security domain in an efficient way with high accuracy and F-measure.

To do that, we implemented the data mining classification experiments, and defense mechanism experiment in sections 1 ,2 respectively.

1. Data Mining Classification Experiments: the EGH dataset used in this case is composed of 15919 profiles, where contained on (50.15%) normal profiles, and (49.84%) DoS profiles which are used in our research, we are classifying instances by applying individual algorithms to test the classification accuracy, then we are classifying examples by applying multi classification algorithms, and choose the best model which accuracy is better than individual algorithms, For more detail see section IV.

2. Defense mechanism experiment: after we detected the DoS attacks in the previous step, we need to prevent the DoS attacks, we can do the prevention by PfSense firewall and snort tool.

E. The Proposed DPDoS Approach

We proposed our approach consisting of two phases,

Detection Mechanism: the first phase based on classification in data mining which is k-nearest and decision tree as a classification, the train of these classifiers is to detect and classify DoS by using EGH dataset as shown in figure 2.

We chose DT. and K-NN with the best accuracy compared with the results obtained from data mining classification experiments, on the same dataset, see section IV in details.

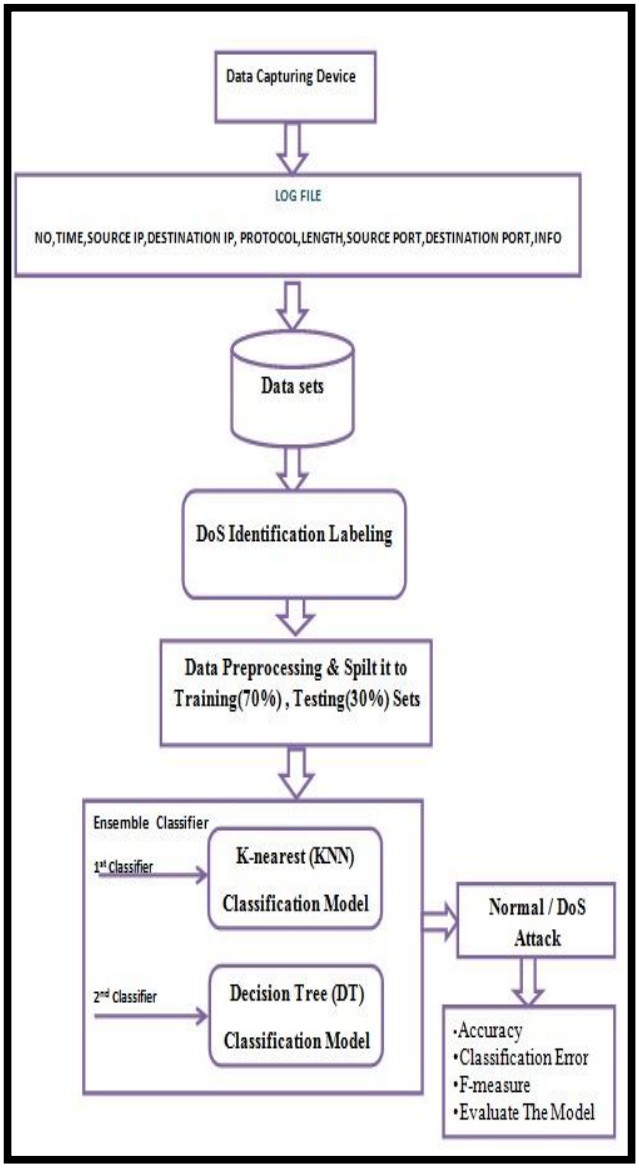

Fig. 2 General View of Proposed "DPDoS" Approach

Defense Mechanism: in the second phase of our approach, we use two defense component, PfSense firewall [17] and snort tool [20] to protect the network from DoS attacks, there are many open-source firewalls, Pf Sense was chosen for the firewall component as it ensures high level of security and performance of the UNIX systems[17], we will be using the defense mechanism depicted in figure 3.

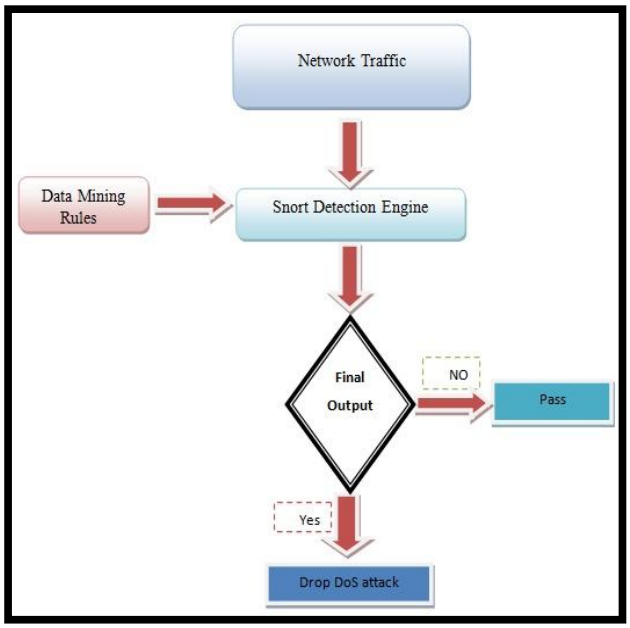

Fig. 3 Defense Mechanism

IV. Experimental Results and Evaluation

We performed experiments on EGH dataset collected from EGH by using the rapidminer program, to implement our approach using data mining methods.

In our experiment, the performance of each detection and classification model is measured and compared by using accuracy, misclassification error rate ,F-measure.

Accuracy: the proportion of correct classification classes (i.e., TP and TN), over the total number of classification attempts.

$$\textbf{Accuracy} = \frac{TN+TP}{TN+TP+FN+FP}$$

Misclassification Error Rate: the proportion of normal traffic flows, that are falsely labeled as (DoS).

$$\textbf{Misclassification Error rate} = \frac{FP}{FP+TN} * 100$$

F-measure is defined as the harmonic mean of recall and precision.

A high F-measure value signifies a high value for both recall and precision, it is evaluated when the learning objective is to achieve a performance between the identification rate (recall), and the identification accuracy (precision) of a specific class [15,18].

$$\textbf{f-measure} = \frac{2 \times Recall \times precision}{Recall + precision}$$

we use accuracy ,misclassification error rate and F-measure to evaluate the performances of our approach.

A. Experiment Scenarios we have two scenario in our experiments.

1. *Scenario I (individual classifier):* the first scenario used by individual algorithms and the results shown in table 2 and figure 4.

Table 2 Experiments Results of Scenario I

Classifier	Overall Accuracy	Classification Error	F-measure
Naïve Bayes	91.88	8.77	88.96
SVM	98.50	1.51	97.79
Decision Tree	98.92	1.08	98.42
K-NN with K=3	99.86	.13	99.79

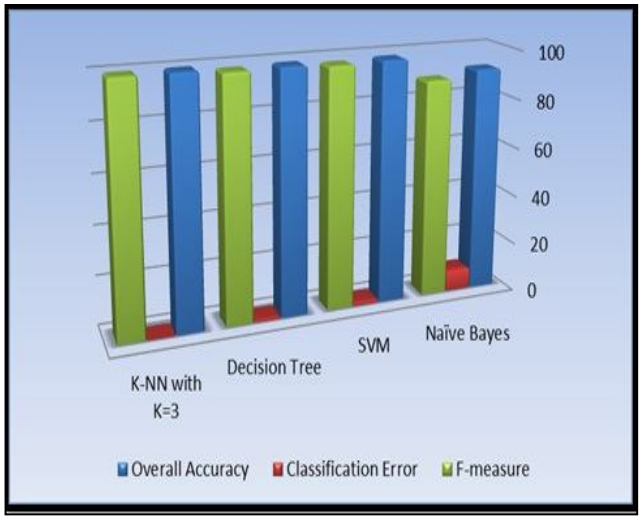

Fig. 4 Experiments Results of Scenario I.

We can summarize accuracy results for experiments scenario I, K-NN gave the highest accuracy result with (99.86%) in detecting DoS attacks, and the lowest accuracy was Naïve Bayes with (91.88%), the classification error was 0.13 in K-NN that means the lowest misclassification rate, also, we noted using scenario I that K-NN had the best value among the other algorithms.

2. *Scenario II (Ensemble Method) :* majority voting is easier, and as well a extremely successful set of scheme for improving the problems of classification, this method can

be applied with classification algorithms as Naive Bayes, k-nearest neighbour , decision tree algorithms , support vector machine.

3.

We classified instances after applying ensemble method (voting) on EGH dataset, this means we used a combination of algorithms as depicted in Table 3 and figure 5, showing the classification experiment results.

Table 3 Experiments Results of Scenario II

Classifier	Overall Accuracy	Classification Error	F-measure
Naïve Bayes +SVM	96.01	4.15	96.13
KNN+ Naïve Bayes	95.71	4.47	95.88
Decision tree + Naïve Bayes	96.82	3.81	96.44
Decision Tree +SVM	99.68	0.31	99.67
KNN+SVM	99.81	0.18	99.79
DPDoS Model	99.96	0.03	99.95

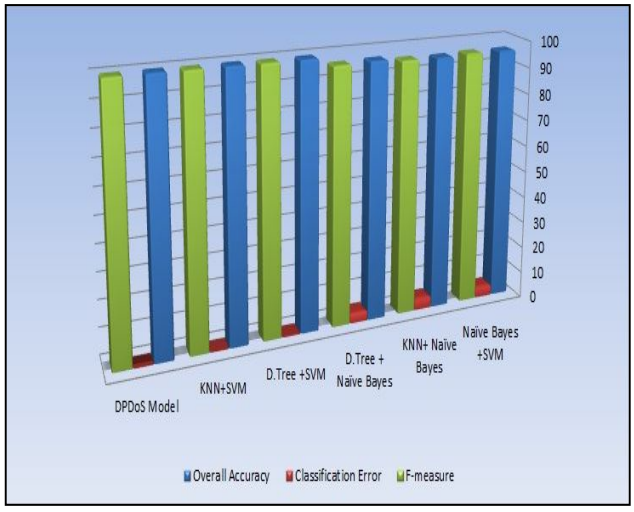

Fig. 5 Experiments Results of Scenario II.

We note that accuracy ranged up to (99.96%), which is considered a good result comparing with the results obtained from data mining classification experiments by individual algorithms on the same dataset.

we chose the best model that was K-NN and Decision tree, to be part of our an approach.

The summary of all experiments results, depicted in figure 6.

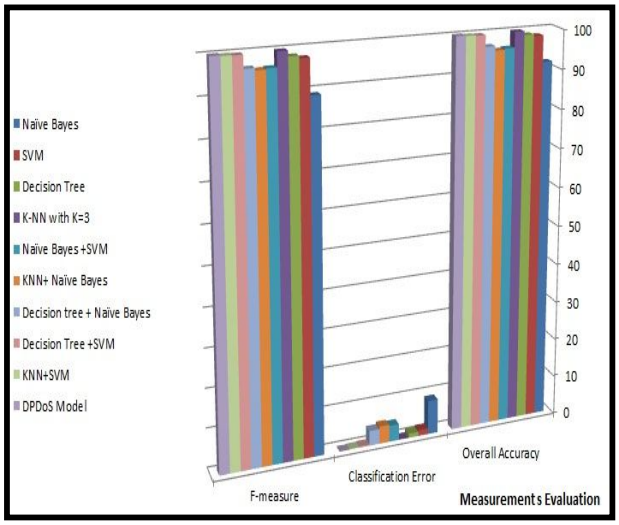

Fig. 6 Summary of All Experiments

B. Defense Mechanism

we applied the snort tool on PfSense firewall by using VMware workstation program, the EGH dataset uploaded by PfSense and Tcpreplay [19] used to replaying the collected network traffic into snort, in addition, we installed and configured snort and used the extracted rule from phase (1), then we tested the system and the results were highly accurate.

The next step to implement the defense mechanism in EGH environment as depicted in Figure 7.

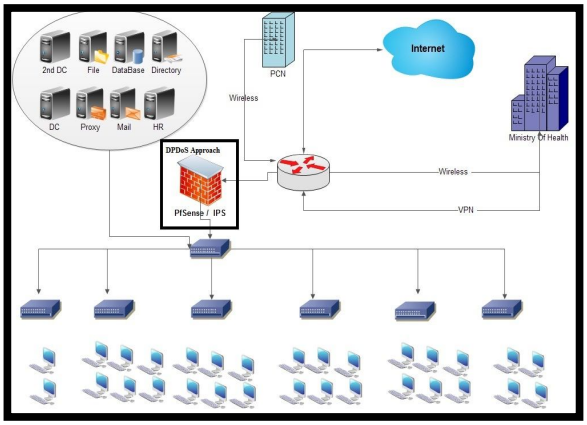

Fig. 7 EGH Infrastructure with Defense Mechanism .

We can summarize our experiments results as follows:

The experiments on EGH dataset of scenario I achieved the highest accuracy (99.86%), F-measure results (99.79%), and lowest misclassification rate (0.13) were in K_NN.

The experiments on EGH dataset of scenario II achieved the highest accuracy (99.96%), F-measure (99.95%), and lowest misclassification rate (0.03) were in our approach.

In general, we can say that our approach has achieved good results among all experiments on EGH dataset, where scenario I and scenario II recorded the highest accuracy that was (99.96%), and lowest misclassification rate (0.03), and F-measure (99.95%) which were in "DPDoS" approach.

The experiment in phase (2), was the prevention of DoS attacks by snort tool, according to the rules extracted from the first phase of our approach (offline detection), we added the signature of DoS attacks by snort tool.

Snort role in our research became as a intrusion, detection and prevention system because of it's corporate with Pf Sense firewall, snort detected attacks according to rules and blocked them immediately, which lead to more security to network.

The experiment in phase (1) data mining classification and phase (2) snort under PfSense firewall, achieved significant results for accuracy, misclassification rate and F-measure.

V. Conclusions

In this paper, we proposed an approach for detecting and preventing DoS attacks, the proposed approach is called DPDoS and based on ensemble method from data mining, the proposed approach structure and components were presented and explained, it involved the following steps for detecting and preventing DoS attacks: data acquisition, DoS identification labeling, preprocessing, processing stage, and finally evaluation of the approach, in addition to a defense mechanism by PfSense firewall and snort tool.

For evaluation purposes, we used confusion matrix method provided by Rapid Miner environment, experimental results show our approach performed significant improvement on F-measure results up to (99.95%), misclassifications (0.03) and accuracy (99.96%).

The experimental results confirm our thesis, which says that the ensemble method has better accuracy than single classification techniques, our approach achieved the best classification accuracy for detecting DoS attacks, and preventing DoS attacks by a defense mechanism (snort tool on PfSense firewall)

Possible directions for future work include applying mining techniques on dataset, to modify the approach to classify many types of DoS attacks on network.

In addition, we will classify new types of attacks such as Probing, User to Root and Remote to User Attacks that can be applied by our approach, moreover, we will use many types of worms or intrusions that can be applied by our approach.

Finally, we will try to make our approach to detect many types of threats, DoS attacks and worms with accepted and sufficient accuracy.

References

[1] S. Rastegari, M. I. Saripan, and M. F. A. Rasid, "Detection of Denial of Service Attacks against Domain Name System Using Neural Networks," *International Journal of Computer Science Issues*, vol. 6, no. 1, pp. 23–27, 2009.

[2] M. O. Schneider and J. Calmet, "Fibered Guard - A Hybrid Intelligent Approach to Denial of Service Prevention," *International Conference on Computational Intelligence for Modelling, Control and Automation*, vol. 1, Nov., pp. 121–127, 2005.

[3] M. Sharma, "Network Intrusion Detection System for Denial of Service Attack based on Misuse Detection," *International Journal of Computational Engineering & Management* , vol. 12, no.4, April, pp. 19–23, 2011

[4] R. J. Jadhav and U. T. Pawar, "Data mining for intrusion detection," *International Journal of Power Control Signal and Computation* , vol. 1, no. 4, pp. 45–48,2005.

[5] S. Sivabalan and P. J. Radcliffe, "A novel framework to detect and block DDoS attack at the application layer," *IEEE 2013 Tencon - Spring, Apr.*, pp. 578–582, 2013.

[6] F. Y. Leu and Z. Y. Li, "Detecting DoS and DDoS Attacks by Using an Intrusion Detection and Remote Prevention System," Fifth *International Conference on Information Assurance and Security*, vol. 2, Aug., pp. 251–254, 2009.

[7] X. Zeng, X. Peng, M. Li, H. Xu, and S. Jin, "Research on an Effective Approach against DDoS Attacks," *International Conference on Research Challenges in Computer Science*, Dec., pp. 21–23, 2009

[8] D. kailashiya and Dr. R.C. Jain , "Improve Intrusion Detection Using Decision Tree with Sampling," *International Journal of Computer Technology & Applications* , vol. 3, no. 3, June, pp. 1209–1216, 2012.

[9] L. Portony, "Intrusion Detection with Unlabeled Data Using Clustering", *In Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001)*, Philadelphia, PA: November 5-8, 2001.

[10] H. Nguyen and Y. Choi, "Proactive Detection of DDoS Attacks Utilizing k-NN Classifier in an Anti-DDoS Framework," *International Journal of Electrical & Electronics Engineering* , vol. 4 no. 4, Nov., pp. 247-252, 2010.

[11] Y. Ling, Y. Gu and G. Wei , "Detect SYN Flooding Attack in Edge Routers," *International Journal of Security and Its Applications (IJSIA)*, vol. 3, no. 1, Jan., pp. 31-45, 2009.

[12] Y. Zhang, Q. Liu and G. Zhao ,"A Real-Time DDoS Attack Detection and Prevention System Based on per -IP Traffic Behavioral Analysis," *3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*,Jul., pp. 163–167, 2010.

[13] J. Biswas, A. "An Insight in to Network Traffic Analysis using Packet Sniffer," *International Journal of Computer Applications* ,vol. 94, no. 11, pp. 39–44, 2014.

[14] J. Han and M. Kamber, Data Mining: Concepts and Techniques. 2nd Edition. Morgan Kaufmann Publishers, San Francisco, USA. (ISBN 1-55860-901-6), 2006.

[15] Ye N., "The Handbook Of Data Mining", Lawrence Erlbaum Associates, 2003.

[16] Rapid Miner 5.1, http://www.rapidminer.com , (2014, October), [last access]

[17] Pfsense,https://www.pfsense.org/about-pfsense/featues.html, (2014,October),[last access]

[18] U. Albalawi, S. C. Suh, and J. Kim, "Algorithms for Effective Intrusion Detection," *International Journal of Computer, Information Science and Engineering*, Vol.8,No. 2, pp. 20–24, 2014.

[19] Tcpreplay. http://tcpreplay.synfin.net/, (November 2014),[Last access].

[20] Snort tool, https://www.snort.org/,(2014,October),[last access]

Majed Tabash, is a network Administrator in Information Technology Department, European Gaza Hospital, Palestine, he holds MSc degree in Information Technology in 2014, from the Islamic university, and BSc degree in Computer Science in 2000, from the Al Azhar University of Gaza, Palestine, his research interest is: Information security. Networks, Data mining.

Tawfiq Barhoom, received his PH.D degree from Shanghai Jiao Tong University (SJTU),in 2004.This author is the Dean of Faculty of IT, Islamic University-Gaza, his current interest research include Secure software, Modeling, XMLs security, Web services and its Applications, and Information retrieving.