ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

Peak Hour Mobile Core Network Data Traffic Analysis to Improve Network Quality Using Flow Based Method: The Case of Ethio-Telecom

By: Mahlet Merid

Advisor: Dr. Yihenew Wondie

A Thesis Submitted to the School of Electrical and Computer Engineering of Addis Ababa Institute of Technology, School of Graduate Studies, in Partial Fulfillment of the Requirement for the Degree of Masters of Science in Communication Engineering

October, 2021

Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

Peak Hour Mobile Core Network Data Traffic Analysis to Improve Network Quality Using Flow Based Method: The Case of Ethio-Telecom

By Mahlet Merid

Approved By Board of Examiners

_____	_____
Dean,	**Signature**
School of Electrical and Computer Engineering	
Dr. Yihenew Wondie	_____
Advisor	**Signature**
_____	_____
External Examiner	**Signature**
_____	_____
Internal Examiner	**Signature**

Declaration

I, the undersigned, declare that this thesis work is my original work, has not been presented for a degree in this or any other universities, and all sources of materials used for the thesis work have been fully acknowledged.

Name: Mahlet Merid Signature: _____

Place: Addis Ababa, Ethiopia

Date of Submission: _____

This thesis is submitted for examination with my approval as university advisor subject to the candidate incorporating the comments given by the advisor.

Advisor: Dr. Yihenew Wondie Signature: _____

Acknowledgment

First of all, I would like to thank the almighty God for enabling me to reach the end of this work. I would like to thank my advisor Dr. Yihenew Wondie for his help in my work. I would also like to extend my deep gratitude to the lecturers and staff at the school of electrical and computer engineering department who gave me a valuable assistance and help with this work.

I would like to deeply thank all my families; my father, mother and sisters for giving me their unconditional support in any way possible through all the ups and downs. In addition I would also like to extend my deep gratitude to Eyob Getachew who has been a great help with everything. I would like to sincerely thank Ethio Telecom employees too who provided me with the necessary input data that is needed for this work and were very helpful as much as possible.

Thank you to everyone who in one way or another assisted and contributed to the completion of this thesis.

Abstract

It is known that the telecom industry is one of the core areas in a country's sustainability and growth. So it is important that great emphasis be given to it on deploying necessary infrastructures in different areas, maintaining the existing available resources and also upgrading the already existing networks as necessary. Once the basic layout is done, it is also equally important that the necessary follow up is done for giving solutions to problems that arise from customers from time to time. One of the biggest reason that lead to customer complaints arise from poor quality of service which results in dissatisfaction of customers' needs. In order to give a solution to this, one of the ways is to do a network traffic analysis.

In this thesis, a data traffic analysis is done in the Ethio Telecom core network. Data captured from its network is used as an input in order to firstly identify the peak hour during the day because this is the time where there is the most communication and transmission. The peak hours of each day are recoded and then finally the average is taken for the purpose of this study. In general over the sampled data the peak hour is found to be at 21:06hr. For this work identification of the peak hour is necessary because this thesis focuses the traffic analysis during the peak hour and for the work to be thorough and to be confirmed, first identification of the busiest hour of the day is necessary. After that by filtering out the data at the peak hour, the Key Performance Indicators, Packet Loss Ratio in percentage (%) and throughput (packet/sec) are studied from the capture data in order to be able to see how exactly the system is working. In order to do so, two approaches are used. First the cumulative distribution functions of the data are fitted against the different traffic analysis distribution models. Out of the selected distribution models, it is seen that our data best fits with the Normal Distribution and the Gamma Distributions. For better accuracy the RMSE (Root Mean Square Error) is calculated for each one of them. Second, the KPI's for the peak hour and the slow hour are compared. From the sample gathered data, for both PLR and throughputs, the number of packets being lost are higher during the peak hour compared to that of the slow hour by 37%. But despite this, when comparing the Packet Loss Ratio recorded for both peak hour and slow hour they are both less than 1% which is the acceptable threshold range.

Similarly the number of packets being received per second that are sent for the downlink and uplink throughputs, during peak hour the minimum downlink and uplink throughputs exceed that

of the slow hour by 15.4% and 11.9% respectively and for the uplink throughput by 16% and 12.5% respectively. So finally from the analysis result, it is seen that the network works fine with a very minor glitch which is expected from a real life operating network.

Keywords: Traffic analysis, peak hour, KPI's, PLR, throughput, RMSE

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BSS	Base Station Subsystem
BTS	Base Transceiver Station
BSC	Base Station Controller
CN	Core Network
GSM	Global System for Mobile communications
MS	Mobile Station
MSC	Mobile Switching Center
NSS	Network Switching Subsystem
NMS	Network Management Subsystem
PSTN	Public Switched Telephone Network
SNMP	Simple Network Monitoring Protocol
RAN	Radio Access Network
RMON	Remote Monitoring
WREN	Watching Resource from the Edge of the Network
SCNM	Self-Configuring Network Monitor
TCP	Transmission Control Protocol
PLR	Packet Loss Ratio
RMSE	Root Mean Square Error

Chapter One

1. Introduction

In Ethiopia, Ethio-Telecom is working towards achieving the best possible infrastructure the country can have by taking into consideration the level of development it reached right now. As the number of population increases, the number of subscribers increases too. As a result, it become more difficult to meet all of their needs effectively if the existing network system isn't upgraded. So in order to solve this problem, different expansion projects are being implemented by the service provider. In addition different types of services are being introduced. Since these expansions need high budget, first additional work needs to be done towards analyzing the existing networks by doing network traffic analysis. This helps to prevent network congestion and unproductive use of the available resources.

In this thesis it is intended to do a traffic analysis on the existing mobile core network during peak hours. The objective of doing this analysis is to see the patterns of the existing network traffic and analyze using different mathematical models and tools. After the analysis is done, it is intended to give a recommendation on how to manage the available network resource so that the subscribing customers get the best possible quality of service for their demand. The reason it is intended to do this thesis is due to the fact that most customers are not satisfied with the service they are provided by Ethio-Telecom. This is because the ever increasing customer demand and the service provided are not proportional. The fact that Ethio-Telecom is the only service providing company in the country will put more pressure to it. Until the other competitive telecom companies introduced in to the market start giving their service, Ethio-Telecom has the sole responsibility for all its customer satisfactions.

Though the government is now making the market open to other telecom service providers, but until they start their work effectively, it will take years and until then Ethio-Telecom has to work day and night to improve the network quality for its customer's satisfaction. Hence network administrators need to closely monitor the network traffic to avoid network congestion problems

1.1. Statement of the Problem

The telecom industry has significant effect on the economic activities of the countries, not only on the economy but also towards the growth of other industries. The problem which initiated to do this thesis is that there was only one telecom service providing company in Ethiopia and the ever increasing demand of customers and its providing capacity are not proportional. Hence frequent problems such as busy network, error in connection, poor network coverage, slow internet service and expensive price of service for expansion projects are some of the main reasons. Therefore, traffic analysis is important for evaluating the performance of existing networks. It enables network administrators to closely analyze and monitor the existing mobile network for utmost satisfaction of subscribers.

1.2. Scope and Limitation

The scope of this thesis is limited to the study of data network traffic at the core network layer with in the Addis Ababa region only. The key performance indicators to be used for analysis are limited to Packet Loss Ratio and throughput and the outcome from the real time data gathered from the service provider network will be compared with the standard acceptable values. One of the expected limitation in this thesis is that the highly increasing number of subscribers.

1.3. Significance of the Study

The significance of doing this thesis is that it enables the network administrators to easily identify and evaluate faulty points between source and destination of data transmission during the peak hour and take corrective actions, if necessary, without losing too much time and resource easily. It will also be helpful as an input in network planning and to manage and utilize the existing resource effectively. In addition to its cost effectiveness, it is also important to upgrade the Quality of Service (QoS) for network subscribing customers. This will be achieved easily because once the trend of the network traffic is known, it will be easy to take actions on more demanding areas.

1.4. Objective

1.4.1. General Objective

The main objective of this thesis is to analyze the peak hour mobile core network data traffic at flow connection level using flow method to improve the quality of the network flow in the case of Ethio-Telecom.

1.4.2. Specific Objectives

The specific objectives of this thesis are:

- To identify peak hour during the day
- Analysis of the Packet Loss Ratio (PLR) and throughput Key Performance Indicators (KPI's) and compare the real time network data output with standard acceptable values
- Study and compare the output values with standard acceptable values and conclude & give recommendation for best network improvement

1.5. Methodology

In this thesis network traffic is analyzed at the core network level. In order to do the analysis, the peak hour during the day is chosen. So sample data is collected from Ethio Telecom network. At first the busiest hour is identified. After that two KPI's are analyzed. For this thesis the KPI's chosen for analysis are Packet Loss Ratio (PLR in percentage) and throughput (packet/sec). The reason these parameters are chosen is in order to be able to see the number of packets that are being sent and received and to identify how many of them are failing to reach their intended destination. From this information, it will be easier to give a recommendation to the problems. The descriptions of methods that will be used are as shown in figure1-1 below.

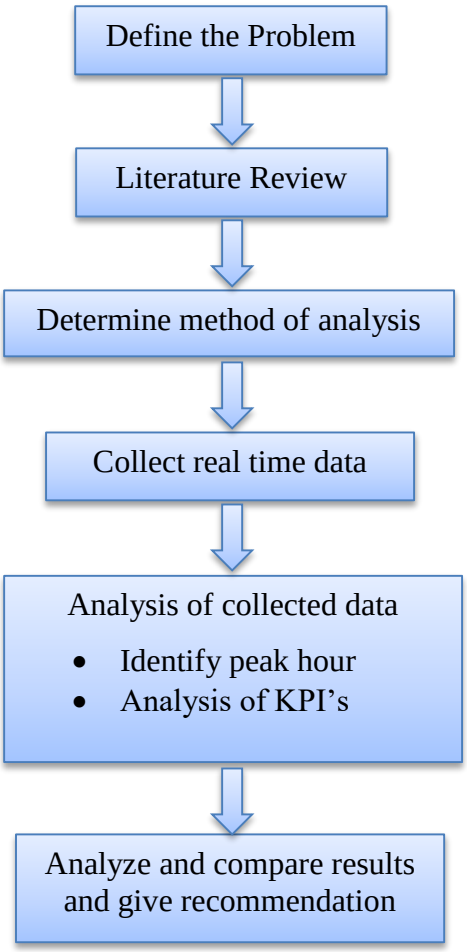

Figure 1-1: General work flow chart

1.6. Thesis Outline

This thesis is organized into five chapters. Chapter one dealt with the overall introduction of this thesis including the statement of the problem, objectives and scope and limitations. Chapter two covers general theoretical background and also related literature reviews are included. Chapter three describes the method used in order to achieve the objective set out in the first chapter. In chapter four the results and analysis are discussed. In the final section, chapter five, conclusion is given and then a recommendation for future work is set out.

Chapter Two

2. Literature Review and Theoretical Background

In different times, there are different researches and analysis done on mobile traffic analysis. They used different analysis models and different key parameters to identify what the existing mobile network is like. Previously they had the problem of collecting the data they needed. But now days, this problem is minimized. Getting the requested data is sometimes difficult in situations related to security issues. In this section some related works that are done previously by other researchers is reviewed to put a point of reference to this study.

In order improve the energy efficiency and resource management of cellular networks, predicting analysis has been made by Xuan Zhou et al. in [1]. In their work they mainly took the theory of entropy. They categorized the traffics to voice, text and data. They made a simple observation that, any type of service traffic is similar to other types and also the type of traffic in one day is also similar to those on other days too. Yet, the regular pattern of traffic led them to a question to what extent can the traffic be predicted with certain prior information like historical traffic types and adjacent cells' traffic. In order to answer their question, they gathered numerous data and then used entropy theory to quantify the information measurement that sequential service and spatial traffic will provide for prediction. Then finally they concluded that voice traffic value is more uniform than data, traffic at one moment can be well predicted when the preceding 15 hours traffic is known, voice service is the easiest to predict for its regularity, voice traffic has so close similarity to text traffic in the same cell and knowledge of adjacent cells traffic can enhance the predictability of voice and text more effectively than data.

An article was written by F. Ricciato et al. in [2] for understanding of the complexity of the 3G wireless network with the protocol dynamics of TCP/ IP (Transmission Control Protocol/ Internet Protocol) networks. To initiate their study, understanding of such an environment was deemed more urgent and at the same time more difficult than for legacy 2G networks, that were intrinsically simpler and subject to slower changes.

The most important features they outlined for the monitoring system that enable the analysis tasks include complete capture (with standard hardware equipment, with no need to resort to packet sampling), user- and control-plane capture (the system captures and parses signaling frames at each layer. This allows for cross-layer analysis and data/ control plane correlation), anonymization and stateful association tracking. Then with these, they detected the congestion in the Core Network, the Radio Access Network and also detected behaviors on the data plane.

A study was made on traffic analysis on busy hours of the day by A. Ozovehe et al., in [3]. This was made with the intention of improving the Quality of Service (QoS). Due to increasing demands of customers, they are continuously rolling out new services. This has resulted in many congested networks and consequently degradation of QoS due to inadequate provision of the needed resources or underutilization of the available resources. For their work, they used busy hour traffic data of access network from a live network to analyze traffic congestion in some macro cells of GSM/GPRS network. Their analysis showed that traffic channel congestion beyond the acceptable 2% threshold for traffic channel occurred. The analysis of the ten most congested showed that slow response to congested cells is the major setback that affects the QoS as some cells congested continuously over one year. The strong correlation showed that the knowledge of call setup success rate and busy traffic can be used to predict traffic channel congestion which is crucial for cellular network optimization and resource management.

These days it is becoming easier to get access to a user's information by using different tools. This intrusion not only threatens the security of our information but it also affects network performance by influencing the bandwidth consumption by attacking the network. Myung-Sup Kim et.al [4] in their work propose a method to detect abnormal network traffic. They proposed a flow-based method for abnormal network traffic detection in two ways; detection from flow header and detection from traffic patterns. The flow header detection part checks the field values of a flow header. Flow is defined as a collection of packets with the same 5-tuple: source IP address, destination IP address, source port, destination port, and protocol number. The flow size and packet count, refer to the total bytes and the number of packets that belongs to the flow, respectively. For detecting attack from traffic pattern, they characterize patterns by parameters such as flow count, flow size, packet count, packet size, total bandwidth and total packet count of traffic based on flow.

Their work was mainly focused on detecting different attacks in a network. They created a comparison algorithm for checking a destination-based traffic pattern data and source-based traffic pattern data. Regardless of the source and destination of traffic pattern data, traffic sent or received from a certain machine is investigated.

A thesis by Meheretu Daka [6] prepared on traffic analysis of IP core networks first divided network layers based on their functionality as access, aggregate and core network layers to simplify and also depict problems very easily. On the thesis he chose the core network layer for his case study and while preparing the modeling, he chose the packet delay modeling because one of the most important performance analysis parameter in a network is the average delay required to deliver a packet from origin to destination. In addition to simulation results from real time data, mathematical modeling has been done in order to compare the Key Performance Indicators (KPI's) with the standard values. His findings in the study, were proposed to be used as a reference for new build as well as expansion of core network.

2.1. Mobile Network Architecture

The overall end-to-end architecture of a carrier network is composed of three big parts: the Radio Access Network (RAN), Core Network (CN), and External Network [7]. M. Vaezi et.al [7] explain about Radio Access Network and the overall architecture briefly as RAN is the first component of any carrier network and provides access and shuttles the voice/data to and from the user equipment. The core network then connects the radio access network to the external network, e.g., the public- switched telephone network or the public Internet. RAN is composed of Radio Base Stations (BS), Base Station Controllers (BSC) and backhaul network. In different generations of mobile networks, i.e. in 2G, 3G and 4G network technologies, the main components of RAN are different. For 2G network, at the base station there is BTS (Base Transceiver Station), at the controller there is BSC (Base Station Controller) and Abis is the Interface. For 3G network, at the base station there is NodeB, at the controller there is RNC (Radio Network Controller) and lub is the interface. In 4G network, at the base station there is eNodeB, at the controller there are eNodeB, MME (mobility management Entity), and SGW (serving gateway) and S1 is the backhaul interface. The other component of the RAN, backhaul network, acts as the link between BSs and BSCs, toward the core.

Mobile backhaul enables to transport mobile data from the end user to mobile networks, traditional telephone networks, and the Internet. It acts as a link between the edge of a network (access nodes) and the mobile core. The core network is responsible for service management including the establishment, termination and reconfiguration of the current communications. It also provides the gateway to other networks. Within the big parts, the mobile network architecture again consists of the Base Station Subsystem (BSS), at the access layer of the network/ RAN/, the Network Switching Subsystem (NSS), at the core network layer, and the Network Management Subsystem. These major components are composed of smaller subunits which are discussed briefly in the coming sections. In this thesis, the 4G network is chosen for analysis. So in the coming sub sections, brief discussion will be made on the components of the 4G network.

2.1.1. Network Switching Subsystem (NSS)

The Network Switching Subsystem (NSS) is the component that carries out call and mobility management functions. The main functions of NSS are call control (i.e switching of calls between the mobile and other fixed or mobile network users), charging, mobility management, signaling and subscriber data handling. It contains the network elements MSC, VLR, HLR, AC and EIR as discussed by Rakibul Hasan et.al. in [8].

MSC (Mobile Services Switching Center) is the central and main component of the NSS. It is responsible for controlling of calls, identifying the origin and destination of a call and as well as what type of call has been initiated. It is also responsible for handovers, updating the location of users and charging data collection. An MSC acting as a bridge between a mobile network and a fixed network is called a Gateway MSC. Every MSC has a unique identification [8].

Visitor Location Register (VLR) is integrated with the MSC. It is a database which contains temporary information about subscribers currently being in the service area of the MSC/VLR such as: Identification numbers of the subscribers, security information for authentication of the SIM card and for ciphering (encoding), services that the subscriber can use. Later, if the mobile station makes a call, the VLR will have the information needed for call setup without having to interrogate the HLR each time [8].

The HLR is a database used for storage and management of subscriptions. It is the most important database, as it stores permanent data about subscribers such as subscriber identity numbers and the subscribed services. For example when an individual buys a subscription in the form of SIM, then all the information about this subscription is registered in the HLR of that operator. In addition to the fixed data, the HLR also keeps track of the current location of its customers. The MSC asks for routing information from the HLR if a call is to be set up to a mobile station (mobile terminated call). The two network elements, Authentication Centre (AC) and Equipment Identity Register (EIR), are located in the HLR [8].

The Authentication Center (AC) is part of the network that provides security and which also stores a protected database storing a copy of secret information in each subscribers SIM card. It is used for Authentication between the mobile station and the VLR and encodes the information transmitted in air interface between the MS and the BTS. The AC protects operators from different types of fraud found in today's cellular world [8].

The Equipment Identity Register (EIR) is a database that is used for security which contains list of valid mobile equipment in a network. The International Mobile Equipment Identity (IMEI) is used to identify the validity of each mobile equipment with a unique identity (IMEI) number which consists of type approval code, final assembly code and serial number of the mobile station. As the IMEI is used for security purposes, it will be marked as invalid if it has been reported stolen or is not approved.

The EIR contains three lists; (1) Mobile equipment in the white list is allowed to operate normally, (2) If we suspect that mobile equipment is faulty, we can monitor the use of it. It is then placed in the grey list, (3) If the mobile equipment is reported stolen, or it is otherwise not allowed to operate in the network, it is placed in the black list.

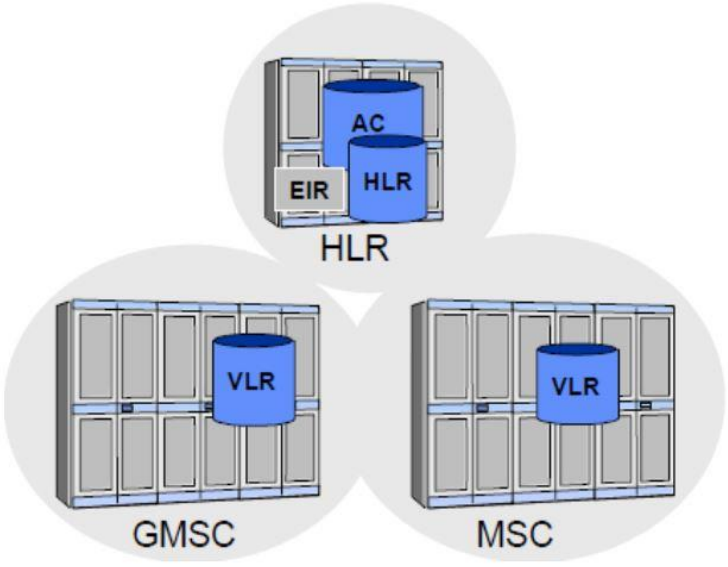

Figure 2-1: Network Switching Subsystem (NSS) [8]

2.1.3. Network Management Subsystem (NMS)

It is discussed that the purpose of the Network Management System (NMS) is to monitor various functions and elements of the network by Rakibul Hasan et.al in [8]. The operator workstations are connected to the database and communication servers via a Local Area Network (LAN). The database server stores the management information about the network. The communications server takes care of the data communications between the NMS and the network elements. The functions of the NMS can be divided into three categories: fault management, configuration management and performance management. In [9] the author discuses about each functions of the categories of the NMS. The fault management provides the network operator with information about the current status of alarm events and maintains a history database of alarms. The configuration management maintains up to date information about the operation and configuration status of network elements such as radio network configuration, software and hardware management of the network elements, time synchronisation, and security operations. In performance management, the NMS collects measurement data from individual network elements and stores it in a database. On the basis of these data, the network operator is able to compare the actual performance of the network with the planned performance and detect both good and bad performance areas within the network.

2.2. Network Monitoring Technique

As discussed by M. Uma et.al in [10] there are various traffic monitoring techniques available based on many concepts and they are classified into four types. Out of the four types discussed, for this thesis, only the Monitoring and Analysis technique and Statistical Method are considered. Further, monitoring and analysis techniques will be discussed in the coming sub sections.

The Monitoring and Analysis technique is again classified as Router-Based Monitoring Technique and Non-Router Based Monitoring Technique. Router Based Monitoring are monitoring functionalities that are built-into the routers themselves and do not require additional installation of hardware or software. It consists of three methods i.e. Simple Network Monitoring Protocol (SNMP), Remote Monitoring (RMON) and Netflow Monitoring. Whereas Non-Router Based Monitoring needs additional hardware and software to be installed to it. In this thesis, Non-Router Based Monitoring technique is used. The Non-Router Based Monitoring is further classified as follows.

2.2.1. Active monitoring

Active monitoring injects additional traffic into the network in order to make its measurement. It is suitable for making controlled measurements that are not possible with passive monitoring. The author in [10] discusses that active monitoring shows the way to gather the dimensions between two endpoints in a particular network. Active measurement systems deal with metrics such as availability, routes, packet delay, packet reordering, packet loss, packet inter-arrival jitter and bandwidth measurements. Interfering into the network to examine its performance is the problem that exists in active monitoring due that the normal traffic information seems to be questioning the validity of the network information. The problem that exists with active monitoring is that introducing probes into the network can be an interference to the normal traffic on the network. Often times the active probes are treated differently than normal traffic as well, which causes the validity of the information provided from these probes to be questioned. As a result of the information detailed above, active monitoring is very rarely implemented as a stand-alone method of monitoring as a good deal of overhead is introduced. On the other hand passive monitoring does not introduce much of any overhead into the network [10-11].

2.2.2. Passive monitoring

Unlike active monitoring, passive monitoring has an advantage over active monitoring because it does not inject traffic into the network or modify the traffic that is already on the network. But on the other hand it has its own downfall in that it has a problem with the post processing. It will require more time and measurements can only be analyzed off-line and not as they are collected. This creates another problem with processing the huge data sets that are collected. Also unlike active monitoring, passive monitoring collects information about only one point in the network that is being measured rather than between two endpoints as active monitoring measures. Passive monitoring system has a simple setup in a single link between two end point and monitors traffic as it passes along the link and it can be achieved with the assistance of any packet sniffing program [10-11]. In this thesis, passive monitoring techniques is used.

2.2.3. Hybrid monitoring

Though both active and passive monitoring have their own merits, it is clear that using only one of the methods for network monitoring has a drawback. As a result using the combination of both active and passive methods enables us to utilize the best aspects of both methods. The two hybrid monitoring techniques include; Watching Resource from the Edge of the Network (WREN) and Self-Configuring Network Monitor (SCNM).

2.2.3.1. Watching Resource from the Edge of the Network (WREN)

Watching Resource from the Edge of the Network (WREN) is one of the hybrid network monitoring method which combines both active and passive methods. When traffic is low, it uses the active monitoring method, while when traffic is high it uses the passive monitoring method. It gives a more accurate measurements because it monitors at both the source and destination end host. WREN uses packet traces from existing application traffic to measure the available bandwidth. WREN is split into two levels, the kernel level packet trace facility and the user level trace analyzer.

The kernel level packet trace facility is responsible for capturing the information associated with incoming and outgoing packet. One call starts the trace and provides the information needed to conduct it while another call retrieves the trace from the kernel. The packet trace facility will coordinate the packet messages sent from one machine to the other machine by ensuring the same range of packets have been flagged for transmission by tracing. This coordination ensures

that the information about the same packets is stored at each end of the connection regardless of what happens in between.

The user level trace analyzer is the other level in the WREN environment. It is the component that begins any packet traces and collects and processes the data returned from the kernel level trace facility. By design the user-level components are not required to read the information from the packet trace facility at all times. It can be analyzed immediately after the trace is completed to make runtime decisions or stored for future analysis [11].

2.2.3.2. Self-Configuring Network Monitor (SCNM)

Self-Configuring Network Monitor (SCNM) is another technique of hybrid monitoring method. This environment consists of hardware and software components. The hardware being installed at critical points in the network is responsible for passively collecting the packet headers whereas the software runs on the endpoints of the network. The software is responsible for creating and sending the activation packets that are used to start the monitoring of the network [11].

2.3. Network Traffic Analysis Methods

Network administrators always require gathering new information and learning something new about the state of the network. For this, Probes are inserted at key points in a network for the purpose of monitoring or collecting data about network activity.

Routers and switches are the computer networking devices that allow one or more computers to be connected to other computers, networked devices, or to other networks. They function as embedded monitoring probes and do not have enough computing power to monitor network traffic on high-speed links as their primary concern is packet forwarding.

When one starts to do a network traffic analysis, there are a number of parameters that will be taken into consideration. In order to do traffic analysis, there are two different methods that can be used, Flow Analysis and Packet Analysis. Generally flow analysis can help to determine traffic statistics overall. But it has a drawback in that when it is intended to analyze a specific detail in depth. For this packet analysis has an advantage on this rather than the flow analysis. If we only want to see IP address and how much data they are transferring, flow analysis will be used. However in order to trouble shoot performance problems we will use packet analysis [12].

We will see in brief about packet analysis and then further continue to see a well in depth discussion about flow based analysis in coming sections.

2.3.1. Packet Analysis

Packet analysis will be used when we want to see a specific problem and also when we want to identify specific websites, users, applications, files etc. With this method, we can identify individuals and their access and usage of resources. Packet analysis is normally associated with (Switched Port Analyzer) SPAN or mirror ports, which are available on most managed network switches. "Port mirroring" is used on a network switch to send a copy of network packets seen on one switch port (or an entire VLAN) to a network monitoring connection on another switch port [12].

2.3.2. Flow Analysis

A flow is a traffic stream with a common set of identifiers. Typically, a flow is defined by traffic that has the same source IP, destination IP, protocol, source port, and destination port. If any of these variables change, then a new flow is defined. For example, when a client is connecting to a server, several flows might be created because the client might establish several connections to the server, involving new source ports. Each one of these connections would be a separate flow [12].

As defined by Anna Sperotto et.al in their work in [5], defined flow as a unidirectional stream of packets that share common characteristics, such as source and destination addresses, ports and protocol type. In addition, a flow includes aggregated information about the number of packets and bytes belonging to the stream, as well as its duration. Flows are often used for network monitoring, permitting to obtain a real time overview of the network status.

Flows can also be used to measure the exact delays end-users experience. This can be calculated by recording the time between the pair of TCP SYN and TCP ACK packets. Measuring these delays with flows is not often performed because of the major adjustments that must be made to the probe [13].

2.4. Traffic Analysis Models

In order to do traffic analysis, different traffic models are used by network designers to make assumptions about the networks being designed based on past experience and also enable them to predict the performance for future requirements. As per a survey done by the author in [14], traffic models are used in two fundamental ways: (1) as part of an analytical model or (2) to drive a Discrete Event Simulation (DES). In the coming sections, we will see the basic traffic models that can be used while doing traffic analysis.

2.4.1. Exponential Distribution

The Exponential distribution is the type of distribution that is widely used for events that happen very randomly. It has a very similar characteristics with that of Poisson distribution. The most important characteristics of the Exponential distribution is that it is "Memory less", that means time has no effect on future outcomes. In addition, the Exponential model is also most appropriate to describe a chance of failure rate [20].

The probability density and cumulative distribution functions of the Exponential model are given as follows:

$$F(x) = 1 - e^{-\lambda x} \quad \ldots\ldots (2.1)$$

$$f(x) = \lambda\, e^{-\lambda x} \quad \ldots\ldots (2.2)$$

where lambda, λ is the mean of the distribution data.

2.4.2. Weibull Distribution

Unlike the exponential distribution, the Weibull distribution can assume different distributions based on its shape and scale parameters. It is a very flexible type of distribution. When the shape parameter is equal to 1, it becomes identical to Exponential distribution. But when its shape parameter is less than 1, the Weibull distribution becomes a steeply declining curve. [20].

The Weibull distribution has probability density and cumulative distribution functions defined as follows [15]:

$$f(x) = \frac{\alpha}{\beta^{\alpha}}(x)^{\alpha-1}e^{-\left(\frac{x}{\beta}\right)^{\alpha}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.3)$$

$$F(x) = 1 - e^{-(\frac{x}{\beta})^{\alpha}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.4)$$

Where x ≥ 0, α is shape parameter and β is scale parameter.

2.4.3. Normal Distribution

One of the main distributions in the probability theory and which is also used widely is the Normal distribution. It is very helpful to describe uncertain values in different phenomenon. The main three conditions underlying in the Normal distribution are: (1) some value of the uncertain variable is most likely the mean of the distribution, (2) the uncertain variable could as likely be above the mean as it could be below the mean (symmetrical about the mean) and (3) the uncertain variable is more likely to be in the vicinity of the mean than further away [20].

The Normal distribution has probability density and cumulative distribution functions as follows:

$$f(x) = \frac{1}{\sigma\sqrt{2\mu}}e^{-(\frac{(x-\mu)^2}{2\sigma})} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.5)$$

$$F(x) = \int_{-\infty}^{x}\frac{e^{-x^{2}/2}}{\sqrt{2\pi}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.6)$$

Where μ is the mean and σ is the standard deviation.

2.4.4. Gamma Distribution

The Gamma distribution can be used to measure the time in between occurrence of events and when the event process is not completely random [20].

The Gamma distribution has probability density and cumulative distribution functions defined as follows:

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\, x^{\alpha-1}e^{-(\frac{x}{b})} \quad \text{..... (2.7)}$$

$$F(x) = \frac{\Gamma_x(\alpha)}{\Gamma(\alpha)} \quad \text{..... (2.8)}$$

Where Γ(α) is the gamma function defined as

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}e^{-t}dt \quad \text{..... (2.9)}$$

2.5. Parameters for Mobile Traffic Analysis

In order to do a network traffic analysis, after being connected to the network and gathered the available data needed, a study will be made on the basic parameters (Key Performance Indicators (KPI's). We will discuss in short these KPI's in the coming section.

2.5.1. Packet loss

As the name indicates, packet loss is the number of packets lost during transmission. When there is a packet loss, the packet data sent from transmitter side will not be delivered at their intended destination. Though due to packet loss any application can be disrupted, but those with real time packet processing such as video and audio programs are the main victims. Packet loss displays itself in the form of network disruption, slow service and even total loss of network connectivity. It occurs if network discards packets when a router or other network device is overloaded and cannot accept additional packets at a given moment.

There are a number of factors that cause packet loss, they include congestive loss (i.e. occurs when the network is unable to support the amount of traffic that it receives), Device Performance (i.e. when network equipment such as Router, Switch Firewall etc. performance is unable to meet the requirement of the network flow), when there are software issues on a network device, faulty hardware or cabling and loss due to transmission bit errors (i.e. bit errors resulting from transmission channel noise, distortion, signal weakness, bit synchronization, or attenuation) [6]. In order to avoid poor network quality and maintain good customer satisfaction, it is recommended for the packet loss ratio to be within the range1% to 1.5% [8].

2.5.2. Throughput

Network throughputs are the data or packets transmitted through a link from source to destination. Since it is the amount data transmitted form source and received at the destination, for any network to be more efficient, it must have more throughput value. It is calculated in terms of packets/sec or bits/sec. It tells us how much data was transferred from a source to a destination at a given period of time. Throughput measures how many packets arrive at their destination. It is the rate of successful message delivery over communication channel. It can be said that the value of the throughput of the network is directly proportional to the performance of the network [17]. In order to have an efficiently working network, we must work on increasing the throughput. The acceptable threshold value for throughput is 11Mbps [9].

2.6. Network Switching Systems

There are two switching methods that are used to connect multiple communicating devices with one other, they are Circuit Switching and Packet Switching. While Circuit Switching was particularly designed for voice communication, since it was less suitable for data transmission, a better solution evolved for data transmission called Packet switching.

In Circuit Switching, a dedicated communication channel (circuit) is established in a network before the nodes may communicate. The circuit will remain connected for the duration of the communication session. It has a drawback in that the line might remain idle even when there is no communication unless it is reconnected with another node for another communication. In Circuit Switching, message is received in the order sent from the source and it is implemented at the physical layer.

If we see the billing of such communication services, circuit switching is characterized by a fee per unit of connection time, even when no data is transferred, while packet switching may be characterized by a fee per unit of information transmitted, such as characters, packets, or messages [18], [19].

In contrast to Circuit Switching, the other switching method is Packet Switching. Unlike Circuit Switching, Packet Switching is flexible, because a route is created for each packet to travel to the

destination. Packets of a message are received out of order and assembled at the destination and it has two approaches Datagram Approach and Virtual Circuit Approach. Packet Switching is implemented at network layer level [18], [19].

Packet switching is a method of grouping data that is transmitted over a digital network into packets. Packets are made of a header and a payload. Header refers to supplemental data placed at the beginning of a block of data being stored or transmitted. In data transmission, the data following the header is sometimes called the payload or body. The payload is the part of transmitted data that is the actual intended message. Headers and metadata are sent only to enable payload delivery. Data in the header are used by networking hardware to direct the packet to its destination where the payload is extracted and used by application software. Packet switching is the primary basis for data communications in computer networks worldwide.

Packet switching is used to optimize the use of the channel capacity available in digital telecommunication networks, such as computer networks, and minimize the transmission latency (the time it takes for data to pass across the network), and to increase robustness of communication [19].

2.7. Interfaces in the Mobile Network

The mobile network had different interfaces at the different layers. Cisco defines Gi/SGi as an LTE interface to the Packet Data Network (PDN) to enable subscribers protect themselves from threats and the public internet. The Gi interface is the connection between a GGSN and the Internet or destination networks connected to a PLMN (Public Land Mobile Network) while the SGi interface is the connection between a P-GW (Packet Data Network Gateway) and the Internet or destination networks connected to a PLMN in 4G networks. In general the Gi interface is in 3G network whereas the SGi interface is in a 4G network.

On a different note on an article written by [21] Gn/Gp interface is the interface between the SGSNs and the GGSNs. "Gn interface" is the interface between SGSNs and the GGNs when they exist within the same mobile network. The other case is when the user is travelling to another mobile network, which is the case of roaming. In this case, the SGSNs and the GGSNs will be in different mobile networks, in this situation, the interface between the SGSNs and the

GGNs will be "Gp interface". The protocol stack for the Gn/Gp interface shows that it uses TCP/IP for packet routing between different nodes. While it uses the GTP Protocol (GPRS Tunneling Protocol) in the application layer. The GTP protocol includes "GTP-C" (GPRS Tunneling Protocol for the Control Plane) to tunnel signaling messages between the SGSNs and the GGSN and "GTP-U" (GPRS Tunneling Protocol for the User Plane) to tunnel user data between the SGSNs and the GGNs. The difference between the control plane and the user plane is that at the control plane, signaling messages are exchanged in order to initiate, modify or terminate the data session of the use. But at the user plane the nodes exchanged user data (sent & received) between the user equipment and the external equipment while the data session is active.

In a cellular network, the primary components that employ GPRS (General Packet Radio Service) are SGSN and GGSN. GPRS is a technology for the support of packet switching traffic in a GSM network. GPRS enables high-speed wireless internet and other data communications in GSM [22]. The GPRS core network is the central part of the GPRS which allows 2G, 3G and WCDMA mobile networks to transmit IP packets to external networks such as the internet. The GPRS system is an integrated part of the GSM network switching subsystem. The network provides mobility management, session management and transport for Internet Protocol packet services in GSM and WCDMA networks. The core network also provides support for other functions such as billing and lawful interception [23].

SGSN (Serving GPRS Support Node) is a main component of the GPRS network, which handles all packet switched data within the network, e.g. the mobility management and authentication of the users. The SGSN performs the same functions as the MSC for voice traffic. The SGSN and the MSC are often co-located [24]. The SGSN overall performs tasks which are mainly connected to the subscriber including collection of different billing data of each mobile subscriber. On the other hand GGSN (Gateway GPRS Support Node) converts incoming data traffic from mobile users (via the SGSN) and forwards it to the relevant network, and vice versa [25].

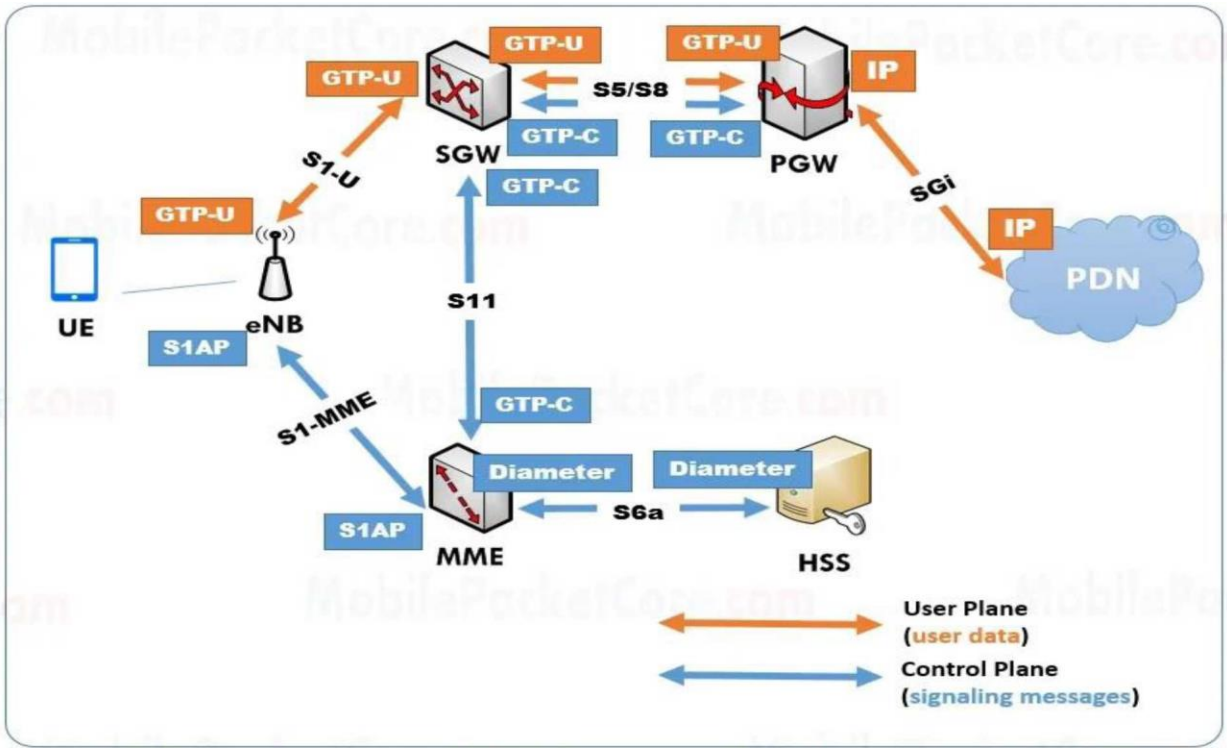

Figure 3-2: SGi Interface Location [21]

Chapter Three

3. Methodology

The basic steps taken are discussed in this chapter and it is seen how the analysis will be done in consecutive section. First let's see the system model in order to be able to visualize the process at the beginning.

3.1. System Model

With a system model, it can be seen the basic steps and procedures taken and the tools that will be used. It clearly portrays what procedures will be taken to achieve the final result. Figure 3-1 below shows the system model.

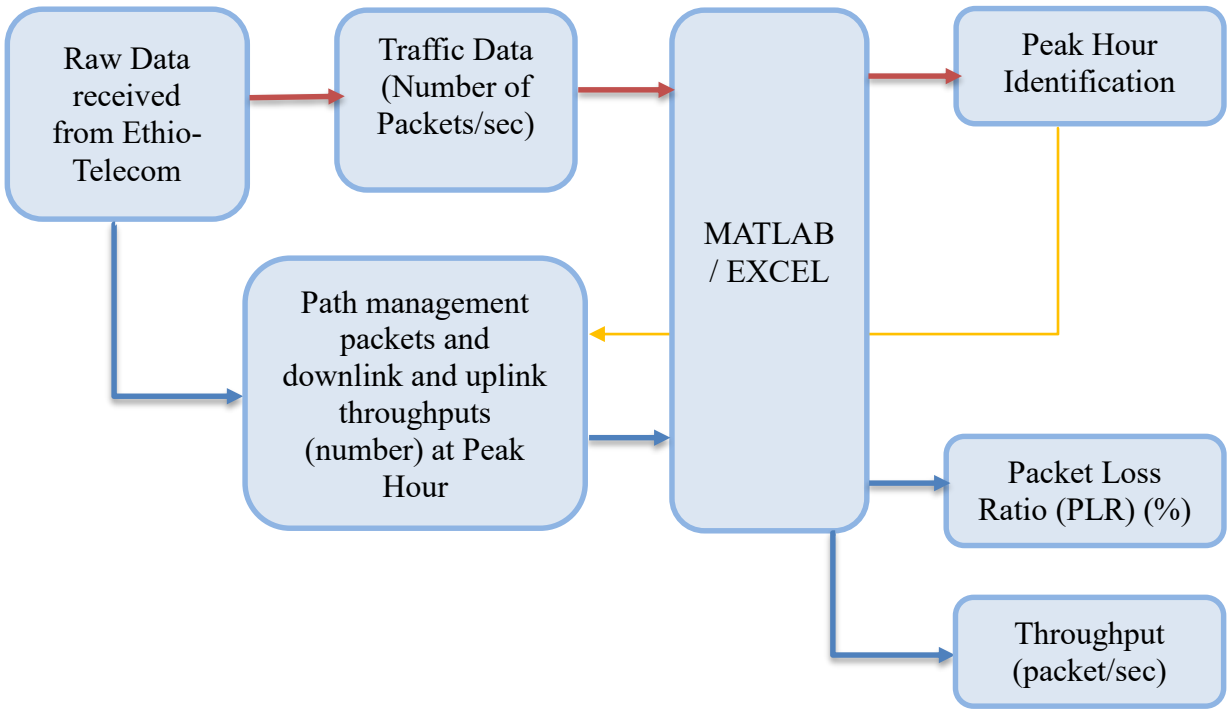

Figure 4-1: System Model

For this thesis, the main components from the system model are the MATLAB and Excel tools which are used for the analysis of the network traffic. With the Excel tool, the raw data that has been captured will be filtered and sorted out for the specific data that are necessary for this work. The extracted data are then reorganized to be exported into MATLAB. All the result analysis, i.e. the identification of peak hour and traffic model identification are done on the MATLAB tool.

3.2. Data Gathering

Since this thesis is the case study for the Ethio-Telecom network, the first step is gathering of data from its network. There are two ways to capture data traffic from a network system; the first one is mirroring a port whereby the active packets being transmitted are copied onto a mirroring port. The second one is by the use of additional hardware devises called Network Test Access Point (TAP) which is used in our case. These external hardware devices are used for the purpose of network traffic analysis and to identify a problem which is what we are doing in this thesis. They are very important because if we are to make captures of the data directly from the network system, the performance will decrease because we are introducing additional request into the system and this will make it to toggle between either capturing the data we requested it or to transmit the data from the sender to receiver. So this division will interrupt data transmission which causes unnecessary additional traffic congestion which the system already has from transmission of its data. So in order to avoid this, the service provider already has TAP device in place. The data collected from this access point is the exact copy which is 100% similar to the one in the original network system.

3.3. Peak Hour Identification

After the captured data is gathered from Ethio-Telecom, the next step is identification of the peak hour because that is the time where there is high traffic congestion within the network. In order to determine the peak hour, the sampled data is gathered for a period of one moth every day for 24hrs with in each day. From that, the traffic data in packets per second is chosen for analysis because it clearly shows the number of transmissions with in the network. The data is filtered and classified for each day, divided into per hour. After that the raw data is imported to MATLAB one by one for each day for identification of the time where there is high traffic within the network. This enables us to see the pattern clearly and compare it for each day by identifying where there is the highest activity and lowest activity.

From this, the average peak hour can be determined which is helpful for next step Key Performance Indicators (KPI's) analyses.

3.4. Analysis of KPI’s

Once the peak hour is identified, then the full raw data is filtered and organized in such a way that enables us to be able to analyze the Key Performance Indicators. The KPI's that are chosen for analysis in this thesis are Packet Loss Ratio (PLR) in percentage and Throughput in packets/sec.

From the filtered and extracted raw data, number of packets sent and number of packets received are used in order to calculate the PLR. From the raw data path management packets are chosen to be used because they are closer to help with the analysis of flow of the packets. For the analysis of throughput, the downlink peak throughput and the uplink peak throughput in packet per second are used for each day during the peak hour.

The other approach that will be taken to make analysis of KPI's is by identifying which of traffic analysis models best fit with that of our networks. For this the different Cumulative Distribution Functions for different traffic analysis models will be seen against that of our data and to see which one it fits closest too. In addition of this the root mean square error (RMSE) will also be calculated for each one of them for a better accuracy.

Chapter Four

4. Result and Analysis

In this chapter, by using the raw data collected from the core network, first the peak hour is identified during the day. After that, by using the mathematical formulas (2.1 to 2.9) indicated in chapter two, the KPI's for the peak hour data are analyzed. Different traffic analysis models are also seen in order to identify which model best fits with the collected network data. The results obtained are then analyzed by using compare and contrast method with the highly acceptable standard thresholds for the maximum or minimum values. In addition to that, a comparison between the peak hour and the slow hour is shown in order to be able to depict the differences.

4.1. Identification of Peak Hour

The first step in this thesis is to identify when the peak hour is during the day time. This is done because the focus of this work is analysis of network traffic during the busiest time of the day and it has to be thorough shall be confirmed. The data collected is classified on hourly basis, in order to enable clearly show when the peak hour is during each day. The numbers of packets per second versus time of the day are drawn. Figure 4-1 to 4-4 are sample graphs that show the network traffic pattern for each day for the period of time when the data has been collected.

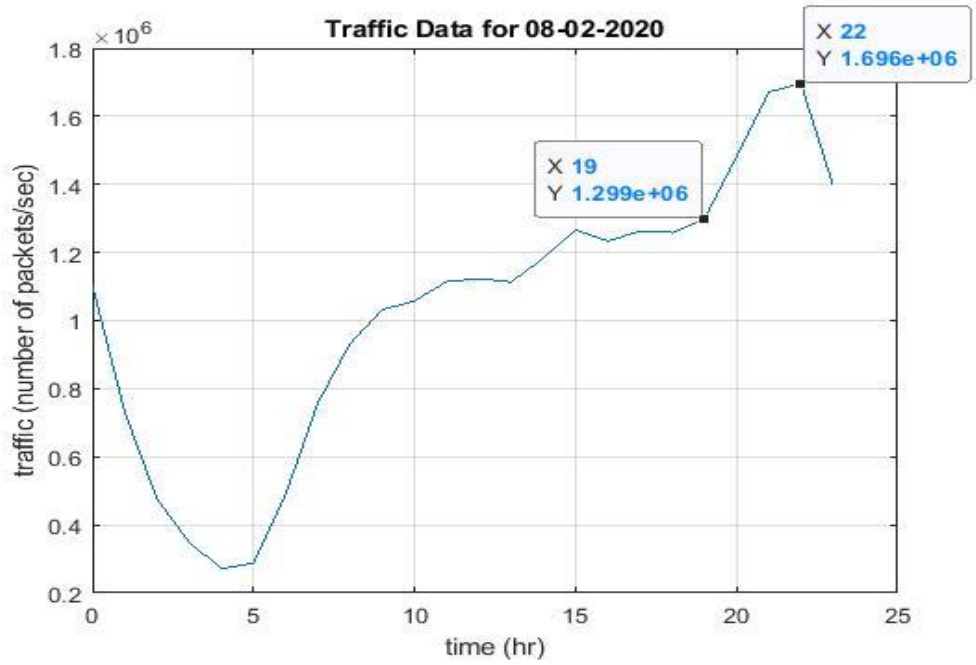

Figure 5-1: Sample Graph of Traffic Data for the day 08-02-20

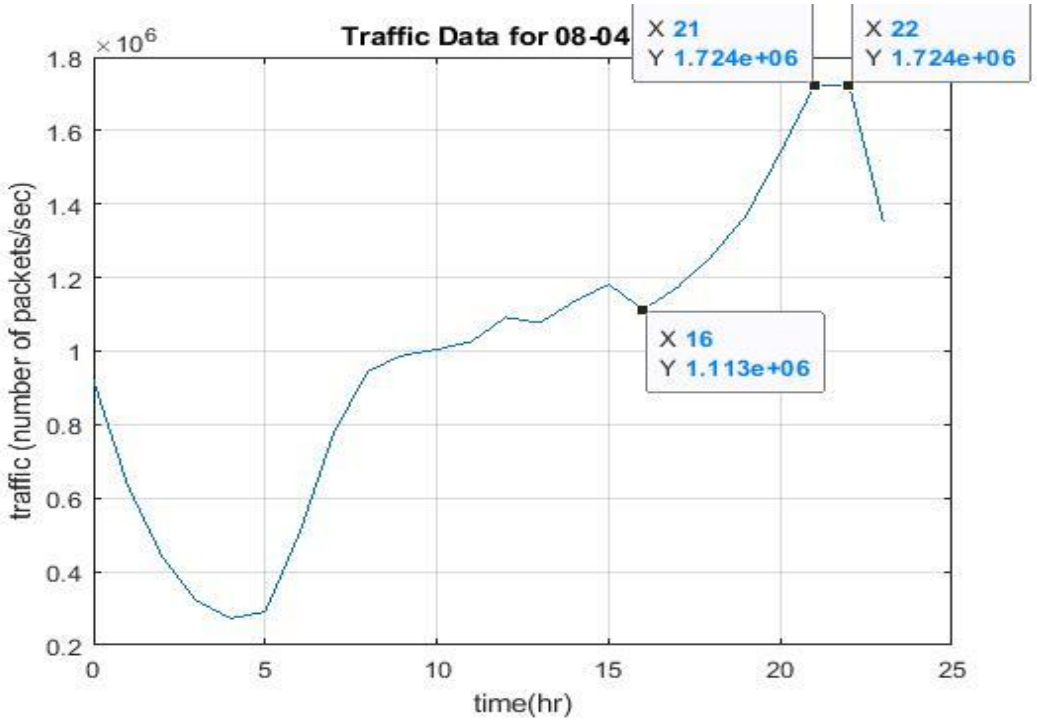

Figure 6-2: Sample Graph of Traffic Data for the day 08-04-20

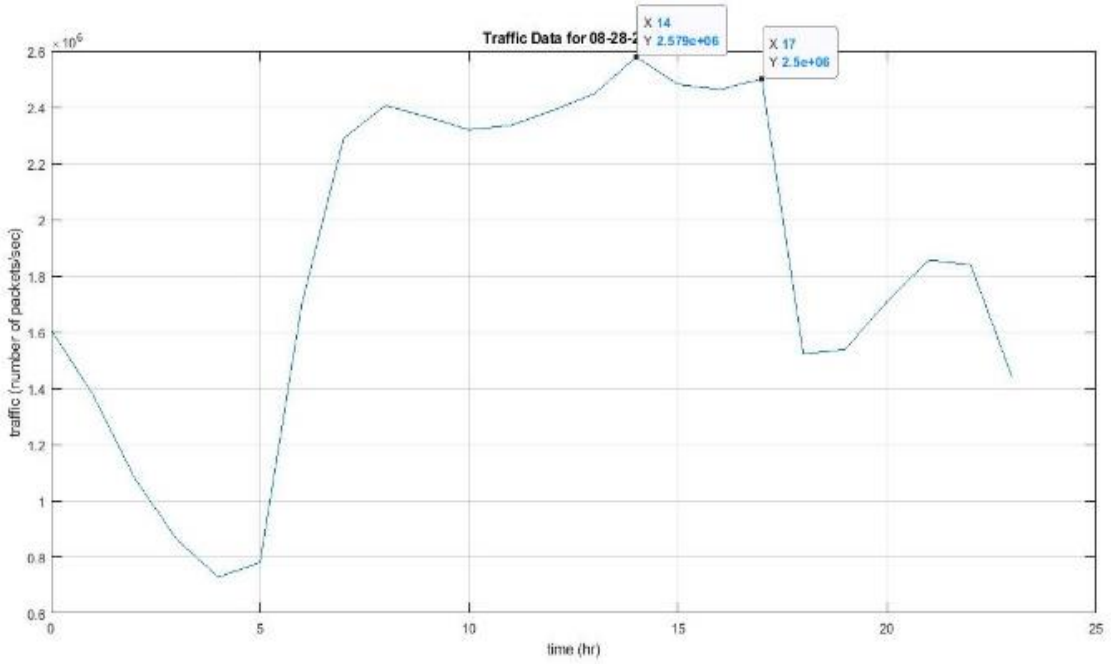

Figure 7-3: Sample Graph of Traffic Data for the day 08-28-20

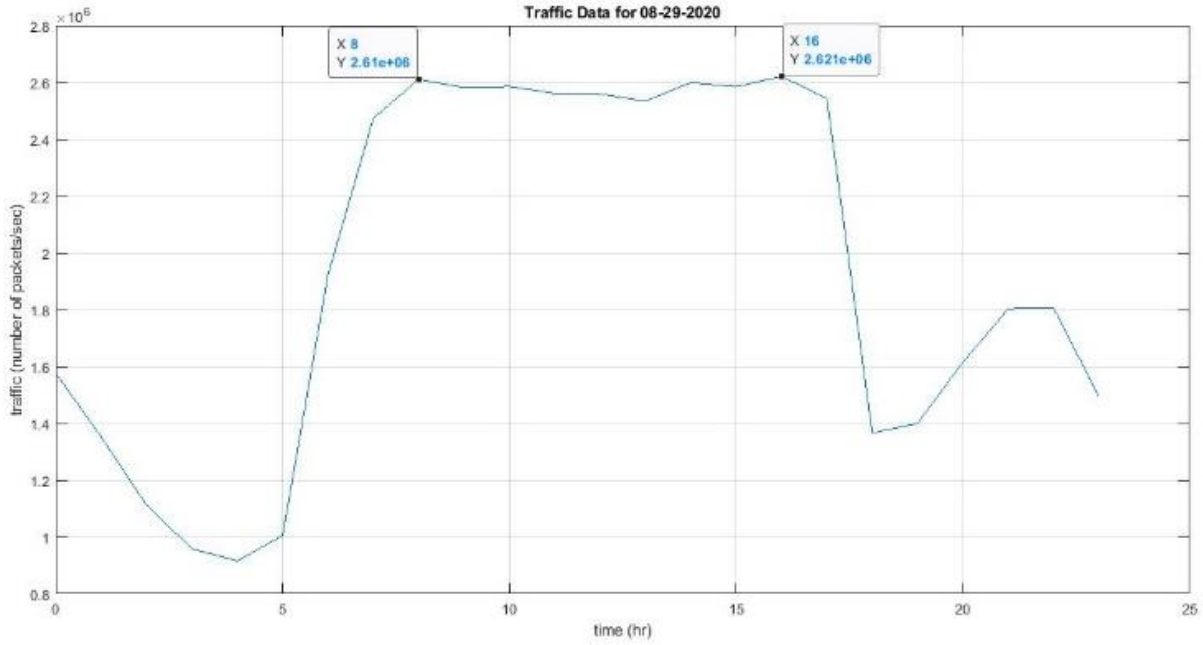

Figure 8-4: Sample Graph of Traffic Data for the day 08-29-20

From the result, the peak hour for each day is summarized in Table 4-1 below.

Table 4-1: Peak hour for each day with in the sampled period of time

Day	**Peak Hour Time during the day**
08/02/2020	22:00:00
08/03/2020	22:00:00
08/04/2020	21:30:00
08/04/2020	
08/05/2020	22:00:00
08/06/2020	22:00:00
08/07/2020	22:00:00
08/08/2020	21:00:00
08/09/2020	22:00:00
08/10/2020	21:00:00
08/11/2020	22:00:00
08/12/2020	22:00:00

08/13/2020	22:00:00
08/14/2020	22:00:00
08/15/2020	22:00:00
08/16/2020	22:00:00
08/17/2020	21:00:00
08/18/2020	21:30:00
08/18/2020	
08/19/2020	22:00:00
08/20/2020	22:00:00
08/21/2020	22:00:00
08/22/2020	21:00:00
08/23/2020	22:00:00
08/24/2020	21:00:00
08/25/2020	22:00:00
08/26/2020	21:00:00
08/27/2020	22:00:00
08/28/2020	14:00:00
08/29/2020	16:00:00
08/30/2020	17:00:00
08/31/2020	21:00:00
Average	**21:06:00**

From the above table, it can be seen that the peak hour during all the days is almost constant with the exception on the dates August 28th, 29th and 30th 2020. Out of the thirty days that have been analyzed, there is a difference only on three of them. This could be due to various reasons such as additional actions, customers might have been expecting to view specific information in those days and the announcement might have been at that time, there might have been additional events or these days might have been holidays. So the average peak hour is at 21:06hr.

4.2. Key Performance Indicators analysis at the identified peak hours

4.2.1. Analysis of KPI's with traffic analysis models

In this section we will see and compare which of the traffic analysis models best fit with the network gathered data. For this, first the Cumulative Distribution Function (CDF) of the data is plotted. Then this value is fitted against each of the different models for comparison. This helps to analyze which model best describes it. In this thesis, the selected models are Normal, Exponential, Gamma and Weibull distributions. Now let's see which of the above distributions our network traffic characteristics best fits to.

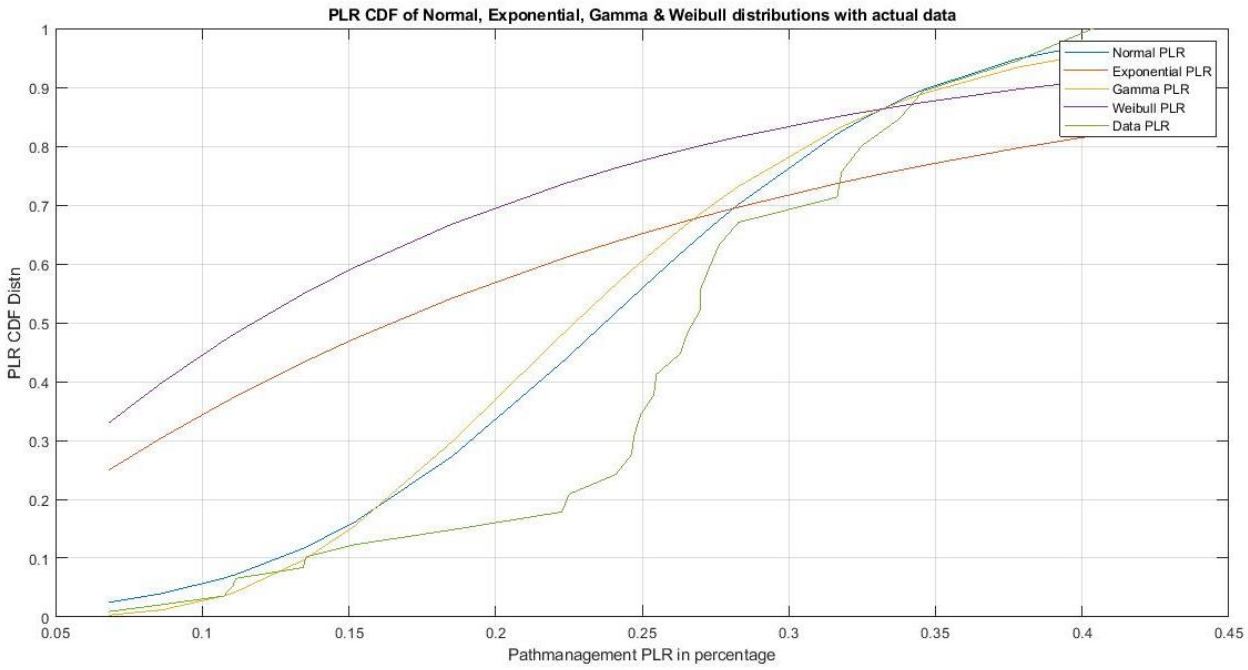

Figure 9-5: Traffic Model Analysis for Packet Loss Ratio

When we see on the above figure 4-5, the packet loss ratio analysis fitted against the four models, the CDF of the data is very close to its Normal and Gamma distributions. The Exponential and the Weibull distributions have the worst fit to the data. The Normal and Gamma distributions have the most accurate fits. Now in order to be able to choose from the two distributions i.e. from the Normal and Gamma, let's calculate the root mean square error (RMSE) for each of them. The RMSE for each of the distributions is as follows:

RMSE for Normal distribution is 0.002832526, for Gamma distribution 0.00106722, for Weibull distribution is 0.082767 and for Exponential distribution is 0.0432514. From this it can be easily seen that out of the four distributions, the one with the lowest RMSE is the Gamma distribution. Hence it can be concluded that for the PLR, the distribution that best fits it is the Gamma distribution.

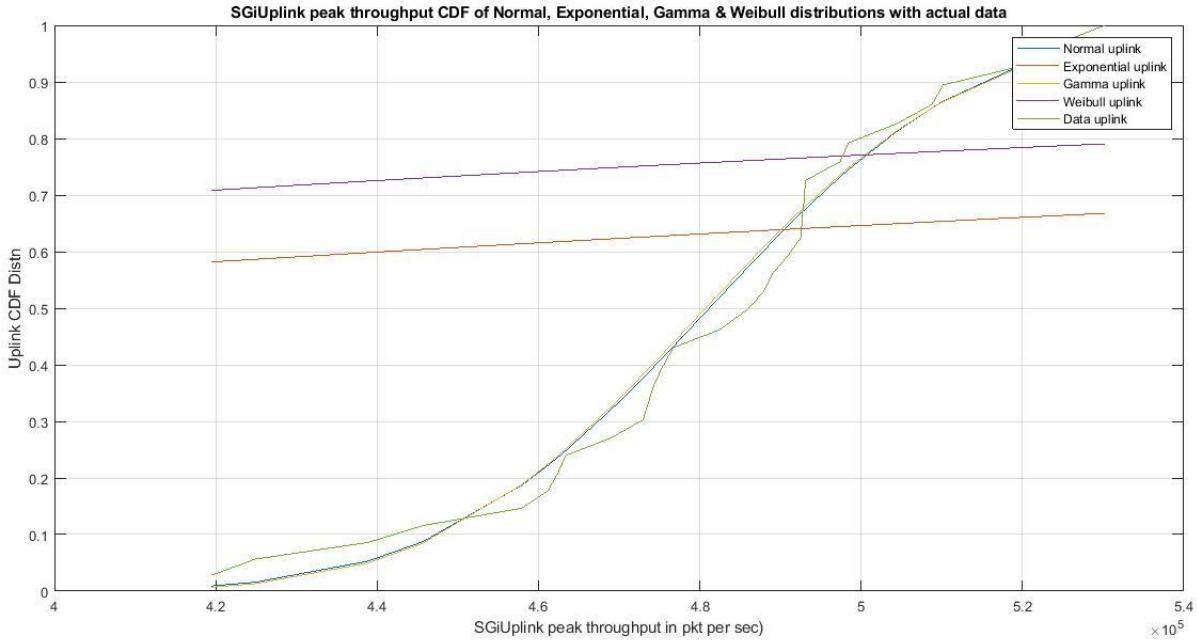

Figure 10-6: Traffic Model Analysis for Uplink Peak Throughput

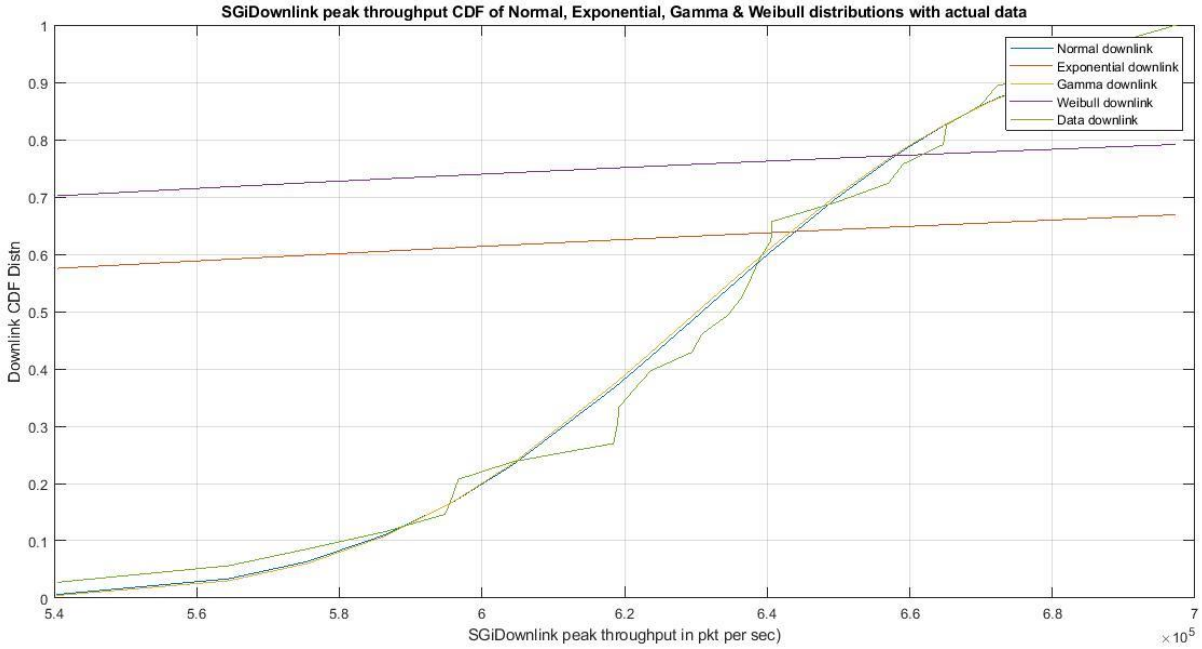

Figure 11-7: Traffic Model Analysis for Downlink Peak Throughput

When we come to see figure 4-6 and figure 4-7 for uplink and downlink throughputs respectively, again the CDF of the data more accurately fits with the Normal and Gamma distributions. Similarly the Exponential and Weibull distributions have the worst fit. In a similar way, in order to be able to choose from the two distributions i.e. from the Normal and Gamma, let's calculate the root mean square error (RMSE) for each of them. The RMSE for each of the distributions is as follows:

RMSE for uplink throughput: for Normal distribution is 0.00340, for Gamma distribution 0.00376, for Weibull distribution is 0.12216 and for Exponential distribution is 0.09944. RMSE for downlink throughput: for Normal distribution is 0.003811, for Gamma distribution 0.004124, for Weibull distribution is 0.12110 and for Exponential distribution is 0.09838. From this it can be easily seen that out of the four distributions, the one with the lowest RMSE is the Normal distribution for both downlink and uplink throughput. Hence it can be concluded that for the throughput, the distribution that best fits it is the Normal distribution.

4.2.2. Packet Loss Ratio (PLR) of data

After identifying the peak hour, the data is extracted for those at the peak hour time during the day and by using that data, the packet loss ratio is calculated by the use of the following formula (4.1):

$$Packet\ loss\ ratio = \frac{\#of\ lost\ packets}{\#of\ sent\ packets}$$

$$Packet\ loss\ ratio = \frac{\#of\ sent\ packets - \#of\ recieved\ packets}{\#of\ sent\ packets} \quad \text{…….. (4.1) [6]}$$

From the collected real time data, path management packets sent and received are used since the use of flow method focuses on the flow packets which are sent at the source and received at the destination, hence the path of the packets.

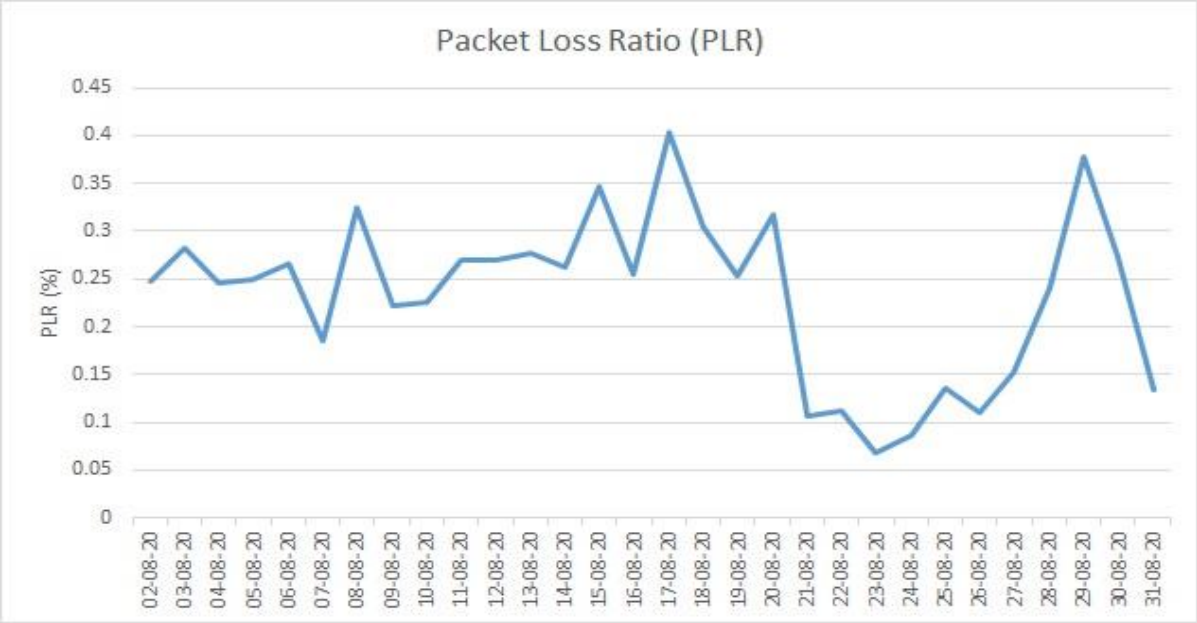

Figure 12-8: Peak Hr Packet Loss Ratio for peak hour

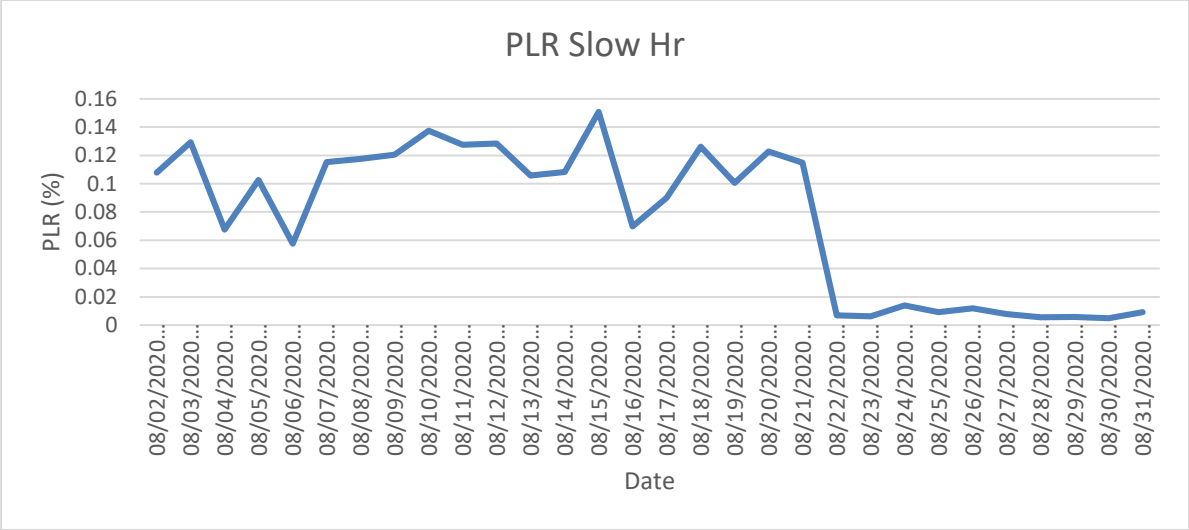

Figure 13-9: Peak Hr Packet Loss Ratio for slow hour

From the above figure 4-8, it can be seen that the maximum percentage of PLR for the peak hour is 0.4037% which is recorded on 17/08/20. Similarly from figure 4-9, it can be seen that the maximum PLR recorded for the slow hour is 0.15% on 15/08/20. From this comparison, it can be confirmed that the PLR at the peak hour is clearly higher than that of the slow hour. It is evident that when there is a high traffic rate, the number of packets that will be lost will also increase. In general, the acceptable PLR threshold is 1% for a data in a very critical transmissions. This value might increase up to 5% range when we are dealing with applications such as voice transmissions. For our case it can be seen that the PLR value is actually less than 1% value which shows that the network is working a safe range.

4.2.3. Uplink and Downlink Throughputs of data

The second KPI that will be simulated is throughput. From the collected data, the uplink and downlink throughput are as shown below.

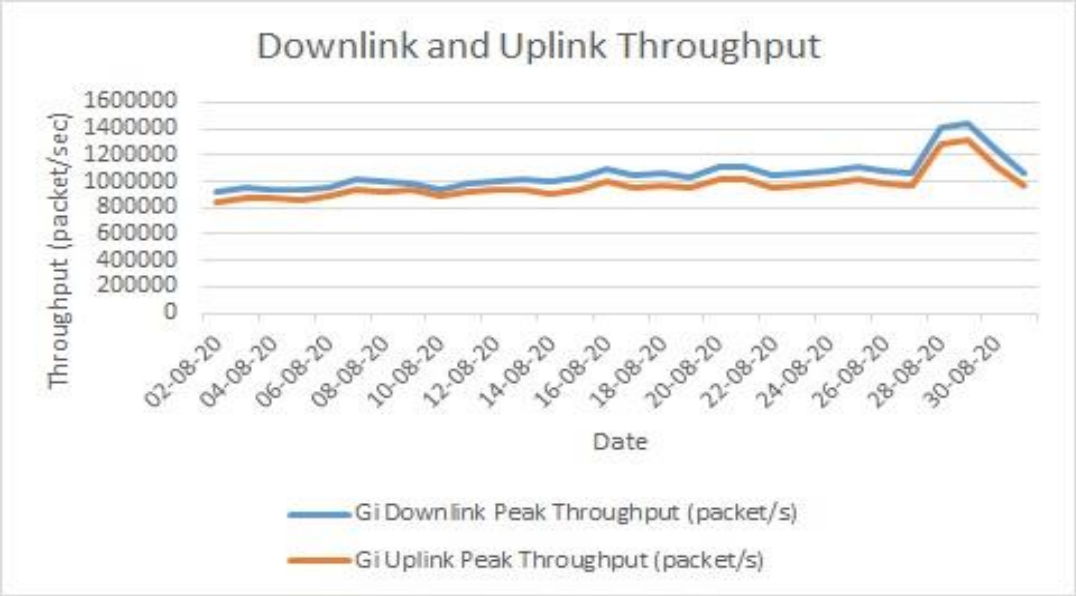

Figure 14-10: Downlink and Uplink Throughput for peak hour

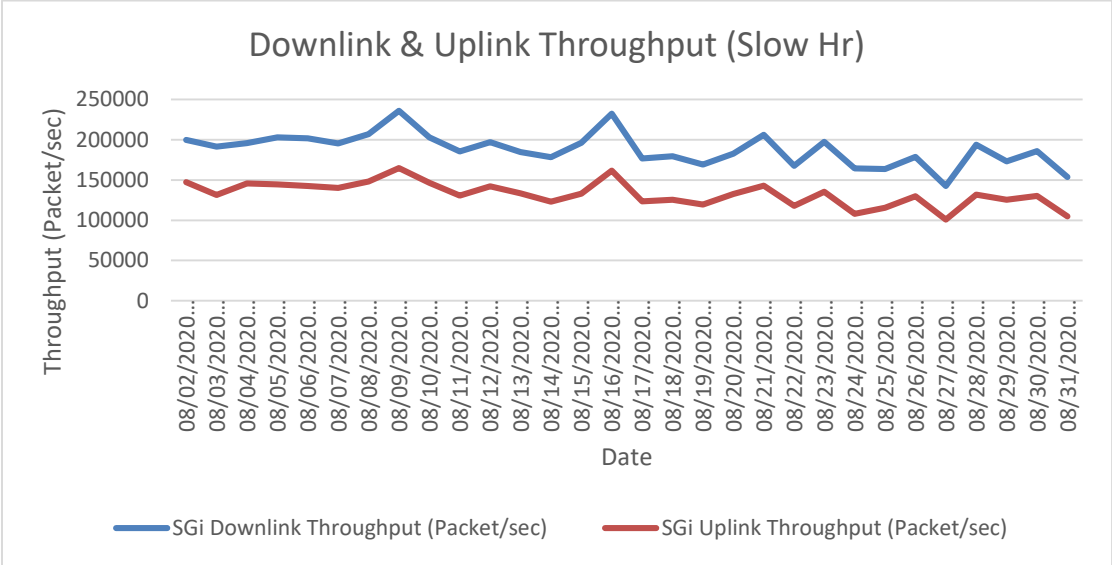

Figure 15-11: Downlink and Uplink Throughput for slow hour

From the above figure 4-10 for peak hour, it can be seen that the maximum downlink peak throughput value obtained is 1,411,292 packets/sec and the uplink throughput is 1,315,315 packets/sec on 29/08/2020. The minimum downlink and uplink throughput recorded are 924,357 packets/sec on 02/08/2020 and 845,943 packets/sec on 02/08/2020 respectively. Similarly from figure 4-11 during slow hour, it can be seen that the maximum and minimum downlink throughputs recorded are 236,152packets/sec on 08/09/20 and 142,667packets/sec on 08/27/20 respectively. The maximum and minimum uplink throughputs recorded are 164,690packets/sec on 08/09/20 and 100,678packets/sec on 08/27/20. When it comes to throughput, the standard values are dependent on the type of service used (either if it is VoIP, video or other streaming) being consumed. Other factors that affect it include the network technology, i.e. 3G or 4G, or whether a copper cable or fiber cable, are used affects the throughput achieved. In general, for a throughput 11Mbps is an acceptable range. So for our case it can be compared that for the downlink throughput, the maximum and minimum number of packets per second being transmitted exceed the slow hour by 16% and 15.4% respectively. Similarly for the uplink throughput, the maximum and minimum number of packets being transmitted exceed the slow hour by 12.5% and 12% respectively. When we see closely the graphs of the downlink and uplink throughputs, they are very symmetric. This could be due to the video streaming process, which creates the major asymmetry on network transmission, might be saved offline instead of the users streaming it online. This reduces the asymmetry to the greater extent and could be the reason why it shows a very symmetrical downlink and uplink throughput graphs.

Chapter Five

5. Conclusion, Recommendation and Future Work

5.1. Conclusion

In this thesis basic traffic analysis is done at the peak hour by the use of raw data collected from the current internet service provider, Ethio Telecom's network. First identification of peak hour is done because that is the time where there is a lot of traffic in the transmission and this enables to see more activity. After identification of peak hour, two KPI's are analyzed in order to give a more appropriate and meaningful analysis.

The average peak hour during the day is found to be at 21:06hr over a sampled data for a period of a month every day. The number of internet users during the day time increases almost linearly starting from 05:00 hr. and especially after 18:00 hr. it increases drastically. Then on average at 21:06hr it reaches its peak and after that it starts to decline again linearly. This can be clearly seen from the number of packets vs time graph. In order to be able to see the difference, the slow hour is also identified and the packet loss ratio and throughput values are compared with that of at the peak hour. The slowest hour during is constant which is at 04:00hr.

After identification of peak hour, the data is analyzed to see which of the traffic analysis models it best fits too. For this the CDF of the raw data is fitted against the CDF of Normal, Exponential, Gamma and Weibull Distributions. From this, it is seen that the PLR best fits with the Gamma Distribution model with a minimum RMSE of 0.00106722 and both the uplink and downlink throughputs best fit with the Normal Distribution model with the minimum RMSE of 0.00340 and 0.003811 respectively.

Then after modeling is done, the KPI values are analyzed and further the peak hour values are compared with the slow hour. The KPI's analyzed in this thesis are Packet Loss Ratio (PLR) and throughput. Analysis of PLR enables us to see out of the packets that are sent from the source to destination, the number of packets being lost is within the acceptable maximum range. In addition to that, the maximum packet loss ratio recorded during the slow hour is 0.15%. This confirms that there is more loss during the busiest network traffic.

As stated in previous chapters the standard maximum acceptable range of PLR is 1%, though 1% to 2% is acceptable in most cases. In this thesis the maximum Packet Loss Ratio recorded is 0.4037% at the peak hour. Therefore, this shows that when the network is transmitting messages, there is no significant loss of packets which disrupts normal communication. This is one of the basic requirements in maintaining the Quality of Service towards customers.

The second parameter taken for analysis is throughput. Studying the throughput enables us to see how many packets are being transmitted every day. In this way, we can identify if there is too much traffic congestion and identify and take preventive measures easily before the problem occurs. For a better analysis, again the throughputs during the peak hour and slow hour are compared. From the result it is seen that that the maximum and minimum number of packets per second being transmitted exceed the slow hour by 16% and 15.4% respectively. Similarly for the uplink throughput, the maximum and minimum number of packets being transmitted exceed the slow hour by 12.5% and 12% respectively. In addition when the graphs of the downlink and uplink throughputs are seen, they are very symmetrical. This could be due users could be downloading video offline rather than streaming it online. This reduces the asymmetry greatly because video is one of the highest bandwidth consuming applications.

5.2. Recommendation

The demand of telecom service users will always be increasing. Instead of holding the telecom service by only the government, sharing it with private companies is one of the huge steps the government is taking recently which already reliefs it from the congestion it is already in. Since Ethio-Telecom is already distributed throughout the country ahead of the new non-governmental telecom service providers, it might have more credibility until the others are well known by each customer. But these competitions will force Ethio Telecom to upgrade its current infrastructure because that is its significant drawback.

Improvement in the current Ethio Telecom infrastructure can be made by:

- Use of tagged priority routing for labeled packets facilitates network routing decisions especially for very critical transmissions. These can be in either of the following two ways
    - The first can be identifying data flows that are considered sensitive such as video applications and voice applications and give priority for those for a better communication.
    - Second one allocation of a bandwidth dynamically for the prioritized applications will come in handy especially when a customer is looking to use simultaneous applications at once.
- Use of selective discarding of packets becomes handy when the network traffic becomes so congested more than its capacity.

5.3. Future Work

In this thesis a great deal of time and effort is spent in order to achieve the objective. So for the future the following works are recommended to be done:

- Analysis of traffic for voice transmissions
- Analysis of the network traffic outside of Addis Ababa in the regional cities

In order to improve the Quality of Experience of customers, traffic analysis study will always be necessary from time to time. It cannot be relied up on a study done once and expect it to give accurate remedial solutions to existing network congestion problems.

References

[1] Xuan Zhouy et al., "*The Predictability of Cellular Networks Traffic*", Laboratory for Cognitive Radio and Green Communications Department. of Information Science and Electronic Engineering, Zhejiang University, October 2012

[2] F. Ricciato et al., "*Traffic monitoring and analysis in 3G networks: lessons learned from the METAWIN project*" Article in "e & I, Elektrotechnik und Informationstechnik" August 2006

[3] A. Ozovehe et al., "*Busy Hour Traffic Congestion Analysis in Mobile Macrocells*", Nigerian Journal of Technology. (NIJOTECH) Vol. 36, No. 4, October 2017, pp. 1265 – 1270

[4] Myung-Sup Kim, Hun-Jeong Kang, Seong-Cheol Hong, Seung-Hwa Chung, and James W. Hong, *"A Flow-based Method for Abnormal Network Traffic Detection"* Department of Computer Science and Engineering, POSTECH, Gyeongsangbuk-do, South Korea

[5] Anna Sperotto, RaminSadre, and Aiko Pras, "*Anomaly Characterization in Flow-Based Traffic Time Series"* University of Twente, Centre for Telematics and Information Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, P.O. Box 217, 7500 AE Enschede, The Netherlands

[6] Meheretu Daka, "*Traffic Analysis of IP Core Networks: The Case of Ethio-Telecom*", Master's Thesis, School of Electrical. & Computer Engineering, Addis Ababa Institute of Technology, Addis Ababa, Ethiopia, 2017

[7] M. Vaezi and Y. Zhang,Cloud, "*Mobile Networks, Wireless Networks"* Southeast University, Springer International Publishing AG 2017, pp. 67-86

[8] Rakibul Hasan et al., "*Handover Management in GSM Cellular System",* Department of Electrical, Electronics & Telecommunication Engineering of Dhaka International University, Bangladesh

[9] GSM Theory, "*Network Management Subsystem*" Retrieved March 2021 *https://sites.google.com/site/gsmtheory/5-network-management-subsystem-nms*

[10] M. Uma, Ph.D and G. Padmavathi, Ph.D, "*An Efficient Network Traffic Monitoring for Wireless Networks"* Research Scholar and Professor, Department of Computer Science Avinashilingam Institute for Home Science and Higher Education for Women, Volume 53– No.9, September 2012

[11] Alisha Cecil, "*A Summary of Network Traffic Monitoring and Analysis Techniques*", Computer Science and Engineering, Washington University, St. Louis, Washington, USA 2017

[12] NetFort Technologies, "*Flow Analysis Versus Packet Analysis. What Should You Choose?*", Retrieved February 2021, *http://www.netfort.com*

[13] Michiel Uithol, Vincent van Kooten, Aiko Pras, Pieter-Tjerk de Boer, "*Network monitoring based on flow measurement techniques*", University of Twente, 2011

[14] Balakrishnan Chandrasekaran, "*Survey of Network Traffic Models",* Computer Science and Engineering, Washington University, St. Louis, Washington, USA 2017

[15] Muhammad Asad Arfeen, K. Pawlikowski, D. McNickle, A. Willig, "*The Role of the Weibull Distribution in Internet Traffic Modeling"* University of Canterbury Christchurch, New Zealand.

[16] Sidney Tyrrell, "*The Poisson Distribution*", from Mathematics Education Innovation, AS Stats book Z2, 5th Draft Ch 8, Coventry University, UK, July 2008

[17] Rajalakshmi Krishnamurthi, Prakash Kumar, and Hima M. Bindu Jaypee, "*Solving Base Station Subsystem Assignment Problem in Mobile Communication Networks Using Hybridized Heuristic Algorithm"* Institute of Information Technology, Noida, Uttar Pradesh, India

[18] TechDifferences, "*Difference Between Circuit Switching and Packet Switching*", Retrieved February 2021, *https://techdifferences.com/difference-between-circuit-switching-and-packet-switching.html*

[19] Wikipedia, the free encyclopedia, "Packet Switching", Retrieved March 2021, *https://en.wikipedia.org/wiki/Packet_switching*

[20] Johnathan Mun, "*Understanding and Choosing the Right Probability Distributions*" *Advanced Analytical Models: Over 800 Models and 300 Applications from the Basel II Accord to Wall Street and Beyond,* pp. 899-917, 2008

[21] Mohammed Salem, "*Gn Interface and Gp Interface*", Posted October 2018, Retrieved March 2021, *https://mobilepacketcore.com/gn-gp-interface/*

[22] Telecomabc, "*GPRS*", Retrieved February 2021, *http://www.telecomabc.com/g/gprs.html*

[23] Wikipedia, the free encyclopedia, "*GPRS Core Network*", Retrieved March 2021 *https://en.wikipedia.org/wiki/GPRS_core_network*

[24] Telecomabc, "*SGSN*", Retrieved February 2021, *http://www.telecomabc.com/s/sgsn.html*

[25] Dan Jones, "*Gateway GPRS Support Node (GGSN)*", Posted September 2010, Retrieved March 2021, *https://www .lightreading.com/document.asp ?doc_id=679949*

Appendix

Publishable Material