<div dir="rtl">

إقـــــرار

أنا الموقع أدناه مقدم الرسالة التي تحمل العنوان:

</div>

An Approach for Detecting and Preventing DoS Attacks in LA N

<div dir="rtl">

أقر بأن ما اشتملت عليه هذه الرسالة إنما هو نتاج جهدي الخاص، باستثناء ما تمت الإشارة إليه حيثما ورد، وإن هذه الرسالة ككل أو أي جزء منها لم يقدم من قبل لنيل درجة أو لقب علمي أو بحثي لدى أي مؤسسة تعليمية أو بحثية أخرى.

</div>

DECLARATION

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted elsewhere for any other degree or qualification

Student's name: Majed I. M. Tabash

اسم الطالب: ماجد إسماعيل محمد طبش

Signature:

التوقيع:

Date: 25/1/2015

التاريخ: 2015/1/25م

Islamic University of Gaza
Deanery of Higher Studies
Faculty of Information Technology
Information Technology Program

An Approach for Detecting and Preventing DoS Attacks in LA N

By:
Majed I. M. Tabash

Supervised By:
Dr. Tawfiq S. Barhoom

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master in Information Technology

Rabi'ul-Awwal 1436H − January 2015

الجامعة الإسلامية – غزة
The Islamic University - Gaza

مكتب نائب الرئيس للبحث العلمي والدراسات العليا   هاتف داخلي: 1150

الرقم: ج س غ/35/   Ref

التاريخ: 2015/01/25م   Date

نتيجة الحكم على أطروحة ماجستير

بناءً على موافقة شئون البحث العلمي والدراسات العليا بالجامعة الإسلامية بغزة على تشكيل لجنة الحكم على أطروحة الباحث/ **ماجد اسماعيل محمد طبش** لنيل درجة الماجستير في كلية **تكنولوجيا المعلومات** برنامج تكنولوجيا المعلومات وموضوعها:

آلية اكتشاف ومنع هجوم الحرمان من الخدمة في الشبكات المحلية
An Approach for Detecting and preventing DoS Attacks in LAN

وبعد المناقشة التي تمت اليوم الأحد 05 ربيع الآخر 1436هـ، الموافق 2015/01/25م الساعة العاشرة صباحاً، اجتمعت لجنة الحكم على الأطروحة والمكونة من:

د. توفيق سليمان برهوم	مشرفاً ورئيساً	....
أ.د. علاء مصطفى الهليس	مناقشاً داخلياً	....
أ.د. سامي سليم أبو ناصر	مناقشاً خارجياً	....

وبعد المداولة أوصت اللجنة بمنح الباحث درجة الماجستير في كلية **تكنولوجيا المعلومات**/ برنامج تكنولوجيا المعلومات.

واللجنة إذ تمنحه هذه الدرجة فإنها توصيه بتقوى الله ولزوم طاعته وأن يسخر علمه في خدمة دينه ووطنه.

والله ولي التوفيق ،،،

مساعد نائب الرئيس للبحث العلمي والدراسات العليا

أ.د. فؤاد علي العاجز

بسم الله الرحمن الرحيم

Dedication

To my beloved father…
To my beloved mother…

To my wife…

To my sons …….
To sisters and brothers…
To my best friends…

Acknowledgement

Praise is to **Allah**, the Almighty for having guided me at every stage of my life.

Many thanks and sincere gratefulness go to my **supervisor Dr. Tawfiq Barhoom**, without his help, guidance, and continuous follow-up; this research would never have been.

In addition, I would like to extend my thanks to **the academic staff of the Faculty of Information Technology** who helped me during my master's study and taught me different courses.

I would like to thank my **colleagues and classmates** for making my study a great experience, useful, enjoyable, and full of a warm atmosphere.

Lastly, I am greatly indebted to my **family** for their support during my course studies and during my thesis work.

Majed I. M. Tabash

January , 2015

Abstract

Nowadays, Denial of service (DoS) attacks, have become a major security threat to networks and to the Internet, DoS is harmful to the networks as it delays legitimate users from accessing the server, usually services such as in the Medical field, E-business field, etc. are out. In critical cases, may cause the server shut down, wasting valuable resources, therefore, leading to financial loss and in worst cases, loss of patient life due to delays in medical tests.

Moreover, the DoS detection problem is complex because attackers always invent new methods that can't be recognized easily, so many traditional approaches were used, such as, intrusion detection to detect intrusions through their signatures, but these techniques were unable to protect networks and servers before the appearance of their signatures.

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) to detect and prevent DoS attacks, and therefore increasing network security, yet, detecting and preventing DoS attacks is still a challenging task, especially in LAN.

In this research, we proposed an approach using data mining techniques by combination of classifiers (decision tree and k-nearest neighbor) to detecting and preventing DoS attacks. Our work is based on European Gaza Hospital (EGH) Dataset that is collected from EGH network ,then Labeled dataset manually. In addition preprocessing and processing stages, our approach is implemented using Rapidminer and exploits data mining algorithms to identify DoS attacks. The experimental results showed that the proposed approach is effective in identifying DoS attacks, our designed approach achieves significant results.

In the average case, our accuracy is up to 99.96%, we used defense mechanism and compared our approach with other approaches, and we found that our approach achieved best results in accuracy.

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

.

عنوان البحث

الية اكتشاف ومنع هجمات الحرمان من الخدمة في الشبكات المحلية

الملخص

في الوقت الحاضر، هجمات الحرمان من الخدمة أصبحت تشكل تهديدا أمنيا كبيرا على الشبكات والإنترنت، الحرمان من الخدمة ضار للشبكات كما أنه يؤخر المستخدمين الشرعيين من الوصول إلى الخادم، وعادة خدمات في مجال الطب، مجال الأعمال الإلكترونية وغيرها تكن خارج الخدمة.

وفى الحالات الحرجة، قد يتسبب في إيقاف الخادم، وإهدار الموارد الثمينة، مما يؤدي إلى خسارة مالية كبيرة جدا، وفى وأسوأ الحالات قد تؤدى الى فقدان حياة المريض بسبب التأخير في الاختبارات الطبية، وعلاوة على ذلك، فإن مشكلة الكشف عن هجمات الحرمان من الخدمة تكون معقدة لأن المهاجمين دائما يخترعون الأساليب الجديدة التي لا يمكن التعرف عليها بسهولة،

استخدمت العديد من الأساليب التقليدية، مثل الكشف عن التسلل من خلال توقيعاتهم، ولكن هذه التقنيات لم تتمكن من حماية الشبكات والخوادم قبل ظهور التوقيعات، بشكل عام، قد أجريت بعض الأبحاث لكشف ومنع هجمات الحرمان من الخدمة التي تحدث في الشبكات العريضة، ولكن الأبحاث على الشبكة المحلية قليلة وغير كافية لرصد ومنع هجمات حجب الخدمة، وبدورة يؤدى الى زيادة أمن الشبكة، بالإضافة إلى أن كشف ومنع هجمات حجب الخدمة لا يزال مهمة صعبة، خاصة في الشبكة المحلية.

في هذا البحث اقترحنا نهجا استخدام طرق تنقيب البيانات بواسطه عدد من المصنفات مثل (DT,K–NN)لاكتشاف ومنع هجمات حجب الخدمة ، يستند عملنا على قاعدة بيانات تم جمعها من شبكة مستشفى غزة الأوروبي بالإضافة الى عمل الية التسميه الى قاعدة البيانات بشكل يدوى والانتقال الى مرحلة الاعداد والمعالجة. نهجنا استخدام (RapidMiner) واستغل خوارزميات تنقيب البيانات في اكتشاف هجمات الحرمان من الخدمة وتظهر النتائج التجريبية أن النهج المقترح فعال في دقه كشف هجمات حجب الخدمة، فالنموذج المصمم حقق نتائج هامة، حيث في المتوسط وصلت الدقة الى٩٩.٩٦٪، واستخدمنا الية لحماية الشبكة من هجمات الحرمان من الخدمة، ثم قارنا نهجنا مع أعمال أخرى ووجدنا كذلك أن نهجنا حقق نتائج أفضل في الدقة.

الكلمات المفتاحية: استخراج البيانات، وهجمات حجب الخدمة، كشف التسلل، وكشف سوء استخدام ، تصنيف متعدد.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

DM	Data Mining
DoS	Denial of Service
DDoS	Distributed Denial of service
DPDoS	Detect and Prevent Denial of Service Attack in LAN (our proposed Approach)
DT	Decision Tree
EGH	European Gaza Hospital
HIDS	Host-Based Intrusion Detection System
IDS	Intrusion Detection System
K-NN	K-nearest Neighbor
LAN	Local Area Network
NB	Naïve Bayes.
NIDS	Network-Based Intrusion Detection System
SVM	Support Vector Machine.
WAN	Wide Area Network

Chapter 1
Introduction

This chapter is an introduction to the thesis, first it gives a brief description of DoS attacks, In addition, it states the thesis problem, the research objectives, the significance of the thesis, the scope and limitation of the thesis work, and the research methodology.

1.1 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 system.

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.

In October 2002, nine out of thirteen root servers that provide the Domain Name System (DNS) service to Internet users around the world have shut down for an hour of time, because of a distributed denial of service (DDoS) flooding attack [26].

Malicious applications and scripts are also used to launch DoS attacking computer systems over a network, furthermore, malicious attempts congests network systems, and causes very large damages like server crashes and networks drop down.

In addition attackers establish the most sophisticated and recent type of DoS attacks known as amplification attacks to increase the effect of normal DoS attacks[41].

Recently, there were many approaches proposed to detect DoS attacks, such as using an intrusion detection system, this approach was able to detect the attacks, but it has some drawbacks, for example ,the inability to detect new attacks and the need for signatures update [25].

Another approach is the routing structure, used to detect and prevent many types of denial of service, but have proven drawbacks in detecting DoS attacks, such as in [19]. Integration process of storing the packet information, where the source and destination IP addresses are difficult to integrate. In the case of congestion in the network flow, this leads to values not correct in the mapping table of bloom filter data structure.

So data mining has an important and essential role in intrusion detection, such as DoS attacks, by using different data mining techniques[8].

2

We propose detecting and preventing DoS attack in LAN Approach (DPDoS) based on multi classifiers, that is capable to detect DoS with sufficient accuracy and lowest classification error.

1.2 Research Motivation

In our work and daily life, we are using networks so there are a lot of risks facing it, may be intrusion, virus, worm, etc., and specially DoS attack. It may cause network damage, so it is very important to know these threats, protect network and keep it stable and available.

DoS attacks increase of risks to networks, led to an increasing in challenges of the security issues related to network systems, due to the increasing amount of new types of attacks, any activity which is malicious may not be identified.

DoS is one of the top threats where there are thousands of DoS attacks launched yearly across networks around the world, there is an urgent need to use an effective approach to detect and prevent DoS attacks.

We need to reach maximum level of network protection and sufficient DoS attacks detection accuracy .

1.3 Problem Statement

LAN is suffering from many attacks, one of them is DoS that makes servers and networks not to work efficiently, DoS leads to stop services resulting in a negative effect to businesses.

1.4 Research Objectives

1.4.1 Main Objective

The main objective of this research is to develop an approach that detects and prevents DoS attacks in the LAN(DPDoS), that can be valid for some proper domain in an accurate way.

The proposed approach based on data mining methods in order to achieve an acceptable accuracy .

1.4.2 Specific Objectives

There are many specific objectives extracted from the main objective:

1. Building dataset of the selected domain, our case study is EGH network.
2. Using real network traffic behavior to be able to detect, estimate the behavior of example traffic, then labeling the instance traffic manually .
3. Finding the most suitable data preprocessing steps such as cleaning, transformations, and reduction.
4. Applying a classification technique for DoS based on appropriate technique such as k-nearest neighbor, Decision tree, support vector machine, Naïve Bayes.
5. Evaluating the obtained classification results using different classification measures accuracy.
6. Applying defense mechanism by using snort and firewall.
7. Improving network security and protecting them from threats of DoS attacks.
8. Comparing our proposed approach with other existing models.

1.5 Significance of the Thesis

The significance of this thesis is:

1. Adding a significant contribution to scientific research in the field of finding effective solutions in DoS detection and prevention .
2. Constructing effective approach for detecting, preventing DoS attacks in LAN.
3. Studying the issue of DoS deeply in information security.
4. Many researches were conducted on DoS detection, but this is the first research in DoS attacks in EGH, so this research participates in enhancing this research area.
5. Maintaining EGH network from DoS attacks.
6. Protecting the Health Care System and other systems at the EGH from DoS attacks.

1.6 Research Scope and Limitation

This research proposes DPDoS approach by multi classification techniques, which is able to detect and prevent DoS attacks.

The work is applied with some limitations and assumptions such as:

1. The data set used in this research is collected from "European Gaza Hospital Network " in real time.
2. We started to capture packet and collected datasets over a period of three months (180 hours in working days).
3. The research is limited to detecting and preventing DoS in LAN.
4. We use the Wireshark program to capture data from EGH network.
5. The research based on behavior anomaly detection technique, to build our DoS detection and prevention approach.
6. The research, based on supervised learning algorithms to detect DoS with single class label.
7. Applying the data mining techniques using the open source machine learning tool RapidMiner software in our research.
8. Using combination of multi classification techniques in data mining to detect DoS attacks.

1.7 Research Methodology

In our research, we devote our study on detecting and preventing DoS attacks, our research methodology consists of 6 main phases as follows:

- *Research and Survey:* Include reviewing Recent papers, books, articles, websites and the recent researches closely related in the thesis problem statement. After analyzing the existing methods in DoS attacks, identifying the drawbacks. we formulate the strategies and solutions how to use or extend in order to be overcome in our research.
- *Data Acquisition:* In this step, we collect dataset from EGH network ,with different characteristics and sizes over three month, we explain this dataset in details in chapter 4.
- *DoS Identification Labeling :* In this step, we identify DoS attack in the dataset, and label each record with DoS or normal manually.

5

- *Preprocessing:* In this step, we apply a number of preprocessing techniques to deal with missing, noisy and inconsistent data. There are a number of preprocessing techniques such as cleaning, transformations, reduction.

  We use preprocessing techniques in details in Chapter 4.

- *Processing Stage:* to do this step, we implement the following steps:
  - **Data Mining Classification Experiments**: In this step, the data enter into classification process, the output of this step is to classify whether traffic is DoS or normal. Then we apply each previous step by more than one classification method and make comparisons between them according to accuracy, the structure of our approach, training, testing, and extract the results. Details in chapter 5.
  - **Preventing DoS Attacks:** In this step, we analyze the obtained results then using defense mechanism according to rules resulting from the previous step to prevent DoS attacks .Details in chapter 5.

- *Evaluate an Approach:* In this step, we analyze the obtained results and justify the feasibility of an approach by comparing it with other approaches.

1.8 Outline of the Thesis

This thesis has been divided into six major chapters, which are structured around the objectives of the research.

The thesis is organized as follows:

- *Chapter 2, Theoretical Foundation:*

  Describes the theory needed for thesis work, Knowledge Discovery in Databases (KDD) and its disciplines, namely: Data Mining (DM).Major kinds of classification algorithms which are used in our research: Naïve Bayes (NB),K-Nearest Neighbor (K-NN),Support Vector Machine (SVM) and decision tree.

  In addition, this chapter describes details about DoS attacks and EGH system.

- *Chapter 3, Related Works:*

  Presents some related work of detection and prevention DoS attack.

  In addition the main advantages and shortages were highlighted and discussed.

- *Chapter 4, An Approach for Detecting and preventing DoS in LAN :*

  It includes the methodology steps and the architecture of the "DPDoS" approach.

  An explanation about the dataset used in the experiments, identification dataset label (normal and DoS), preprocessing of this dataset, and the experiment cases are included as well.

- *Chapter 5, Experimental Results and Evaluation:*

  Give the details about the sets of experiments, and analyze the experimental results. Also discussion for each experiment, produced some experiments to comparison goals.

- *Chapter 6, Conclusions and Future Work:*

  Draw the conclusion and summarize the research achievement of experiments, and suggests future work.

Chapter 2
Theoretical Foundation

In this chapter, we talk about EGH system, obstacles and infrastructure of EGH network, identify denial of service attack, DoS Taxonomy, finally, we explain the use of data mining, especially classification techniques, major kinds of classification algorithms, which are used in our research: decision tree(DT) , naïve bayes (NB),k-nearest neighbor (K-NN)and support vector machine (SVM), and clarify their effectiveness in the detection of DoS attacks.

2.1 EGH System Overview

LAN is used in limited areas such as, homes, companies, schools ,etc., it is used to connect many computers and servers, as internal networks that utilizes from many media such as switches, hubs and others.

Computerization of EGH began 2002, and included: clinics, emergencies and childbirth systems, those gates are considered the basis of the data entered into a patient's medical file and emergency record, then the process moved to integrate the management of medical records department and the dates of clinics[46]. This phase lasted for three months during which all staff have been trained on running systems that have been installed. Then, the computerizing process moved to analyzing the admission and discharge points for three internal departments.

After the completion of the first phase of the process, the following departments were as well integrated, leading to the whole Hospital's integrated computerized system these are: radiology, pharmacy, laboratory and the rest of internal departments.That period lasted for several months, then, computerized work attended in hospital departments, began issuing computerized reporting system and contributed significantly to the arrival of the results and decision making processes.

The computerized system also contributed to facilitate the scheduling, booking and record attendance for patients in the outpatient departments and enabled staff at the medical records of control over the excessive repetition of medical files and controlling the health insurance data and identification number automatically.

The names of lab tests, medicines and drugs and all the data has been adopted into the computerized system, according to the data of the Palestinian Ministry of Health.

The EGH is the first hospital in Palestine to implement computerized systems like healthcare, human resources, finance and store system[46].

2.1.1 Infrastructure of EGH Network

We mentioned above an overview about EGH system, benefits also, now we present the EGH network Infrastructure (see Figure 2.1).

There are 8 servers and at least 200 desktop computers distributed between hospital departments, the servers names are: domain controller, file, database, directory, secondary domain controller ,proxy ,mail and human resource all connected together by a main switch.
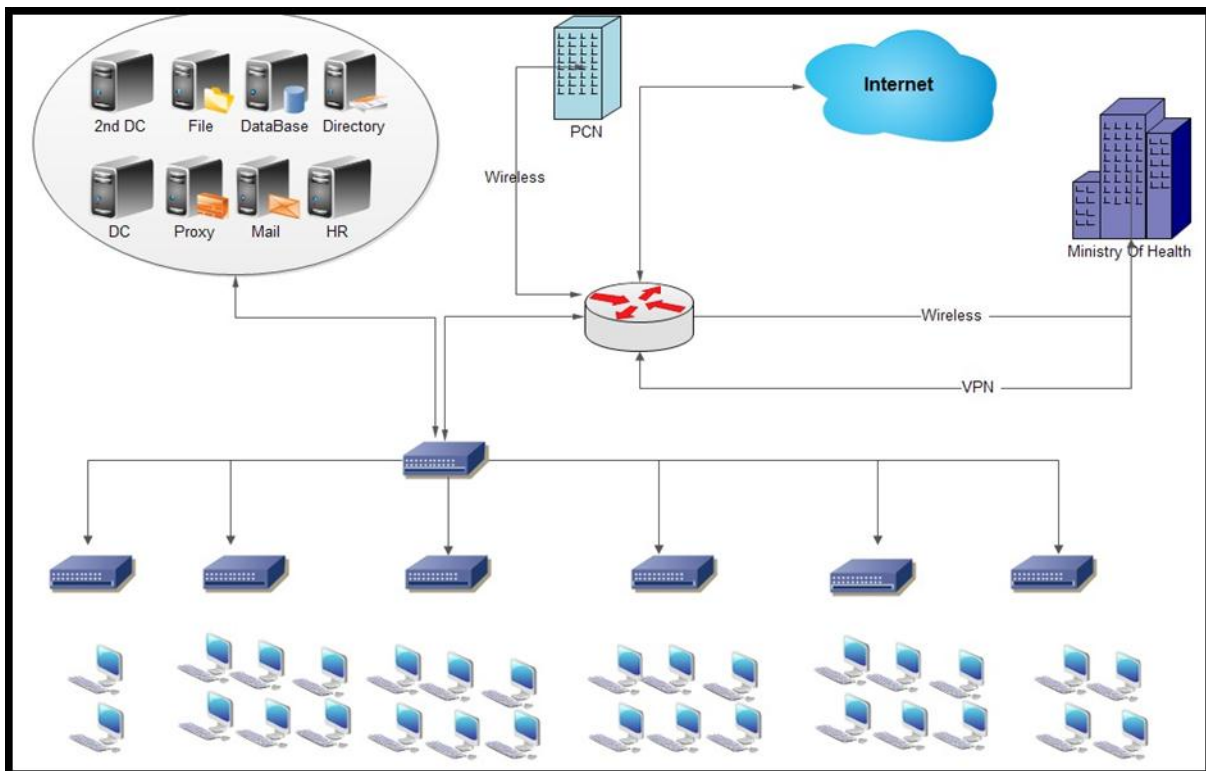
Figure 2.1: EGH Network Infrastructure.

There are six sub blocks connected with a main switch, also the systems are managed by central database, the hospital network is connected to the Ministry of Health, by two routes: VPN, and wireless connection, in addition connected with the Palestine Nursing College.

10

2.1.2 EGH Systems Problem

There is one problem that makes network and servers, not work efficiently so the information systems which depend on the network does not continue with services. This situation makes us to search about solutions to deal with this situation. So servers and network misfunctions at EGH network matches results of DoS attacks, we proposed a solution in our research to handle this problem in an effective way .In next section we talk about Denial of Service(DoS) in more details .

2.2 Denial of Service (DoS) Overview

In 2000, during the period 7–11th of Feb., many DoS attacks in the new category were launched on the Internet, the attacking was to many well-known sites, including Yahoo, Buy, eBay, and Amazon.

Since then , DoS attacks became real and a serious threat to computer networks, and can cause billions of dollars of damages within a few minutes, and cause many of problems[42]., as slowing down or hanging the server to the point where it becomes impossible to work.

2.2.1 DoS Attack Definition

DoS is a major threat to network security, as the name implies its goal is to deny the service of network resources to legitimate users [5,6,13,43,35].DDoS uses many of machines to overwhelm targeted network devices such as routers and servers, definitely interrupting or suspending services of a legitimate host connected to the network [5,6,13].

The main difference between DoS and DDoS attacks, that is DoS attacks use one attack machine to generate malicious traffic, while DDoS attacks use many of attack machines.

2.2.2 DDoS Attack Taxonomy

There are two main categories of DDoS attacks:

- *Bandwidth Depletion*

Designed to flood the victim network with unwanted traffic that prevents legitimate traffic from reaching the primary victim [23].

- *Resource Depletion*

An attack that is designed to tie up the resources of a victim system and making the victim unable to process legitimate requests for service [23].

In our research we work on bandwidth depletion that is result of DoS attacks. In next section we talk about detection methods.

2.3 Intrusion Detection Method

Intrusion detection system (IDS), is the process of detecting and identifying unauthorized activity on the system [10].

2.3.1 Detection Approaches:

An IDS that monitors computer systems, networks and analyze them for signs of security policy violation, then responds based on one of these approaches:

- *Anomaly Based An IDS :* which is anomaly based will monitor network traffic and compare it against an established baseline. The baseline will identify "normal" for that network so when traffic is detected which is significantly different, than the baseline is anomalous[1,3,9,11 ,10,24,35].
- *Signature Based :* A signature based IDS will monitor packets on the network, in addition compare them against a database of signatures or attributes from known harmful threats, this is similar to the way most antivirus software detects malware [1,3,9,10,11,24,35].

2.3.2 Types of Intrusion Detection System

There are three main different types of IDS :

- **Host based approach** *:* HIDS monitors specific files, logs and registry settings on a single PC and can alert on modifications, detection, any access, and copying of the monitored object. It's to flag any tampering with a specific PC and can automatically replace altered files when changed to ensure data integrity [1,7,9,11,24,30,43].
- **A Network based approach:** NIDS collects data at the network level, transparent to the other hosts, their sensors are located somewhere in the network to monitor network traffic. It's role is to flag and to stop an attack

12

before it gets information assets or causes damage, capture network traffic and compare the traffic with a set of known attack signatures.

NIDS devices compare these signatures every single packet that they see, to catch intruders in the act [7,9,11,24,43].

- **Hybrid agents**: Combine the functionality of host based agent with a network based sensor technology, that is limited to analyzing only the network traffic addressed to the specific host, where the hybrid agent is installed [10].

2.4 Data Mining (DM)

DM is a multidisciplinary field, including database technology, machine learning, statistics, pattern recognition, information retrieval, neural networks, knowledge-based systems, artificial intelligence, high-performance computing, and data visualization[8].

Data mining refer to the analysis of large quantities of data that are stored in computers, and is defined as knowledge discovery, which is the process of extracting useful patterns from large volumes of data using special algorithms [27].

Many terms carry a similar or slightly different meaning to data mining, such as knowledge mining from data, data/pattern analysis, knowledge extraction, data archaeology, and data dredging [8], Data Mining is essentially a process of data drive extraction of not so obvious but useful information from large databases that is interactive and iterative, knowledge discovery as a process consists of an iterative sequence of the following steps:

١) **Data Cleaning**: is removing the noise and inconsistent data.

2) **Data Integration**: where multiple data sources may be combined, these sources may include multiple databases, data cubes, or flat files.

٣) **Data Selection**: where data relevant to the analysis task are retrieved from the database, so, irrelevant, weakly relevant or redundant attributes may be detected and removed.

4) **Data Transformation**: where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance.

5) **Data Mining**: an essential process where intelligent methods are applied on data, to extract data patterns for decision making.

6) **Pattern Evaluation**: to identify the truly interesting patterns based on some interestingness measures, a pattern is considered interesting if it is: Valid, Novel, Actionable, Understandable.

7) **Knowledge Presentation**: is the framework that converts a large amount of data into a particular data or procedure that human being can figure out based on an intention.

In knowledge representation visualization, tools and knowledge representation techniques are used to present the mined knowledge to the user, Figure 2.2 illustrates data mining as a step in the process of knowledge discovery.
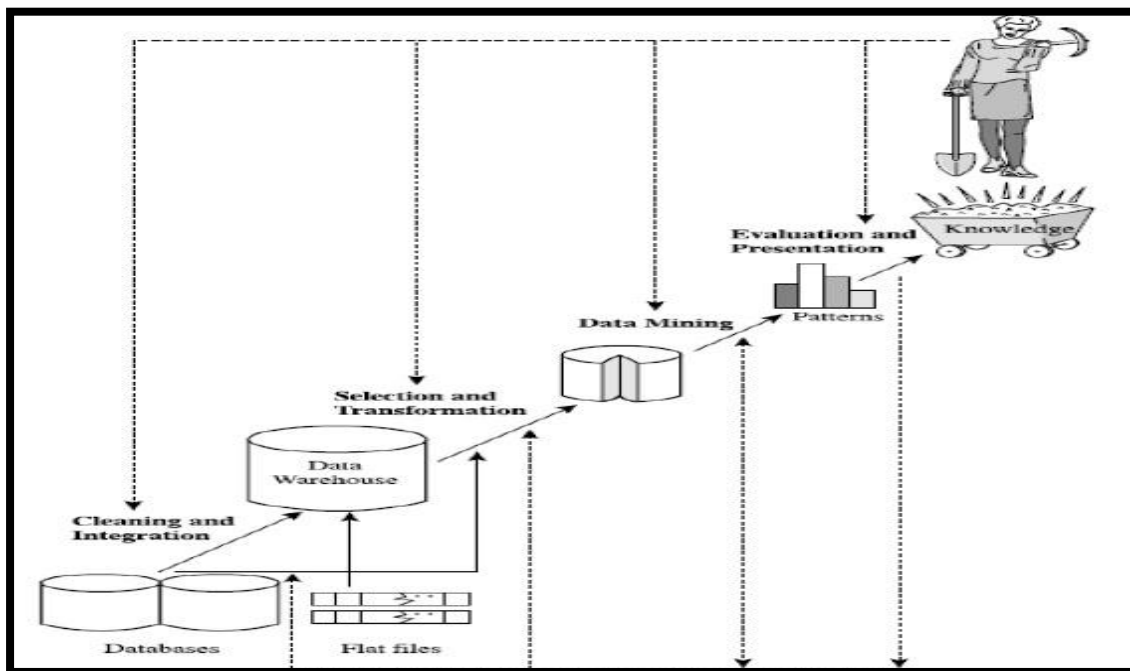
Figure 2.2: Data mining as a step in the process of knowledge discovery [8].

Data Mining functionalities are used to specify the type of patterns to be found in the data mining tasks, In general, data mining tasks can be classified into two main categories: descriptive and predictive.

Descriptive mining tasks characterize the general properties of the data.

14

Predictive mining tasks perform inferences on the current data, in order to make predictions [8], most of data mining tasks can be one or combination of the following:

1) **Classification:**

Known as supervised classification, the classification uses given class labels to order the objects in the data collection. Classification approaches use a training set where all objects are already associated with known class labels, the classification algorithm learns from the training set and builds a model, the model is used to classify new objects. [7,8,12]

2) **Prediction:**

used for predictive mining tasks**,** analysis is related to regression techniques, the key idea of prediction analysis is to discover the relationship between the dependent and independent variables.

For example, by using historical data from both sales and profit, either linear or nonlinear regression techniques can produce a fitted regression curve that can be used for profit prediction in the future [27].

3) **Association Rules:**

used for descriptive mining tasks, it aims to find out the relationship among valuables in database, and produce a set of rules describing the set of features that are strongly related to each other's, so that the relationship of a particular item in a data transaction on other items in the same transaction, is used to predict patterns [8].

4) **Clustering:**

used for descriptive mining tasks, It is unsupervised, and does not require a learning set, it shares a common methodological ground with Classification, it ungroups data and uses automatic techniques to put this data into groups [27].

In other words, finds groups of data pointes (clusters) so that data points that belong to one cluster are more similar to each other than to data points belonging to different cluster.[8,12]

5) **Outlier Analysis:**

used for predictive mining tasks**,** discovers data points that are significantly different than the rest of the data, such points are known as exceptions or surprises.While outliers can be considered noise and discarded in some applications, they can reveal important knowledge in other domains, and thus can

15

be very significant and their analysis are valuable, and important in identifying the outliers [8].

2.5 Classifiers

There are many classification algorithms in data mining, we describe some of those algorithms in order to be used in our research such as NB, DT,SVM, and K-NN.

2.5.1 Naïve Bayes (NB) Classifier

Bayesian classifiers are statistical classifiers based on Bayes theorem, they can predict class membership probabilities, such as the probability that a given tuple belongs to a particular class [8].

A NB classifier assumes that the effect of an attribute value of a given class is independent of the values of the other attributes. This assumption is called class conditional independence. The NB classifier, works as follows [8] :

- Let **D** be training set of tuples and their associated class labels. As usual, each tuple is represented by a n-dimensional attribute vector, $X = (x1, x_2,…..x_n)$, n measurements made on the tuple from **n** attribute, respectively, $A_1, A_2, ……, A_n$.

- Assume that there are m classes, $C_1, C_2,…, C_m$. Given a tuple, **X**, the classifier will predict that **X** belongs to the class having the highest probability, conditioned on **X**. That is, the NB classifier predicts that tuple **X** belongs to the class C_i if and only if

$$P(C_i|X) > P(C_j|X) \quad \text{for } 1 \leq j \leq m, j \neq i \text{ ..……. 2.1}$$

Thus we maximize P(Ci|X). The class Ci for which P(Ci|X) is the maximized is called the maximum posteriori hypothesis. By Bayes' theorem (Equation 2.2),

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{p(X)} \text{……………………2.2}$$

- As **P(X)** is constant for all classes, only **P(X|C_i) P(C_i)** needs maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equal.

- Based on the assumption is that attributes are conditionally independent (no dependence relation between attributes), **P(X|C$_i$)** using Equation 2.3.

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i)\ldots\ldots\ldots\ldots\ldots\ldots\ldots.2.3$$

Equation 2.3 reduces the computation cost, only counts the class distribution. If **A$_k$** is categorical, **P(X$_k$|C$_i$)** is the number of tuples in **C$_i$** having value **x$_k$** for **A$_k$** divided by **|C$_i$, D|** (number of tuples of C$_i$ in D). And if **A$_k$** is continuous-valued, **P(x$_k$|C$_i$)** is usually computed based on a Gaussian distribution with a mean μ and standard deviation **σ** and **P(X$_k$|C$_i$)** is

$$P(X|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i})\ldots\ldots\ldots\ldots\ldots\ldots\ldots..2.4$$

$$g(x_k, \mu_{C_i}, \sigma_{C_i}) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}}\ldots\ldots\ldots\ldots\ldots\ldots.2.5$$

Where **μ** is the mean and **σ^2** is the variance. If an attribute value doesn't occur with every class value, the probability will be zero, and a posteriori probability will also be zero.

NB classifier is fast, accurate, simple, and easy to implement, thus chosen to be one of the classifiers in this case, it is based on a simplistic assumption in real life and is only valid to multiply probabilities when the events are independent, despite its naïve nature. NB classifier actually works well on actual data sets [8], it is chosen to be used in this thesis.

2.5.2 K-Nearest Neighbor (K-NN) Classifier

K-Nearest Neighbor (K-NN) algorithm, is one of the supervised learning algorithms that have been used in many applications in the field of data mining, statistical pattern recognition and many others, it follows a method for classifying objects based on closest training, examples in the feature space [8].

An object is classified by a majority of its neighbors. **K** is always a positive integer. The neighbors are selected from a set of objects for which the correct classification is known. The K-NN algorithm is as follows :

1. Determine the parameter **K** i.e., the number of nearest neighbors beforehand.

17

2. The distance between the query-instance and all the training samples is calculated using Euclidean distance. Euclidean distance between two points, $X = (x1, x_2,\ldots x_n)$ and $Y = (y1, y_2,\ldots y_n)$ is:

$$d(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}\ldots\ldots\ldots\ldots\ldots\ldots2.6$$

3. Distances for all the training samples are sorted and nearest neighbor based on the K-th minimum distance is determined.
4. Since the K-NN is supervised learning, get all the categories of your training data for the sorted value which fall under **K**.
5. The predicted value is measured by using the majority of nearest neighbors.

K-NN works well even when there are some missing data.

2.5.3 Decision Tree (DT) Classifier .

Decision Tree is a common method used in statistics, data mining and machine learning, where it is an efficient method for producing classifiers from data, it is considered as a tree-structured plan of a set of attributes to be tested in order to predict the output, in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels,

Moreover, it is a type of tree diagram used in determining the optimum course of action, in situations having several possible alternatives with uncertain outcomes.

A decision tree classifier is modeled in two phases: tree building and tree pruning, In tree building, the decision tree model is built by recursively splitting the training data set and assigning a class label to leaf by the most frequent class, pruning a sub tree with a leaf or a branch if lower training error obtained, Table (2.1) presents decision tree algorithm [8].

Table 2.1: Basic structure of Decision Tree algorithm [8].

Input :

- Data partition, D, which is a set of training tuples and their associated class labels;

- attribute list, the set of candidate attributes;

- Attribute_selection_method , a procedure to determine the splitting criterion that "best" partitions the data tuples into individual classes, this criterion consists of a splitting attribute and, possibly, either a split point or splitting subset .

Output: A decision tree. Method :

(1) create a node N;

(2) if tuples in D are all of the same class, C then

(3)return N as a leaf node labeled with the class C;

(4) if attribute_list is empty then

(5)return N as a leaf node labeled with the majority class in D; // majority voting

(6) apply Attribute selection method(D, attribute-list) to find the "best" splitting criterion;

(7) label node N with splitting_criterion;

(8) if splitting_attribute is discrete-valued and multiway splits allowed then // not restricted to binary trees

(9) attribute_list ← attribute_ list – splitting_attribute; // remove splitting attribute

(10) for each outcome j of splitting criterion

// partition the tuples and grow subtrees for each partition

(11) let D be the set of data tuples in D satisfying outcome j; // a partition

(12) if Djj is empty then

(13) attach a leaf labeled with the majority class in D to node N;

(14)else attach the node returned by Generate decision tree(Dj, attribute list) to node N; end for

(15) return N;

2.5.4 Support Vector Machine (SVM) Classifier

The support vector machines (SVMs) as one of the most popular, state-of-the-art tools for DM and knowledge discovery, with high generalization ability and substantial theory, the learning algorithms of SVM can be used either a classification or a regression [27].

The basic form of SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, for classification, SVM performs classification by finding the separating hyperplane between two classes in a high dimension feature space that maximize the separation margin between the two classes .

The vectors that define the hyperplane are the support vectors, as shown in figure 2.3 there are many linear classifiers (hyperplanes) that is able to separate data into multiple classes, however, only one hyperplane achieves maximum separation, if such a hyperplane exists it's known as the maximum-margin hyperplane and such a linear classifier is known as a maximum margin classifier [8,27].
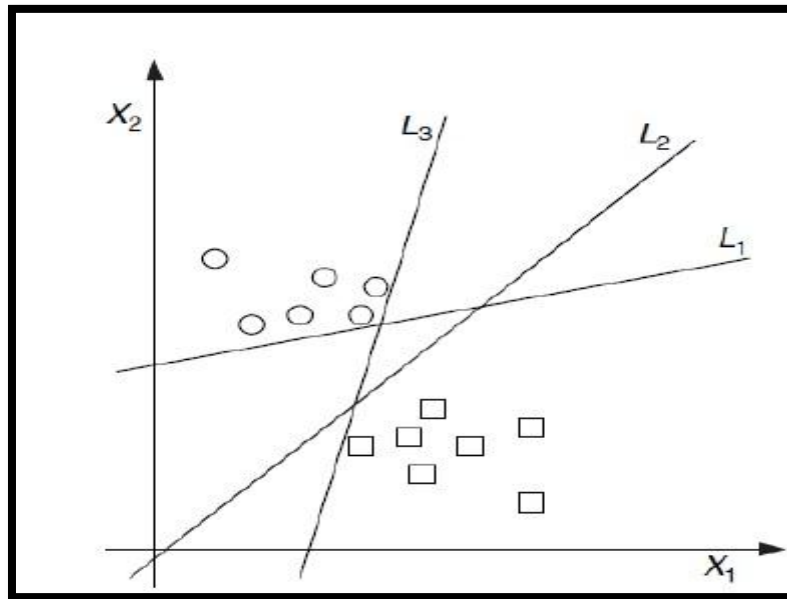
Figure 2.3: Many Linear Classifiers (Hyperplanes) may Separate the Data.

Given a training dataset (x1, y1), (x2, y2)… (xn, yn) where each xi is a vector of real numbers and yi is the corresponding class either 1 or -1, then the training algorithm attempts to place a hyperplane between points yi = 1 and points where yi = -1, any hyperplane Eq. 2.4 [27] can be written as the set of points X satisfying.

$$y = \text{w.X} - \text{b} \dots\dots\dots\dots\dots\dots\dots\dots\dots 2.7$$

The vector w points vertical to the separating hyperplane, adding the offset parameter b allows us to increase the margin.

We are interested in the maximum margin; we are interested in the support vectors and the parallel hyperplanes closest to these support vectors in either class, figure 2.4 shows maximum separation hyperplanes, these hyperplanes can be described by the equations:

$$\text{w.X} - \text{b} = 1 \dots\dots\dots\dots\dots\dots\dots\dots 2.8$$

$$\text{w.X} - \text{b} = -1 \dots\dots\dots\dots\dots\dots\dots\dots 2.9$$

By using geometry, we find the distance between the hyperplanes is 2/|w|, so we want to minimize |w|, to exclude data points, we need to ensure that for all *i* either

$$\text{w.x } i - \text{b} \geq 1 \text{ Or} \dots\dots\dots\dots\dots\dots\dots\dots 2.10$$

$$\text{w.x } i - \text{b} \leq 1 \dots\dots\dots\dots\dots\dots\dots\dots 2.11$$

Figure 2.4 show maximum separation hyperplanes

This can be rewritten as:

$$\text{Yi}\big(\text{w.x} - \text{b}\big) \geq 1, 1 \leq i \leq n \dots\dots\dots\dots\dots\dots 2.11$$
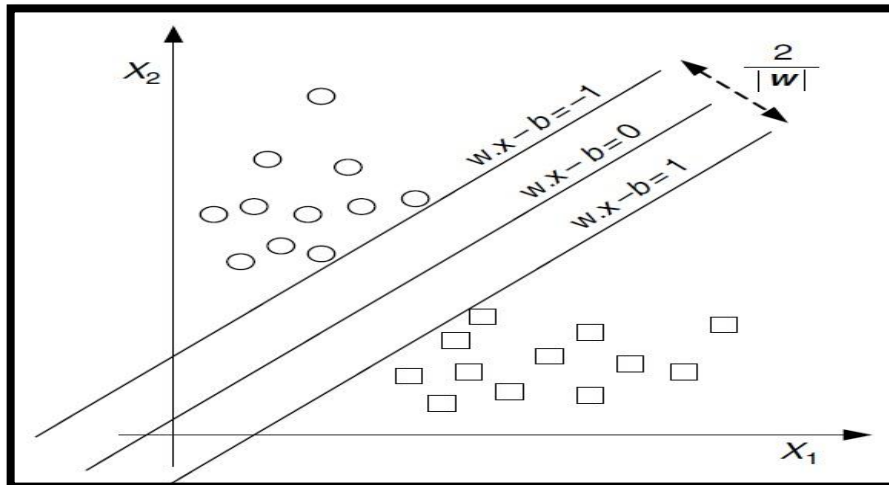
Figure2.4 :Maximum Separation Hyperplanes

2.6 Multi Classifier System

Multi classifier combination is a technique of combining the decisions of different classifiers which are trained to solve the same problem. So combination of multiple classifiers should produce more reliable results better than any of the individual classifiers.

There are many of existing combination methods, such as majority voting, weighted majority voting, linear combination of confidence value and combination neural network [47].

We use multi classifier system which are combined to formulate higher accuracies. The aim of multi classifier combination that can allow to overcome some known limitations of the traditional approach to classifier design means that using a monolithic classifier chosen as the best one for the application at hand, among a given set of available classification algorithms[12].

Multi classifier combines the outcomes of the individual classifier, the performance of the whole system is improved, these methods are very flexible and able to produce more accurate and stable results.

In "DPDoS" approach we used two classifiers algorithms which are (Decision Tree and K-NN), were combined to generate the final decision .

2.7 Prevention Tools

During our study we see many tools are used to prevent attacks like DoS ,worm ,virus .. etc., as in [39].

 We use snort tool and firewall in our research to prevent DoS attacks.

2.7.1 Snort Tool

 Snort is a free and open source network intrusion prevention system (NIPS), and network intrusion detection (NIDS) is capable of performing packet logging and real-time traffic analysis on IP networks.

 Snort was originally written by Martin Roesch and is now developed by Sourcefire [31,39,40,35].

Snort performs protocol analysis, and content searching/matching, it is commonly used to actively block or passively detect a variety of attacks and probes, such as buffer overflows, stealth port scans, web application attacks, etc.

It is mostly used for intrusion prevention purposes, by dropping attacks as they are taking place [40].

Snort can be combined with other software such as SnortSnarf, sguil and the Basic Analysis and Security Engine (BASE) to provide a visual representation of intrusion data.[31]

Snort is the most widely deployed intrusion detection and prevention technology worldwide, it has the most numerous and active community in the open source NIDS field today.

2.7.2   PfSense Firewall

The PfSense project is a free network firewall distribution, based on the FreeBSD operating system with a custom kernel and including third party free software packages for additional functionality [39], through this package system. PfSense software is able to provide most of the functionality of common commercial firewalls. It has successfully replaced every big name commercial firewall you can imagine in numerous installations around the world, including Check Point, Cisco PIX, Cisco ASA, Juniper, Sonicwall, Netgear, Watchguard, Astaro, and more [32].

PfSense software includes a web interface for the configuration of all included components, there is no need for any UNIX knowledge, no need to use the command line for anything, and no need to ever manually edit any rule sets [32].

2.8 Sniffing Tool

There are a lot of tools make sniffing like Wireshark, TCPdump ..etc., so in our research, we use Wireshark tool earlier known as Ethereal[22], it is an open source packet sniffer and analyzer and licensed by GNU GPI (General Public License), it works with the FreeBSD, UNIX, Linux, Solaris, OpenBSD, and Windows platforms [36,37], it is user friendly , also use to capture, filter and analyze packets, this tool is flexible and its log files are in many format [37].

Tcpreplay is a suite of BSD , UNIX and Win32 operating systems which gives you the ability to use previously captured traffic in libpcap format to test a variety of network devices, it allows you to classify traffic as client or server, rewrite Layer 2, 3 and 4 headers and finally replay the traffic back onto the network and through other devices such as IPS's, switches, routers, firewalls, NIDS.

 Tcpreplay supports both single and dual NIC modes for testing both sniffing and inline devices.

Tcpreplay is used by many firewall, IDS, IPS and other networking vendors, labs and open source projects, enterprises, universities.[44].

2.9 Summary

In this chapter, we presented the details of EGH system ,DoS attack, intrusion detection system approaches used in DoS detection.

Data mining techniques and its use in DoS detection have been explained as well. Furthermore, a brief description has been proposed about classifiers algorithms (NB, DT, Support Vector Machine and K-NN).

We used in applying our DPDoS approach DT. And K-NN Algorithms, finally we explained the importance of multi classifier and important tools: snort ,PfSense firewall and Wireshark program.

Chapter 3
Related Works

In order to improve detection and prevention of DoS attacks , some researches have been conducted in this area and they can be classified into three categories as an efficient ways to increase the security of networks.

In the next section, we focus on Detecting and Preventing DoS Attacks Based on , Intrusion Detection, Data Mining Technique and Router Method.

3.1 Intrusion Detection

Sivabalan and Radcliffe in [13], suggested sets of detection algorithms for DDoS detection and blocking at the application layer , their method allows legitimate user to use his signatures that are calibrated, by using occasional CAPTCHAs or AYAH (Are You A Human) page.

Also, each user has signature and all signature are stored in the database, so the AYAH results dynamically in determining signature is an attack, or a legitimate user.

The researcher proposed technique is discriminate legitimate flash traffic from attack traffic, However, Their technique uses AYAH or CAPTCHAs is an annoying manner for users and that is time-wasting.

Leu and YangLi in [14], suggested an intrusion prevention system(IPS),named cumulative-sum-based intrusion prevention system(CSIPS),which detects malicious behaviours, and detects as well DoS attacks and DDoS 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.

Packet Analyzer analyzes the packet header and classifies the packet into one of the three types (inbound, outbound and forwarded), Intrusion Detector, taking charge of packet analysis.

packet detection and notification, is composed of three detection subsystems, including ID-inbound, ID-outbound and ID-forwarded, which as their names receive corresponding classified packets from Packet Analyzer to detect corresponding malicious

behaviors, once an attack is detected, detection subsystem sends an alarm message to response manager which by receiving the message either alerts the local administrator to respond properly.

The experiment of that researcher tested in Tung hai university shows the detection accuracy in the CSIPS performed (98.4%).

Their results introduces a high security level for the environment, but the drawback of the researcher's technique appears detection accuracy is not sufficient compared with the great danger by DoS attacks negative effects on the network and its resources. In our thesis the detection accuracy better than CSIPS and the experiment and results showed in this thesis.

Xiao-hui et. al. *in* [15], introduced a new defense mechanism against DDoS attacks based on the three way handshake process of discarding the aggressive handshake requests .

The researcher proposed method through experiment present mitigate against DDoS attacks and the average time a SYN packet stays in the half-connection queue is 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.

 Bhirud and Katkar *in* [16] , present detection and prevention mechanism of a SYN flood attack using main memory database management System (MMDBMS) Approach so it stores information about the data flow directly in accessing memory .

In this model, three threads are used by IPS for its operations so the first thread is used to create entries for every incoming SYN packet. Second thread is used to delete expired entries of SYN packets and the third thread is used to keep track of time slots used by IPS.

 In addition, two time slots are used by IPS which one time slot is used to make SYN packet entries and another for deleting expired SYN packet entries moreover IPS creates

'N' number of tables store in MMDBMS and every table is associated with one time slot of IPS .

The researcher's method results in experiment present the performance of prevention mechanism using MMDBMS under heavy load is consistent over a period of time and CPU utilization of this mechanism is low.

However, it consumes the memory and CPU when overhead attacks launched from multiple places, knowing that 'backlog' it is another resource vulnerable to attack.

3.2 Data Mining Technique.

Kailashiya and Jain *in* [17] introduced 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 Error rate, they used supervised learning with preprocessing step for intrusion detection, the preprocessing step to the (KDD CUP 99) dataset, which is classified into three stages, data preprocessing stage, fusion decision stage and data callback stage.

Fusion Decision Phase to filter false rates and improves detection rates, and data callback stage to be 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 a decision tree algorithm to classify the records are 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 *in* [18] Introduced a method for clustering similar data examples 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

28

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 intrusion 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 and consequently the system accuracy will fall and give a high false alert.

Nguyen and Choi in [4] proposed 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 doesn't achieve the sufficient accuracy in detection of DoS attacks , as DoS attacks represents a critical threat to networks, systems and resources.

3.3 Router Method.

Ling et. al. in [19] suggested 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 in database.

The researcher's suggested technique guarantees that each packet sent by the client is valid mainly divides 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 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.

Zhang et. al. in [20] proposed detection and prevention mechanism based on per-IP traffic behavior analysis, it uses the concept of CUSUM algorithm to analyze behavior of every IP address.

 The researcher's approach divided into three layers, application layer, network layer and driver layer, so 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 determining whether the traffic behavior of each IP is abnormal (Detection Module) and updating the data buffer.

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 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.

 Mirkovic et al. in [21] introduced an effective defense D-WARD against DDoS attacks which monitors the asymmetry of two-way packet rates and to identify attacks in edge routers ،

D-WARD consists of two main modules; observation module and throttling module. The D-WARD system is installed on the source router that serves as a gateway between the deploying network (source network) and the rest of the Internet.

D-WARD able to monitor traffic in both directions, those traffic flows are periodically compared against predefined 'normal' traffic models, if any anomaly is detected, the throttling module is invoked.

The throttling module dynamically controls the outgoing traffic limit for each of the clients based on the feedbacks from the observation module.

If the observation module reports an attempt of an attack, the throttling module imposes an outgoing rate limit on that particular client.

If the subsequent observations confirm the attack, the outgoing rate is further reduced, thus effectively preventing a DDoS attack in its origin.

The drawback of that researcher method is takes a long time to detect increasing rate attacks and also discard packets of legitimate traffic during a DDoS attack where many malicious flows are present.

3.4 Summary

In this chapter we highlighted and discussed some of related works conducted in detecting and preventing DoS attack, where presented the DoS detection and prevention based on data mining as an efficient way to improve the security of networks like classification techniques was the most widely used for many recent researches, we explained methods used in related work which were problems related to accuracy[17,4] and misclassification error[17] .

Chapter 4
An Approach for detecting and preventing DoS attack in LAN (DPDoS)

In this chapter, we explain our proposed Detecting and Preventing DoS attacks in LAN (DPDoS) approach, which we followed in this research.

This chapter is organized into seven sections, Section 4.1 presents methodology steps of "DPDoS" approach, Section 4.2 will give a description of the collected dataset for designing experimental data, Section 4.3 perform identification of the DoS label, Section 4.4 presents preprocessing steps, Section 4.5 presents the processing stage, Section 4.6 applies the approach by using data mining method, Section 4.7 evaluates the approach using accuracy, misclassification error and F-measure.

To implement and evaluate this approach, various steps have to be performed, the main required steps are shown in Figure 4.1, and stated below:

- **Data Acquisition**: from European Gaza hospital network.
- **DoS Identification Labeling**: different types of the DoS will be identified, and connections are also labeled as belonging to one out of two classes, i.e., normal traffic, DoS attacks.
- **Preprocessing:** we apply a number of preprocessing steps to deal with noisy, missing, and inconsistent data.
- **Processing Stage:** the processing stage implemented based on the following steps:
    - **Data Mining Classification Experiments**.
    - **Defense Mechanism Experiment.**

    Then we applied each previous step by using classification algorithms: naïve bayes (NB), decision tree (DT),k-nearest Neighbor (K-NN) and support vector machine (SVM) as multi classification.

- **Evaluate the Approach:** to evaluate the classification performance of our approach, we use accuracy ,misclassification error and F-measure.
- **Comparing Phase:** we apply two comparisons:
    - Compare performance by using each classifier algorithm independently.
    - Compare performance by using multi classifier algorithms, applied on dataset.
    - Compare performance between our proposed approach and other published work, which has been used for detecting DoS attack.
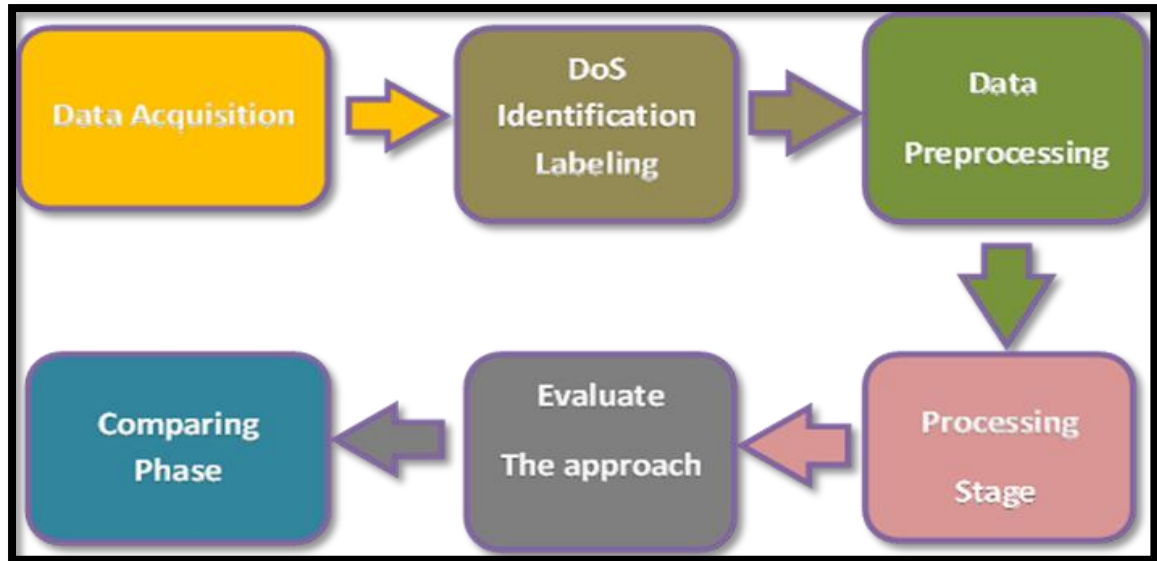
33

Figure 4.1: Methodology Steps [9].

4.1 Detecting and Preventing DoS attacks in LAN (DPDoS) Approach

To achieve an approach, we used multi classifier from data mining more than classifier to get better results in detection DoS attacks and then using the defense mechanism .

4.2 Data Acquisition

We collected the dataset from EGH network by capturing live packet data from a network interface (Wireshark Program), there are many ways to capture traffic we choose 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 (RPCAP) is necessary to execute a server program (rpcapd) on server and authentication, in addition authorized client lists to connect to the server [22].

4.2.1 EGH Dataset

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 4.1 presents, the attributes and their description of the EGH dataset, We added new attribute called attack-type that will be used to label attribute in next steps.

34

Table 4.1: 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		√

4.2.2 Log file samples

Samples of EGH dataset profiles is shown in Table 4.2, This table shows that the log file consists of the following: The attribute named "No.", which is the sequence number of packet. The attribute named "Time" indicates when the packet was captured.

The attribute named "Source", which is the source IP address that send request like source IP is 192.168.3.107 to destination IP " 192.168.3.24".

The attribute named "destination", which is the destination IP address that receives the request, The attribute named "protocol" , transport-layer protocol (i.e., TCP or UDP or SMB (applications protocols)) of the packet; the frequent traffic uses TCP protocol.

The attribute named "length", which is the length of the packet, The attribute named "Source port " which is the source of packet port, the attribute named "destination port" which is the destination of packet port and the attribute named "Info", this is the connection description.

Table 4.2: Samples of Data Profile.

No	Time	Source IP	Destination IP	Source Port	Destination Port	Packet Length	Protocol	Info
2	0.001486	192.168.3.17	192.168.3.107	TCP	80	4899	51145	4899 > 51145 [PSH, ACK] Seq=1 Ack=43 Win=65425 Len=26
3	0.014083	192.168.3.107	192.168.3.17	TCP	96	51145	4899	51145 > 4899 [PSH, ACK] Seq=43 Ack=27 Win=16139 Len=42
4	0.015549	192.168.3.17	192.168.3.107	TCP	80	4899	51145	4899 > 51145 [PSH, ACK] Seq=27 Ack=85 Win=65383 Len=26
5	0.025297	192.168.3.107	192.168.3.17	TCP	96	51145	4899	51145 > 4899 [PSH, ACK] Seq=85 Ack=53 Win=16132 Len=42

4.3 DoS Identification Labeling

There was no labeled EGH dataset to evaluate our method, we needed 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 dataset, we have done this step after attack lunched from internal attackers.

4.4 Preprocessing

Today's real-world databases are highly sensitive to noise, missing, and inconsistent data due to their typically large size and their likely origin from multiple, different sources, low-quality data will lead to low-quality mining results, preprocessing is a necessary step for serious, effective, real-world data mining, there are a number of preprocessing techniques such as: cleaning, reduction, etc.[33].

In order to detect and prevent DoS attack in LAN by applying data mining method, the data should first be preprocessed to get better input data for data mining techniques [33][8], 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, the attributes marked as selected as seen in Table 4.1 are processed via the Rapid Miner environment [28] 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.

4.5 Processing Stage

In this section, 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 (phase1) and Defense mechanism experiment (phase 2) in section 4.5.1 ,4.5.2 respectively.

In our experiments, the following steps are performed as part on the data classification:

- We started by classifying instances according to individual algorithms as depicted in Table 4.3 on the EGH dataset to test the classification accuracy.

Table 4.3: Individual Algorithms.

No	Algorithms Name
1	K-NN
2	DT
3	SVMs
4	NB

- We are classifying instances after applying individual algorithms, we apply multi classification on EGH dataset to test the classification accuracy by using multi classifier .

4.5.1 Data Mining Classification Experiments

In our experiments, the following steps are performed as part on the data classification.

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 algorithm to test the classification accuracy, details in section 5.2.1, then we are classifying instances by applying multi classification algorithms and choose the best model which its accuracy is better than individual algorithms, details in section 5.2.2.

4.5.2 Defense Mechanism Experiment

After we detected the DoS attacks in the previous step, we need to prevent the DoS attacks, so we can do the prevention by two component defense , PfSense firewall and snort tool, for more details see section 5.3.

4.6 Apply the DPDoS Approach

This section describes the major kinds of classification algorithms, which are used in our research: decision tree (DT), naïve bayes (NB), k-nearest neighbor (K-NN), support vector machines (SVM), which are provided by RapidMiner environment [28].

In the following subsections, we present these classification algorithms and their settings which are used during experiment results.

The main objective result of this research achieves high accuracies, better detection and classification error rate, for that, we propose "DPDoS" approach for detection and prevention of DoS in LAN, to achieve this, we used multi classification method and defense mechanism..

To achieve the objective of this research, we proposed the following steps shown in Figure 4.2:

Step 1: Capturing and Collecting datasets by Wireshark from EGH, labeling and preprocessing them, then select some attributes which are related to our goal, this dataset have been divided into 2 cases of experiments.

Step II: Applying preprocessing stage and then used data set according to spilt validation (70% training set and 30% testing set), by all classifiers model in "DPDoS" approach .

Step III: We apply the "DPDoS" approach as follows : applying voting mechanism , majority vote here in classification by used k-nearest classifier and decision tree classifier in the same step on the same training set to build the model, and test it on also the same testing set, this step get better classification and classify traffics output as (DoS/Normal) attack.

Step IV: Extraction results to evaluate classification performance by using the final confusion matrix from previous steps to compute overall classification accuracy, misclassification rate, and F-measure.
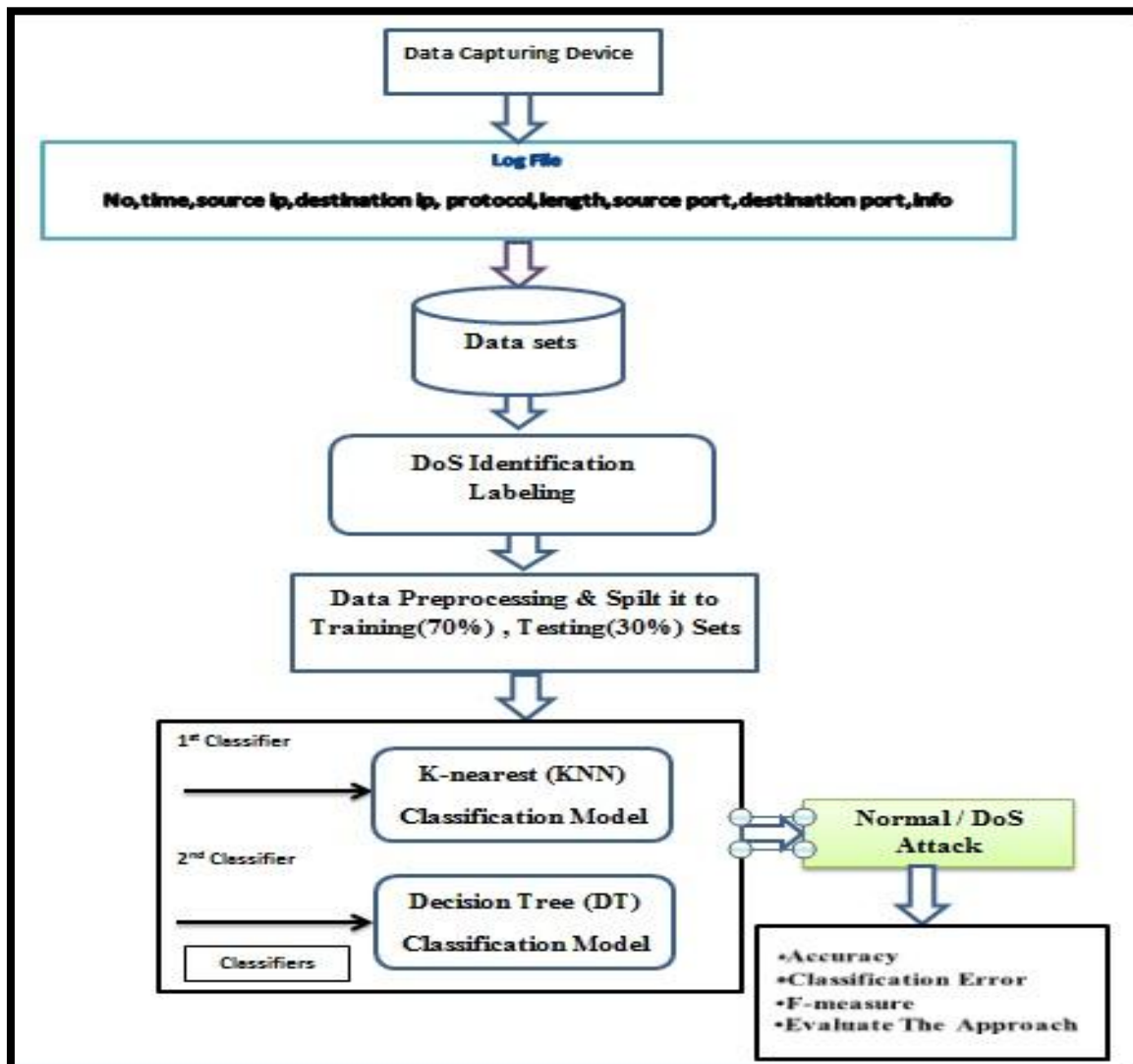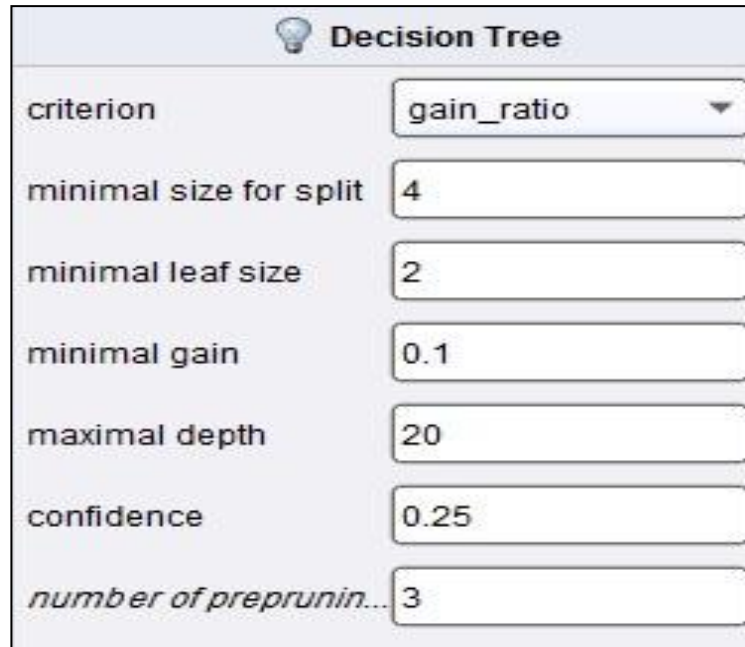

Figure 4.2: General View of Proposed "DPDoS" Approach

4.6.1 Decision tree (DT)

We used decision tree a typical setting, Figure 4.3 explain the setting of decision tree [28] to get the greatest predictive accuracy.

Figure 4.3: Setting of DT.

Criterion: used to specify and select attributes and numerical splits, we chose the gain ratio for the criterion term.

Minimal size for split: The minimal size of a node in order to allow a split = 4.

Minimal leaf size: The minimal size of all leaves = 2.

Minimal gain: which must be achieved in order to produce a split = 0.1.

Maximal depth: The maximum tree depth = 20.

Confidence: used for the pessimistic error calculation of pruning = 0.25.

Number of pre pruning alternatives: The number of alternative nodes tried when pre pruning would prevent a split = 3.

4.6.2 Naïve Bayes (NB)

We used naïve bayes classifier in our research, it is one of the most widely used classifiers, it is based on estimating probabilities of individual variable values, and given a class from training data, and then allow the use of these probabilities to classify new entities[8].

Figure 4.4 explain the setting of naïve bayes classifier [28].We use Laplace correction to prevent high influence of zero probabilities.
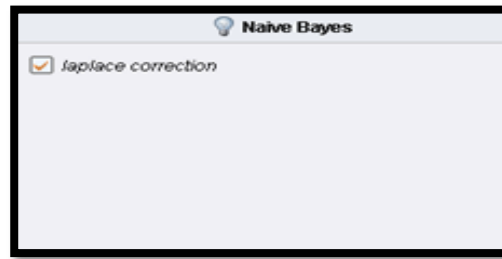
Figure 4.4: Setting of NB.

4.6.3 K-Nearest Neighbor (K-NN)

We used K-NN in our research, which is a supervised learning algorithm, the purpose of this algorithm is to classify a new object based on attributes and training samples [8].

We start by setting k=1 in the parameter setting and try a series of increasing k's with (K=1, 3, 5, 7, 9) and take the highest accuracy.

also we set the measure types appropriately (we chose mixed measures, in this case we have numeric predictors and a nominal label) as depicted in Figure 4.5.
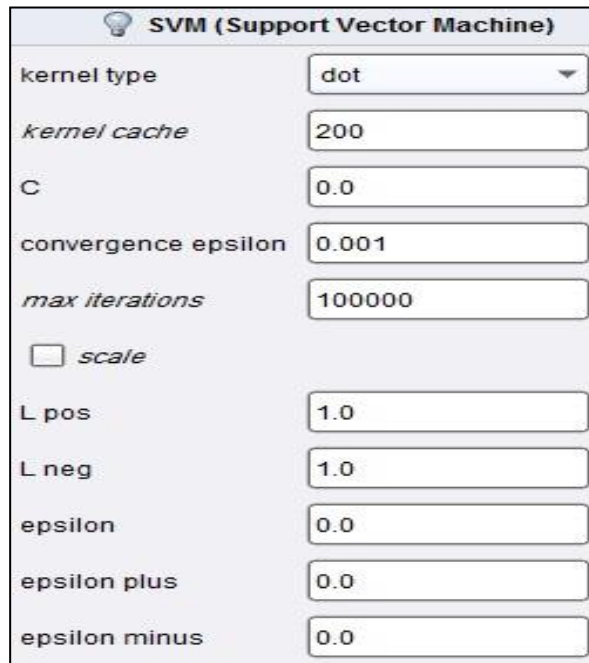
Figure 4.5: Setting of KNN.

4.6.4 Support Vector Machine (SVM)

We used SVM in our research, which is a supervised learning model with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis, the basis of this algorithm is to take a set of input data and predicts, for each given input, which of two possible classes forms the output, and making it a non-probabilistic binary linear classifier [28].

41

We used SVM a typical setting ,figure 4.6 depicted the settings of support vector machine.

Figure 4.6: Setting of SVM.

4.7 Evaluating the Approach

Performance evaluation plays an important role to evaluate classification performance. We use the confusion matrix that are commonly evaluation measures were visualization tools are used in supervised learning, and created for each classifier.

Each column of the confusion matrix represents the instances in a predicted class, while each row represents the examples in an actual class.

The following define accuracy ,misclassification error rate and F-measure which are considered the most commonly to evaluate classification performance.

In our research, three important measures are commonly used that can be defined as follows:

- **Confusion Matrix:** was created for each classifier using the actual and predicted responses [33][8], the following four estimates define the members of the matrix (show Table 4.4).

42

Table 4.4: Simple Confusion Matrix

		True Class	
		Positive	**Negative**
Predicted Class	**Positive**	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

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

$$\text{Accuracy} = \frac{TN+TP}{TN+TP+FN+FP} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{ 4.1}$$

- *Misclassification Error rate* : the proportion of normal traffic flows, that are falsely labeled as DoS.

$$\text{Misclassification Error rate} = \frac{FP}{FP+TN} * 100 \ldots \text{ 4.2}$$

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, F-measure which is shown in Equation 4.3, [33,34].

$$\text{F-measure} = \frac{2 \times \text{Recall} \times \text{precision}}{\text{Recall}+ \text{precision}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{ 4.3}$$

In our experiments, we use accuracy ,misclassification error rate and f-measure and evaluate the performances of our approach.

4.8 Summary

This chapter describes the methodology used for our research, it presents our processing strategy which we followed to achieve our goal in more detail, In addition, it explains the classification algorithms which are used during experiment results, in the next chapter we will be discussing the results of our experiments using our DPDoS approach and the described methodology.

.

Chapter 5
Experimental Results and Evaluation

In this chapter we present and analyze the experiments results, a combination of different data mining classifiers were used as multi and individual classifiers for our experiments, which are naïve bayes, decision tree, support vector machine and K-NN, we explained the machine environment and tools used in our research.

Also we present the evaluation measurements for classifications model during sets of experiments, by using the equation of accuracy, classification error, and f-measure which are illustrated in section 4.7 .

We apply sets of experiments in Section 5.2, we classified instances based on data mining experiment by using individual algorithm, and then classified instances based on data mining experiments using a multi classifier.

The details about these experiments and their results that have been presented and explained in this chapter.

5.1 Experiments Setup

In this section, a description about the experimental environment, tools used in experiments, measures of performance, evaluation of classifiers and DPDoS approach has been provided.

5.1.1 Experimental Environment and Tools

We applied experiments on a machine with properties that are: Intel (R) Core i3 CPU, 4.00 GB RAM, 500 GB hard disk drive and Windows 7 operating system installed. To carry out our thesis (including the experimentation), special tools and programs were used:

1. **Rapid Miner application program:** used to build our approach, and conduct practical experiments and extract the required results.
2. **Microsoft Excel:** used to organize and store datasets in tables, and do some simple preprocessing and analyze the results.
3. **VMware Workstation 9.0.2 build-1031769:** used to install PfSense software and configure it as firewall to prevent DoS attack.

45

5.1.2 Measurements for Experiments

The measures of evaluating the performance of classification are a confusion matrix. In addition to perform the comparisons of the tested algorithms, through the performance of each classifier evaluated using accuracy, classification error (misclassification) rate, and F-measure, based on the equations in section 4.7, we extract our experiments results in the next section.

5.2 Data Mining Classification Experiments (Phase 1)

We apply set of experiments, in the Subsection 5.2.1, we classify instances in the EGH dataset according to individual algorithms.

Finally, in the Subsection 5.2.2. We classified instances by applying multi classifier on EGH dataset.

5.2.1 Experiment Scenario I (individual classifier)

The dataset of this experiment is divided into 70% training and 30% testing dataset, In addition to (50.15%) normal profiles and (49.84%) DoS profiles attack.

We perform four experiments with individual algorithms to classify instances in the EGH dataset, then we calculated the accuracy, misclassification rate, and F-measure, Table 5.1 and figure 5.1 illustrates experiments results in this case.

Table 5.1: Experiments Results of Scenario I

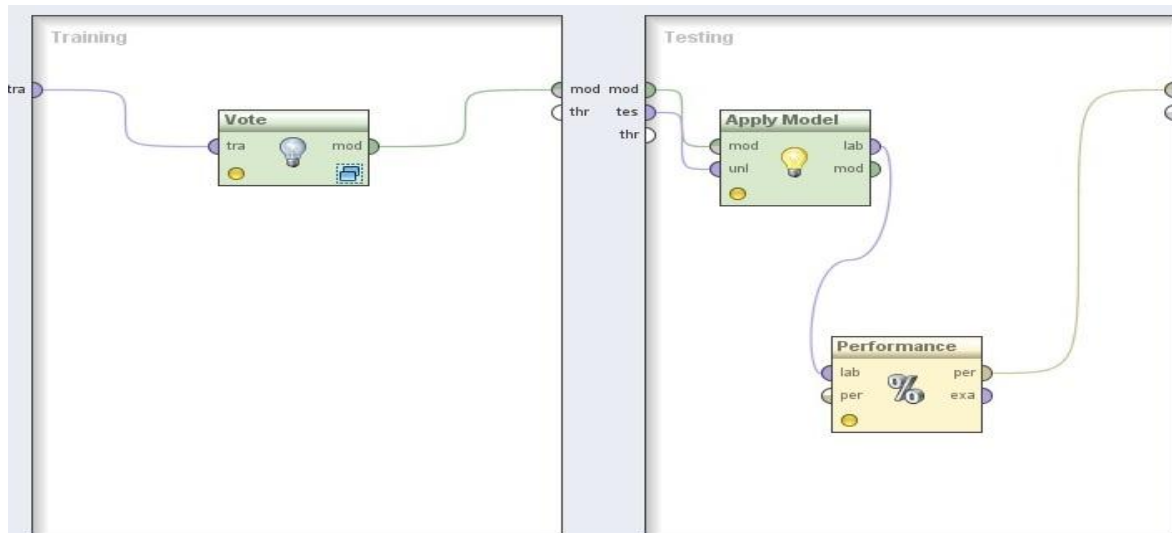| Classifier	Overall Accuracy	Classification Error	F-measure
Naïve Bayes	91.88	8.77	88.96
SVM	98.50	1.51	97.79
Decision Tree	98.92	1.08	98.42
K-NN with K=3	99.86	.13	99.79
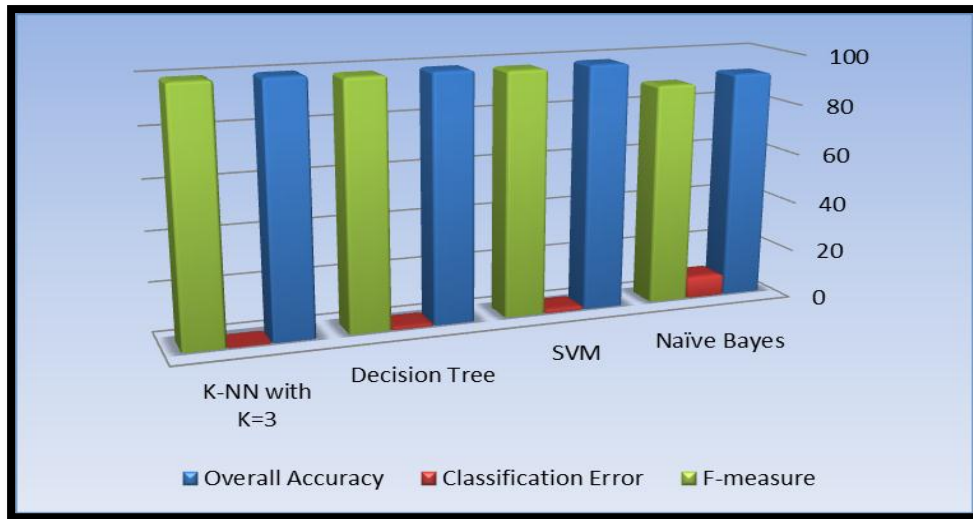
Figure 5.1: Experiments Results of Scenario I.

We can summarize accuracy results for experiments scenario I results as in K-NN algorithm, the highest accuracy result (99.86%), and the lowest accuracy was Naïve Bayes (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.

5.2.2 Experiment Scenario II (Multi Classifier)

The EGH data set is loaded using the rapid miner and used multi classifier methods by Decision Tree, K-nearest, Naïve Bayes and SVM. We classify instances after applying multi classification methods (voting) on EGH dataset as depicted in figure 5.2.


Figure 5.2: Multi Classifier by Vote Mechanism

47

We used a combination of algorithms (majority vote) as depicted in table 5.2 and figure 5.3 show the classification experiment results like accuracy ,classification error rate and f-measure for the EGH dataset.

We note that accuracy range up to 99.96%, which is considered a good result comparing with the results obtained from data mining classification experiments on the same dataset, we choose the best model that is K-NN and decision tree model to be offline detection.

Table 5.2: 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 Approach	99.96	0.03	99.95
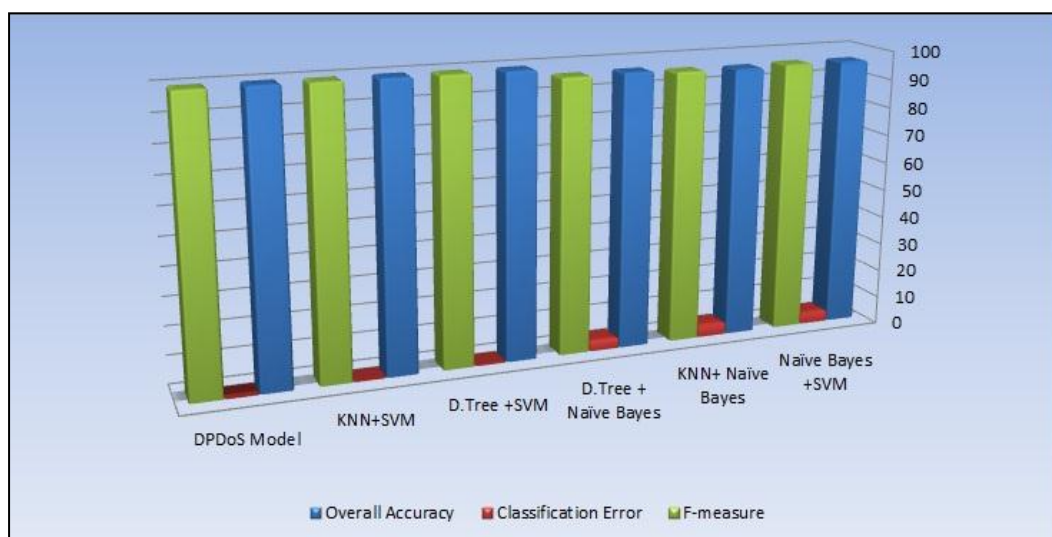
Figure 5.3: Experiments Results of Scenario II.

5.3   Defense Mechanism Experiment (phase 2)

The objective of the defense mechanism is to drop DoS attack packets while allowing legitimate packets (Normal) to reach their destination.

After we analyze the data by using data mining techniques, the result was ICMP and flooding attacks by clients attacking servers, which leads DoS attacks in our network.

We discussed related work, there are methods to detect and prevent DoS attacks[15,16,20] with different techniques ,Moreover N. Akhyari and S. Fahmy [39] used four layer namely Firewall(*PfSense)*, Network Intrusion Detection (Snort) , Vulnerability Scanner(*Nmap)* and Exploit Tool (Metasploit) to protect networks from DoS attacks or intrusions in general.

Gulay Oke and Georgios Loukas [45] used Distributed Defense Against Denial of Service Attacks depend on a detection mechanism combining statistical approach.

We use PfSense firewall and snort tool to protect the network from DoS attacks, there are many open-source firewalls, PfSense is chosen for the firewall component because it ensures the highest level of security and the best performances of the UNIX systems[32]. It allow or deny certain packets according to protocols, ports, IP addresses, payloads, connection states, etc.

The Snort tool [39] capable of performing packet logging and real-time traffic analysis on networks, it can carry out protocol analysis; content searching/ matching; and is commonly used to actively block or passively detect a variety of DoS attacks.

After we've talked about snort and firewall, both of them together make stronger defense mechanism, we did not use more than two techniques because the network performance is very important to us.

So many defense components effects the network performance by delaying or slowing the systems as a result of filtering scheme.

There are information systems used by patients and employees which did not work efficiently and with bad performance, so we used the defense mechanism depicted in figure 5.4.
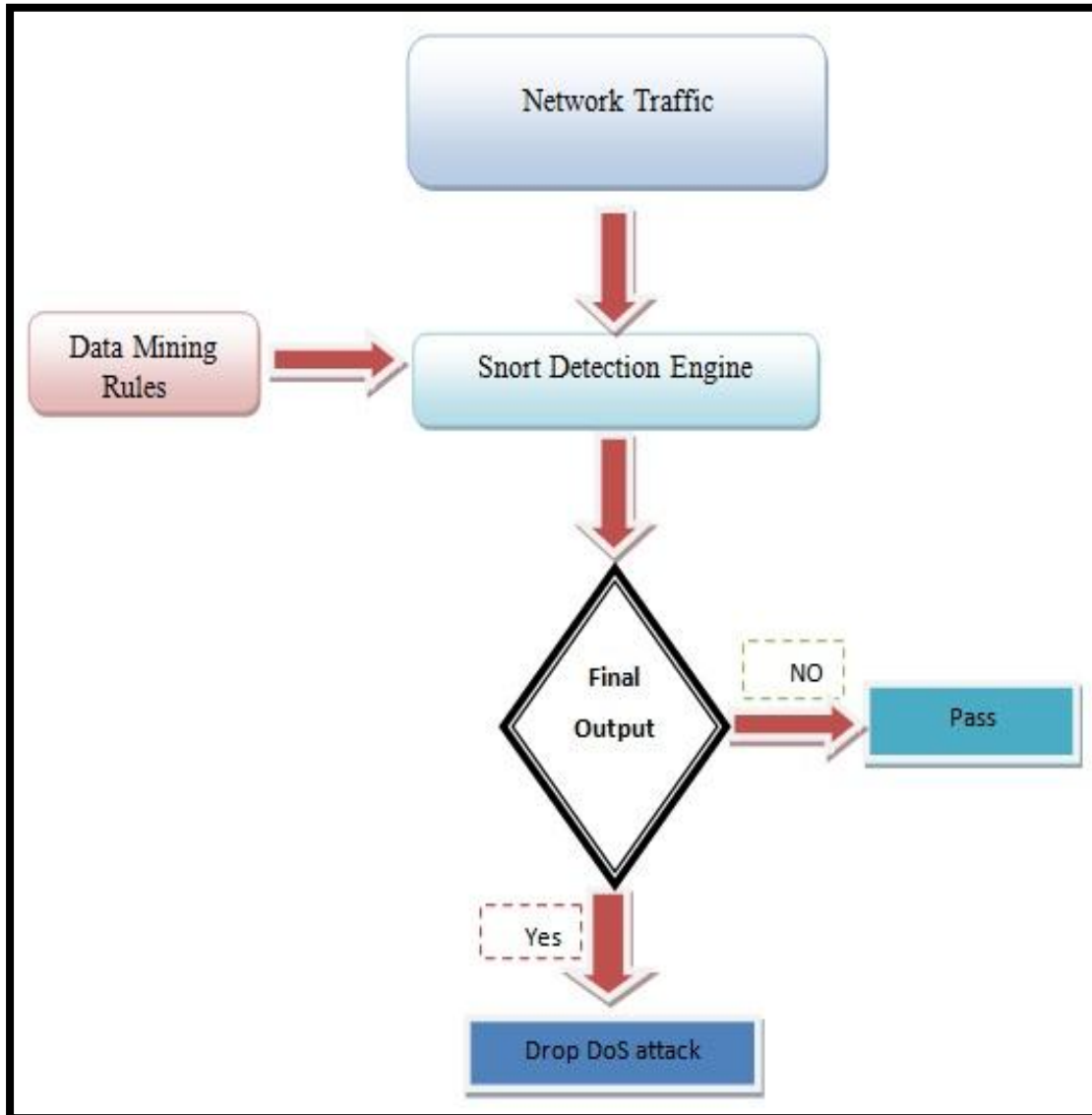
Figure 5.4: Defense Mechanism .

We implemented the snort tool on PfSense by using VMware workstation, the EGH data set was uploaded by PfSense and Tcpreplay [44] used to replaying the collected network traffic into snort, in addition we installed , configured snort and used the extracted rule from data mining model, then we tested the system and the results were highly accurate.

Data Mining Rules

We used KNN and DT in an approach. DT is an efficient method for producing classifiers from data. It can produce useful rules. These rules can be utilized as prediction statements. We clarify sample from extracted rules as follow :

    ! Source =192.168.3.142 : DoS(normal=0,DoS=3)

    Source =192.168.3.73

    ! Length >114: DoS   (normal=0,DoS=2)

    ! Length<114: Normal (normal=1,DoS=0)

Length <= 68

    ! Destination port=1323 : DoS (normal=0 , DoS=1)

To interpret the rules in the DT, as an example; the first branch of the tree says that, if length >68 ,source ip =192.168.3.142, the attack type can be predicted as "DoS" and so on for the rest of the tree.

The defense mechanism at EGH environment will be as depicted in figure 5.5.
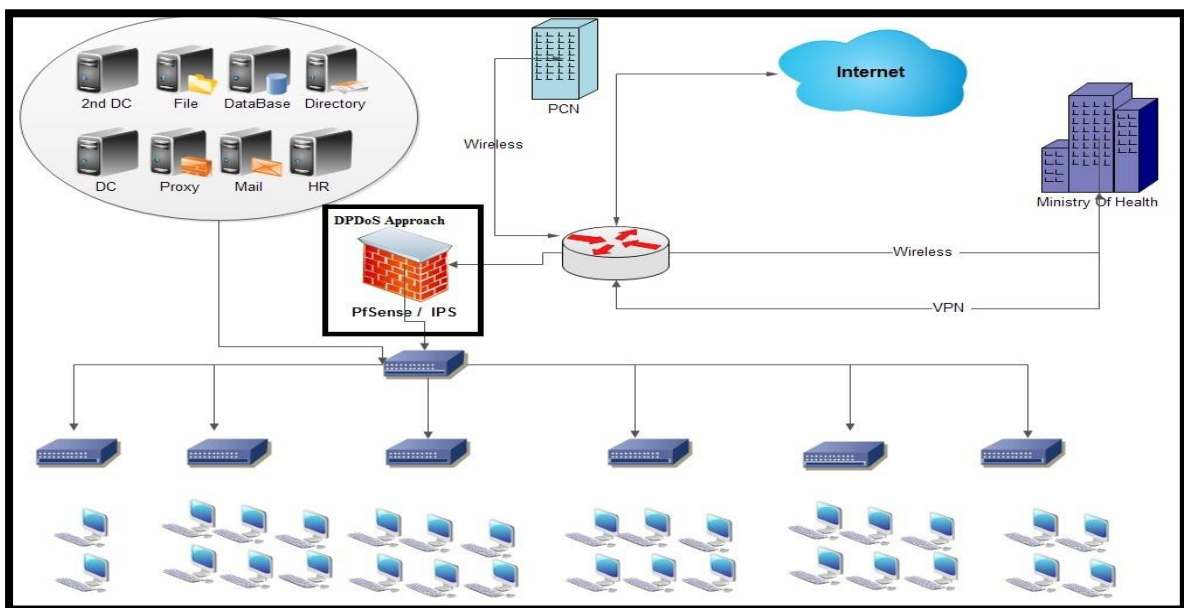

Figure 5.5: EGH Infrastructure with Defense Mechanism .

5.4 Discussion and summary

The following Table 5.3 and Figure 5.6 shows the summary of all experiments results.

Table 5.3: The Accuracy, Misclassification Rate, and F-measure Comparison of our Approach "DPDoS" and   Other Scenarios Mentioned Above.

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
DPDoS Approach	99.96	0.03	99.95
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 approach	99.96	0.03	99.95
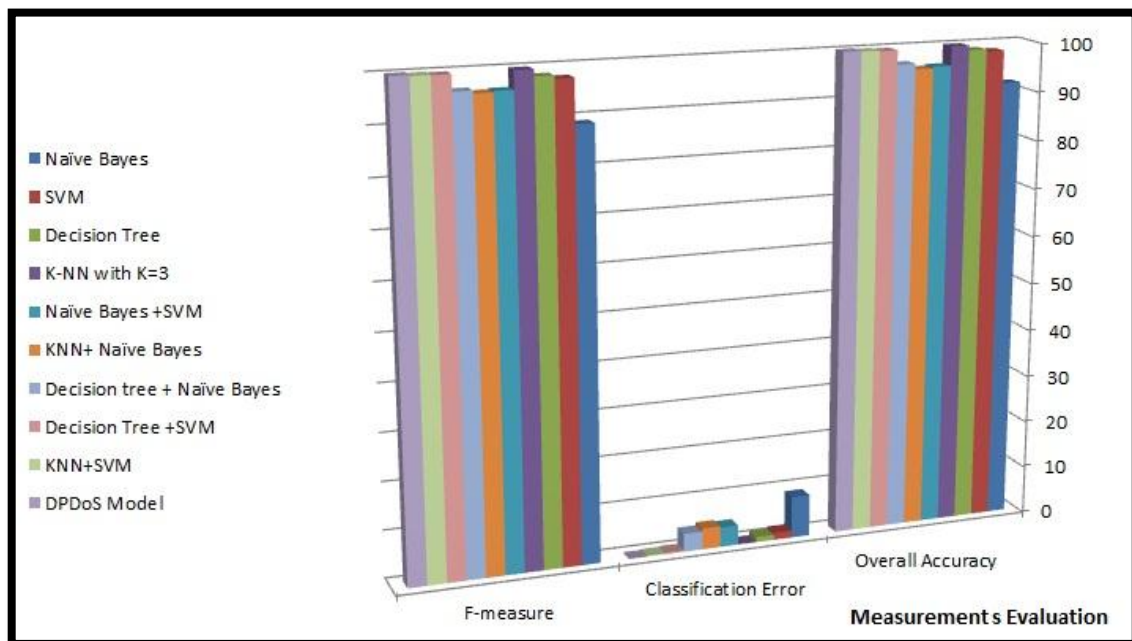
Figure 5.6: Summary of All Experiments

From our experiment, we consider our classification results in terms of accuracy, misclassification rate and f-measure, four different data mining algorithms (Naïve Bayes, Decision Tree ,K-nearest and Support vector machine) are evaluated one by one, our approach has been also evaluated. From Table 5.1, 5.2 and 5.3, each of data mining classification algorithms, and our approach can classify normal and DoS profile in accuracy, misclassification, and F-measure.

We summarize the experiment results as follows:

a) 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 algorithm.

b) The experiments on EGH dataset of scenario II, achieved the highest accuracy (99.96%), F-measure (99.95%), and lowest misclassification rate (.03) were in our approach.

c) In general, we can say that our approach has achieved good results from all experiments on EGH dataset, where scenario I and scenario II recorded the highest accuracy that was (99.96%), and lowest misclassification rate (.03), and F-measure (99.95%) which were in "DPDoS" approach .

d) 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.

e) Snort is a packet filtering rules engine, providing intrusion detection and prevention, allowing for policy enforcement and IP blocking, with custom and regularly updated dynamic rules, so after detecting the attack dropping it immediately by PfSense firewall .

f) The experiment in phase (1) and phase (2), achieved good results for accuracy, misclassification rate and F-measure .

The above experimental results confirm our thesis, which says that the multi classifier method has better accuracy than single classification techniques, our approach achieved the best classification accuracy for detecting DoS attacks, and preventing DoS attacks by defense mechanism (snort tool and PfSense firewall), discussed and mentioned above in section 5.3.

Chapter 6
Conclusion and Future Work

.

.

This chapter draws a conclusion, which includes its results, discussion, then gives some suggestions for future work.

6.1 Conclusion

In this thesis, we proposed an approach for detecting and preventing DoS Attacks, the proposed approach is called DPDoS and based on combining methods from data mining. The proposed approach structure and components were presented and explained, the purpose of using multi classifier method was to obtain the highest accuracy, f-measure, and reducing misclassification rates.

This thesis is composed of two phases (detection and prevention).

First phase, we captured data from EGH network by Wireshark program, then labeled and classified data into Normal / DoS, we dealt as well with noisy, missing, inconsistent data by some of the preprocessing techniques and selection of some attributes such as Protocol, Source Port, Destination Port, Time, source address, destination address and packet length columns to be used as our dataset.

The collected EGH dataset were used for two scenarios of experiments .

We used RapidMiner program to apply our approach, we have conducted a series of experiments to determine the two classifiers used in our "DPDoS" approach which are DT. and K-NN.

Finally, we implemented our strategy which we followed to achieve our goal, which is to develop an approach to detect DoS in LAN that can be valid for the security domain with high accuracy ,lowest misclassification error and F-measure.

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

Second phase (online detection) ,the rule set of the phase (1) used in defense mechanism by snort and PfSense firewall together, in addition we used tcpreplay to send the same EGH dataset, and the result of detecting and preventing DoS attacks was high accuracy and performance .

To confirm our experimental results, Table 6.1, compares our work with some other published work on the field of detecting and preventing DoS attacks.

Table 6.1: Comparison Between our Approach and Some Other Research Related to detect DoS attacks

Research	Method	accuracy	misclassification error	F-measure
DPDoS approach	K-NN and DT	99.96	0.03	99.95
D. kailashiya and Dr. Jain in [17].	DT	94.74	2.81	-
H. Nguyen and Y. Choi in [4].	K-NN	91	-	-
F. Leu and Z. Yang Li in [14]	Cumulative Sum algorithm	98.04	-	-
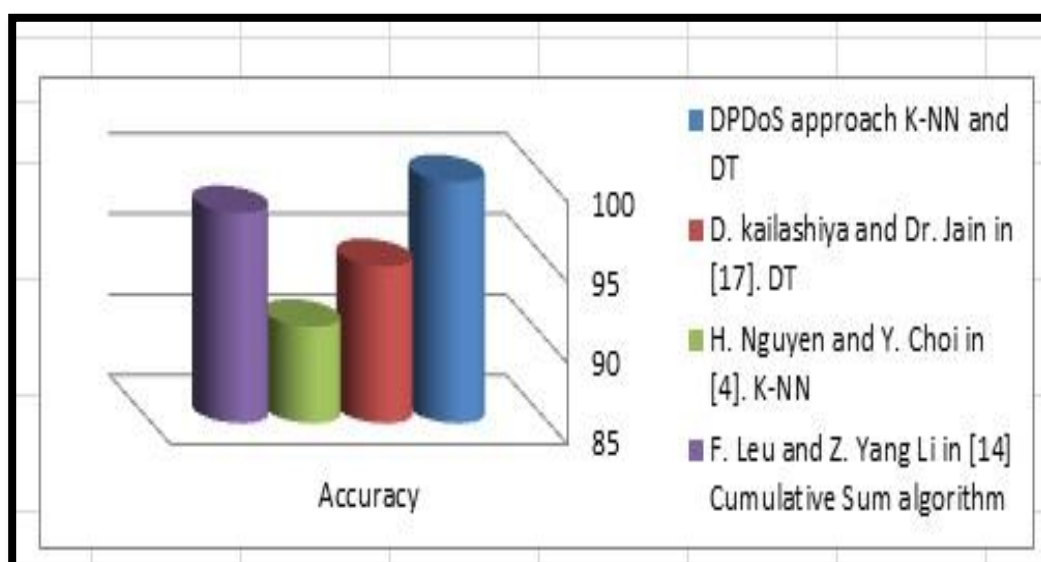
Figure 6.1: Comparison Between our approach and Related Work Models According to Accuracy.

We can conclude that "DPDoS" approach achieved the best results for performance measurements which are accuracy, classification error, and F-measure, comparing by other experiments as depicted in Table 5.4.

There are related work methods used to detect and prevent DoS attacks, based on router techniques, it's high in performance but our approach DPDoS is better than their models .

56

6.2 Future Work

Possible directions for future work include:

- Modification an approach to classify many types of the DoS attacks on network.
- Many types of worms or intrusions can be applied by an approach .
- New types of attacks such as Probing, User to Root and Remote to User Attacks can be applied by an approach.
- Modification an approach to detect many types of threats, DoS and Worms with accepted and sufficient accuracy

References

[1] P. K. Sree, "Exploring a Novel Approach for providing Software Security Using Soft Computing Systems," *International Journal of Security and its Applications* vol. 2, no. 2, April ,pp. 51–58, 2008

[2] H. F. Tipton,M. Krause, *Information Security Management Handbook*, Sixth Edition, 2007.

[3] M. R. Norouzian and S. Merati, "Classifying Attacks in a Network Intrusion Detection System Based on Artificial Neural Networks," *13th International Conference on Advanced Communication Technology (ICACT)*,Feb., pp. 868–873,2011.

[4] 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.

[5] 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.

[6] K. M. Elleithy, D. Blagovic, W. Cheng and P. Sideleau "Denial of Service Attack Techniques: Analysis, Implementation and Comparison," *Journal of Systemics, Cybernetics and Informatics,* vol. 3, no. 1, pp. 66–71, 2000.

[7] V. Jaiganesh, Dr. P. Sumathi and A. Vinitha, "Classification Algorithms in Intrusion Detection System: A Survey," *International Journal of Computer Technology and Applications* , vol. 4, no. 5, Sept., pp. 746–750, 2013.

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

[9] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in databases," AI Magazine , pp. 37–54, 1996.

[10] 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

[11] 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.

[12] H. A. Qeshta, "Adaptive Worms Detection Model Based on Multi Classifiers," M. S. Thesis, Islamic University, Gaza, 2012.

[13]   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.

[14]   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.

[15]   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.

[16]   S. G. Bhirud and V. Katkar, "SYN flood attack prevention using main-memory database management system," *Second Asian Himalayas International Conference on Internet (AH-ICI),* Nov., pp. 1–6, 2011.

[17]   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.

[18]   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.

[19]   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.

[20]   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.

[21]   J. Mirkovic, G. Prier, and P. Reiher, "Attacking DDoS at the source," *10th IEEE International Conference on Network Protocols* , Nov., pp. 312 – 321, 2002.

[22]   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.

[23]   H. Prajapati, "Distributed denial of service attacks," *International Journal For Technological Research In Engineering*, vol. 1, no. 4, Dec., pp. 183–186, 2013.

[24]   D. Ndumiyana, R. Gotora, and H. Chikwiriro, "Data Mining Techniques in Intrusion Detection: Tightening Network Security.," *International Journal of Engineering Research and Technology (IJERT)*, vol. 2, no. 5, pp. 2237–2248, 2013.

[25]   A. Youssef and A. Emam, "Network Intrusion Detection Using Data Mining and Network Behavior Analysis," *International Journal of Computer Science & Information Technology (IJCSIT)* vol. 3, no. 6, Dec., pp. 87–98, 2011.

[26]   R. Kumar and M. Nene, "A Survey on Latest DoS Attacks: Classification and Defense Mechanisms," *International Journal of Innovative Research in Computer and Communication Engineering*,vol. 1, no. 8, Oct., pp. 1847–1860, 2013.

[27]   D Olson., and D. Delen , "*Advanced data mining techniques*", Springer-Verlag Berlin Heidelberg, 2008.

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

[29]   I. Ahmad , A. B. Abdullah, A. S. Alghamdi, "Application of Artificial Neural Network in Detection of Probing Attacks," *IEEE Symposium on Industrial Electronics and Applications (ISIEA 2009)*, October 4-6, 2009.

[30]   S. Chaurasia and A. Jain, "Ensemble Neural Network and K-NN Classifiers for Intrusion Detection," *International Journal of Computer Science and Information Technologies* ,vol. 5, no. 2, pp. 2481–2485, 2014.

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

[32]   Pfsense firewall , https://www.pfsense.org (Oct. ,2014,),[last access].

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

[34]   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.

[35]   M. S. Hoque, Md. A. Mukit,Md. A. Bkas , "An Implementation of Intrusion Detection system using Genetic Algorithm," International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.2 , pp. 109–120, March 2012

[36]   U. Banerjee, A. Vashishtha, M. Saxena ,"Evaluation of the Capabilities of Wireshark as a tool for Intrusion Detection*," International Journal of Computer Applications (0975 – 8887)* vol. 6, No. 7, pp. 7–11, September 2010.

[37]   D. A. Antony, G. Singh, and E. J. Leavline , "Data mining in network security – Techniques & Tools : A research Perspective," *Journal of Theoretical and Applied Information Technology* ,vol. 57, no. 2, pp. 269–278, 2013.

[38]   M. Yao and B. Vocational, "An Effective Intrusion Detection System Based on Multi-layers Mining Methods," *International Journal of Security and Its Applications* , vol. 8, no. 5, pp. 311–322, 2014.

[39]   N. Akhyari and S. Fahmy, "Design of a Network Security Tool Using Open-Source Applications," *Australian Journal of Basic and Applied Sciences*,pp.٤٦-٤٠ , 2014.

[40]   S. Bansal ,M. Singh and S.Mittal , "Developing rules or Signatures to Detect Novel Attacks using open Source IDS : Snort," *International Journal of Advanced Research in Computer Science and Software Engineering* , vol. 4, no. 3, pp. 578–583, 2014.

[41]   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.

[42]   DoS , http://www.irchelp.org/irchelp/nuke/,(Jan.,2014), [Last access].

[43]   V. Balaji and K. Varalakshmi , "Differentiating Network Attacks using C4.5 Algorithm with Multiboosting," *International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)*, pp. 231–235, 2013.

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

[45]   G. Oke and G. Loukas, "Distributed Defence Against Denial of Service Attacks : A Practical View," BCS *International Academic Conference 2008 – Visions of Computer Science* , pp. 153–162, 2008.

[46]   EGH website ,http://www.egh.gov.ps/,(Nov.,2014)[last access].

[47]   W. Wang ,A. Brakensiek, G. Rigoll, "Combination of multiple classifiers for handwritten word recognition," *Proc. Eighth Int. Work. Front. Handwriting Recognition.*, pp. 117–122, 2002.