International Journal of Mechanical and Thermal Engineering

E-ISSN: 2707-8051
P-ISSN: 2707-8043
IJMTE 2023; 4(2): 41-54
Received: 15-06-2023
Accepted: 17-07-2023

Mohammad Taleghani
Associate Professor, Department of Industrial Management, Rasht Branch, Islamic Azad University, Rasht, Iran

Ataollah Taleghani
Ph.D. Student of Mechanical and Industrial Engineering, Toronto Metropolitan University, Toronto, Canada

Corresponding Author:
Mohammad Taleghani
Associate Professor, Department of Industrial Management, Rasht Branch, Islamic Azad University, Rasht, Iran

Providing a framework for classifying hybrid software in internet and virtual space applications

Mohammad Taleghani and Ataollah Taleghani

Abstract

Traffic classification refers to the processes used to categorize traffic based on features in the traffic and in accordance with specific classification objectives. In general, there are numerous forests, fields, and aims in the classification of internet traffic. Internet service providers must be aware of the types of traffic that are sent across their network Classifying Internet applications that are carried by computer networks is the specific goal, create a hybrid software classifier that can be applied to the classification of Internet traffic. Internet service providers (ISPs) attempt to boost bandwidth as a result of rising Internet traffic, but at the same time, more bandwidth is needed for Internet applications. Problems arise from the exponential growth of new internet applications using unregistered ports. The new programs may also contain a lot of viruses and dangerous code. In recent years, traffic classification has drawn more and more attention. With the help of direct and passive observation of the individual packets or stream of packets moving over the network, it seeks to provide the capability of automatically identifying the program that created a specific stream of packets. Data mining (DM) is a method for sifting through enormous databases in search of fresh, obscured, and practical information patterns. The knowledge discovery process includes the DM idea. This study employs a variety of techniques and functions as a hybrid classification. The foundation of many essential network monitoring and controlling jobs, such as billing, quality of service, security, and trend analyzers, is the classification of networks flows by their application type. Identification of Internet traffic is a crucial tool for network management. It enables operators to more accurately forecast upcoming traffic patterns and demand, and it enables security staff to spot unusual conduct.

Keywords: Classifying, hybrid software, internet application, virtual space application, software engineering

Introduction

Internet traffic classification is a technique that identifies applications and protocols of the Internet traffic [1] which is defined as the flow of data across the Internet [2].

Traffic classification describes the methods of classifying traffic by observing features in the traffic, and in line to particular classification goals. There might be some that only have a vulgar classification goal, such as bulk transfer and peer to peer file sharing. Some others will set a finer-grained classification goal, for instance the exact number of applications represented by the traffic. Traffic features included port number, application payload, and packet size, inter arrival time and other characteristic of the traffic. There are a large range of methods to allocate Internet traffic including exact traffic, for instance port (computer networking) number, payload, heuristic or statistical machine learning [3].

Network Traffic Classification Methods

Traffic classification is the process that categories computer network traffic according to various parameters [4]. There are four basic methods that were used in the internet traffic classification:

- Port number method.
- Deep packet inspection method (dpi).
- Signatures of network method.
- Machine learning method [5].

In fact, each of those methods has advantages and disadvantages. The following paragraphs give a brief description of these methods.

Port Number Method

This method is basically used the port number to classify internet traffic [6]. As known, some of internet application use a fixed number [7].

(Table 1) shows port numbers of some internet applications [8].

Table 1: Port numbers of some internet applications.

Applications	Port numbers
http	80
https	443
DNS	53
FTP-data	20
FTP-control	21

Machine Learning Approach

Have many Positives: and is the most famous methods commonly used and a strong method to extract the data and describe the structural data pattern and it used Wide applications, including search engine (Google). That's all about the Internet traffic classification definition and its methods now let's moving to the next topic Traffic classification was used to describe methods of classifying traffic based on features passively observed in the traffic, and according to specific classification goals. One might only have a coarse classification goal, i.e., whether it's transaction-oriented, bulk-transfer, or peer-to-peer file sharing. Or one might have a finer-grained classification goal, i.e., the exact application represented by the traffic. Traffic features could include the port number, application payload, or temporal, packet size, and addressing characteristics of the traffic. Methods to classify include exact matching, e.g., of port number or payload, heuristic, or machine learning (Statistics) [13].

Internet applications

In the field of traffic classification, the term application is used for both protocols (such as http) and internet application (such as WhatsApp). The common applications in the world is a social media applications, chatting, searching, find out what is happening in the world such as (YouTube, Whatsapp, http browser, twitter, etc.). In this project three of these applications (Whatsapp, http, and twitter) which commonly used were considered. Due to the widely used in any community, the ease of capture, and the availability for most customers and users. The following paragraphs briefly describe the internet applications.

Whatsapp

Whatsapp is a proprietary, cross-platform, encrypted instant messaging client for smartphones. It uses the Internet to make voice calls, video calls; send text messages, documents, PDF files, images, GIF, videos, user location, audio files, phone contacts and voice notes to other users using standard cellular mobile numbers. Whatsapp had a user database of over one billion, making it the most popular messaging application at the time [14].

Twitter

Since its creation in 2006, the micro blogging site Twitter has accumulated more than 554 million active registered users with 58 million tweets per day. Twitter provides users a communication platform to initiate and develop connections in real time with thousands of people with shared interests. It is also a way to get to know strangers who share the details of their daily lives. Every second, on average, around 6,000 tweets are tweeted on Twitter (visualize them here), which corresponds to over 350,000 tweets sent per minute, 500 million tweets per day and around 200 billion tweets per year [15].

HTTP

HTTP (Hypertext Transfer Protocol) is the set of rules for transferring files (text, graphic images, sound, video, and other multimedia files) on the World Wide Web. As soon as a Web user open their Web browser, the user is indirectly making use of HTTP. HTTP is an application protocol that runs on top of the suite of protocols such as TCP/IP (the foundation protocols for the Internet) [16].

HTTP provides a general framework to access control and authentication, via an extensible set of challenge-response authentication schemes, which can be used by a server to manage a client request and by a client to provide authentication information.

The HTTP authentication also provides an arbitrary, implementation specific construct for further dividing resources common to a given root URL. The realm value string, if present, is combined with the canonical root URL to form the protection space component of the challenge [17].

Problem Statement

With the increasing of Internet usage, Internet Server Provider (ISP) tries to increase the bandwidth, at the same time Internet applications requirement increases. The rapid increase in the new internet applications that have no registered ports creates a problem. In addition, the new applications can content many of malicious code and viruses.

The problem is how to classify internet traffic? Is this packet that we are interested in? The Internet traffic detection has been made more challenging because certain application uses dynamic port-negotiation mechanism and payload encryption. In addition, another challenge is identifying the most appropriate parameters to use in the classification phase.

Project Objectives

Generally, the internet traffic classification has many forest and fields and many goals. Internet service provider need to know which traffic carried in their network. In addition, Internet traffic classification was important in traffic engineering, and security issues.

The specific objective of this project can be summarized in the following points:

- To classify Internet applications which are carried by computer network.
- To analyze and understand the traffic specifications of the considered Internet applications (http, Whatsapp, and Twitter).
- To design hybrid software classifier which can be used in Internet traffic classification.

Project Questions

This project tries to answer some questions which are used to evaluate the results of the project. These questions are highlighted as following:

- What are the appropriate algorithms within machine learning (Weka) program to classify the considered applications (http, Whatsapp, and Twitter) and give a high performance?
- What are the traffic characteristics of the considered internet applications?
- What are suitable traffic features (packet size, inter arrival time, etc.) that can be used to identify considered application?
- Can we able to design a hybrid classifier base on port number and traffic statistical features to classify internet applications?

Existing System

At Saudi Arabia, in most case, traffic classifier not used. Only normal router and switches are fixed in the network entrances.

The existing classifier system can be divided into two-part DiffServ classifier and port number method classifier.

DiffServ classifier

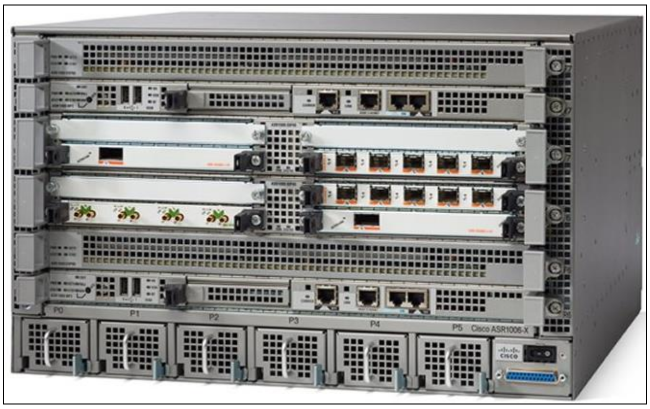


Fig 1: Cisco 1006-X Router (Support DiffServ feature) [19]

Differentiated services or DiffServ is a computer networking hardware architecture that specifies a simple, scalable and coarse-grained mechanism for classifying and managing network traffic and providing quality of service (QoS) on modern IP networks [18].

Problems in DiffServ classifier

When looking deeply in the mechanism of DiffServ, Per Hop Behavior (PHB) marks the packet to satisfy DiffServ function. The question arises here; what is adequate PHB used to achieves the quality of services. In other word, PHB was used DiffServ Code Point (DSCP) to mark the packet, the question is how to select DSCP to reach End-to-End quality of service. Another disadvantage of using hardware classifier is the high cost of these devices [20].

Port Number Classifier

Other existing classifier is Port Number Classifier which uses port number method as traffic features (parameter) to classify Internet applications. The main advantage of this method is the easy of the used. In addition, the speed of port number method was very high.

Problems in Port Number Classifier

- Cannot classify all Internet applications because Internet traffic includes more and more applications which use dynamic port numbers.
- Many Internet applications can potentially run on non-good port numbers.

System Analysis

Traffic classification has received increasing attention in the last years**.** It aims at offering the ability to automatically recognize the application that has generated a given stream of packets from the direct and passive observation of the individual packets, or stream of packets, flowing in the network.

Data Analysis

Data Flow Diagram

Figure2 illustrates the data flow diagram, which includes input, the system, and output. The input is internet flow packets which can collected from any internet entrance such as access point, router, Ethernet, etc. The packets is the unit of data that is routed between an origin and a destination on the Internet or any other packet-switched network. When any file (e-mail message, HTML file, Graphics Interchange Format file, Uniform Resource Locator request, and so forth) is sent from one place to another on the Internet, the Transmission Control Protocol (TCP) layer of TCP/IP divides the file into "chunks" of an efficient size for routing. Each of these packets is separately numbered and includes the Internet address of the destination.

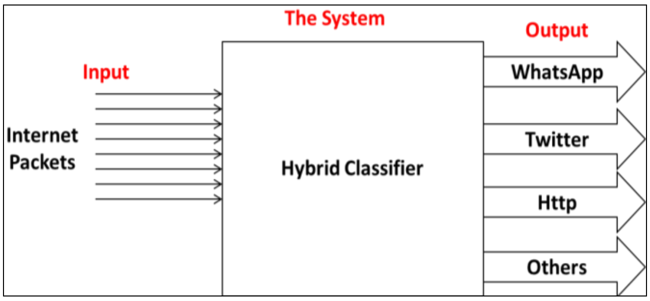


Fig 2: Data Flow Diagram

The second stage is the system which is the proposed hybrid classifier. This classifier is called hybrid because it based on two difference classification methods, statistical and port methods.

The output is expected to be the three considered applications (Whatsapp, http, and twitter) plus the other unconsidered applications which classified as one group.

System Requirements

Usually we pick Traffic from Wi-Fi or router and then entering a period of capturing and we used in these are phase program Wire shark which analytical program for traffic started our work with capturing and when we process the capturing looked brighter large amount of detail, for example, time, source, destination, protocol, length and information if we opened one package will display detailed information about the package and the information is displayed at this stage in the form of OSI (Open Systems Interconnect) layers and can be expanded or closed, and then move on to the stage traffic analytical study which is doing rejecting to traffic that we do not need and then move on to the stage classification that use the Weka, which is a package that open source software contains a set of algorithms that help us to analyze and extract data, Weka used for data exploration and classification which application belong to (Whatsapp, http, Twitter).

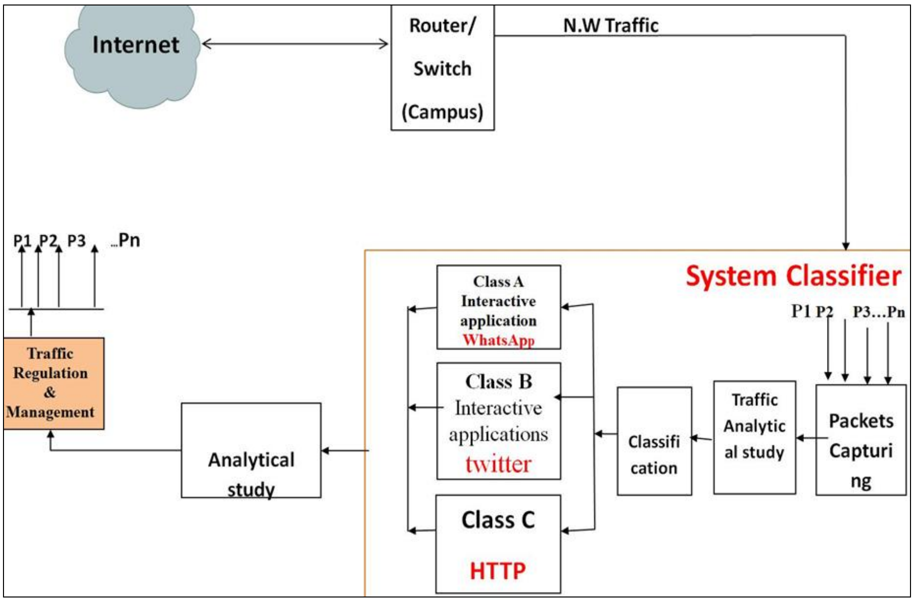


Fig 3: Proposed System Framework

Traffic Measurement (Capturing)

Measurement of network traffic allows network managers and analysts to make decisions about operations and plans. Its start to collect or capture data from the applications (Whatsapp, twitter, and http) by the capturing program Wireshark and the data will be collected in packet its possible reach thousands of packets in this phase these packets storing in the appropriate flow in the last step it compares the number of packets to the information's of capture.

Traffic Classification

The classification will be divided into three phases

- Manual classification that will depend on the previous classification and select the features contained.
- Training and testing on the packets of data will executed in offline classification examines the provided data (called the training dataset) and constructs (builds) a classification model. The model that has been built in the training phase is used to classify new unseen instances.
- Final testing on these packets.

Non-functional requirements

Performance requirement

From performance point of view, the proposed classifier can achieves high classification accuracy. This because the two used traffic features (inter-packet arrival time and packet length) always gives a high classification accuracy.

Usability requirements

The proposed system will be available all the time that it will be run. Moreover, the collected traffic can easily save and filtering. One type of packet capturing is filtering, in which filters are applied over network nodes or devices where data is captured. Conditional statements determine which data is captured. For example, a filter might capture data coming from our Wi-Fi network.

Design Constraints

The hybrid classifier includes some software and hardware requirements. The following paragraphs discuss these software and hardware.

Hardware and Software Environment

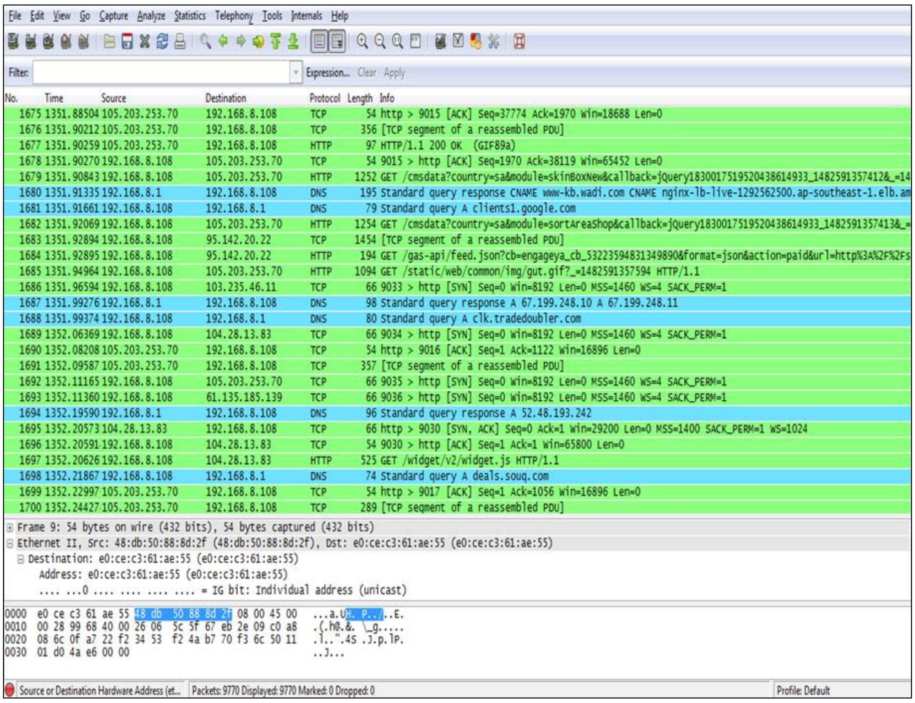
Wi-Fi Access Point

Plug your phone line into your router and data will be sent through it. This is converted into radio signals which are then picked up by devices with Wi–Fi capability, such as PCs, tablets, smartphones and games consoles. To connect to your router, you will be prompted to enter a password on the device you're using. This is usually supplied by your ISP and can be found on the bottom of the router. This tends to be a series of numbers and letters, designed to offer heightened security and protect your network from being used by others illegally [20].

Packet Analyzer (Wireshark)

A packet analyzer (also known as a network analyzer, protocol analyzer or packet sniffer, for particular types of networks, an Ethernet sniffer or wireless sniffer) is a computer program or piece of computer hardware that can intercept and log traffic that passes over a digital network or part of a network [23]. As data streams flow across the network, the sniffer captures each packet and, if needed, decodes the packet's raw data, showing the values of various fields in the packet, and analyzes its content according to the appropriate RFC or other specifications. Packet capture is the process of intercepting and logging traffic [24].

Wireshark is a data capturing program that understands the structure of different networking protocols. It's an open source tool for profiling network traffic and analysis packets. Typically, will display information in a panel. Is doing a capture to all packets that pass through specific network card and coping the contents that passing then analysis it [25].


File Edit View Go Capture Analyze Statistics Telephony Tools Internals Help
Filter:
Expression... Clear Apply

```
No.   Time       Source          Destination     Protocol Length Info
 1675 1351.88504 105.203.253.70  192.168.8.108   TCP        54 http > 9015 [ACK] Seq=37774 Ack=1970 Win=18688 Len=0
 1676 1351.90212 105.203.253.70  192.168.8.108   TCP       356 [TCP segment of a reassembled PDU]
 1677 1351.90259 105.203.253.70  192.168.8.108   HTTP       97 HTTP/1.1 200 OK  (GIF89a)
 1678 1351.90270 192.168.8.108   105.203.253.70  TCP        54 9015 > http [ACK] Seq=1970 Ack=38119 Win=65452 Len=0
 1679 1351.90843 192.168.8.108   105.203.253.70  HTTP     1252 GET /cmsdata?country=sa&module=skinBoxNew&callback=jQuery1830017519520438614933_1482591357412&_=14
 1680 1351.91335 192.168.8.1     192.168.8.108   DNS       195 Standard query response CNAME www-kb.wadi.com CNAME nginx-lb-live-1292562500.ap-southeast-1.elb.am
 1681 1351.91661 192.168.8.108   192.168.8.1     DNS        79 Standard query A clients1.google.com
 1682 1351.92069 192.168.8.108   105.203.253.70  HTTP     1254 GET /cmsdata?country=sa&module=sortAreaShop&callback=jQuery1830017519520438614933_1482591357413&_=
 1683 1351.92894 192.168.8.108   95.142.20.22    TCP      1454 [TCP segment of a reassembled PDU]
 1684 1351.92895 192.168.8.108   95.142.20.22    HTTP      194 GET /gas-api/feed.json?cb=engageya_cb_53223594831349890&format=json&action=paid&url=http%3A%2F%2Fs
 1685 1351.94964 192.168.8.108   105.203.253.70  HTTP     1094 GET /static/web/common/img/gut.gif?_=1482591357594 HTTP/1.1
 1686 1351.96594 192.168.8.108   103.235.46.11   TCP        66 9033 > http [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=4 SACK_PERM=1
 1687 1351.99276 192.168.8.1     192.168.8.108   DNS        98 Standard query response A 67.199.248.10 A 67.199.248.11
 1688 1351.99374 192.168.8.108   192.168.8.1     DNS        80 Standard query A clk.tradedoubler.com
 1689 1352.06369 192.168.8.108   104.28.13.83    TCP        66 9034 > http [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=4 SACK_PERM=1
 1690 1352.08208 105.203.253.70  192.168.8.108   TCP        54 http > 9016 [ACK] Seq=1 Ack=1122 Win=16896 Len=0
 1691 1352.09587 105.203.253.70  192.168.8.108   TCP       357 [TCP segment of a reassembled PDU]
 1692 1352.11165 192.168.8.108   105.203.253.70  TCP        66 9035 > http [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=4 SACK_PERM=1
 1693 1352.11360 192.168.8.108   61.135.185.139  TCP        66 9036 > http [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=4 SACK_PERM=1
 1694 1352.19590 192.168.8.1     192.168.8.108   DNS        96 Standard query response A 52.48.193.242
 1695 1352.20573 104.28.13.83    192.168.8.108   TCP        66 http > 9030 [SYN, ACK] Seq=0 Ack=1 Win=29200 Len=0 MSS=1400 SACK_PERM=1 WS=1024
 1696 1352.20591 192.168.8.108   104.28.13.83    TCP        54 9030 > http [ACK] Seq=1 Ack=1 Win=65800 Len=0
 1697 1352.20626 192.168.8.108   104.28.13.83    HTTP      525 GET /widget/v2/widget.js HTTP/1.1
 1698 1352.21867 192.168.8.108   192.168.8.1     DNS        74 Standard query A deals.souq.com
 1699 1352.22997 105.203.253.70  192.168.8.108   TCP        54 http > 9017 [ACK] Seq=1 Ack=1056 Win=16896 Len=0
 1700 1352.24427 105.203.253.70  192.168.8.108   TCP       289 [TCP segment of a reassembled PDU]

⊞ Frame 9: 54 bytes on wire (432 bits), 54 bytes captured (432 bits)
⊟ Ethernet II, Src: 48:db:50:88:8d:2f (48:db:50:88:8d:2f), Dst: e0:ce:c3:61:ae:55 (e0:ce:c3:61:ae:55)
  ⊟ Destination: e0:ce:c3:61:ae:55 (e0:ce:c3:61:ae:55)
      Address: e0:ce:c3:61:ae:55 (e0:ce:c3:61:ae:55)
      .... ...0 .... .... .... .... = IG bit: Individual address (unicast)

0000  e0 ce c3 61 ae 55 48 db  50 88 8d 2f 08 00 45 00   ...a.UH. P../..E.
0010  00 28 99 68 40 00 26 06  5c 5f 67 eb 2e 09 c0 a8   .(.h@.&. \_g.....
0020  08 6c 0f a7 22 f2 34 53  f2 4a b7 70 f3 6c 50 11   .l..".4S .J.p.lP.
0030  01 d0 4a e6 00 00                                  ..J...
```

Source or Destination Hardware Address (et...
Packets: 9770 Displayed: 9770 Marked: 0 Dropped: 0
Profile: Default

Fig 4: Wireshark Interface.

Machine Learning Tool

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can change when exposed to new data.

In this project, Waikato Environment for Knowledge Analysis (Weka) are used as machine learning tools. Weka is a data mining/machine learning tool developed by Department of Computer Science, University of Waikato, New Zealand. In addition, Weka is open source software, so there is no need to pay any cost to use this tool.

Smart Devices

A smart device is an electronic device, generally connected to other devices or networks via different wireless protocols such as Bluetooth, NFC, Wi-Fi, 3G, etc., that can operate to some extent interactively and autonomously. Several notable types of smart devices are smartphones,

Smart watches, smart bands and smart key chains. The term can also refer to a device that exhibits some properties of ubiquitous computing, including-although not necessarily-artificial intelligence**.**

Smart devices can be designed to support a variety of form factors, a range of properties pertaining to ubiquitous computing and to be used in three main system environments: physical world, human-centered environments and distributed computing environments [22].

Network Router

A router is a networking device that forwards data packets between computer networks. Routers perform the traffic directing functions on the Internet. A data packet is typically forwarded from one router to another router through the networks that constitute the internetwork until it reaches its destination node.

A router is connected to two or more data lines from different networks. When a data packet comes in on one of the lines, the router reads the address information in the packet to determine the ultimate destination. Then, using information in its routing table or routing policy, it directs the packet to the next network on its journey. This creates an overlay internetwork.

The most familiar type of routers are home and small office routers that simply pass IP packets between the home computers and the Internet. An example of a router would be the owner's cable or DSL router, which connects to the Internet through an Internet service provider (ISP). More sophisticated routers, such as enterprise routers, connect large business or ISP networks up to the powerful core routers that forward data at high speed along the optical fiber lines of the Internet backbone. Though routers are typically dedicated hardware devices, software-based routers also exist [22].

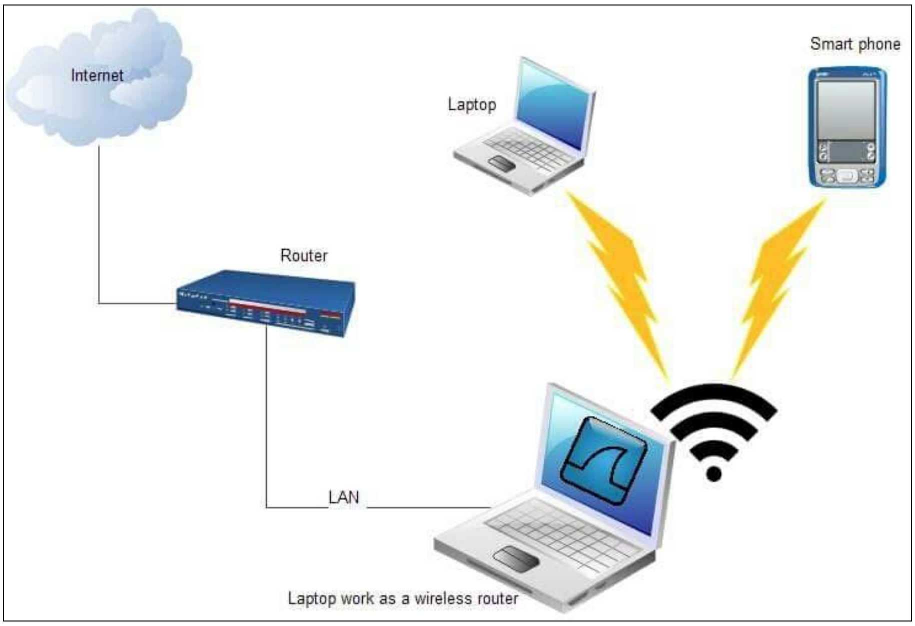


Fig 5: Network environment of proposed system

Figure 5 illustrates the network environment of the proposed system which show that the access point (Wi-Fi) received the internet and distributed to the surrounding devices.

The data was passing through the AP router which supports capture process. This AP can connect to different networks.

The data we capturing by using Wireshark program. This project considers and analyze the traffic of three types of applications (What's app, Twitter and Http, In the next stage Weka program was used to be classify the considered application in two stages training and testing (see section)

Algorithms to be used

Data Mining (DM) is a technique which is used to find new, hidden and useful patterns of knowledge from large databases. From statistics, artificial intelligence and data warehouses, it is very easy to design methods and procedures to classify the data for the use of real-world applications. DM concept is actually part of the knowledge discovery process [26]

Many of classification algorithms has been proposed by several researchers in the field of classification applications and investigated data using decision tree algorithms. They used these algorithms to predict classification of breast cancer data. They selected classification algorithm to find the most suitable one for predicting cancer [27].

This project used ten algorithms to training the hybrid classifier. These algorithms come from two categories Tree and Rules.

Tree Category

Decision tree learning is a method commonly used in data mining [27]. The goal is to create a model that predicts the value of a target variable based on several input variables. The following paragraphs explain the seven algorithms of the tree category which used in this project.

J48 Algorithm

J48 tree algorithm: Every aspect of the data is to split into minor subsets to base on a decision. J48 examine the normalized information gain that actually the outcomes the splitting the data by choosing an attribute.

The splitting methods stop if a subset belongs to the same class in all the instances. J48 constructs a decision node using the expected values of the class. J48 decision tree can handle specific characteristics, lost or missing attribute values of the data and differing attribute costs. Here precision can be increased by pruning [28].

Simple Cart Tree Algorithm

Simple Cart method is CART (Classification and Regression Tree) analysis. CART is abbreviated as Classification and Regression Tree algorithm. It was developed by Leo Breiman in the early 1980s. It is used for data exploration and prediction also. Classification and regression trees are classification methods which in order to construct decision trees uses historical data. CART uses learning sample which is a set of historical data with pre-assigned classes for all observations for building decision tree. Simple Cart (Classification and regression tree) is a classification technique that generates the binary decision tree [29].

Random Tree Algorithm

Random Tree is a supervised Classifier; it is an ensemble learning algorithm that generates many individual learners. It employs a bagging idea to produce a random set of data for constructing a decision tree. In standard tree each node is split using the best split among all variables. In arandom forest, each node is split using the best among the subset of predicators randomly chosen at that node. Random trees have been introduced by Leo Breiman and Adele Cutler. The algorithm can deal with both classification and regression problems. Random trees is a collection (ensemble) of tree predictors that is called forest.

RepTree Algorithm

RepTree uses the regression tree logic and creates multiple trees in different iterations. After that it selects best one from all generated trees. That will be considered as the representative. In pruning the tree, the measure used is the mean square error on the predictions made by the tree. Basically, Reduced Error Pruning Tree ("REPT") is fast decision tree learning and it builds a decision tree based on the information gain or reducing the variance [29-30].

Random Forst Tree Algorithm

Random forests or random decision forests [30-31] are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (Classification) or mean prediction (Regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set [32].

J48graft Tree Algorithm

Although simple methods of selecting appropriate trees have typically been considered advisable, tree grafting differs in that this process works on the assumption that similar objects are very likely in the same class. Tree grafting therefore attempts to generate superior classification models at the expense of producing highly complex trees [33].

BFTREE Algorithm

Judeay described the best-first search which searches up to the collected point and additional knowledge about the problem domain. It expands the most promising node chosen based on specified rules [34].

Rules Category

The following paragraphs explain the three algorithms of the tree category which used in this project.

JRIP rules classifiers

JRip (RIPPER) is one of the basic and most popular algorithms. Classes are examined in increasing size and an initial set of rules for the class is generated using incremental reduced error JRip (RIPPER) proceeds by treating all the examples of a particular judgment in the training data as a class, and finding a set of rules that cover all the members of that class [35].

PART Classifier

PART (Partial Decision Tree) [36] is an indirect technique for constructing classification rules.

RIDOR Classifier

Ripple down Rule learner (RIDOR) is also a direct classification method. First and foremost, it constructs the default rule and then produces the exceptions for the default rule with lowest error rate [36].

Project Management Strategies

Identification of network traffic is crucial in network management and monitoring purposes. Nowadays port based and payload-based classification methods have become inadequate as many applications use dynamically allocated port numbers, masquerade to be another application by using some standard port number or use encryption to avoid detection [37-38]. Internet traffic management, also known as application traffic management, refers to tools that monitor the flow of Web application traffic over a network. These tools route traffic among multiple devices within a network, limiting delays and freeing bandwidth,[Classification management is the way in which decisions relating to organizational structures and the work value of jobs are managed [39-40].

In this project we have applications (http, watts, twitter) we work on them to capture the traffic and we need to install the watts and twitter on the computer so that the capture process on the same device, and the capture process at the same time as it is working on it, And also http when we look at websites, and the location of the capture is at the university or home so that there is a certain number of people connected online.

Each application has a different capacity (the amount of data transmitted) by capture per second, computes the data each application carries, and in the capture process we capture a large number of traffic.

We control and manage these applications during the capture process and when editing the data that was captured.

In our project, we defined two types of data to work on, the length and time of each application.

Development Method

Network traffic classification can be used to identify different applications and protocols that exist in a network. (Shaikh and Prof. Dr. D.G. Harkut)

This research works as a hybrid classification and uses many methods as follows:

First method is by using port numbers. This method is Fast and low resource-consuming. It is supported by many Network devices. It does not implement the application-layer Payload, so it does not compromise the users' privacy. It is Useful only for the applications and services, which use fixed Port numbers since easy to cheat by changing the port Number in the system.

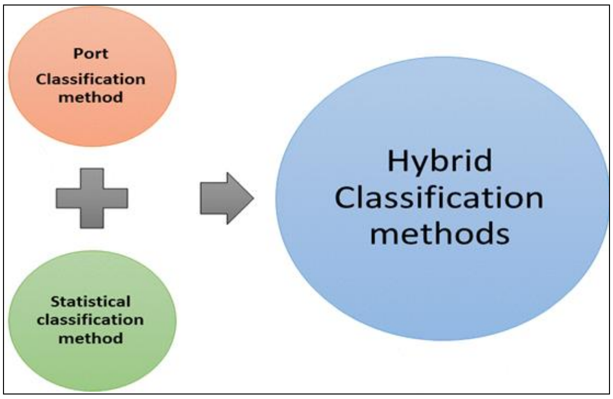

Port Classification method
Statistical classification method
Hybrid Classification methods

Fig 6: Hybrid Classification Methods.

Second method based on statistical classification which relies on statistical analysis of attributes such as packet sizes and packet length and packet inter-arrival times. It often uses Machine Learning Algorithms as Part, J48Grft, J48, Random Tree, or Random Forest. This technique is fast technique compared to port-based classification. It can detect the class of yet unknown applications. (Shaikh and Prof. Dr. D.G. Harkut)

Future Enhancements plans

Offline Classification

In this section we will capturing the data on the wireshark program and modify the packets (delete & underestimated) then the classify and run the algorithms will done in the weka program and each stage is managed in a specific way and each stage in this section takes place at different times and this task is called the training task.

Online Classification

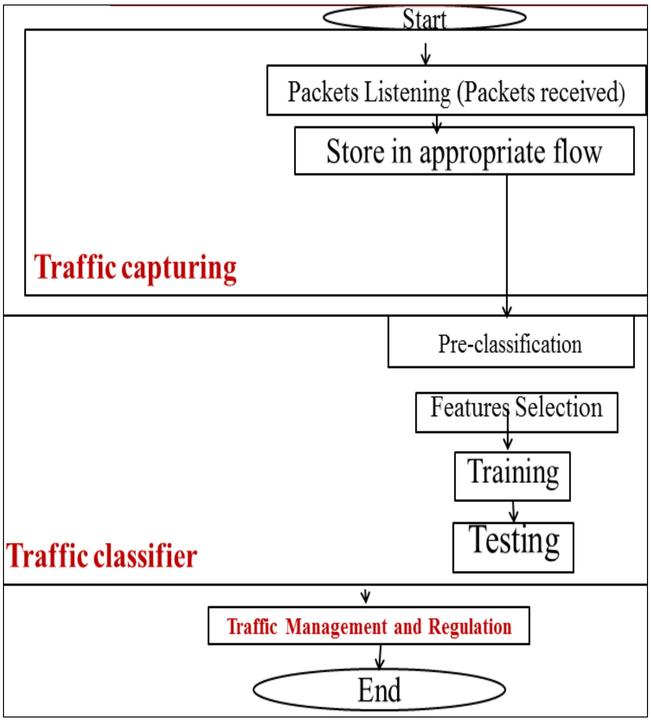
This section is different from the previous section because the two stages of capture and classification will be done simultaneously and will be in the Matlab program, which will classify the data that was captured on the wireshark program. Matlab will know the applications that used and compare between their data and distinguish each application from the other through its special port number. The task is called a testing task.

The obstacles that will face us in online section

Several obstacles will face online classification. First: it is very difficult for wireshark program in the online section to do the data capture and then classification at the same time because the large number of packets and information that contained. The Matlab program cannot be modified or underestimated these packets as we did in the section of the offline. Here we want to compare the data of both sections. This hybrid classifier can enhance to be used as online classifier.

System design

System Architecture and Program Flow



Start
Packets Listening (Packets received)
Store in appropriate flow
Traffic capturing
Pre-classification
Features Selection
Training
Testing
Traffic classifier
Traffic Management and Regulation
End

Fig 7: System Architecture of Hybrid Classifier

Classifying network flows by their application type is the backbone of many crucial network monitoring and controlling tasks, including billing, quality of service, security and trend analyzers. Internet traffic identification is an important tool for network management. It allows operators to better predict future traffic matrices and demands, security personnel to detect anomalous behavior, and researchers to develop more realistic traffic models.

Classification have several shortcomings. These limitations have motivated the study of classification techniques that build on the foundations of learning theory and statistics.

Our method is based on a hybrid. The proposed classifier is both fast and accurate, as implied by our feasibility tests, which included implementing and intergrading statistical classification into a real time embedded environment.

The data testing phase is a MATLAB stage where it is online. When the process is picking up the tracks and the process of rating if this path is (http, twitter, WhatsApp) at the same time. If the whole classification process goes through the previous two stages and in the same order of stages.

Major Modules: This project includes three types of modules, capturing, classification, and regulation

Modules 1: Network Traffic Capturing: In this project the data was collected from the applications that's manually runs through network and from located device. Three types of Internet application (http, twitter, WhatsApp) will be captured. This means in our machine (computer or mobile) only the considered applications will be run. Based on this filtering the collected data will be known.

Modules 2: Network Traffic Classification

Characterization of Internet traffic has become over the past few years one of the major challenging issues in telecommunication networks. Network traffic measurement has recently gained more interest as an important network-engineering tool for networks of multiple sizes.

In this project a hybrid classifier was proposed. This classifier is based in two common classification methods port and statistical methods. In statistical classifier, two traffic features (inter-packet arrival time and packet length) was used as parameter to classify the three considered applications (http, twitter, WhatsApp). The common port was used (such as 80 for http) was used in port method classifier.

Modules 3: Network Traffic Management

Network traffic management is applied to ensure that networks operate efficiently, fixed-line and mobile internet service providers (ISPs) can restrict or ration traffic on their networks, or give priority to some types of traffic over others during peak periods or more generally.

The three considered applications can be easily controlled. This means, network administrator can close any one of this application or can increase/decrease the bandwidth.

Sub modules: The three major modules discussed in this section can be divided in some sub modules. The following paragraphs discuss these sub modules

Environment of Capturing of the considered applications: As mentioned before, this project considers three type of applications (http, twitter, WhatsApp)

Http Traffic was capturing by preparing appropriate environment. This means only http traffic was run during traffic capturing.

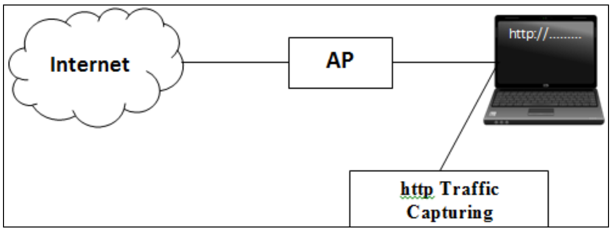


Fig 8: Http traffic capturing environment

WhatsApp traffic was captured by using Whatsapp for web software. This software was installed in a laptop which used as the first part of the communication. The second part of the WhatsApp chatting/call is normal smart phone. The generated traffic between these two parts was captured and collected.

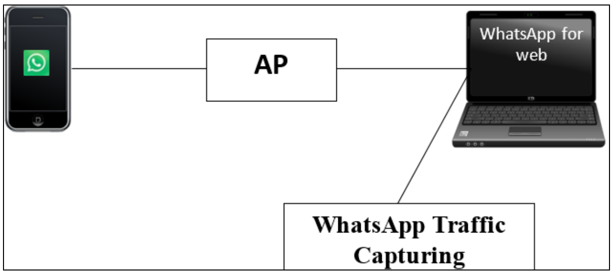


Fig 9: WhatsApp Traffic capturing Environment

Twitter traffic was captured on the laptop by using twitter application directly on the smart phone which log into internet through the laptop that works as a hot spot wireless router.

Features selection: In the first step of classification, Packet Inter arrival time and packet length were selected as features for Weka machine learning classifier. The advantage of selecting of only two features is reducing.

Training data: The data here will be simple, modified and sometimes few because the two phases of capture and classification are done at different times. And we will delete some columns and reduce the number of pacifiers to facilitate the classification in the Weka program and the important columns at this stage is (time - length - app-name)

Testing data

At this stage, the data will be more detailed and many packets because the capture and classification will be performed at the same time **.**So, that the data is taken after the capture process in the Wireshark program immediately participate in the MATLAB program to classify and arrange and compare data and identify applications by the special port number. So, the data will be many without deleting or reducing the number of columns and packets.

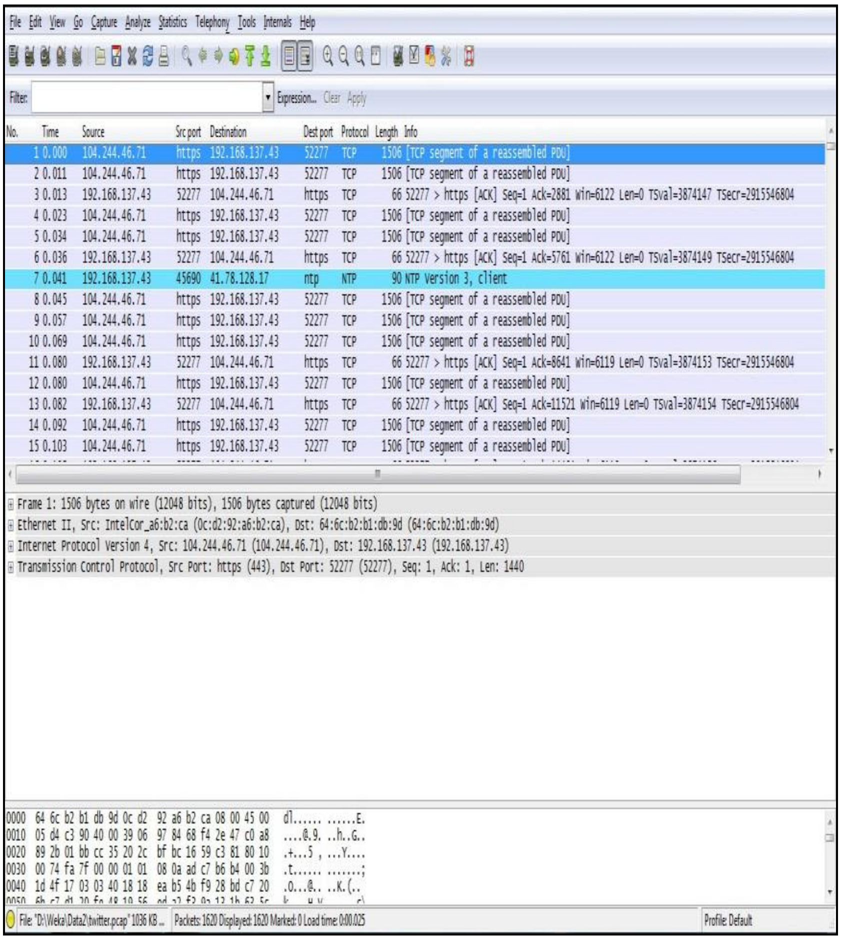
Validation

This project designed hybrid software classifier which was able to identify three types of Internet application (http, twitter, WhatsApp). The proposed hybrid classifier was implemented and tested by using real internet traffic.

Traffic Capturing

Figure 10 illustrate the capturing of the three considered applications (http, twitter, WhatsApp). Wireshark software was used to capture and analysis these applications traffic. The capture file contains a large number of packets, carrying information and data about the captured application (source, time, length, etc.). In the end we have 3 files from the Wireshark, one for each application.

During the capture process, manually only the needed traffic was generated and captured. This means, all the other Internet traffic was prevented from generated traffic. Even the windows and applications update were closed.



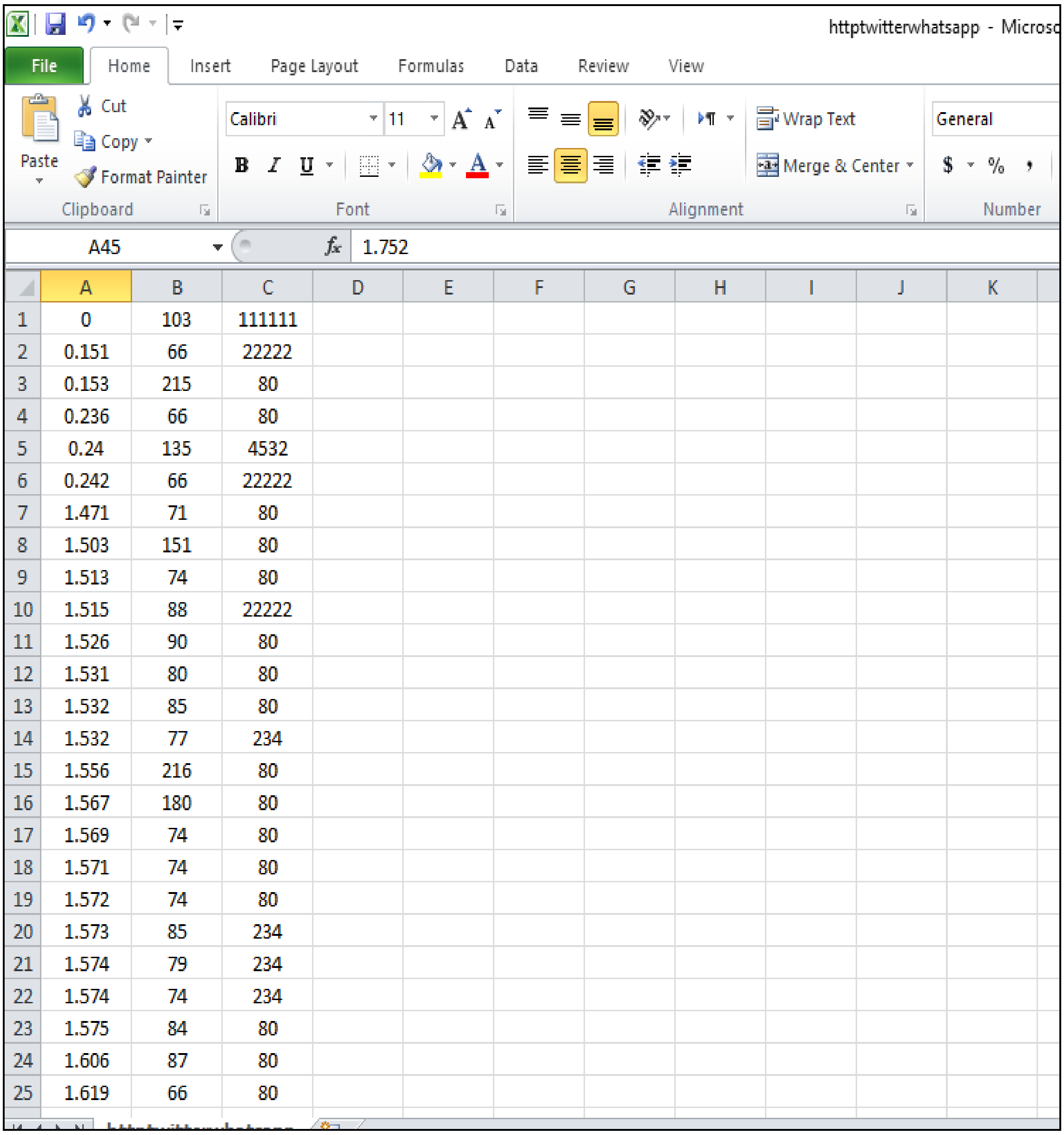
```
No.   Time   Source          Src port  Destination      Dest port  Protocol  Length  Info
   1 0.000   104.244.46.71   https     192.168.137.43   52277      TCP       1506 [TCP segment of a reassembled PDU]
   2 0.011   104.244.46.71   https     192.168.137.43   52277      TCP       1506 [TCP segment of a reassembled PDU]
   3 0.013   192.168.137.43  52277     104.244.46.71    https      TCP         66 52277 > https [ACK] Seq=1 Ack=2881 Win=6122 Len=0 TSval=3874147 TSecr=2915546804
   4 0.023   104.244.46.71   https     192.168.137.43   52277      TCP       1506 [TCP segment of a reassembled PDU]
   5 0.034   104.244.46.71   https     192.168.137.43   52277      TCP       1506 [TCP segment of a reassembled PDU]
   6 0.036   192.168.137.43  52277     104.244.46.71    https      TCP         66 52277 > https [ACK] Seq=1 Ack=5761 Win=6122 Len=0 TSval=3874149 TSecr=2915546804
   7 0.041   192.168.137.43  45690     41.78.128.17     ntp        NTP         90 NTP Version 3, client
   8 0.045   104.244.46.71   https     192.168.137.43   52277      TCP       1506 [TCP segment of a reassembled PDU]
   9 0.057   104.244.46.71   https     192.168.137.43   52277      TCP       1506 [TCP segment of a reassembled PDU]
  10 0.069   104.244.46.71   https     192.168.137.43   52277      TCP       1506 [TCP segment of a reassembled PDU]
  11 0.080   192.168.137.43  52277     104.244.46.71    https      TCP         66 52277 > https [ACK] Seq=1 Ack=8641 Win=6119 Len=0 TSval=3874153 TSecr=2915546804
  12 0.080   104.244.46.71   https     192.168.137.43   52277      TCP       1506 [TCP segment of a reassembled PDU]
  13 0.082   192.168.137.43  52277     104.244.46.71    https      TCP         66 52277 > https [ACK] Seq=1 Ack=11521 Win=6119 Len=0 TSval=3874154 TSecr=2915546804
  14 0.092   104.244.46.71   https     192.168.137.43   52277      TCP       1506 [TCP segment of a reassembled PDU]
  15 0.103   104.244.46.71   https     192.168.137.43   52277      TCP       1506 [TCP segment of a reassembled PDU]

⊞ Frame 1: 1506 bytes on wire (12048 bits), 1506 bytes captured (12048 bits)
⊞ Ethernet II, Src: IntelCor_a6:b2:ca (0c:d2:92:a6:b2:ca), Dst: 64:6c:b2:b1:db:9d (64:6c:b2:b1:db:9d)
⊞ Internet Protocol Version 4, Src: 104.244.46.71 (104.244.46.71), Dst: 192.168.137.43 (192.168.137.43)
⊞ Transmission Control Protocol, Src Port: https (443), Dst Port: 52277 (52277), Seq: 1, Ack: 1, Len: 1440

0000  64 6c b2 b1 db 9d 0c d2  92 a6 b2 ca 08 00 45 00   dl...... ......E.
0010  05 d4 c3 90 40 00 39 06  97 84 68 f4 2e 47 c0 a8   ....@.9. ..h..G..
0020  89 2b 01 bb cc 35 20 2c  bf bc 16 59 c3 81 80 10   .+...5 , ...Y....
0030  00 74 fa 7f 00 00 01 01  08 0a ad c7 b6 b4 00 3b   .t...... .......;
0040  1d 4f 17 03 03 40 18 18  ea b5 4b f9 28 bd c7 20   .O...@.. ..K.(..

File: "D:\Weka\Data2\twitter.pcap" 1036 KB ...   Packets: 1620 Displayed: 1620 Marked: 0 Load time: 0:00.025   Profile: Default
```

Fig 10: Traffic capturing of the considered applications

CSV File Editing

In the next stage, the Wireshark file was saved in CSV (Comma separated values). Microsoft Excel was used to open and edit this file. Two types of file were prepared training and testing file. In the end of the preparation of training data, only three columns were saved which include packet length, time, and application named.

On the testing file preparation, the third column includes port number instead of application name. Figure 11 shows the testing file.



	A	B	C	D	E	F	G	H	I	J	K
1	0	103	111111								
2	0.151	66	22222								
3	0.153	215	80								
4	0.236	66	80								
5	0.24	135	4532								
6	0.242	66	22222								
7	1.471	71	80								
8	1.503	151	80								
9	1.513	74	80								
10	1.515	88	22222								
11	1.526	90	80								
12	1.531	80	80								
13	1.532	85	80								
14	1.532	77	234								
15	1.556	216	80								
16	1.567	180	80								
17	1.569	74	80								
18	1.571	74	80								
19	1.572	74	80								
20	1.573	85	234								
21	1.574	79	234								
22	1.574	74	234								
23	1.575	84	80								
24	1.606	87	80								
25	1.619	66	80								

Fig 11: CSV testing file editing

Weka Machine Learning

Weka open source was used as machine learning tool. The CSV file which prepared in the previous step was used to prepare Weka file. Notepad++ was used to add Weka header which is definition of the variables and data that used in Weka. Note that this file combines the datasets of three applications (http, twiter, whatsapp), each application includes three attributes (length, time and class).

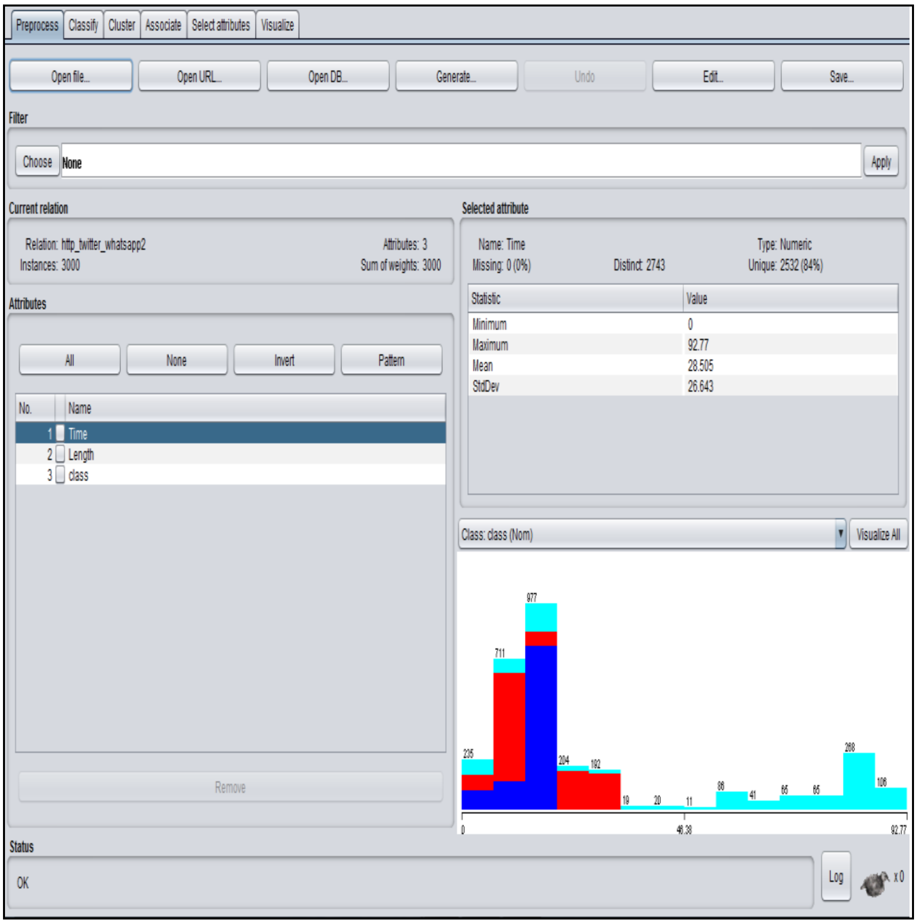
Arff file was opened using Weka software. In Weka the following operations was applied:

- Select the training option out of the other options
- Apply the classification using the ten selected algorithms
- Record the results of the classification of each algorithm
- Copy the rules of the algorithm with the best accuracy (Random Tree Algorithm) and saved into Word file.

MATLAB Hybrid Classifier

Based on step four in previous section, the rule of Random Tree was copied and saved in Microsoft Word file (figure 13). This rule was prepared to be used by MATLAB which are involved in if else statement. The code of hybrid software classifier was written in MATLB and the prepared rules were added to this code. Figure 14 shows the hybrid classifier code in MATLAB software.

The hybrid classifier is directly saved the classification results of testing file in Excel file. This is done by using of xlswrite function. Figure 5.8 shows the classification result in Excel file.



Filter: None

Relation: http_twitter_whatsapp2 Attributes: 3
Instances: 3000 Sum of weights: 3000

Name: Time Type: Numeric
Missing: 0 (0%) Distinct: 2743 Unique: 2532 (84%)

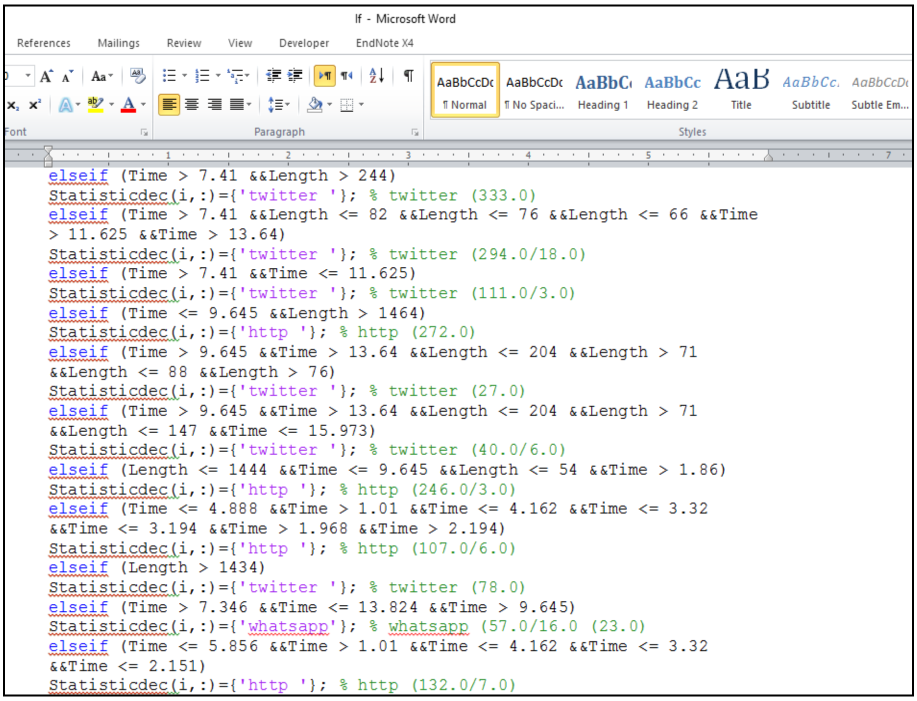
Statistic	Value
Minimum	0
Maximum	92.77
Mean	28.505
StdDev	26.643

No.	Name
1	Time
2	Length
3	class

Class: class (Nom)

Status: OK

Fig 12: Weka software environments



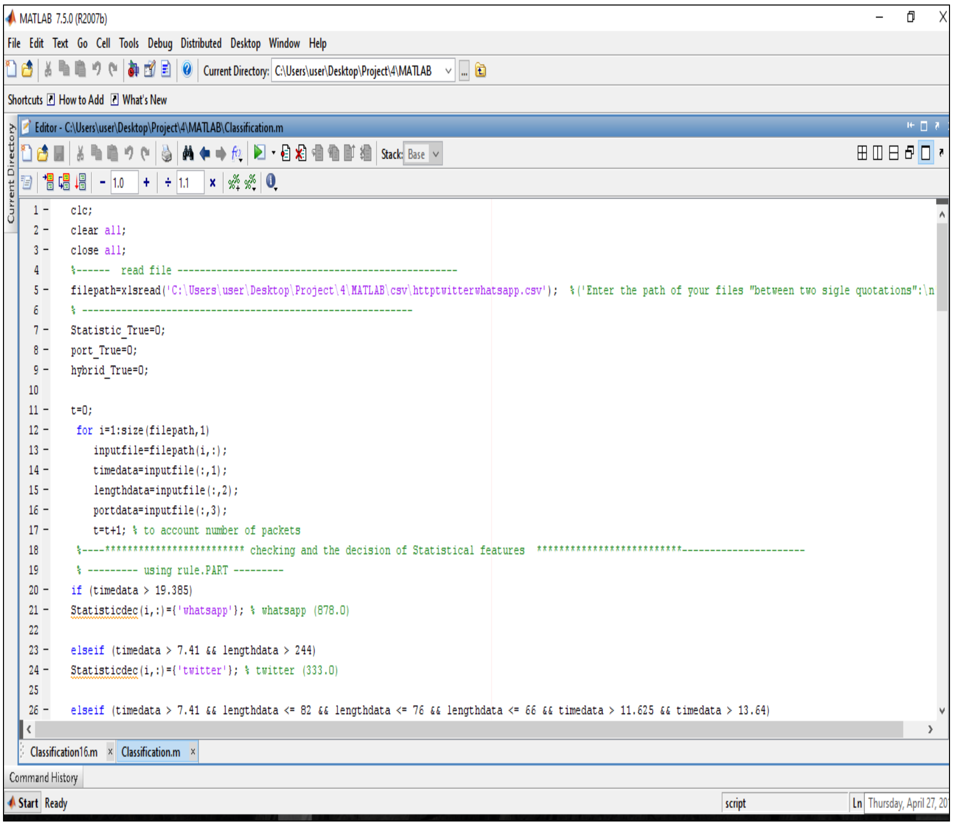
```
elseif (Time > 7.41 &&Length > 244)
Statisticdec(i,:)={'twitter '}; % twitter (333.0)
elseif (Time > 7.41 &&Length <= 82 &&Length <= 76 &&Length <= 66 &&Time
> 11.625 &&Time > 13.64)
Statisticdec(i,:)={'twitter '}; % twitter (294.0/18.0)
elseif (Time > 7.41 &&Time <= 11.625)
Statisticdec(i,:)={'twitter '}; % twitter (111.0/3.0)
elseif (Time <= 9.645 &&Length > 1464)
Statisticdec(i,:)={'http '}; % http (272.0)
elseif (Time > 9.645 &&Time > 13.64 &&Length <= 204 &&Length > 71
&&Length <= 88 &&Length > 76)
Statisticdec(i,:)={'twitter '}; % twitter (27.0)
elseif (Time > 9.645 &&Time > 13.64 &&Length <= 204 &&Length > 71
&&Length <= 147 &&Time <= 15.973)
Statisticdec(i,:)={'twitter '}; % twitter (40.0/6.0)
elseif (Length <= 1444 &&Time <= 9.645 &&Length <= 54 &&Time > 1.86)
Statisticdec(i,:)={'http '}; % http (246.0/3.0)
elseif (Time <= 4.888 &&Time > 1.01 &&Time <= 4.162 &&Time <= 3.32
&&Time <= 3.194 &&Time > 1.968 &&Time > 2.194)
Statisticdec(i,:)={'http '}; % http (107.0/6.0)
elseif (Length > 1434)
Statisticdec(i,:)={'twitter '}; % twitter (78.0)
elseif (Time > 7.346 &&Time <= 13.824 &&Time > 9.645)
Statisticdec(i,:)={'whatsapp'}; % whatsapp (57.0/16.0 (23.0)
elseif (Time <= 5.856 &&Time > 1.01 &&Time <= 4.162 &&Time <= 3.32
&&Time <= 2.151)
Statisticdec(i,:)={'http '}; % http (132.0/7.0)
```

Fig 13: MATLAB rules preparation

Implementation Results

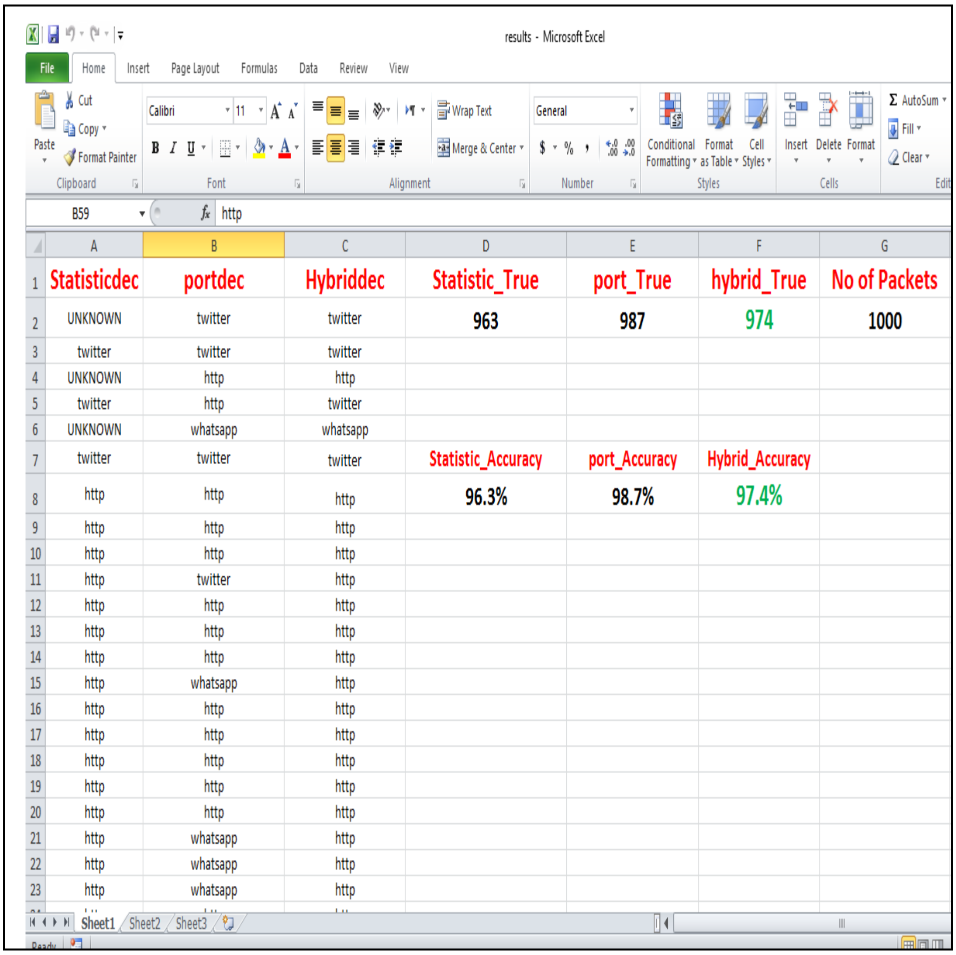
Table 2: Illustrates number of packets which used in training and testing stage for each of the considered application.

Application	Number of training packets	Number of testing packets
http		
WhatsApp		
Twitter		


MATLAB 7.5.0 (R2007b)
File Edit Text Go Cell Tools Debug Distributed Desktop Window Help
Current Directory: C:\Users\user\Desktop\Project\4\MATLAB
Shortcuts How to Add What's New
Editor - C:\Users\user\Desktop\Project\4\MATLAB\Classification.m
Stack: Base
Classification16.m
Classification.m
Command History
Start Ready
script
Ln Thursday, April 27, 20

```
clc;
clear all;
close all;
%------  read file ---------------------------------------------------
filepath=xlsread('C:\Users\user\Desktop\Project\4\MATLAB\csv\httptwitterwhatsapp.csv');  %('Enter the path of your files "between two sigle quotations":\n
% -----------------------------------------------------------

Statistic_True=0;
port_True=0;
hybrid_True=0;

t=0;
 for i=1:size(filepath,1)
    inputfile=filepath(i,:);
    timedata=inputfile(:,1);
    lengthdata=inputfile(:,2);
    portdata=inputfile(:,3);
    t=t+1; % to account number of packets
 %----************************ checking and the decision of Statistical features  *************************---------------------
 % --------- using rule.PART ---------
if (timedata > 19.385)
Statisticdec(i,:)={'whatsapp'}; % whatsapp (878.0)

elseif (timedata > 7.41 && lengthdata > 244)
Statisticdec(i,:)={'twitter'}; % twitter (333.0)

elseif (timedata > 7.41 && lengthdata <= 82 && lengthdata <= 76 && lengthdata <= 66 && timedata > 11.625 && timedata > 13.64)
```

Fig 14: Hybrid classifier code in MATLAB


results - Microsoft Excel
File Home Insert Page Layout Formulas Data Review View
Cut Copy Format Painter Paste
Calibri 11
Wrap Text Merge & Center
General
Conditional Formatting Format as Table Cell Styles
Insert Delete Format
AutoSum Fill Clear
Clipboard Font Alignment Number Styles Cells Edit
B59 fx http
Sheet1 Sheet2 Sheet3
Ready

	A	B	C	D	E	F	G
1	Statisticdec	portdec	Hybriddec	Statistic_True	port_True	hybrid_True	No of Packets
2	UNKNOWN	twitter	twitter	963	987	974	1000
3	twitter	twitter	twitter				
4	UNKNOWN	http	http				
5	twitter	http	twitter				
6	UNKNOWN	whatsapp	whatsapp				
7	twitter	twitter	twitter	Statistic_Accuracy	port_Accuracy	Hybrid_Accuracy	
8	http	http	http	96.3%	98.7%	97.4%	
9	http	http	http				
10	http	http	http				
11	http	twitter	http				
12	http	http	http				
13	http	http	http				
14	http	http	http				
15	http	whatsapp	http				
16	http	http	http				
17	http	http	http				
18	http	http	http				
19	http	http	http				
20	http	http	http				
21	http	whatsapp	http				
22	http	whatsapp	http				
23	http	whatsapp	http				

Fig 15: Classification result in Excel file.

As mentioned before the hybrid classifier makes his decision based on the decision of both statistical and port classifier. Table 3 show rules how the hybrid classifier makes his decision.

Table 3: Hybrid classifier decision

Statisticdec	Portdec	Hybriddec
xx	xx	xx
xx	yy	xx
xx	Unknown	xx
Unknown	yy	yy

Table 4: Example of classifier decision

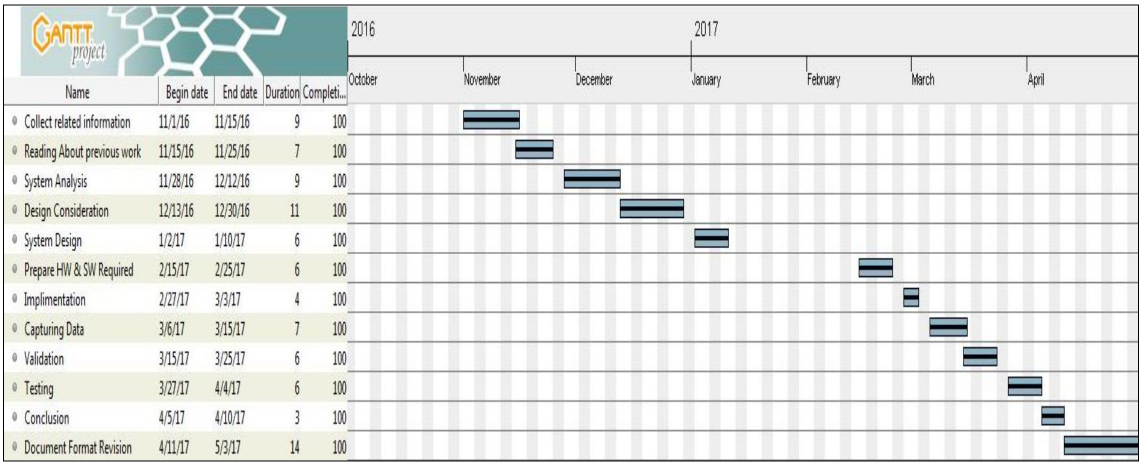
Statisticdec	Portdec	Hybriddec
http	http	http
http	WhatsApp	http
http	Unknown	http
Unknown	WhatsApp	WhatsApp

Table 5: Illustrate a comparison between ten Weka algorithms

Algorithms	Http		Whatsapp		Twitter		Accuracy
	TP	FP	TP	FP	TP	FP	
PART	96.4%	4.4%	90.5%	0.3%	92.6%	5.7%	93.1667%
SimpleCart	95.2%	4.4%	92.2%	0.9%	91.9%	5.1%	93.1%
REPTREE	96%	5.1%	92.1%	0.9%	91.1%	4.5%	93.0667%
Randomforst	95.3%	3.9%	94.3%	2.3%	89.4%	4.3%	93%
BFTREE	96.2%	4.9%	92.4%	1.3%	90.1%	4.5%	92.9%
J48GRFT	96.1%	5.1%	91%	0.6%	91%	5.3%	92.7%
J48	95.9%	5.2%	91%	0.7%	90.9%	5.3%	92.6%
JRIP	96.7%	6%	91.4%	1%	89.8%	4.2%	92.6333%
Randomtree	93.7%	3.7%	94.4%	3%	89.4%	4.7%	92.5%
Ridor	95.5%	4.9%	92.1%	1.8%	89.7%	4.7%	92.4333%

The comparisons done between two of Weka categories tree and Rules. In addition, the True Positives (TP) and False Positive (FP) of each application were recorded.

Gantt chart



Name	Begin date	End date	Duration	Completi...
Collect related information	11/1/16	11/15/16	9	100
Reading About previous work	11/15/16	11/25/16	7	100
System Analysis	11/28/16	12/12/16	9	100
Design Consideration	12/13/16	12/30/16	11	100
System Design	1/2/17	1/10/17	6	100
Prepare HW & SW Required	2/15/17	2/25/17	6	100
Implimentation	2/27/17	3/3/17	4	100
Capturing Data	3/6/17	3/15/17	7	100
Validation	3/15/17	3/25/17	6	100
Testing	3/27/17	4/4/17	6	100
Conclusion	4/5/17	4/10/17	3	100
Document Format Revision	4/11/17	5/3/17	14	100

Fig 16: Gantt chart.

References

1. Internet traffic From Wikipedia, the free encyclopedia.
2. Internet traffic classification. National Science Foundation; c2013. Retrieved 18 October 2014.
3. IETF RFC 2475 An Architecture for Differentiated Services" section 2.3.1 - IETF definition of classifier.
4. Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012, Proceedings.
5. Internet traffic classification An Enhancement in performance using classifiers combination, Indra Bahn Arya *et al*,/(IJCSIT) International Journal of Computer Science and Information Technologies. 2011;2(2):663-66.
6. Internet Application Traffic Classification Using Fixed IP-Port, Sung-Ho Yoon, Jinn-Wan Park, Jun-Sang Park, Young-Seek Oh, Myung-Sup Kim, Dept. Of Computer and Information Science, Korea University, Korea.
7. Network Working Group, J. Reynolds, J. Postal, ISI, October; c1994.
8. Internet traffic From Wikipedia, the free encyclopedia.
9. Patrick Schneider Division of Applied Sciences, Harvard University, 29 Oxford Street, Cambridge, MA 02138, USA Patrick.Schneider@Switzerland.ORG
10. Deep Content Inspection vs. Deep Packet Inspection, Wedge Networks Inc; c2011. August 2, accessed August 23, 2011.
11. https://en.wikipedia.org/wiki/Traffic_classification
12. Carela-Espaoll V, Barlet-Ros P, Sole-Simo M, Dainotti A, de Donato W, Pescape
13. Kotsiantis SB. Supervised machine learning: A review of classification techniques. In Proceeding of the 2007 conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies, Amsterdam, The Netherlands, The Netherlands, IOS Press; c2007. p. 3-24.
14. https://en.wikipedia.org/wiki/WhatsApp
15. Twitter via SMS FAQ" Retrieved April 13, 2012.4"
16. https://www.w3.org/Protocols/HTTP/AsImplemented.html

17. https://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol
18. whatis.techtarget.com ›. › IT standards and organizations
19. www.cisco.com/c/en/us/products/./models-comparison.html
20. Blake S, Black D, Carlson M, Davies E, Wang Z, Weiss W. An Architecture for Differentiated Services, RFC 2475; c1998.
21. https://en.wikipedia.org/wiki/Router_(computing).
22. https://en.wikipedia.org/wiki/Smart_device
23. www.boosla.com
24. Q & A with the founder of Wireshark and Ethereal. Interview with Gerald combs. protocol Testing.com 24-7-2010.
25. www.wireshark.org
26. Syed SS, Shanthi S, Chitra VM. Application of Data Mining techniques to model breast cancer data. International Journal of Emerging Technology and Advanced Engineering. 2013;3(11):362-9.
27. Bellaachia A, Guven E. Predicting breast cancer survivability using data mining techniques. Society for Industrial and Applied Mathematics. 2006;58(13):1–4.
28. Kaur G, Chhabra A. Improved J48 classification algorithm for the prediction of diabetes. International Journal of Computer Applications. 2014;98(22):13-7.
29. Ian H Witten, Frank E, Mark A Hall. Data MiningPractical Machine Learning Tools and Techniques, Third Edition. Morgan Kaufmann Publishers is an imprint of Elsevier.[4]
30. Dr. B Srinivasan, Mekala P. Mining Social Networking Data for Classification Using REP Tree, International Journal of Advance Research in Computer Science and Management Studies. 2014, 2(10).
31. The Random Subspace Method for Constructing Decision Forests (PDF). IEEE Transactions on Pattern Analysis and Machine Intelligence. 20(8):832-844. doi:10.1109/34.709601.
32. Trevor H, Robert T, Jerome F. The Elements of Statistical Learning (2nd ed.). Springer. ISBN 0-387-95284-5; c2008.
33. http://fiji.sc/javadoc/weka/classifiers/trees/J48graft.htm.
34. PG and Research Department of Computer Science, D.G. Vaishnav College, Chennai-600106, Tamil Nadu, India; venkatelumalai12@yahoo.co.in, velmurugan_dgvc@yahoo.co.in
35. Asstt. Prof., Department of Mathematics and Computer Science, Govt. P.G. College Bareli (M.P.), 464668, India
36. M. Thangaraj, Ph.D. Associate Professor Madurai Kamaraj University, Madurai Tamil Nadu, India
37. Muppala, Suresh. Coordinated environment for classification and control of network traffic. U.S. Patent No. 7,742,406. 22 Jun. 2010.
38. Floyd, Sally, and Kevin Fall. Promoting the use of end-to-end congestion control in the Internet. IEEE/ACM Transactions on networking 7.4 (1999): 458-472.
39. Matti H, Jukka-Pekka Laulajainen. Two-phased network traffic classification method for quality of service management. Consumer Electronics, 2009. ISCE'09. IEEE 13th International Symposium on. IEEE; c2009.
40. Pierangela S, Capitani de Vimercati S. Access control: Policies, models, and mechanisms. International School on Foundations of Security Analysis and Design. Springer Berlin Heidelberg; c2000.