MIDDLE EAST TECHNICAL UNIVERSITY

COMPUTER ENGINEERING DEPARTMENT

CENG491

Requirement Analysis Report

By

EsSoft

INDEX

1. INTRODUCTION

Network is the critical part of the computer systems. Especially, after the widely usage of internet, network becomes more important compared with the past. Every person, whether interested in computer systems or not, begins using network systems widely in daily life. Therefore, network programs, such as online chat, e-mail, are made more effective to attract users. These programs are designed by using different techniques, different algorithms, different protocols ,if we look at the concept of network systems. However, these differences cause some problems. Since there are lots of differencies, identification and classification phase for computer is a big problem.

1.1 PROBLEM DEFINITION

Our Project will be used for analysing, identifying, and classifying incoming network data in terms of their protocols.

1.2 PROJECT SCOPE

Firstly, since our project basically concerns about protocols, we want to give brief explanation of which protocols will be identified in our project:

Simple Mail Transfer Protocol (SMTP) is a relatively simple, text-based protocol, in which one or more recipients of a message are specified (and in most cases verified to exist) and then the message text is transferred.

- Hypertext Transfer Protocol (HTTP) is a communications protocol used to transfer or convey information on intranets and the World Wide Web.

- File Transfer Protocol (FTP) is used to transfer data from one computer to another over the Internet, or through a network.

- Post Office Protocol version 3 (POP3), an application-layer Internet standard protocol, to retrieve e-mail from a remote server over a TCP/IP connection.

- The Session Initiation Protocol (SIP) is an application-layer control (signaling) protocol for creating, modifying, and terminating sessions with one or more participants. It can be used to create two-party, multiparty, or multicast sessions that include Internet telephone calls, multimedia distribution, and multimedia conferences.

If our project's definition is extended, basic details are:

- Our program should identify as much as possible protocols, in order to increase effectiveness.

- Our program should consist of interface that's usage is easy for the people using first time our program.

- Our program should satisfy the feature that is version increasing.

- Our program should be effective in terms of time.

Our project consists of three phases. First phase is taking PCAP files without classifying. After that, in classification phase that is second and most significant phase, according to protocol structs, program classifies the data, additionally in these phase in order not to lose the main data, it uses database system and data that belongs to the protocol syntax is inserted in database which will be used in last phase. Third phase of our project is giving the output of the data. After the second phase system should determine the right protocol of data, and in third phase this classification is used in order to call the right programs.

2. TEAM ORGANIZATION

2.1 TEAM STRUCTURE

We decided our team structure to be Controlled Decentralized (CD). In Controlled Decentralized team structure, the team has a team leader who coordinates the team. Moreover, the team leader assigns tasks to group members in the team and each person is responsible for some subtasks. The team takes decisions together and communication between the team members is very important. Therefore we thought that a Controlled Decentralized team structure is the most suitable one for our team.

2.2 MEMBER ROLES

In order to provide the equivalence in the group in terms of mission, the roles are distributed according to the table below:

Müjdat BAYAR	Team leader, coding process
Akif Gencay DEMIREL	Spokeperson, testing phase
Kerem OZARKAN	Time keeper, testing phase
Arda GÜÇLÜ	Recorder, coding process

2.3 TIMELINE

Our project's Gannt chart can be seen in the Appendix part of our report.

3.PROCESS MODEL

In our project we will develop our work by analyzing, designing, making implementation, and testing. So it is obvious that we will develop our project as a step by step manner. We know that the project phases are distinct. The requirements are well defined and understood. But there is a possibility that in some phases we may do mistakes and sometimes we may need to return to a specific phase and make changes. So during the project a feedback mechanism is needed.

“Waterfall model with feedback” suits our goal best, because the steps are distinct as we needed. In addition to this between each phase feedback loops exist so it allows us to make modifications easily.

4.MARKET OBSERVATION

4.1 Literature Survey

Market research is one of the most important parts of the requirement analysis in the sense that it provides us to have general information on the similar systems and helps us to determine the requirements in appropriate way. With the help of market research, we had the chance of examining the similar projects to reconsider the features of our project and we had specified the features that users expect from such a system.

In order to deepen our research, we divided our subject into two subtopics namely defining protocols (HTTP, SMTP, POP3, IMAP, FTP) and examining sniffers(wireshark). From the market search, we have seen that there is not an exact example of our product; but only there are some applications that has some similar parts of our project. For this purpose, we examined wireshark.

4.1.1 Protocols

- HTTP:

HTTP is the abbreviation of Hypertext Transfer Protocol. HTTP is the network protocol of the Web. It is used to transmit resources. HTTP uses the client-server model: An HTTP client opens a connection and sends a request message to an HTTP server; the server then returns a response message, usually containing the resource that was requested.

-SMTP:

SMTP is the short form of Simple Mail Transfer Protocol, a protocol for sending e-mail

messages between servers. Most e-mail systems that send mail over the Internet use SMTP to send messages from one server to another; the messages can then be retrieved with an e-mail client using either POP or IMAP. In addition, SMTP is generally used to send messages from a mail client to a mail server. This is why you need to specify both the POP or IMAP server and the SMTP server when you configure your e-mail application.

-POP3 & IMAP:

POP3 (Post Office Protocol 3) is the most recent version of a standard protocol for receiving e-mail. POP3 is a client/server protocol in which e-mail is received and held for you by your Internet server. Periodically, you (or your client e-mail receiver) check your mail-box on the server and download any mail, probably using POP3. This standard protocol is built into most popular e-mail products, such as Eudora and Outlook Express. It's also built into the Netscape and Microsoft Internet Explorer browsers.

POP3 is designed to delete mail on the server as soon as the user has downloaded it. However, some implementations allow users or an administrator to specify that mail be saved for some period of time. POP can be thought of as a "store-and-forward" service.

An alternative protocol is Internet Message Access Protocol (IMAP). IMAP provides the user more capabilities for retaining e-mail on the server and for organizing it in folders on the server. IMAP can be thought of as a remote file server.

POP and IMAP deal with the receiving of e-mail and are not to be confused with the Simple Mail Transfer Protocol (SMTP), a protocol for transferring e-mail across the Internet. You send e-mail with SMTP and a mail handler receives it on your recipient's behalf. Then the mail is read using POP or IMAP.

The conventional port number for POP3 is 110.

-FTP :

File Transfer Protocol (FTP), a standard Internet protocol, is the simplest way to exchange files between computers on the Internet. Like the Hypertext Transfer Protocol (HTTP), which transfers displayable Web pages and related files, and the Simple Mail Transfer Protocol (SMTP), which transfers e-mail, FTP is an application protocol that uses the Internet's TCP/IP protocols. FTP is commonly used to transfer Web page files from their creator to the computer that acts as their server for everyone on the Internet. It's also commonly used to download programs and other files to your computer from other servers.

As a user, you can use FTP with a simple command line interface (for example, from the Windows MS-DOS Prompt window) or with a commercial program that offers a graphical user interface. Your Web browser can also make FTP requests to download programs you select from a Web page. Using FTP, you can also update (delete, rename, move, and copy) files at a server. You need to logon to an FTP server. However, publicly available files are easily accessed using anonymous FTP.

Basic FTP support is usually provided as part of a suite of programs that come with TCP/IP. However, any FTP client program with a graphical user interface usually must be downloaded from the company that makes it.

-SIP:

Session Initiation Protocol - is a network communications protocol commonly employed for Voice over IP (VoIP) signaling. In VoIP networking, SIP is an alternative approach to signaling using the H.323 protocol standards.

SIP is designed to support the calling features of traditional telephone systems. However, unlike the traditional SS7 technology for telephone signaling, SIP is a peer-to-peer protocol. SIP is also a general-purpose protocol for multimedia communications not limited to voice applications.

4.1.2 Sniffers

-**Wireshark**

In computing, Wireshark (formerly known as Ethereal) is a free software protocol analyzer, or "packet sniffer" application, used for network troubleshooting, analysis, software and protocol development, and education. It has all of the standard features of a protocol analyzer.

The functionality Wireshark provides is very similar to tcpdump, but it has a GUI front-end, and many more information sorting and filtering options. It allows the user to see all traffic being passed over the network (usually an Ethernet network but support is being added for others) by putting the network card into promiscuous mode.

Its features:

- Data can be captured "from the wire" from a live network connection or read from a capture file.
- Live data can be read from Ethernet, FDDI, PPP, token ring, IEEE 802.11, Classical IP over ATM, and loopback interfaces (at least on some platforms; not all of those types are supported on all platforms).

- Captured network data can be browsed via a GUI, or via the terminal (command line) version of the utility, tshark.
- Captured files can be programmatically edited or converted via command-line switches to the "editcap" program.
- Display filters can also be used to selectively highlight and color packet summary information.
- Data display can be refined using a display filter.
- Hundreds of protocols can be dissected.

-Tcpdump

Tcpdump is a command-line tool for monitoring network traffic. Tcpdump can capture and display the packet headers on a particular network interface or on all interfaces. Tcpdump can display all of the packet headers, or just the ones that match particular criteria.

Tcpdump works on most Unix-like operating systems: Linux, Solaris, BSD, Mac OS X, HP-UX and AIX among others. In those systems, tcpdump uses the libpcap library to capture packets. There is also a port of tcpdump for Windows called WinDump; this uses WinPcap, which is a port of libpcap to Windows.

Common uses of tcpdump:

Tcpdump is frequently used to debug applications that generate or receive network traffic. It can also be used for debugging the network setup itself, by determining whether all necessary routing is occurring properly, allowing the user to further isolate the source of a problem.

It is also possible to use tcpdump for the specific purpose of intercepting and displaying the communications of another user or computer. A user with the necessary privileges on a system acting as a router or gateway through which unencrypted traffic such as

TELNET or HTTP passes can use tcpdump to view login IDs, passwords, the URLs and content of websites being viewed, or any other unencrypted information.

We will use these sniffers to understand the network flow, PCAP file format and how to start our prototype progress. Since there isn't an exact example of our project we will use these sniffers while implementing CLASSIM.

Diffrences of our project from these sniffers are:

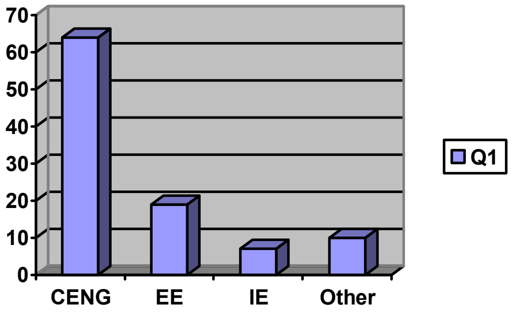
- There will be Autosensing in CLASSIM which will classify protocols by examining the PCAP file format. Furthermore, it will not look only its port number. If there is a PCAP file which comes form a non-standard port, it will check its headers and keywords to understand which protocol it belongs. In wireshark, there is also a network admin that tries to do this job but its results aren't satisfaible.

- In CLASSIM, we can go into network flow at any time i.e. while someone reading his e-mail. CLASSIM will also detect which protocol is used with this process.

4.2 Questionnaire

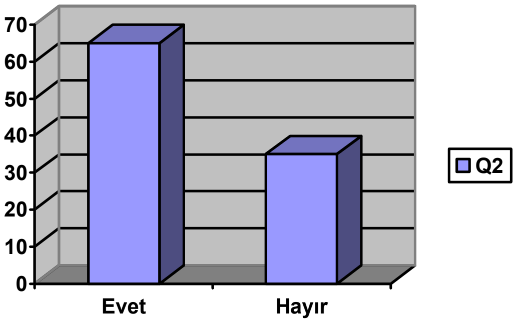
Questionnaire is one of the most significant analyses for specifying user needs on a system that will be developed. Since questionnaire involves ideas of large number of people, analyzing these data will provide us to consider our previous decisions again, according to user needs.

1. Hangi bölüm öğrencisisiniz / mezunusunuz?



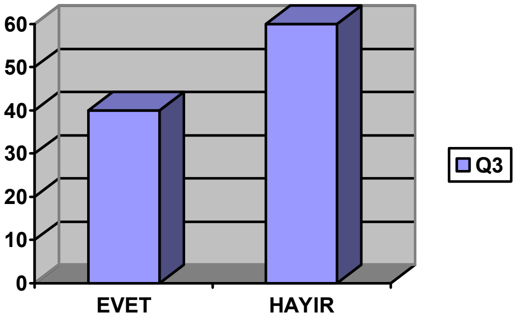
We have asked this question for specifying possible users of sniffers and try to find people which is interested in network. Since CENG and EE students are most familiar with network applications, it is reasonable to conduct our questionnaire on them.

2. Herhangi bir sniffer kullandınızmı veya kullanmak istemisiniz?



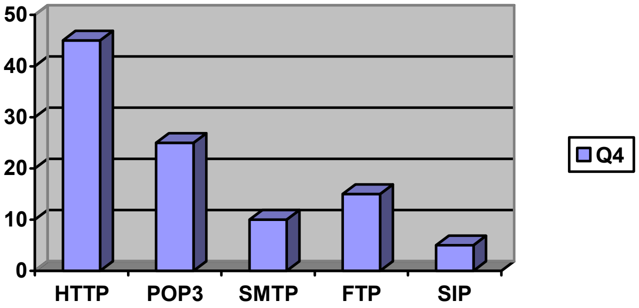
We tried to measure the interest of people in sniffers to ask more complicated questions about our project. We continued our questionnaire with people asnwered 'yes'.

3. Kulladığınız snifferları yeterli görüyormusunuz?



Since the people, we asked questions, are has knowledge about classifying protocols , most of them doesn't find wireshark and tcpdump satisfaible with capturing files and detecting its protocols.

4. Hangi protocollerin ayırt edilebilmesini isterdiniz ?



Most of the users chose protocols that we will classify in autosensing process.Therefore our CLASSIM project will also be usable for people who use sniffers in daily life.

5.PROJECT REQUIREMENTS

The most important part of the project works is determining the requirements. The requirements analysis must be done carefully to get project maintenance. Therefore, during the development of the project, it must be guarantied that no unexpected events and requirements should occur. In order to avoid failure, requirements must cover all our needs. For our project, requirements are divided in three main issues.

5.1 SYSTEM REQUIREMENTS

5.1.1-Hardware Requirements

Minimal hardware requirements for our project are: A PC with the following configuration will be needed:

>> For Developer
A minimum of 512 MB DDRAM
A minimum of 5 GB free space on hard disk, for database storage and server applications
A Pentium IV processor
Internet Connection
Local or Wide Area Network

>>For User
A minimum of 1 GB DDRAM
A minimum of 50 GB free space on hard disk, for huge database storage
A Pentium IV processor
Internet Connection

5.1.2-Software Requirements

We will use several tools for different phases during development of the project. These
development phases are;

>> Documentation Part
- Analysis
- Design

>> Develeopment Part
- Implementation
- Testing

Documentation Tools
Microsoft Office 2003 Word
Adobe Acrobat Professional
SmartDraw

Development Tools
Windows XP
C++
WireShark(WireShark is a sniffer program as it mentioned before.Usage of wireshark's purpose is testing and also to create PCAP files)

5.2 Functional Requirements

- In our Project, required protocols that should be identified are:
  - HTTP
  - SIP
  - POP3
  - FTP
- The data which is captured from network, which will be identified, is to be fed in the PCAP file format.
- Our identification process will also determine when the identified protocol is no longer available in the flow through the identified port.
- When we identificate protocols our outputs will be like these:
  - HTTP : If connected HHTP server is Yahoo! Mail and user downloads a mail, then RFC 822 complied e-mail messages(.eml).
  - POP3 : Download mail messages.
  - FTP : Transferred files.
  - SIP : Voice files in Microsoft ASF format.

- The project must work in real-time. It must identify incoming data in real-time and must give the required output.

5.3 Nonfunctional Requirements

Usability:

As for all softwares, graphical user interface is very important for our project too since easy to use, easy to learn and adaptation is crucial for all softwares. Therefore, we will make the user interface clear and understandable. Moreover, graphical user interface will meet the user's needs as far as it is possible. Hence, while we were deciding our tool's features, we took into consideration both the inexperienced and professionals. As a conclusion, we considered the user satisfaction as the primary goal of our project.

Reliability:

The system should be as bug free as possible. All sub components should work asynchronously, so that any delay caused by one of the components should not block other components work on its own.Moreover, we plan to do many tests after implementation to minimize the bugs on the program.

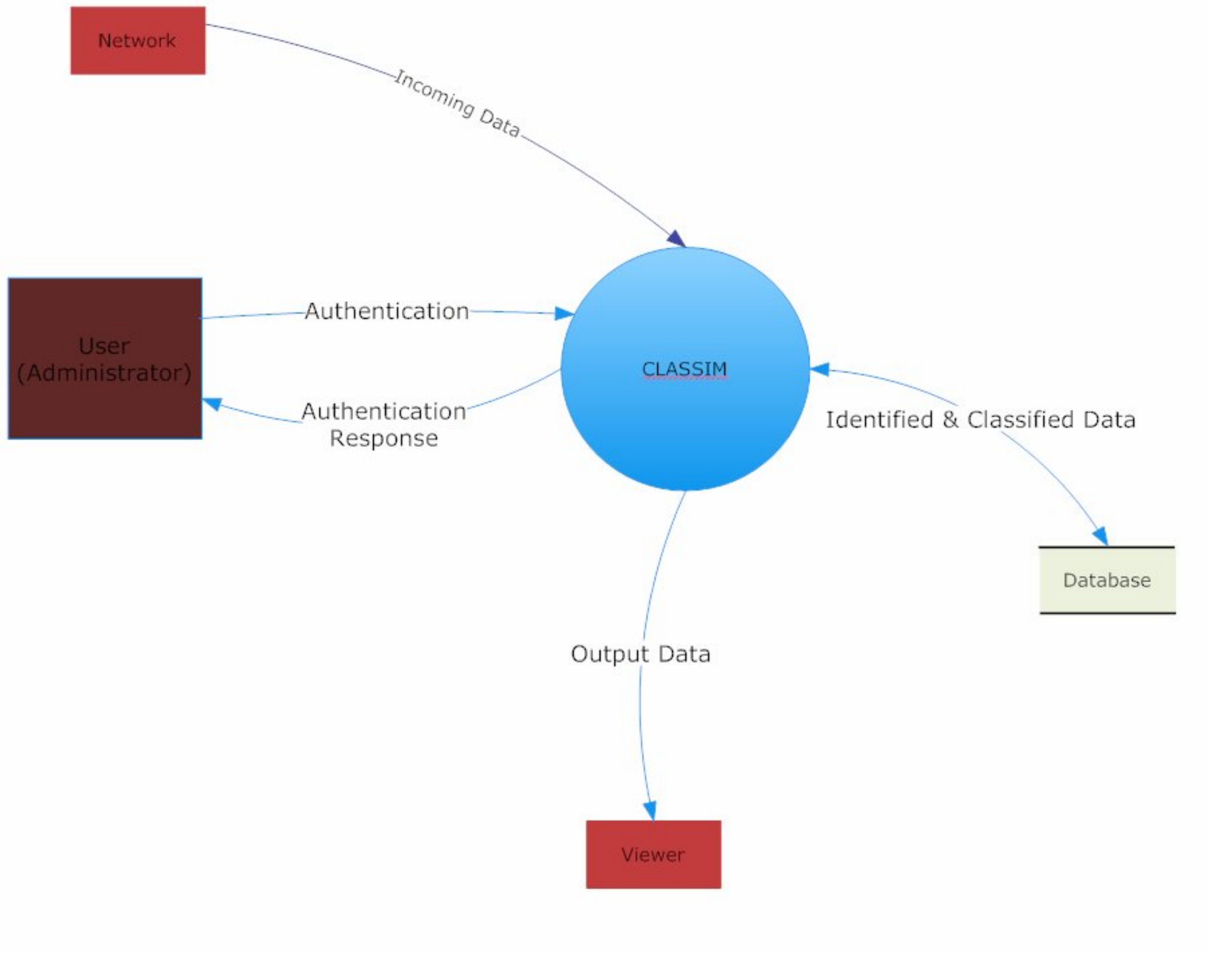
Performance:

Speed of our product becomes an important issue in our design. In general, wrong programming methodologies that are used in the applications slow down the programs, not the complex algorithms used in the applications. Therefore, we will try not to do this mistake. The usage of system resources will be reduced as much as possible to increase the performance of our design. User can run other applications easily while our program working.

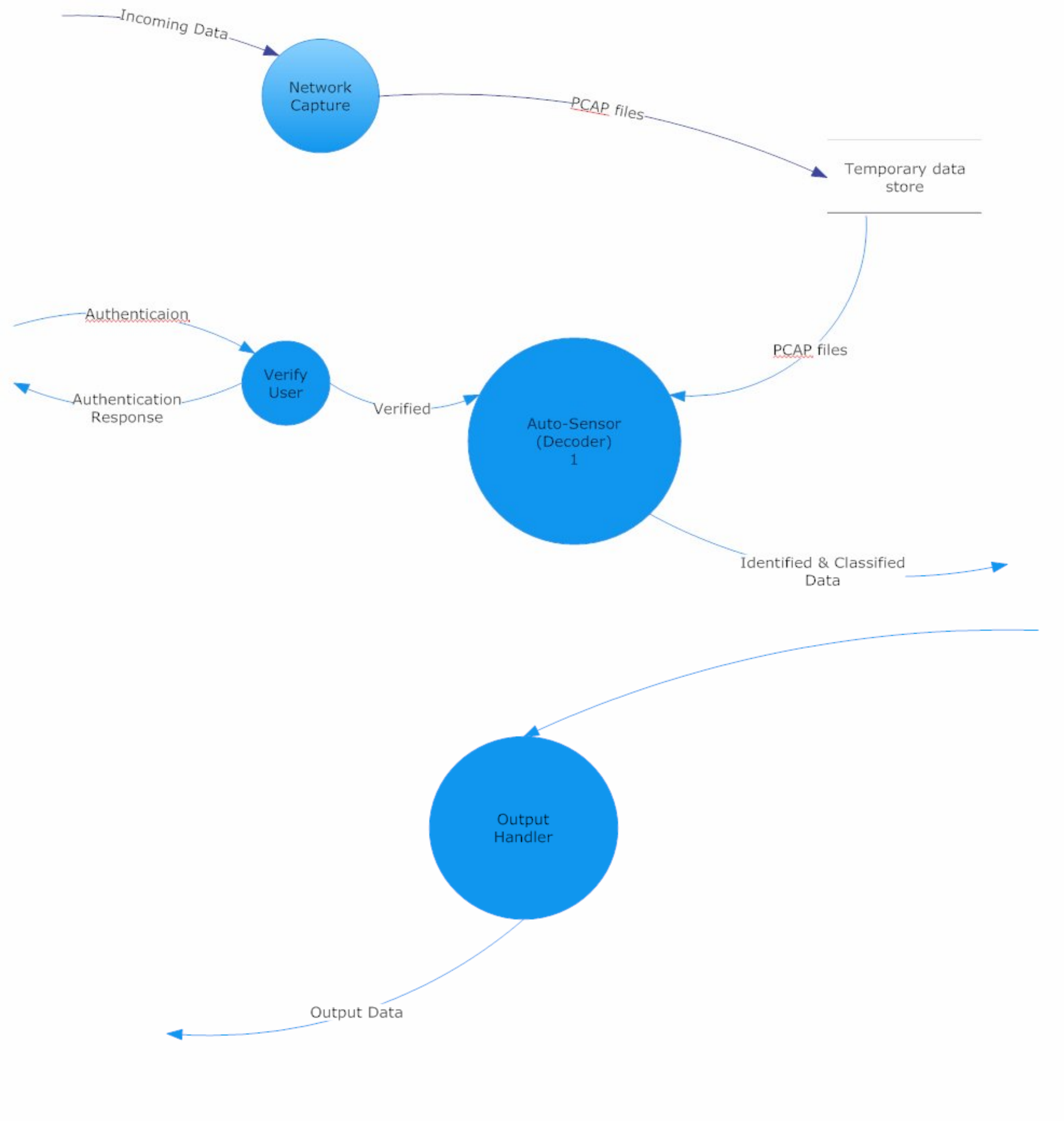
6. MODELING

6.1 – DATA FLOW DIAGRAM (DFD)

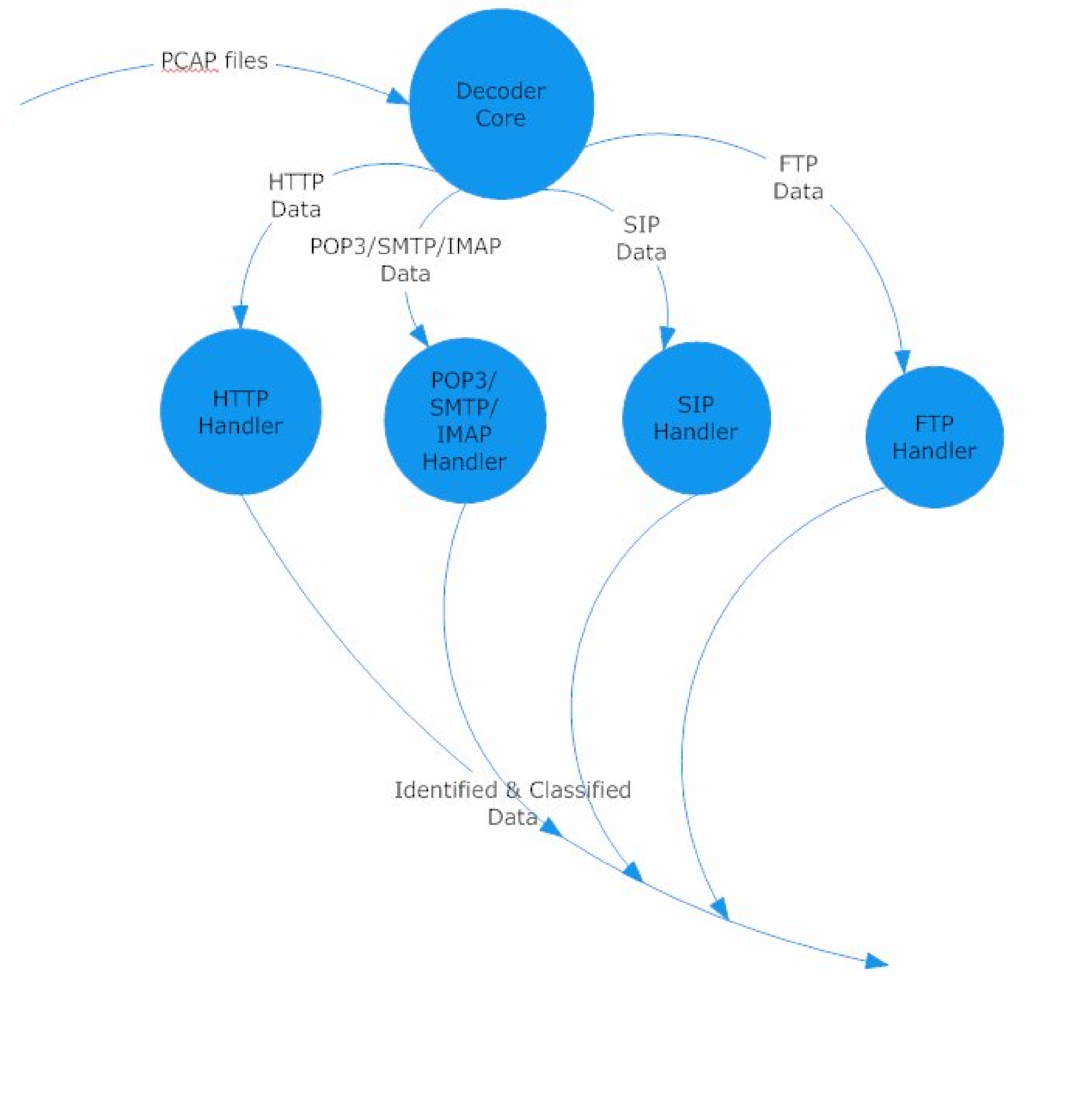
6.1.1 - DFD – Level 0



6.1.2 - DFD – Level 1 – *CLASSIM*



6.1.3 - DFD – Level 2 – *Auto-Sensor*



6.2 - DFD Data Dictionary

Name : PCAP files

Input for: Auto-Sensor , Temporary data store

Output for : Network Capture

Description : The file which has PCAP format that identifying will be done according.

Name : Identified & Classified Data

Input for: Database, output handler

Output for : Auto-Sensor , database

Description : Includes idenified protocol type, source port,destination port, time info, transmitted data etc.

Name : Output Data

Input for: viewer

Output for : Output Handler

Description : data to be outputted in the format determined according to the identified protocol e.g. asf format for SIP protocol.

6.3 – DFD Process Description

Name : Auto-Sensor

Input : PCAP files

Output : Identified & Classified Data

Description : This process takes PCAP files and decodes the information which is in PCAP format. According the decoding, incoming information sent to the protocol handlers. After protocol handlers, data becomes ready for storing and sending to output.

Name : Output Handler

Input : Identified & Classified Data

Output : Output Data

Description : This process gets the identified data. According to the protocol of data, process uses the required output format. Looking the protocol type, outputs the transmitted data in wanted format. For example in pop3 protocol outputs the mail message in 'eml' format. After all it sends the output data to viewer.

7. CONCLUSION

Thanks to development of computer technology, network systems become complex and sometimes make problems for users that have to be dealed. For instance, increase of usage of internet in the world, control of network systems is a critical issue. Hardwares and softwares to connect one computer to the other have to do identification and classification of programs in terms of protocols.

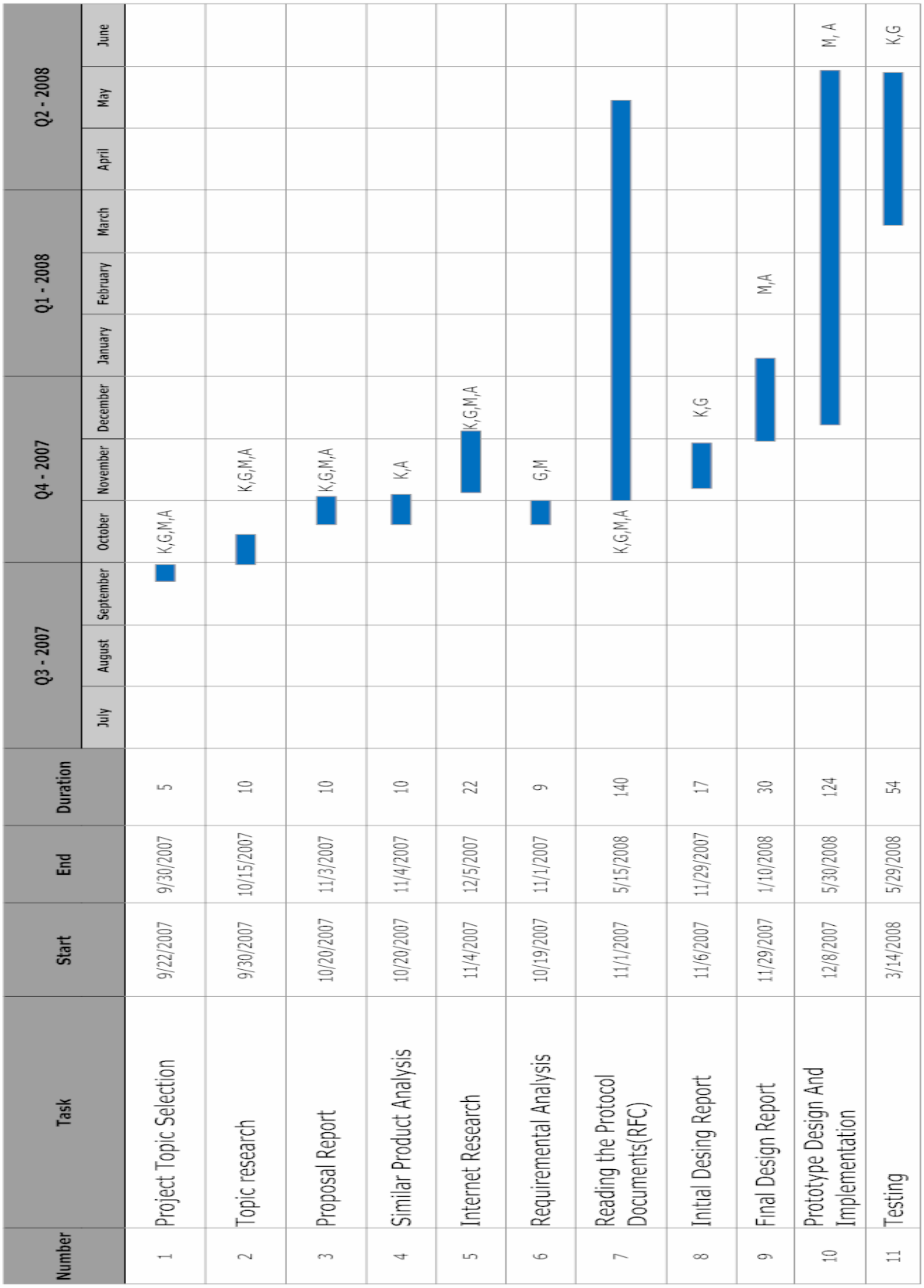
Through the whole process, we appreciate needs about network systems. We put different features to our project in order to make it more efficient in terms of time and easibility.

8.REFERENCES

- http://www.wireshark.org
- http://www.tcpdump.org
- http://en.wikipedia.org
- http://searchvb.techtarget.com/Definitions
- http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/ip.htm
- http://w3school.org

APPENDIX A

GANNT CHART:



Number	Task	Start	End	Duration
1	Project Topic Selection	9/22/2007	9/30/2007	5
2	Topic research	9/30/2007	10/15/2007	10
3	Proposal Report	10/20/2007	11/3/2007	10
4	Similar Product Analysis	10/20/2007	11/4/2007	10
5	Internet Research	11/4/2007	12/5/2007	22
6	Requiremental Analysis	10/19/2007	11/1/2007	9
7	Reading the Protocol Documents(RFC)	11/1/2007	5/15/2008	140
8	Intial Desing Report	11/6/2007	11/29/2007	17
9	Final Design Report	11/29/2007	1/10/2008	30
10	Prototype Design And Implementation	12/8/2007	5/30/2008	124
11	Testing	3/14/2008	5/29/2008	54