



Addis Ababa University

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

**Spatial and Temporal Traffic Distribution Modeling Using Statistical
Techniques: The Case of UMTS Network in Addis Ababa, Ethiopia**

By

Woinshet Taddele Ejigu

Advisor

Dr. -Ing. Dereje Hailemariam

A Thesis Submitted to School of Electrical and Computer Engineering in Partial
Fulfillment of the Requirements for the Degree of Masters of Science in
Telecommunications Engineering

December 2019

Addis Ababa, Ethiopia

Addis Ababa University
Addis Ababa Institute of Technology
School of Electrical and Computer Engineering

By

Woinshet Taddele Ejigu

**Spatial and Temporal Traffic Distribution Modeling Using Sta-
tistical Techniques: The Case of UMTS Network in Addis
Ababa, Ethiopia**

Approval by Board of Examiners

_____	_____	_____
Chairman, school Graduate committee	Signature	Date
	<u>Committee</u>	
<u>Dr. -Ing. Dereje Hailemariam</u>	_____	_____
Advisor	Signature	Date
_____	_____	_____
Internal Examiner	Signature	Date
_____	_____	_____
External Examiner	Signature	Date

Abstract

In cellular mobile networks due to the emerge of different application services, the data traffic generated by these application services is increasing. Moreover, the traffic consumption pattern is changing in time and space. Understanding and characterizing the traffic demands and dynamics associated with different mobile services (e.g., knowing when and where users use dominant application services) is instrumental for operators. By doing so, operators can improve application services usability, optimizing network service quality and use it as an input for technical and business strategies. In the case of ethio telecom, the sole telecom service provider in Ethiopia, knowledge of the application services distribution is not a known practice.

Therefore, the main purpose of this thesis work is modeling the spatial and temporal traffic distribution of observed application services for Addis Ababa Universal Mobile Telecommunication System (UMTS) network. In this regard, to study a temporal and spatial distribution analysis of the traffic density (the traffic load per unit area) for selected mobile application services, data traffic is collected from 739 bases Stations (BSs).

To model the temporal distribution of the selected application services, which are Streaming, Social Networks and Web browsing, four candidate models: Normal, Lognormal, Weibull and Stable are used. Based on maximum log-likelihood and probability plot evaluation criteria, out of four models, a Stable distribution modeling best fit for this particular application services. Similarly, to model spatial distribution, out of the candidate four models, Weibull distribution model best fits for these particular services.

Keywords— *UMTS; Modeling; Application Services; Temporal; Spatial; Statistical Modelling; Weibull; Normal; Lognormal and Stable distribution.*

Declaration

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

Woinshet Taddele Ejigu

Name

Signature

Place: Addis Ababa

Date of Submission: _____

This thesis has been submitted for examination with my approval as a university advisor.

Dr. -Ing. Dereje Hailemariam

Advisor

Signature

Acknowledgment

First of all, I am thankful to GOD that giving me the strength to accomplish my thesis work.

I would like to express my greatest gratitude to Dr. -Ing. Dereje Hailemariam, for his supervision, direction, and unreserved feedbacks. I am very thankful for the weekly meetings and all the advice, time, and finally, your work ethics will be remembered and took as a lesson into my professional life.

I would like to thank my company, ethio telecom, for giving me the chance to study and being part of this successful program and Mr. Dawit Kefale for providing me the required data in doing the thesis.

Finally, I would like to express my deepest gratefulness to my dearest families, my husband and my children for always encouraging me to do better, for your love and priceless cares throughout my postgraduate program.

Table of Contents

Abstract.....	i
Declaration	ii
Acknowledgment.....	iii
Table of Contents.....	iv
List of Figures	vii
List of Tables	viii
Abbreviations and Acronyms.....	ix
Chapter 1: Introduction.....	1
1.1 Statement of the problem.....	3
1.2 Objective.....	3
1.2.1 General objective.....	3
1.2.2 Specific Objectives	4
1.3 Related literature review.....	4
1.4 Methodology.....	6
1.5 Scope and limitation	7
1.5.1 Scope of the thesis.....	7
1.5.2 Limitations of the thesis.....	8
1.6 Contribution of the thesis	8
1.7 Thesis organization.....	9

Chapter 2: Background.....	10
2.1 UMTS network.....	10
2.2 UMTS architecture	11
2.2.1 Universal terrestrial radio access network.....	11
2.2.2 Core network.....	12
2.3 UMTS services and bearer architecture	14
2.4 UMTS QoS basic classes.....	15
Chapter 3: Traffic Distribution Modeling.....	18
3.1 Statistical traffic distribution modeling	18
3.2 Important statistical functions.....	19
3.3 Commonly used continuous probability distributions	21
3.4 Distribution parameters	25
3.5 Parameter Estimation	26
Chapter 4: UMTS Application Services Traffic Distribution Modeling.....	28
4.1 System model.....	28
4.2 Data collection and preparation.....	29
4.3 Feature selection.....	29
4.4 Model Selection and Parameter Estimation	30
4.4.1 Traffic Distribution Modeling in Space	30
4.4.2 Traffic Distribution Modeling in Time	39

4.5 Result and Discussions	46
Chapter 5: Conclusions and Future Work.....	50
5.1 Conclusions.....	50
5.2 Future work	51
References	52

List of Figures

Figure 1 Methodology used.....	7
Figure 2 UMTS network architecture [15]	11
Figure 3 UMTS QoS Classes [14].....	15
Figure 4 System model	28
Figure 5 Mobile application services traffic demand ratio in Addis Ababa.....	30
Figure 6 Streaming spatial traffic distribution.....	32
Figure 7 Web browsing spatial traffic distribution	32
Figure 8 Social networks spatial traffic distribution	33
Figure 9 Streaming traffic PDF and CDF plot.....	34
Figure 10 Social networks traffic PDF and CDF plot.....	35
Figure 11 Web browsing traffic PDF and CDF plots	36
Figure 12 Probability plot of Streaming traffic	38
Figure 13 Probability plot of Social Networks traffic.....	38
Figure 14 Probability plot of Web Browsing traffic	39
Figure 15 Social networks temporal traffic trend	40
Figure 16 Streaming temporal traffic trend.....	40
Figure 17 Web browsing temporal traffic trend	41
Figure 18 Streaming real traffic with candidate distribution models.....	41
Figure 19 Web browsing real traffic with candidate distribution models.....	42
Figure 20 Social networks real traffic with candidate distribution models.....	42
Figure 21 Streaming traffic probability plot.....	44
Figure 22 Social networks traffic probability plot	45
Figure 23 Web browsing traffic probability plot	45

List of Tables

Table 1 Fundamental characteristics of the UMTS QoS classes	17
Table 2 Sample raw data	29
Table 3 List of candidate distributions and their fit to streaming traffic data.....	37
Table 4 List of candidate distributions and their fit to web browsing traffic data	37
Table 5 List of some distributions and their fit to SNS traffic data.....	37
Table 6 List of candidate distributions and their fit to streaming traffic data.....	43
Table 7 list of candidate distributions and their fit to SNS traffic data	43
Table 8 List of candidate distributions and their fit to Web browsing traffic data	44
Table 9 Estimated Parameters and standard error for the temporal distribution	47
Table 10 Estimated parameters and standard error for the spatial distribution.....	49

Abbreviations and Acronyms

1G	First Generation
2G	Second Generation
3G	Third Generation
4G	Fourth Generation
AA	Addis Ababa
ARMA	Auto Regression and Moving Average
BSC	Base Station Controller
CC	Call Control
CN	Core Network
CDF	Cumulative Density Function
CS	Circuit Switched
CSV	Comma Separated Values
E2E	End to End
ETSI	European Telecommunications Standards Institute
GGSN	Gateway GPRS Support Node
GMSC	Gateway MSC
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HLR	Home Location Register
HSPA	High Speed Packet Access
IEEE	Institute of Electrical and Electronics Engineers
IM	Instant Messaging
IMS	IP Multimedia Subsystem

IP	Internet Protocol
ISP	Internet Service Provider
MATLAB	Matrix Laboratory
MB	Megabyte
Mbps	Megabits per second
ME	Mobile Equipment
MLE	Maximum Likelihood Estimation
MSC	Mobile Switching Center
PDF	Probability Density Function
PDN	Packet Data Network
PDP	Packet Data Protocol
PS	Packet Switched
QoS	Quality of Service
RRC	Radio Resource Control
RAN	Radio Access Network
RNC	Radio Network Controller
RLC	Radio Link Control
SGSN	Serving GPRS Support Node
SM	Session Management
SMS	Short Message Service
SPPP	Spatial Poisson Point Process
TE	Terminal Equipment
UE	User Equipment
UMTS	Universal Mobile Telecommunications System

USIM	UMTS Subscriber Identity Module
UTRAN	UMTS Terrestrial RAN
VLR	Visitor Location Register

Chapter 1: Introduction

According to Cisco's forecast on global Internet Protocol (IP) traffic growth, mobile data is experiencing a phenomenal rise with an estimated Compound Annual Growth Rate of 46 percent from 2016 to 2021[1]. There are several reasons for the growth of this mobile data. To mention some of them: advancements in cellular data network technologies, increasing popularity of smartphones and the emergence of diversified mobile applications service [2]. To transport the rapid growing data demand and diverse expectations of mobile data users, service providers need to optimize their radio spectrum resources and evolve their infrastructures [3].

The goal of a mobile network operator is to offer the customer an assured end-to-end Quality of Service, with a variety of service levels and predictable service response. Such a goal requires a set of intra-domain (radio access and core networks) and inter-domain agreements, which is consistent along the end-to-end 'chain'. For that purpose, service providers need to have the capability to measure the service performance as experienced by the customer, guarantee their QoS expectations and succeed in the launch of new mobile services [4][5].

Characterizing the traffic dynamics associated to different mobile services, e.g., knowing dominant application services and when and where users use these dominant application services, is instrumental in improving application services usability, optimizing network service quality and perform proper network dimensioning [6].

Knowledge of the temporal-spatial (i.e., the when and where) traffic distribution in cellular networks mostly concentrates on the statistical distribution of users, base stations (BSc) and traffic.

The common approaches that are explained in [7] are summarized here:

1. Spatial Poisson Point Process distribution model: The spatial point process is a random pattern of points in d-dimensional space and it will be SPPP when the pattern takes poisson distribution. Using the approach differently enables the modeling to capture both uniform and non-uniform distribution of a system.
2. Lognormal distribution model: In probability theory, a log-normal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Models that use this distribution in cellular technology can be applied to generate traffic for simulation, to estimate user distribution in a network and also to determine traffic distribution in a coverage area.
3. Exponential distribution model: The exponential distribution is used to model the time between the occurrence of events in an interval of time, or the distance between events in space.
4. Gaussian distribution model: in this model, the system under study has a normal distribution. The system can be analyzed in d-dimensions and the model will be called the d-dimensional Gaussian distribution model.

Traffic distribution models that capture or describe the statistical characteristics of the actual traffic on the network are preferred for this study.

1.1 Statement of the problem

Core tasks of any network operator include meeting consumers' expectations, in application services quality, and achieving optimum loading of the network. Improving application services quality in a network requires proper resource and traffic management [8].

In the case of ethio telecom's Universal Mobile Telecommunication System (UMTS) network deployed in the city of Addis Ababa, Ethiopia, the demand for various application services is increasing and hence, the operator is continuously facing with bandwidth limitation and service quality degradation [6]. Addressing these challenges faced by the operator, among others, require knowledge of the traffic distribution at application services level.

The aim of the thesis is to use statistical modeling approaches to understand the traffic demand at an application service granularity. The modeling is done both in spatial and temporal domains. Such knowledge will contribute to a variety of factors like understanding the consumption pattern (or behaviors) of users, improving network bandwidth utilization, optimizing the network, improving quality of services offered to users, and, ultimately increasing the company's revenue. To the best of our knowledge, the company does not have the practice of modeling the distribution of application services.

1.2 Objective

1.2.1 General objective

The ultimate goal of this thesis work is to model the spatial and temporal traffic distribution of UMTS mobile traffic in the city of Addis Ababa using statistical distribution modeling techniques at the selected application service granularity.

1.2.2 Specific Objectives

To achieve the above mentioned general objective, the following specific objectives are set:

- Review of related papers, journals, and books.
- Study about statistical modeling.
- Study about mobile application services.
- Analyse how the collected application services data temporal patterns behave.
- Study how the collected application services data over the spatial domain visualized.
- Evaluate the candidate distribution models.

1.3 Related literature review

For modeling, the spatial and temporal distribution of the traffic in mobile cellular network literatures have used different approaches. Some papers make the modeling either in time, space or both in time and space domain. Regarding the analysis of the traffic pattern, modeling techniques, methodology followed, used to model cellular network traffic distribution literature written are explained as follows. In [7] application-level cellular networks traffic data that are: IM, Web browsing and Video are selected. The study proposes α stable modeled property in the temporal domain. In this research, a similar approach as the temporal modeling followed and verified. But the difference is that they provide prediction framework as well as in spatial domain Sparsity & Dictionary Learning" module imposed. The authors in [8] verified that lognormal distribution for spatial traffic modeling which can reflect the spatial characteristic of traffic from real data and a sinusoid superposition model is proposed for describing the temporal traffic variation.

A three weeks' data recording with a 5 minutes period collected in a $6\text{km} \times 2.5\text{km}$ area used. Compute the probability density function of the traffic volume for the area and then the models compared. In this thesis the traffic distribution is modeled spatially followed similar methodology but the difference is that this thesis selected three types of application services rather than considering the whole cellular network. [9] Investigates the traffic behavior of specific services across time and space that unveils a strong heterogeneity in the demand for different mobile services, both in time and space. A method for selection of the services categorized and ranked on the relative traffic volume they generate that covers a large fraction (over 60%) of the overall network traffic. The temporal dynamics of different mobile service patterns observed between weekends and working days. Similarly, this research follows the same method for the selection of the services based on the relative traffic volume they generate. A study in [10] models the spatial traffic patterns for three typical regions by taking the average traffic within the selected study area. The spatial traffic distributions of typical areas show that the parameters of the model of the three regions are different. In dense urban, users are highly accumulated and traffics are highly fluctuant over the space, on the contrary, in suburban users are distributed relatively uniform and traffics are smoother over the space.

Authors in [11] study traffic patterns and predict per-user traffic demand utilizing application usage. The exploration of patterns of large-scale users done by feature extraction, feature reduction, and pattern identification. Predicting per-user traffic demand in application-level using their historical traffic sequences is based on a hybrid model which consists of the wavelet transform and ARMA with fitted mathematical model.

Spatial modeling of the traffic density in cellular networks is presented in [12] uses statistical distribution modeling techniques and modeled by log-normal mixtures.

Performed analysis for the traffic profiling is done by divided target area is into a square grid and each pixel (square) in the grid is assigned to the traffic density of its nearest BS (cell), and thus a matrix of traffic density obtained to understand the real spatial distribution of traffic demand.

Both in time and space domain the paper in [1] model UMTS AA traffic data using Frequency-based methods. Fast Fourier Transform (FFT) is used for modeling the traffic in a time domain and Two Dimensional Discrete Cosine Transform (2D-DCT) and its inverse is used to model the real traffic in space domain.

1.4 Methodology

The methodologies followed to achieve the general and specific objective of this thesis,

- Related literatures in the field of cellular network data traffic distribution modeling are reviewed.
- Review ethio telecom's UMTS network architecture and traffic monitoring system.
- Collected the input datasets from ethio telecom's UMTS network through SMART CARE server in .xlsx format.
- Apply statistical distribution models for selected input datasets.
- Implement distribution models for selected input datasets.
- MatLab software tool was chosen to model the distribution of the collected data, and Microsoft Excel is used to preprocessing of the data.

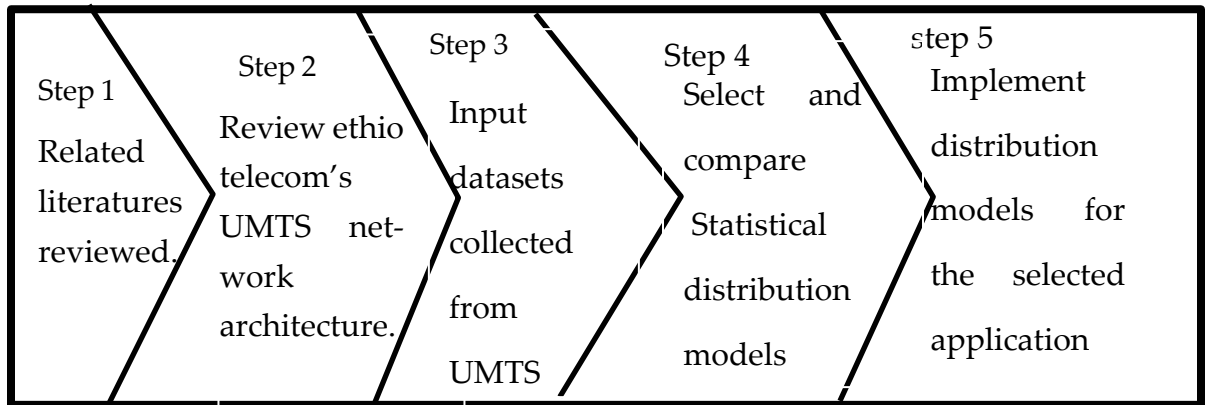


Figure 1 Methodology used

1.5 Scope and limitation

1.5.1 Scope of the thesis

As described above traffic distribution modeling that characterizes the traffic dynamics associated with different mobile application services is instrumental in improving application services usability, optimizing network service. This thesis focuses to model the spatial and temporal traffic distribution of top-three application services that have the highest traffic demand. The conducting case study on ethio telecom's Addis Ababa UMTS network because it covers the wider range and it is most usable in Addis Ababa mobile communication services. UMTS mobile technology offers both voice and data services but due to the number of subscribers of internet services greatly increased and the application services are the main source of traffic this thesis focuses only on application services.

1.5.2 Limitations of the thesis

Due to a large number of services available in different categories, it is not possible to model the distribution for each service that exists in the UMTS network. In addition to this, during data collection, the dataset has sites that miss full traffic information. Hence, the modeling is performed for selected weeks by having full information about the data. The result for temporal trend analysis is only valid for the normal conditions that are a normal week and weekend without any holidays and/or events that could change the demanded traffic.

1.6 Contribution of the thesis

Modeling the traffic for UMTS data service in Addis was done in [1]. However, and as of our knowledge, only the referenced work-study focuses on modeling the distribution of the selected application services for Addis Ababa UMTS network. The main contribution of this thesis work is to represent the observed data by providing visual aids and statistical models that highlight the distribution and behaviors associated with application services. This helps mobile operators to deal with each application services Qos provided to their customers and as an input for further analysis.

1.7 Thesis organization

The rest of the thesis is organized as follows. Chapter two provides a brief description of the fundamental concepts of UMTS and its architectures. The quality of Service classes and popular applications are briefly mentioned. Chapter three discussed the characteristics and mathematical background of statistical models used in the rest of the paper. Chapter four describes the system model used, explanation about the dataset used, procedures in developing the distribution model followed by the referenced work and finally discusses the obtained results. At last, Chapter five presents the conclusion of the work and recommends ideas for future work.

Chapter 2: Background

This section provides an overview of the UMTS network and its bearer service architecture. UMTS is developed in terms of various applications, services and bandwidth that allows provisioning of QoS for various applications and services (e.g. guaranteed error rates and delays), support of QoS differentiation (e.g. between delay-sensitive real-time services and best-effort) [13]. Without a clear understanding of the traffic and the applications that might be using the network, QoS guarantees cannot be provided.

2.1 UMTS network

The development of UMTS network is mainly based on the GSM (Global System for Mobile communication) /GPRS (General Packet Radio System) network that can support different types of applications such as data, voice, and video. . GPRS is a telecommunications technology that enables high speed wireless internet and data communications in GSM mobile networks. The 1st Generation of mobile Telecommunication Networks offered voice service, exclusively. In 2G, support for fax and some data services such as a mobile-adapted web browsing functionality, Short Message Service and Multimedia Message Service were added [6]. 3G UMTS brings an evolution in terms of greatly enhanced capacity, data rates and new service capabilities from 2G systems. Compared to GSM, UMTS offers a much higher throughput in the Megabit range as compared to kilobit and a much larger variety of services [14]. It provides global mobility and high-speed transmissions with a wide range of services including voice telephony, messaging, images, video, Internet access, and broadband data.

2.2 UMTS architecture

UMTS network comprises three main parts: User Equipment (UE), UMTS Terrestrial Radio Access Network (UTRAN), and Core Network (CN). As shown in Figure 2.1, the UMTS network has two major divisions: UTRAN and CN. And, the other one, User Equipment (UE) is also included in the architecture.

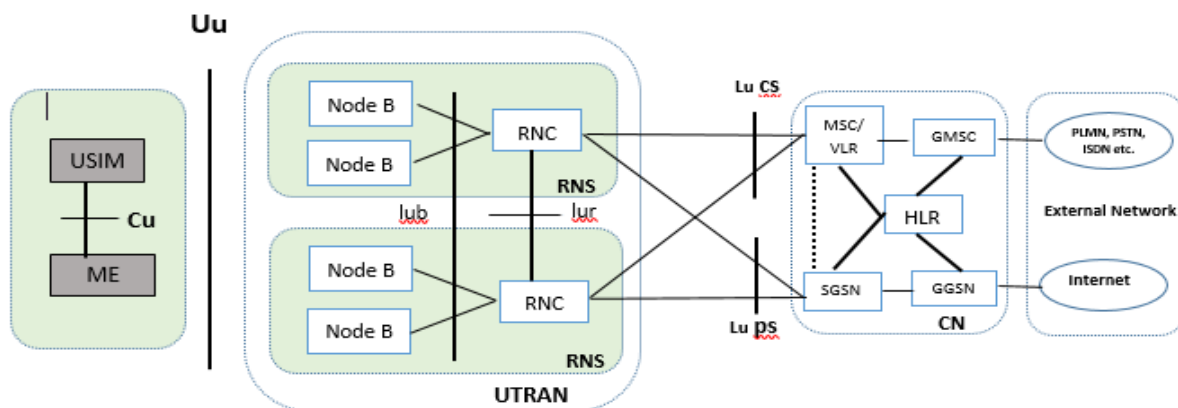


Figure 2 UMTS network architecture [15]

User Equipment (UE) consists of two parts: The Mobile Equipment (ME), which is a radio terminal handling the communications on the Uu interface (radio interface), and the UMTS Subscriber Identity Module (USIM), which is a smart card that contains the subscriber's confidential data such as identity, authentication algorithm, and subscription information. USIM and ME communicate over the internal Cu interface. The UE accesses the fixed network via the Uu radio interface.

2.2.1 Universal terrestrial radio access network

The radio access network of UMTS is called UTRAN, which is responsible for radio resource, data, signaling traffic, an exchange between UE and CN. The main task of

UTRAN is to create and maintain Radio Access Bearers (RAB) for communication between User Equipment (UE) and the CN [16]. UTRAN is a combination of two parts: Radio Network Controller and Node-B, both of them are explained below.

Radio Network Controller (RNC): It is the switching and controlling element of the UTRAN, it performs various tasks and controls all radio resources within RNC. It is responsible for call admission control, radio resource management, data transmission, radio bearer setup and release, code allocation, power control, packet scheduling.

Node-B: It is similar to the Base Transceiver System (BTS) in the GSM network. As like as BTS, Node-B has the capability for both transmission and reception. It is responsible for allocating the channel and radio resources to interact with the user equipment. It also performs error detection and correction, modulation and handover management. Node-B is connected with RNC and User Equipment via Iub and Uu interfaces respectively. Both Node-B and UE verify the status of the connection and level of interference and send the update to connected RNC.

2.2.2 Core network

The UMTS Core Network(CN) can be seen as the basic platform for all communication services provided to the UMTS subscribers. The CN interacts with external networks for different types of communication such as data, voice, and video, etc. The basic communication services include switching of circuit-switched calls and routing of packet-switched data [16]. Packet Switched elements of Core Network has also two basic mobile network specific elements, SGSN (Serving GPRS Support Node) and Gateway GPRS Support Node (GGSN).

Serving GPRS Support Node (SGSN): As the name implies, this entity was first developed when GPRS was introduced, and its use has been carried over into the UMTS network architecture. It is responsible for routing the incoming and outgoing packet data from and to GPRS users via a radio access network. It is also responsible for user authentication, user location update, data encryption and decryption, establishment, maintaining, terminating sessions and mobility management procedure for UMTS users.

Gateway GPRS Support Node (GGSN): Like the SGSN, this entity was also first introduced into the GPRS network. The GGSN is the main component to support the packet-switched network. It acts as a gateway to external Packet Data Network (PDN) such as the internet. This component is connected with an external IP network to route the IP packets. The protocol which is used in PDN is called Packet Data Protocol. GGSN converts the Packet data into PDP format to send the External IP networks. The GGSN is responsible for dynamically generating IP addresses.

Circuit Switched elements of Core Network: has two basic network elements, which can be physically combined. These elements are serving Mobile Switching Centre (MSC)/Visitor Location Register (VLR) and Gateway Mobile Switching Centre (GMSC). The serving *MSC/VLR* element is responsible for circuit-switched connection management activities, mobility management related issues like location update, location registration, paging and security activities.

GMSC element takes care of the incoming/outgoing connections to/from other networks. From the connection management point of view, the GMSC establishes a call path towards the serving MSC/VLR under which the addressed subscriber is to be found. From the MM point of view, the GSMC initiates a location info retrieval procedure, whose aim is to find the correct serving MSC/VLR for call path connection.

2.3 UMTS services and bearer architecture

The 3G network mainly acts as an infrastructure providing facilities, adequate bandwidth and quality for the end-users and their applications. This facility provision, bandwidth allocation, and connection quality are commonly called Quality of Service (QoS). The various parts of the UMTS network contribute to fulfilling the QoS requirements of the services in different ways. The end-to-end service requirements divided into three entities: local bearer service, UMTS bearer service and external bearer service [6].

- Local bearer service contains the mechanisms on how the end-user service is mapped between the terminal equipment and Mobile Termination.
- UMTS bearer service, in turn, contains mechanisms to allocate QoS over the UMTS/3G network consisting of UTRAN and CN. Bearer service is used for the transfer of user and control data between network equipment.
- Since the UMTS network attaches itself to the external network(s), the end-user QoS requirements must be handled towards the other networks, too.

UMTS allows a user/application to negotiate bearer characteristics that are most appropriate for carrying information. The layered architecture of a UMTS bearer service is depicted in below Figure 3, each bearer service on a specific layer offers its services using those provided by the layers below and the UMTS bearer service plays a major role in end-to-end service provisioning.

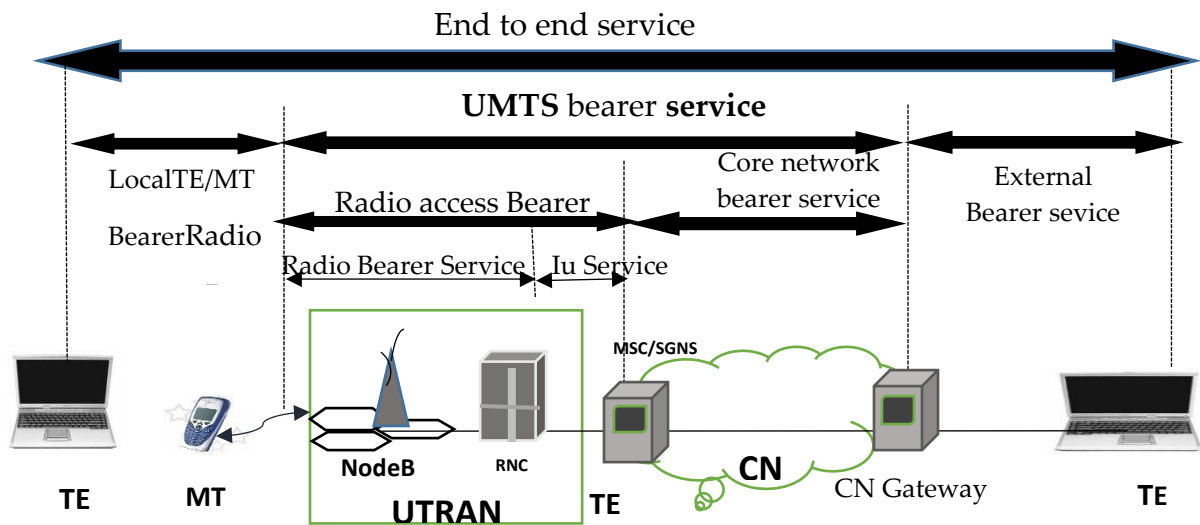


Figure 3 UMTS QoS Classes [14]

2.4 UMTS QoS basic classes

The concept of QoS has been created and defined in order to provide different treatments to traffic from different services and/or users. To clearly understand UMTS QoS classes also called traffic classes, the European Telecommunications Standards Institute (ETSI) describes four UMTS QoS classes [18]. These classes are:

- Conversational,
- Streaming,
- Interactive, and
- Background.

Conversational Class: is the most demanding QoS class that comprises real-time symmetric services with a minimum fixed delay, no buffering, symmetric traffic, guaranteed bit rate

like voice or video telephone. This class has maximum priority over network resources, the maximum transfer delay must be met to guaranty QoS.

Streaming Class: comprises typically one-way real-time services with minimum variable delay, buffering allowed, asymmetric traffic, guaranteed bit rate such as video downloading and news streaming.

Interactive Class: provides an asymmetric non-real-time service with moderate variable delay, buffering allowed, no guaranteed bit rate. Interactive Web and database retrieval are examples of interactive services. The service is provided to either a machine or a human, which requests data from remote equipment; examples of human interaction are web browsing, Social Networking Services (SNS), Instant Messaging (IM), FTP (File Transport Protocol) and an example of machines interaction is automatic database inquiries.

Background Class: in this class, services are characterized by big variable delay, buffering allowed, asymmetric traffic, no guaranteed bit rate. Services in this class include the background delivery of emails, files, and Short Message Service. The UMTS QoS classes and its Fundamental characteristics are defined in [6], explained as follows:

Table 1 Fundamental characteristics of the UMTS QoS classes

Traffic class	Fundamental Characteristics	Examples
Conversational	Preserve time relation (variation) between information entities of the stream conversational pattern (stringent and low delay).	Voice
Streaming	Preserve time relation (variation) between information entities of the stream.	Streaming video
Interactive	Request response pattern and preserve payload content.	Web browsing
Background	The destination does not expect the data within a certain time.	Background download of emails

The UMTS is one of the features of cellular communication which can yield the different types of services as per users' satisfaction and possible to assign QoS classes so as the specific services or access packages take priority over the bandwidth resource and the higher service fees can be charged.

Chapter 3: Traffic Distribution Modeling

This chapter introduces the basic concepts and notations that are employed throughout the thesis. It consists of three parts. First, in Section 3.1, the foundations of random variables and probability distributions are introduced. Section 3.2 explains the important Statistical Functions. In Section 3.3 we present commonly used continuous probability distribution. This section will also introduce the important characteristics of the distribution's PDF (probability distribution function), CDF (Cumulative Density Function) and parameters to define each curve. Finally, Section 3.4 & Section 3.5 summarizes the Parameter Types and Parameter Estimation respectively.

3.1 Statistical traffic distribution modeling

A statistical model deals with uncertainties and therefore it contains a stochastic or random part, that is, a component that describes the uncertain nature of the response variable [17]. The main instrument to deal with uncertainty in mathematics is the concept of probability and all statistical models incorporate some probabilistic assumptions by assuming that response variables of interest have a probabilistic distribution. This probabilistic or stochastic part of a statistical model usually arises from the assumption that the data we observe are a sample from a larger (unknown) population whose properties we are trying to study [18]. Probability distributions are widely used in the area of engineering, actuarial science, computer science, biological science, physics, and other applicable areas of study.

In statistical modeling, the continuous aim is to look for the probability law, which best describes the observations arising from a given field and which should represent the underlying data-generating process. The statistical analysis of what is variously referred to

as lifetime, survival time, or failure time data is an important topic in many areas, including the biomedical, engineering, and social sciences [19].

The modeling and analysis of lifetimes is an important aspect of statistical work in a wide variety of scientific and technological fields. The types of models used in lifetime data analysis range from fully parametric to nonparametric; semiparametric models that have both parametric and nonparametric features are common [20].

Distributions may be classified [21] into families or systems such that the members of a family

- have the same special properties and/or,
- have been constructed according to a common design and/or,
- share the same structure.

Such families have been designed to provide approximations to as wide a variety of observed or empirical distributions as possible. Probability models are now a vital component of every scientific investigation. These models lead to well-known parametric lifetime distributions, such as exponential, Weibull, and gamma distributions, as well as the change-point and mixture models [21].

The goal of modeling is to derive new information based on the data available, to describe the data by summarizing it in some appropriate way or to make predictions about future data values.

3.2 Important statistical functions

A continuous distribution displays the ranges of probabilities for the outcomes of a random variable with *infinite values* and is used to model a continuous random variable.

These types of random variables are uncountable and the probability of a continuous random variable at one specific point is zero. Continuous distributions are typically described by *probability distribution functions*.

The two most important statistical functions of probability distributions are PDF and CDF. A probability density function $f(x)$ can be used to describe the probability distribution of a continuous random variable X and provides a simple description of the probabilities associated with a random variable. Let X be a continuous random variable. Then a probability distribution of X is a function $f(x)$ such that for any two numbers a and b ,

$$Pr [a \leq X \leq b] = \int_a^b f(x)dx \quad (3.1)$$

That is, the probability that takes on a value in the interval $[a, b]$ is the area under the density function from a to b .

A *histogram* is an approximation to a probability density function. For each interval of the histogram, the area of the bar equals the relative frequency (proportion) of the measurements in the interval. The relative frequency is an estimate of the probability that measurement falls in the interval. The important point is that $f(x)$ is used to calculate an area that represents the probability that X assumes a value in $[a, b]$.

The *CDF* describes the probability that a random variable X with a given probability distribution will be found at a value less than or equal to x .

For continuous distributions, it provides the area under the probability density function from $-\infty$ to x , this can be expressed as:

$$F(x) = p(X \leq x) = \int_{-\infty}^x f(y)dy \quad (3.2)$$

$F(x)$ represents the cumulative values of the PDF. For the 'well-behaved' distributions (where $F(x)$ everywhere differentiable) typically encountered in data analysis, the two approaches are equivalent.

Expected Value or Mean: the expected value of a random variable is a measure of the central tendency of the random variable. Another term to describe the expected value is the 'first moment', that one expects a random variable to be.

Variance: each defined random variable has a variance associated with it as well. This is a measure of the concentration of the observations within that random variable. When calculating variance, the idea is to calculate how far each observation of the random variable is from its expected value. The *standard deviation* is equal to the square root of the variance.

3.3 Commonly used continuous probability distributions

Many different lifetime distributions can be used to model sample data but some well-known continuous probability distributions in [22] [23] are *Exponential Distribution*, *Normal distribution*, *Lognormal distribution*, *Gamma distribution*, *Weibull distribution*, and *Stable distributions*. *Exponential Distribution* is the probability distribution of the time between events in a Poisson point process, i.e. a process in which events occur continuously and independently at a constant average rate. The PDF of an exponential random variable is described in Equation 3.3, where $\lambda > 0$ is the rate parameter of the distribution.

The CDF for the exponential distribution is described in Equation 3.4. In Equation 3.5, the mean $E[X]$ and variance $\text{Var}[X]$ are described [24].

$$f(x, \lambda) = \lambda e^{-\lambda x} \quad x \geq 0 \quad (3.3)$$

$$F(x, \lambda) = 1 - e^{-\lambda x} \quad (3.4)$$

$$E[X] = \frac{1}{\lambda} \quad \text{and} \quad \text{Var}[X] = \frac{1}{\lambda^2} \quad (3.5)$$

Normal distribution: is the most widely used distribution type. Most of the data are centered around the mean, but either side of the mean, we find fewer and fewer values [25]. The PDF and the CDF of the normal distribution are described in Equation 3.6 and Equation 3.7 respectively. The related error function $\text{erf}(x)$ gives the probability of a random variable with a normal distribution of mean 0 and variance 1/2 falling in the range $[-x, x]$.

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (3.6)$$

$$F(X) = \frac{1}{2} \left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right) \right] \quad \text{where} \quad (3.7)$$

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (3.8)$$

$$\text{Mean} = \mu \quad \text{and} \quad \text{Variance} = \sigma^2 \quad (3.9)$$

Lognormal distribution: it has been used as a model in diverse applications in engineering, medicine, and other areas [26]. The lognormal distribution is a variant of the normal distribution.

A random variable X is said to be log-normally distributed if $Y = \log X$ is normally distributed [25]. The PDF and CDF of the lognormal distribution respectively is given by:

$$f(x; \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}, \quad x > 0 \quad (3.10)$$

$$F(X) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left[\frac{\ln x - \mu}{\sqrt{2}\sigma} \right] \quad (3.11)$$

$$\text{Mean} = \exp \left(\mu + \frac{\sigma^2}{2} \right) \quad (3.12)$$

$$\text{Variance} = \exp [(\sigma^2) - 1] \exp (2\mu + \sigma^2) \quad (3.13)$$

where the location parameter $\mu \in \mathbb{R}$ and the scale parameter $\sigma > 0$.

Gamma distribution: is a flexible distribution that may offer a good fit for some sets of life data. Sometimes called the Erlang distribution [19]. A continuous random variable X is said to have a Gamma distribution with a shape parameter $\gamma > 0$ and rate of the distribution $\lambda > 0$, its PDF is given by:

$$f_X(x) = \frac{\lambda^\gamma x^{\gamma-1} e^{-\lambda x}}{\Gamma(\gamma)} \quad x > 0; \gamma > 0 \quad \text{where} \quad (3.14)$$

Γ is the gamma function which has the formula:

$$\Gamma(\gamma) = \int_0^\infty t^{\gamma-1} e^{-t} dt \quad (3.15)$$

The CDF of the gamma distribution is given by:

$$F(X) = \frac{\Gamma_x(\gamma)}{\Gamma(\gamma)} \quad x \geq 0; \gamma > 0 \quad \text{where} \quad (3.16)$$

$$\Gamma_x(\gamma) = \int_0^x t^{\gamma-1} e^{-t} dt \quad (3.17)$$

Weibull distribution is a continuous probability distribution named after Swedish mathematician Waloddi Weibull. The PDF of a Weibull random variable is described in Equation 3.18, where $\tau > 0$ is the shape parameter and $\sigma > 0$ is the scale parameter of the distribution. The CDF for the Weibull distribution is described in Equation 3.19. The mean $E[X]$ and variance $\text{Var}[X]$ are described in Equation 3.20 and Equation 3.21 respectively [27].

$$f(x; \sigma, \tau) = \frac{\tau}{\sigma} \left(\frac{x}{\sigma}\right)^{\tau-1} e^{-\left(\frac{x}{\sigma}\right)^\tau}, \quad x \geq 0 \quad (3.18)$$

$$F(x; \sigma, \tau) = 1 - e^{-\left(\frac{x}{\sigma}\right)^\tau}, \quad x \geq 0 \quad (3.19)$$

$$E[X] = \tau \Gamma\left(1 + \frac{1}{\sigma}\right) \quad (3.20)$$

$$\text{Var}[X] = \tau^2 \left[\Gamma\left(1 + \frac{2}{\sigma}\right) - \left(\Gamma\left(1 + \frac{1}{\sigma}\right)\right)^2 \right] \quad (3.21)$$

Stable distributions are a family of probability distributions that permit adjustable levels of heavy tails and skewness and include distributions such as Gaussian, Cauchy, and Levy as particular cases [28]. A Stable distribution is characterized by four parameters:

- the stability index $\alpha \in (0, 2]$, that determines the thickness of the tails.
- the skewness parameter $\beta \in [-1, 1]$, determining the asymmetry
- the scale parameter $\sigma > 0$, the parameter σ expands or contracts around μ and

-
- the location parameter $\mu \in \mathbb{R}$, the parameter μ shifts the distribution to the left or right.

Mean= μ when $\alpha > 1$, otherwise undefined and variance = $2\sigma^2$ when $\alpha > 0$, otherwise undefined.

The major drawback for the application of Stable models are the lack of closed analytical expressions for their PDF or CDF except in particular cases, which makes the application of numerical methods a must, which implies a very high computational cost due to the numerous PDF evaluations required to maximize the likelihood in the four-dimensional parameter space [25] [32].

3.4 Distribution parameters

Generally, properties of the distributions include measures of location, scale, skewness, and kurtosis [19]. For parametric distributions usually limited to a maximum of three parameters. These three parameters are known as the scale parameter, shape parameter and location parameter [26] and explained below.

Scale Parameter: a scale parameter usually represents or is related to the 'spread' of the distribution. The scale parameter defines where the bulk of the distribution lies, or how stretched out the distribution is. Occasionally it is the standard deviation or maybe the coefficient of variation of the distribution. *Shape Parameter:* the shape parameter, as the name implies, helps define the shape of a distribution. Some distributions, such as the exponential or normal, do not have a shape parameter since they have a predefined shape that does not change. In the case of the normal distribution, the shape is always the familiar bell shape. The effect of the shape parameter on distribution is reflected in the shapes of the PDF, the reliability function, and the failure rate function.

Location Parameter: the location parameter, usually denoted as γ , defines the location of the origin of distribution and can be either positive or negative. In terms of lifetime distributions, the location parameter represents a time shift. A location parameter usually represents the 'center' of the distribution. The center of a distribution can be explained by the mean (average of the data set), the median where the value of data set cuts the distribution in two halves with probability at each size 0.50, and the mode (the value of the data set which has the highest value in the probability density function).

3.5 Parameter Estimation

An important task in statistical modeling is to fit the data with a probability distribution by estimating its parameter vector. The term parameter estimation refers to the process of using sample data to estimate the parameters of the selected distribution. There are several techniques available for estimating unknown parameters from data that includes [29]:

- Probability plotting, hazard plotting for the common life distributions.
- Maximum likelihood and probability plot estimates of distribution parameters, percentiles, reliability functions, hazard rates, and hazard functions.
- Kaplan-Meier estimates of nonparametric distribution type.
- Confidence intervals for distribution parameters and percentiles.

Finding suitable values of the parameter vector that maximizes the likelihood function is named Maximum Likelihood (ML) estimation, which is an essential method in estimation theory [26]. To determine the most likely values for the parameters of the assumed distribution under MLE, mathematically formulated as follows.

Given a family of probability distributions D_Θ with the parameter θ and a PDF f_θ , the observed data, $f_\theta(X(x_1, x_2, \dots, x_n | \theta))$. To find a fit for the parameter θ , the likelihood function of θ with given x_1, x_2, \dots, x_n is: $L(\theta) = f_\theta(x_1, x_2, \dots, x_n | \theta)$.

The likelihood function is given by:

$$L(\theta) = \prod_{i=1}^N f_\theta(x_i | \theta) \tag{3.22}$$

The logarithmic likelihood function is given by:

$$\log L(\theta) = \sum_{i=1}^N \log f_\theta(x_i | \theta) \tag{3.23}$$

The maximum of this expression (or parameter values) can be found by solving the partial differential equations:

$$\frac{\partial \log L(\theta)}{\partial \theta} = 0 \tag{3.24}$$

Log Likelihood is the value of the log likelihood function calculated using the maximum likelihood parameter estimates. It is often used as a goodness-of-fit statistic by comparing the log likelihood values achieved by each distribution and select as the best fitting the one with the maximum value.

The *probability plot* is a popular graphical method because allows visual analysis of the goodness of fit of the distribution to the data. The method of probability plotting takes the CDF of the distribution and attempts to linearize it by employing a specially constructed paper [25].

Chapter 4: UMTS Application Services Traffic Distribution

Modeling

This section introduces the necessary steps that are required to model the distribution of the application service traffic. To fit an empirical dataset to an analytical distribution four basic steps are necessary. First, the collected data has to be prepared. Secondly, we have to choose a class of analytical distributions which we think fits our data best. Thirdly, the parameters for the distribution are estimated. And finally, testing the goodness of the fit tells us which distribution is the best approximated the real data.

4.1 System model

The general structure of the proposed system model used to model the spatial and temporal distribution of selected application services traffic in Addis Ababa UMTS network are shown in Figure 4 . Activates under each step are presented in detail below.

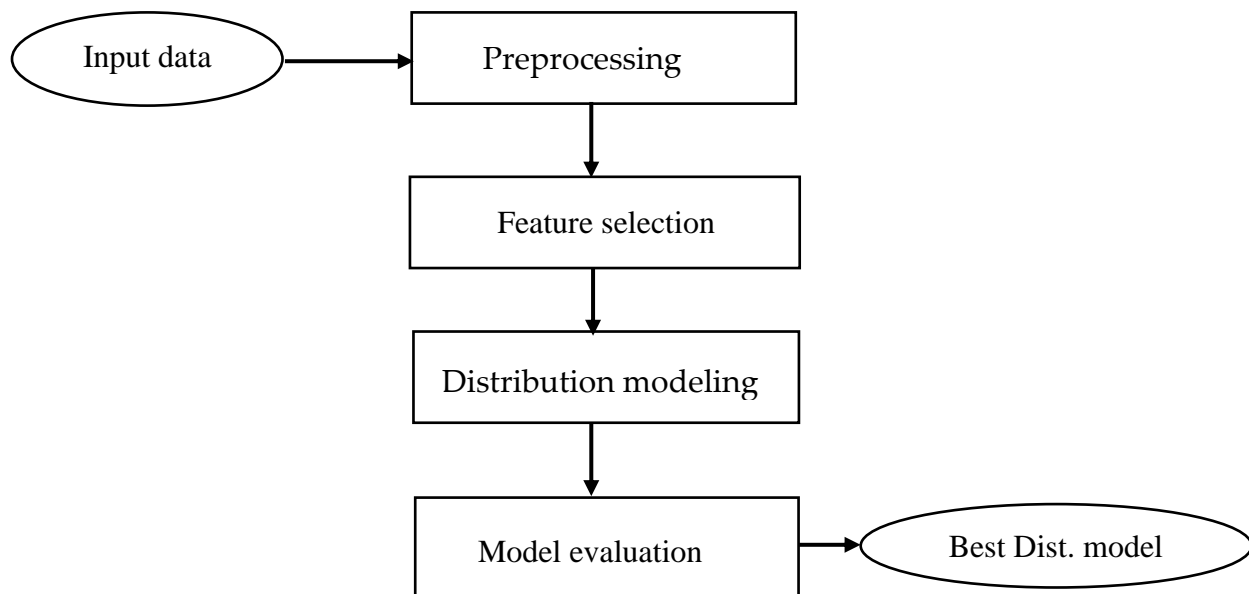


Figure 4 System model

4.2 Data collection and preparation

The input datasets used to model per application services are collected from ethio telecom UMTS network through SMARTCARE server in .xlsx format. The data was collected from 739 cells for every hour of a weak period that contains the data traffic volume (MB) of each service category utilized by each NodeB with their coordinates information. To perform the data analysis, model building and parameter estimation Microsoft Excel and Matlab tools have been used.

4.3 Feature selection

Feature selection is conducted by selecting the only features that are relevant for modeling the temporal and spatial distribution of the selected application services. On the feature selection phase, from the raw data for modeling the temporal domain the attribute time, protocol category and total traffic are selected and for the spatial domain the feature time, protocol category, cell id, and total traffic is selected.

Table 2 Sample raw data

1	1 Hour	PROT CAT	3G Cell	Total Traffic(MB)	Upload Traffic(MB)	Download Traffic(MB)
15	2019-05-01 00:00:00~2019-05-01 01:00:00	IM	111019_W	3.87346	0.31506	3.5584
16	2019-05-01 00:00:00~2019-05-01 01:00:00	VoIP	111019_W	0.03041	0.00972	0.02069
17	2019-05-01 00:00:00~2019-05-01 01:00:00	Web_Brov	111019_W	91.92731	5.15879	86.76852
18	2019-05-01 00:00:00~2019-05-01 01:00:00	Streaming	111019_W	231.5447	8.08002	223.46467
19	2019-05-01 00:00:00~2019-05-01 01:00:00	Email	111019_W	0.1414	0.05196	0.08944
20	2019-05-01 00:00:00~2019-05-01 01:00:00	File_Acce	111019_W	0.47814	0.08326	0.39488
21	2019-05-01 00:00:00~2019-05-01 01:00:00	Other	111019_W	1.55017	0.25092	1.29925
22	2019-05-01 00:00:00~2019-05-01 01:00:00	SNS	111019_W	4.12467	0.39047	3.73419
23	2019-05-01 00:00:00~2019-05-01 01:00:00	Navigatio	111019_W	0.23704	0.0548	0.18224

In this phase among nine application service categories in the dataset, As shown in Figure 5, the first three services whose contribution is more than 73% of total traffic demand are selected.

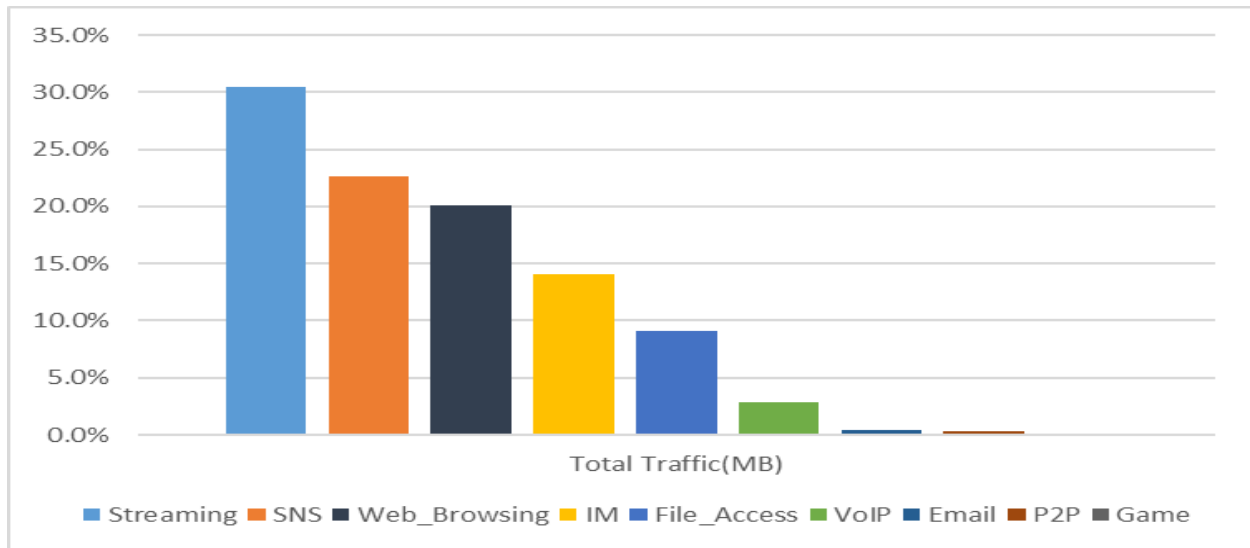


Figure 5 Mobile application services traffic demand ratio in Addis Ababa

Three services that are Streaming, SNS and Web browsing, from the original data collection, are studied and analyzed separately, to obtain models that able to describe and predict the underlying data behaviors.

4.4 Model Selection and Parameter Estimation

Model selection is a necessary step for any practical modeling task. Since the true model behind a real-world process cannot be known, the goal of model selection is to find the best approximation among a set of candidate models.

4.4.1 Traffic Distribution Modeling in Space

After selecting the top application services, to understand the real spatial distribution of traffic demand generated by the area of AA of each service is considered.

To get the traffic density for modeling spatial distribution the steps followed in [1] also implemented here:

1. The dimension of the city area is inscribed in a square of the same dimension.
2. Each side of the square is equally divided into N parts creating a matrix of N by N , which reduced the space domain by a factor of $1/N^2$ and also the average number of sites in each cell by the same factor.
3. NodeB located in the same cell of the matrix is grouped and their data traffic of every hour is summed up so that groups may be represented with a single traffic profile.

After obtaining the traffic density of Streaming, Web browsing, and SNS application services separately. The first step towards the statistical modeling of spatial traffic is visualized the spatial distribution being modeled [30], the real traffic spatial distribution of each service shown in Figures below .

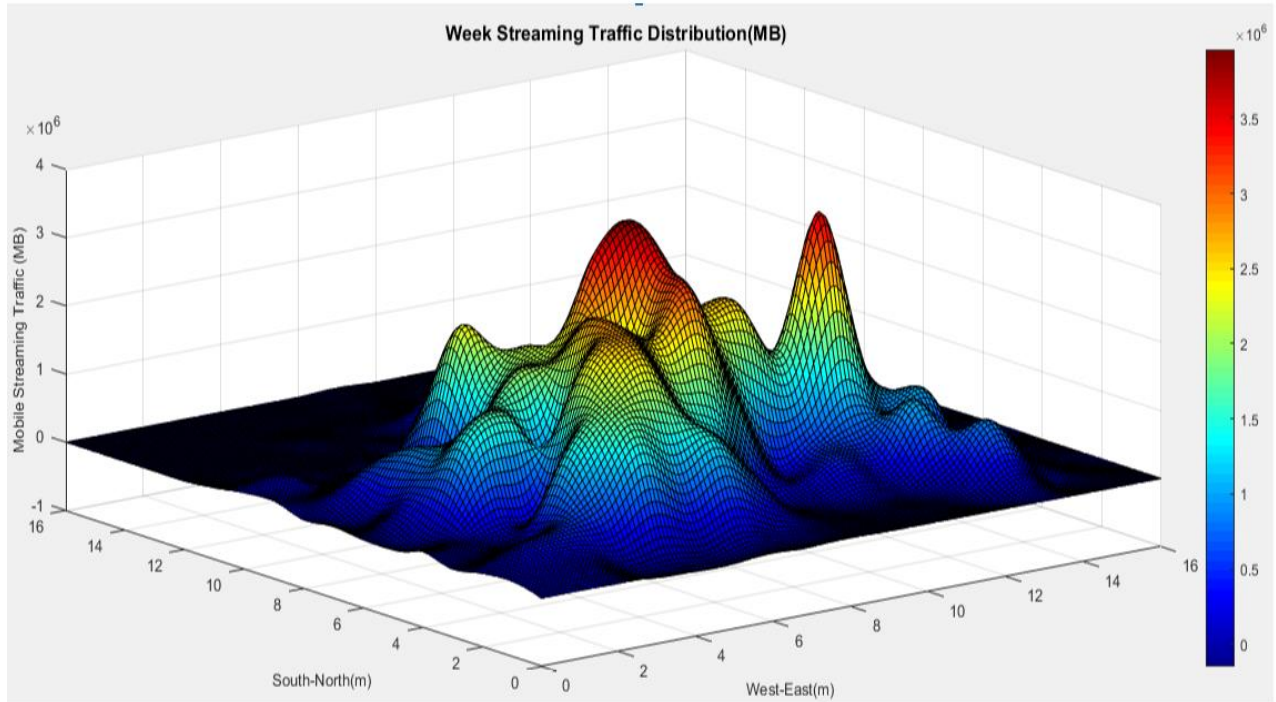


Figure 6 Streaming spatial traffic distribution

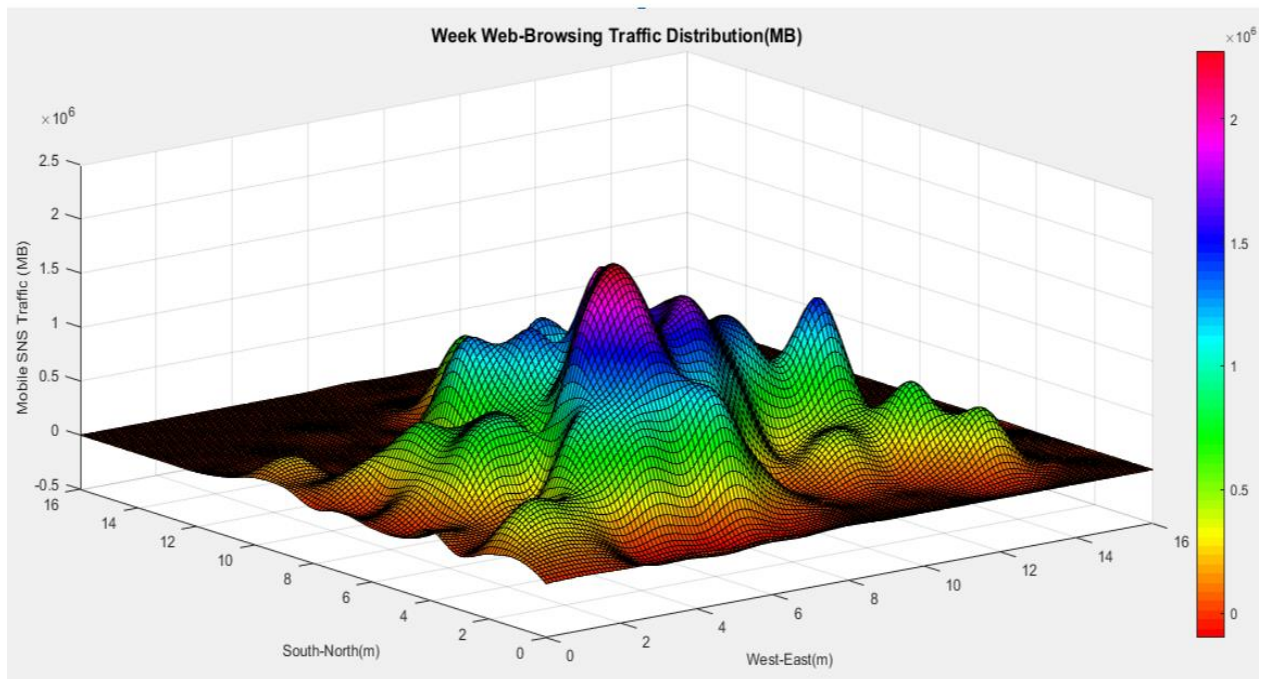


Figure 7 Web browsing spatial traffic distribution

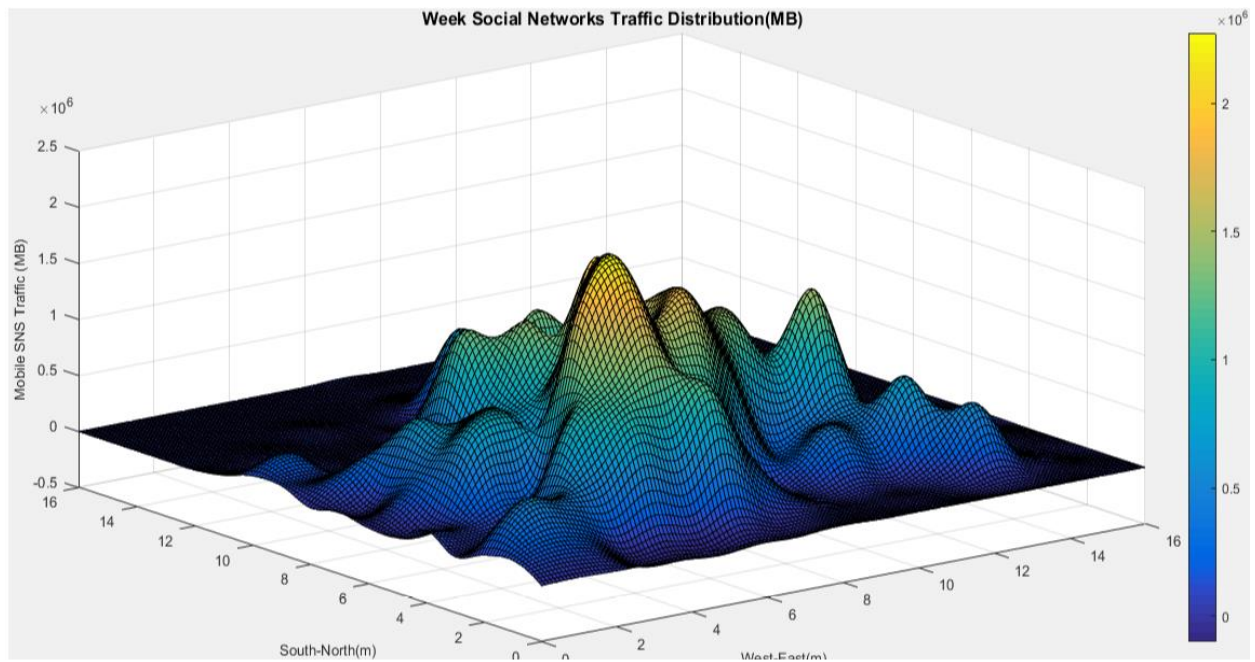


Figure 8 Social networks spatial traffic distribution

To build the best model for each each services, a statistical distribution models Lognormal, Gamma, Weibull and Normal distribution are selected which is the second scenario for modeling the spatial distribution of the three application services. The Plot below shows the fit to the candidate distribution's CDF and PDF based on the datasets of Streaming, Web browsing, and SNS application services.

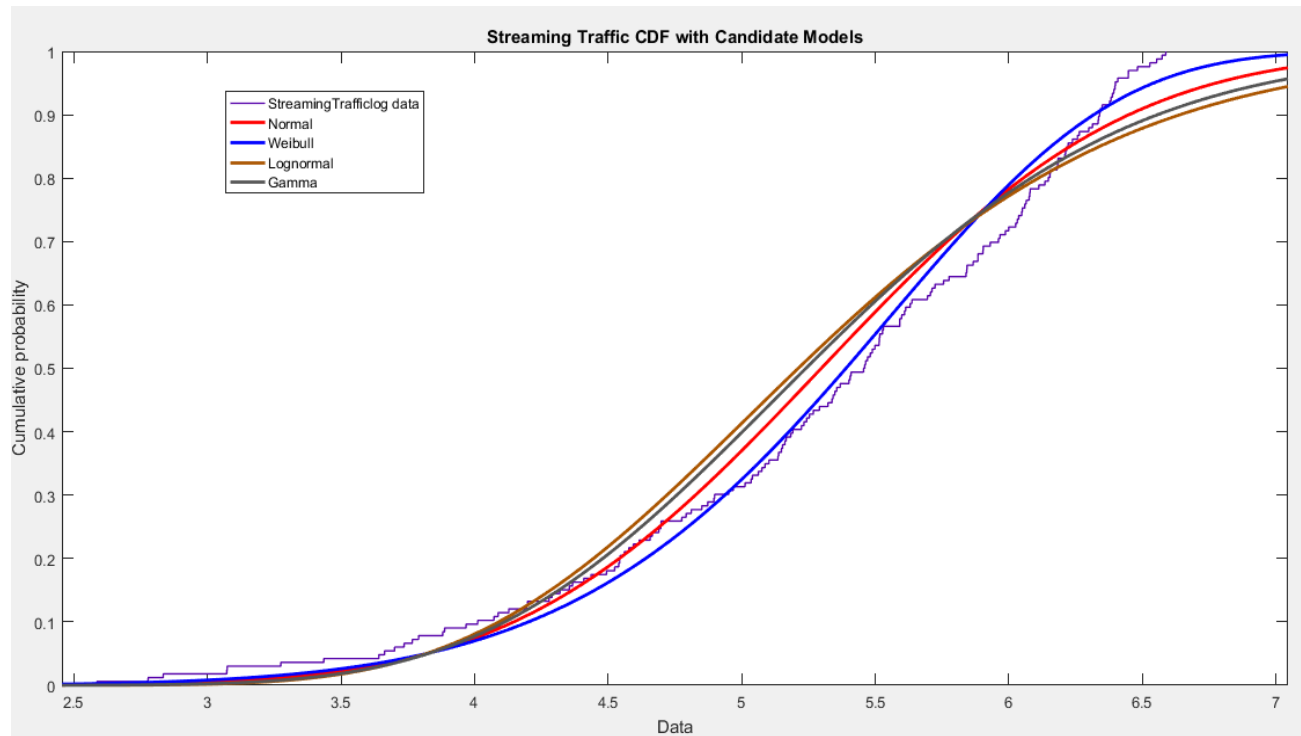
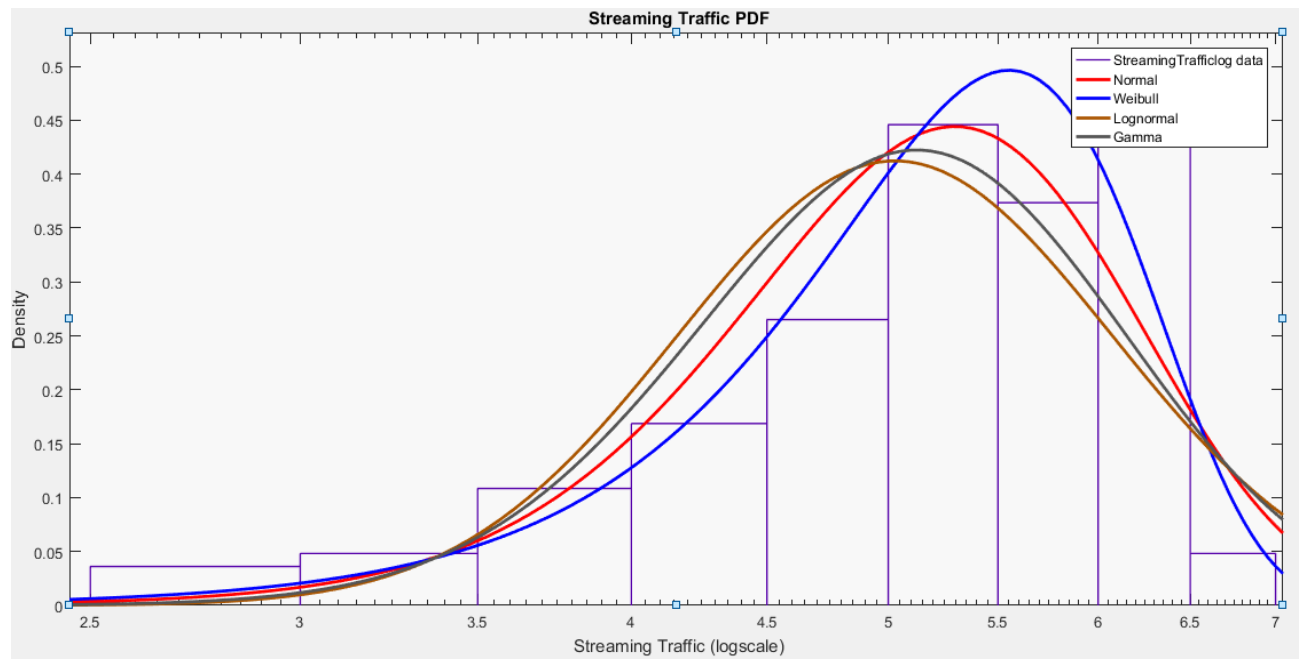


Figure 9 Streaming traffic PDF and CDF plot

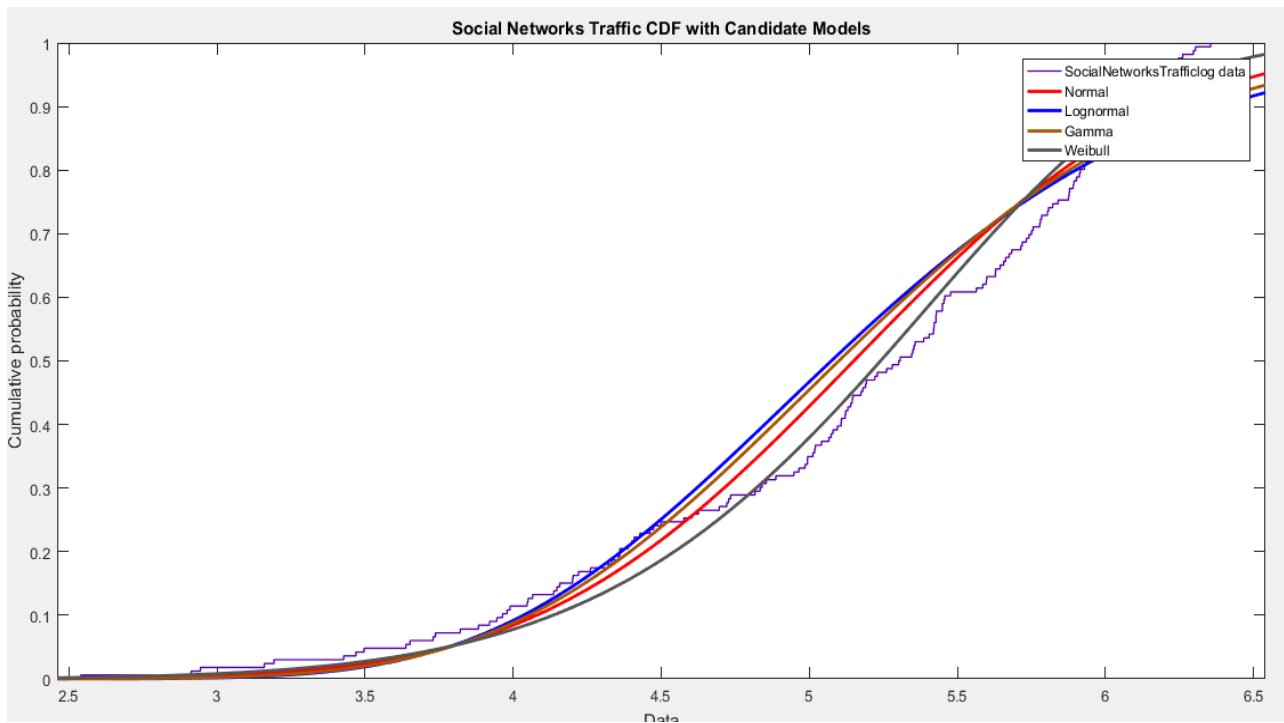
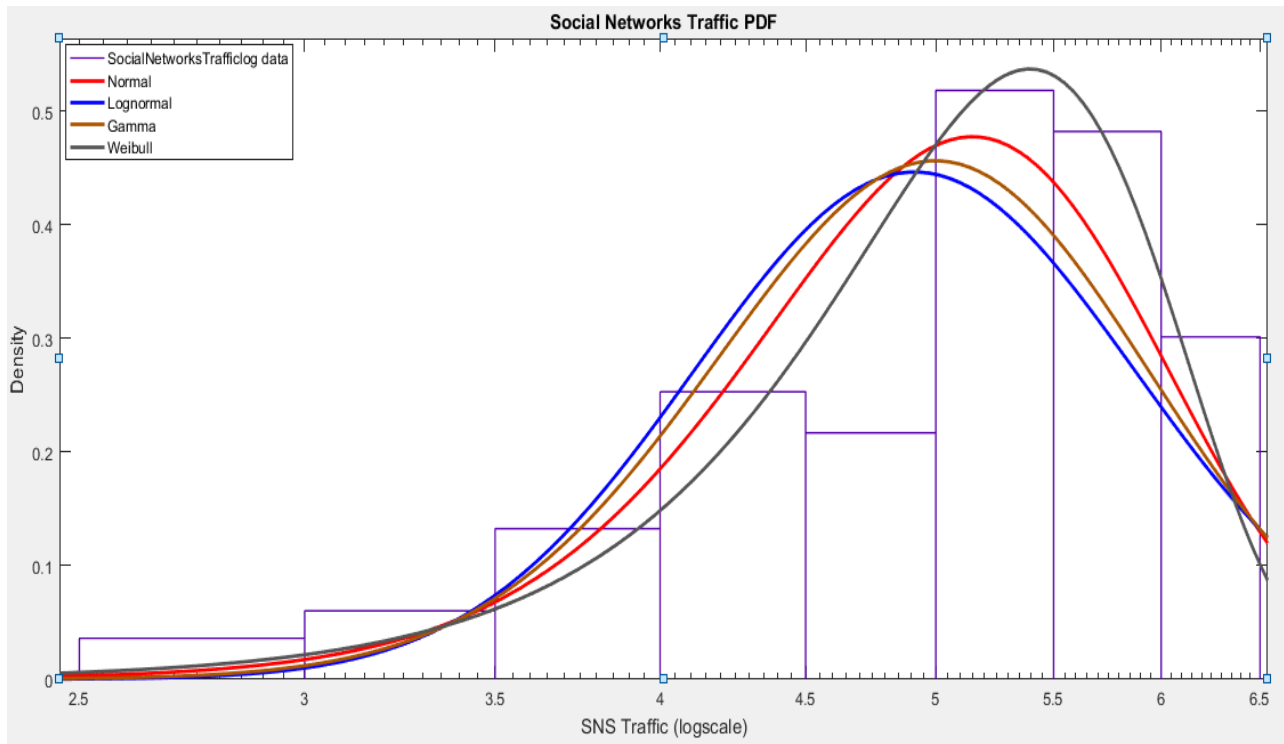


Figure 10 Social networks traffic PDF and CDF plot

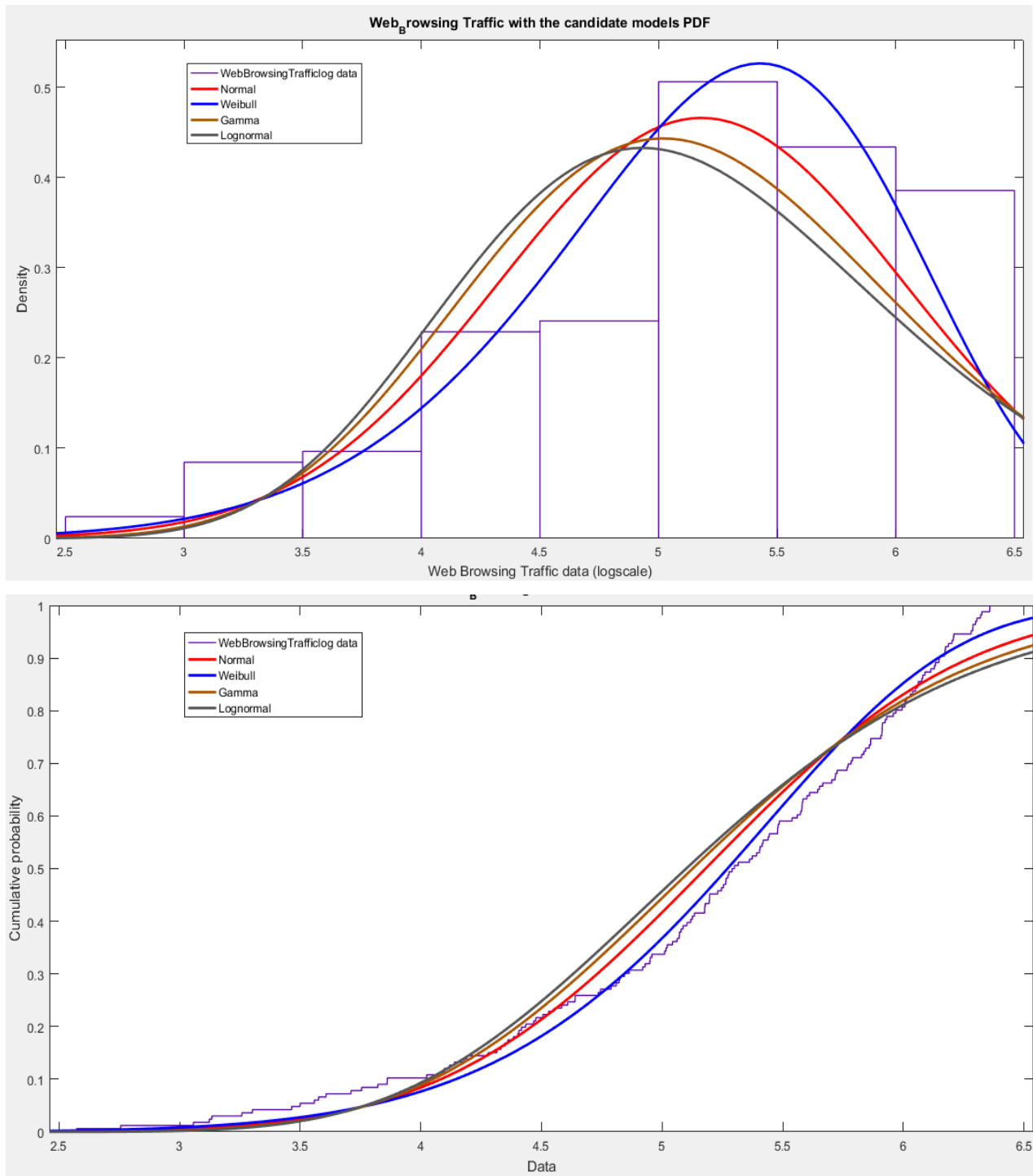


Figure 11 Web browsing traffic PDF and CDF plots

Models Evaluation for the spatial distribution of Application Services

After selecting the candidate models, then for evaluating these models, two approaches are considered. The first Maximum log-likelihood Estimation [31], which is by comparing the log likelihood values achieved by each distribution and select as the best fitting the one with the maximum value, is used.

Table 3 List of candidate distributions and their fit to streaming traffic data

Distribution	Likelihood	Mean	Variance
Normal	-217.234	5.29785	0.80688
Weibull	-207.35	5.31357	0.690119
Lognormal	-232.568	5.30548	1.02412
Gamma	-226.555	5.29785	0.917295

Table 4 List of candidate distributions and their fit to web browsing traffic data

Distribution	Likelihood	Mean	Variance
Normal	-209.274	5.17975	0.733093
Weibull	-198.474	5.19616	0.61788
Lognormal	-224.369	5.18662	0.925107
Gamma	-218.523	5.17975	0.831732

Table 5 List of some distributions and their fit to SNS traffic data

Distribution	Likelihood	Mean	Variance
Normal	-205.558	5.15019	0.698465
Weibull	-194.944	5.16652	0.595253
Lognormal	-219.078	5.15622	0.865671
Gamma	-213.701	5.15019	0.784023

Probability Plot of each Application Services

The second approach to evaluate models for the spatial distribution of the real data with the candidate distribution models, the graphical method for model checking, probability plot is used to select the best model for each application services.

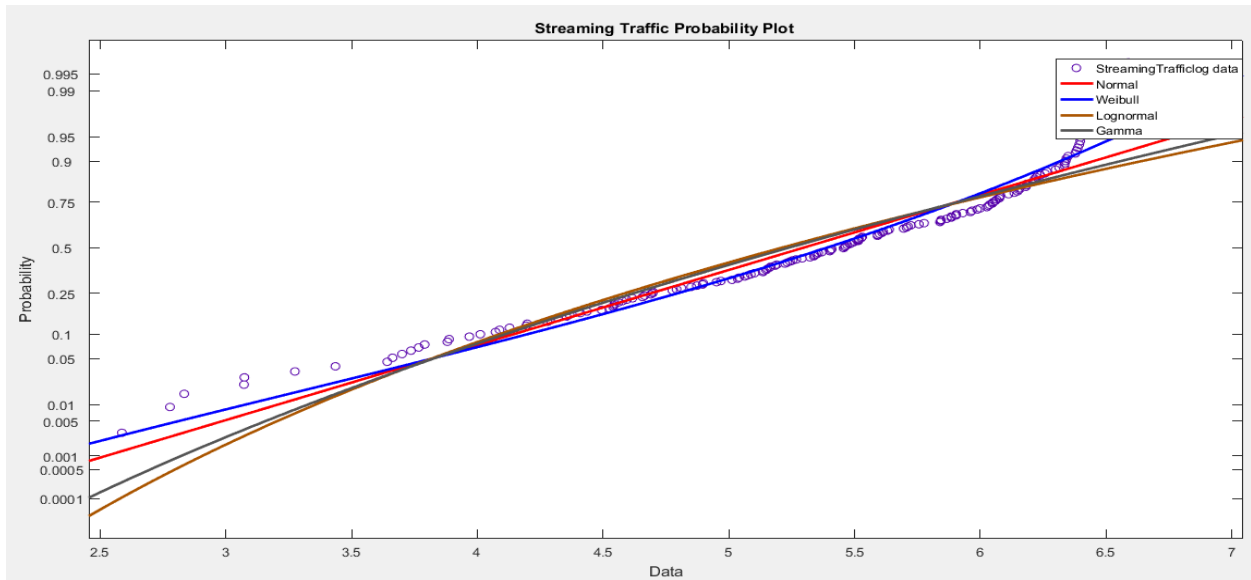


Figure 12 Probability plot of Streaming traffic

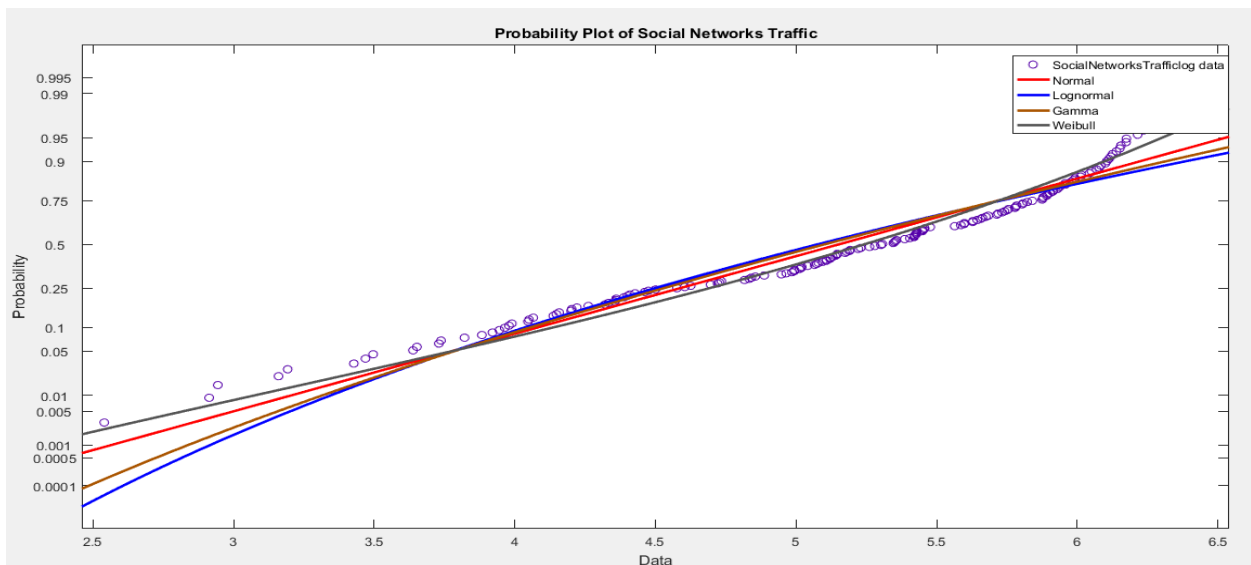


Figure 13 Probability plot of Social Networks traffic

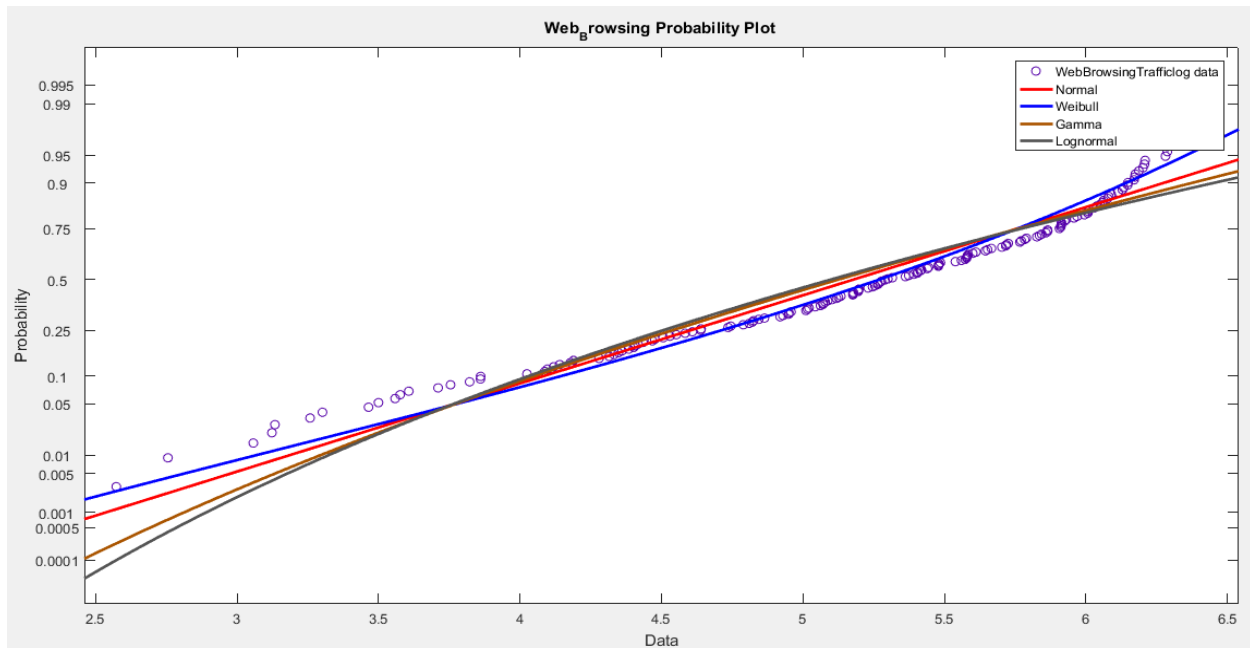


Figure 14 Probability plot of Web Browsing traffic

4.4.2 Traffic Distribution Modeling in Time

To obtain an understanding of the traffic pattern as well as for temporal modeling of each application service of the observed data, we average the traffic data for each application services for every hour of a week for both the Weekdays and Weekend of total traffic separately.

Before diving into modeling the temporal distribution, it's better to understand how the traffic pattern behaves of each service. As shown in the below figures temporal traffic trend for both weekdays and weekends of each service shows that they have nearly constant hourly traffic pattern.

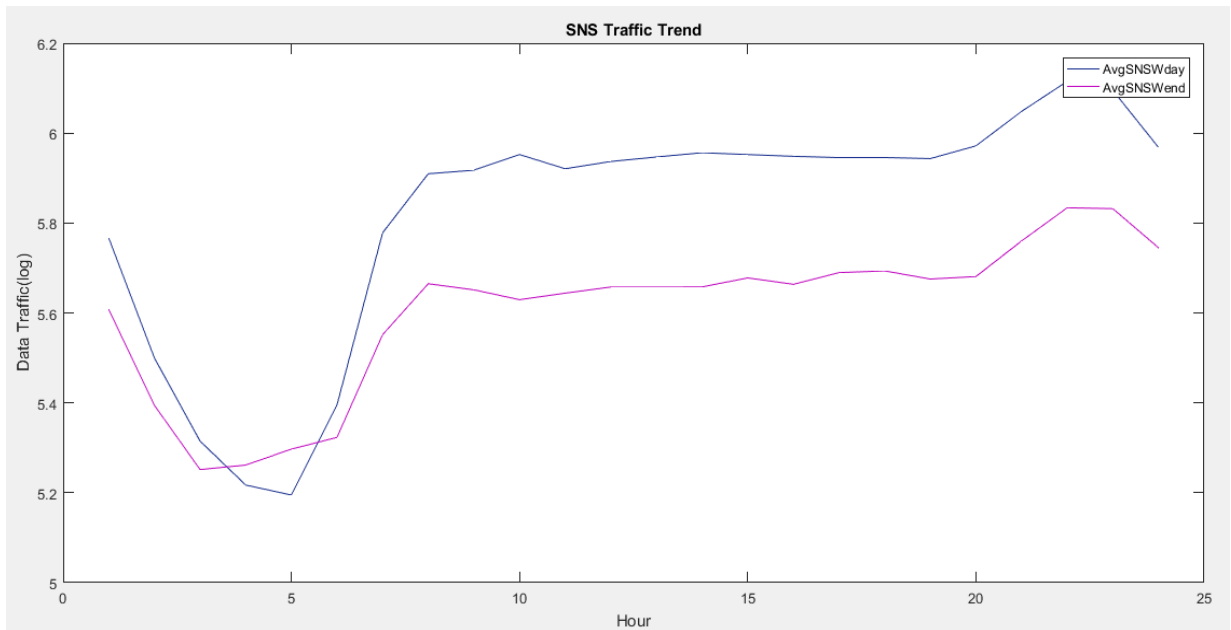


Figure 15 Social networks temporal traffic trend

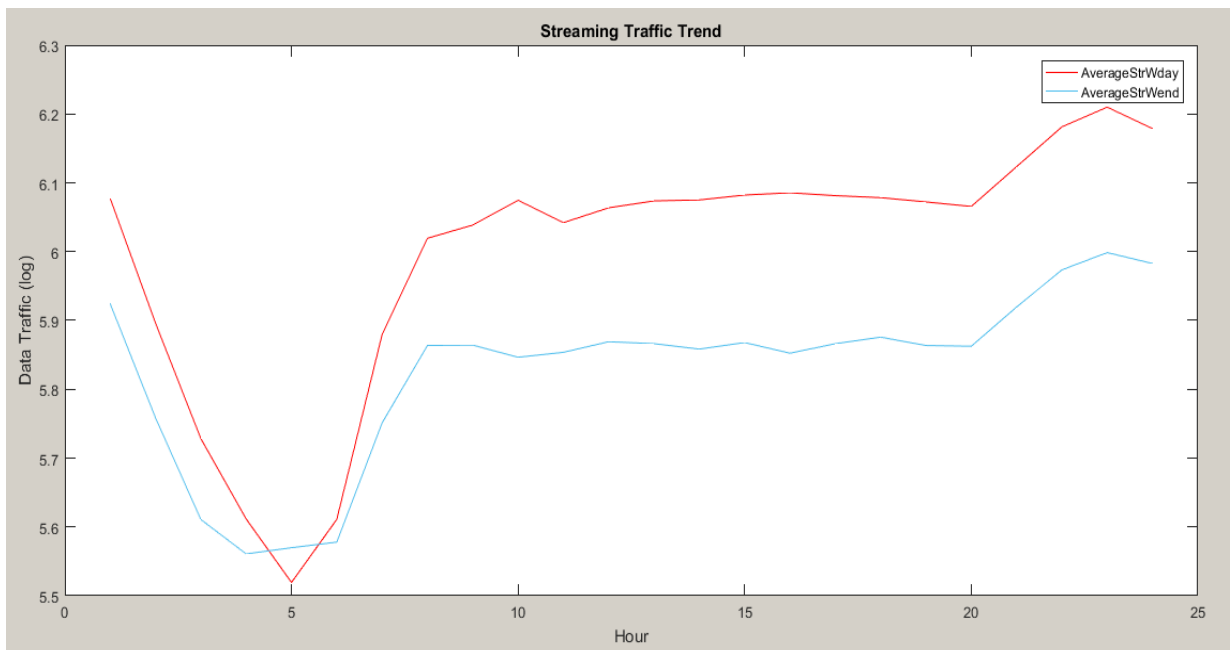


Figure 16 Streaming temporal traffic trend

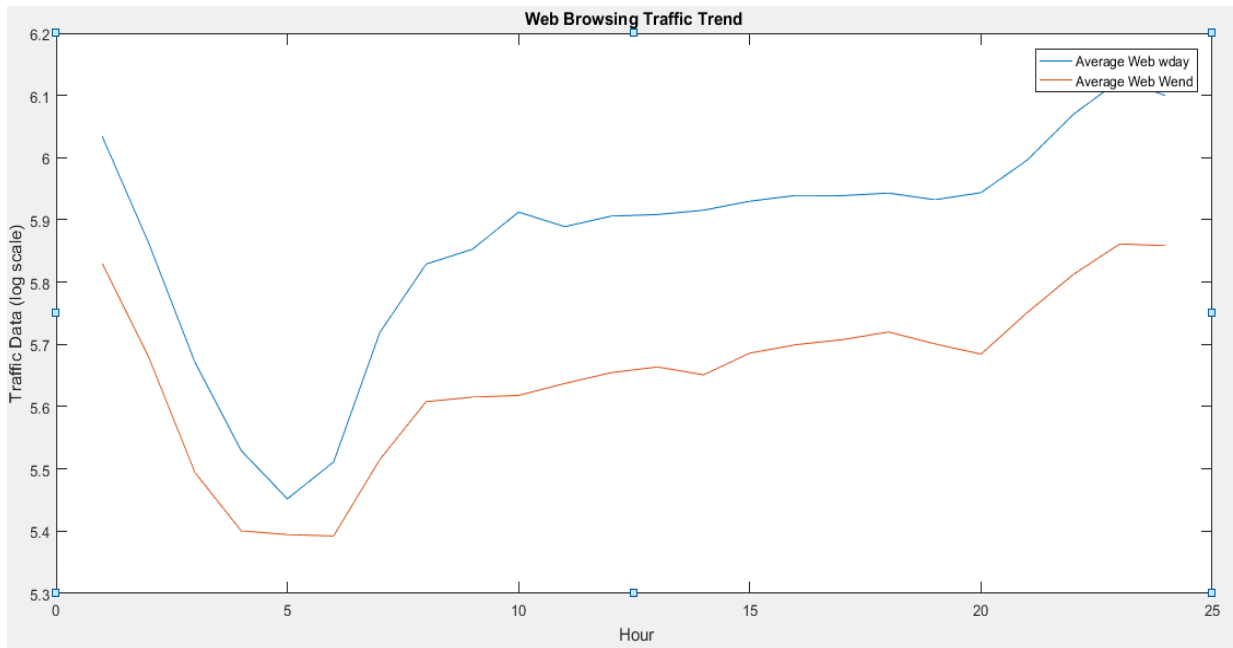


Figure 17 Web browsing temporal traffic trend

The candidate traffic distribution models with the real traffic for temporal variation of each application services of weekend and weekdays shown in the following Figures.

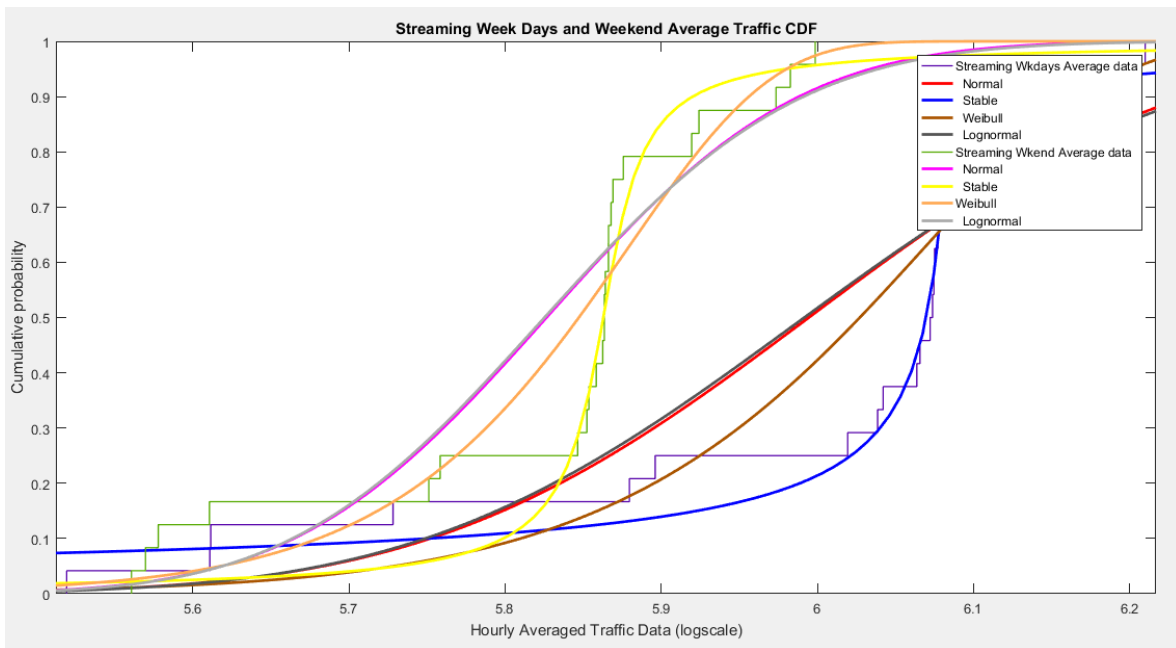


Figure 18 Streaming real traffic with candidate distribution models

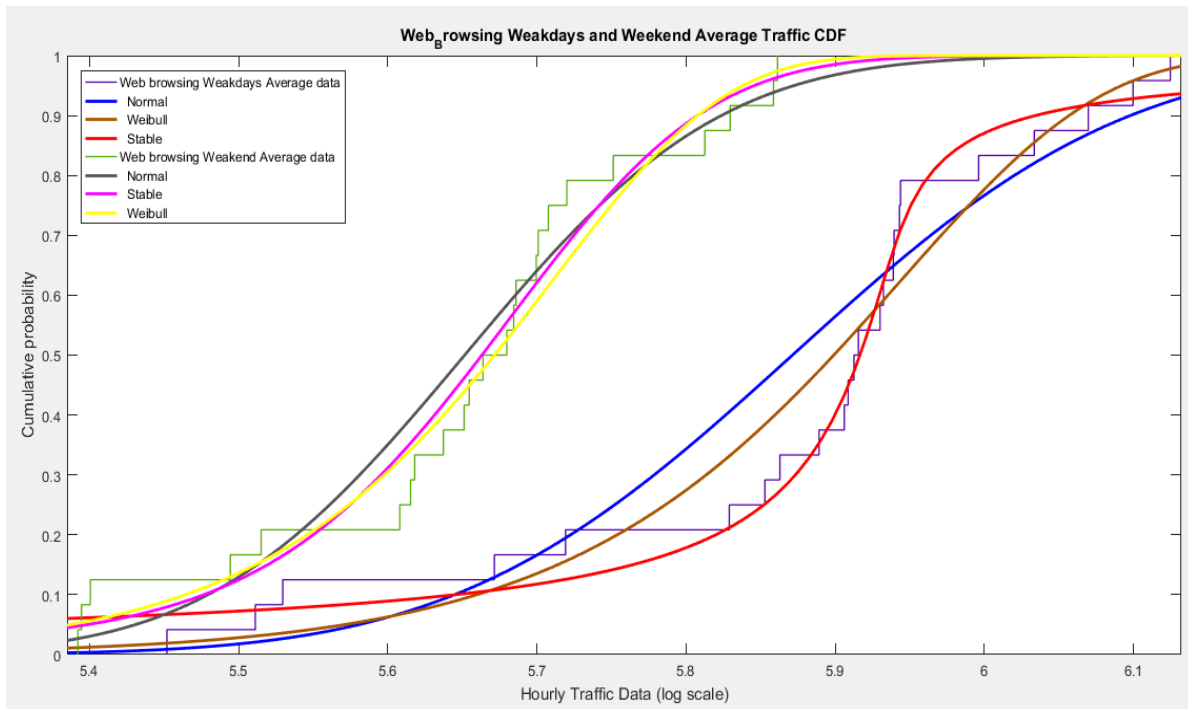


Figure 19 Web browsing real traffic with candidate distribution models

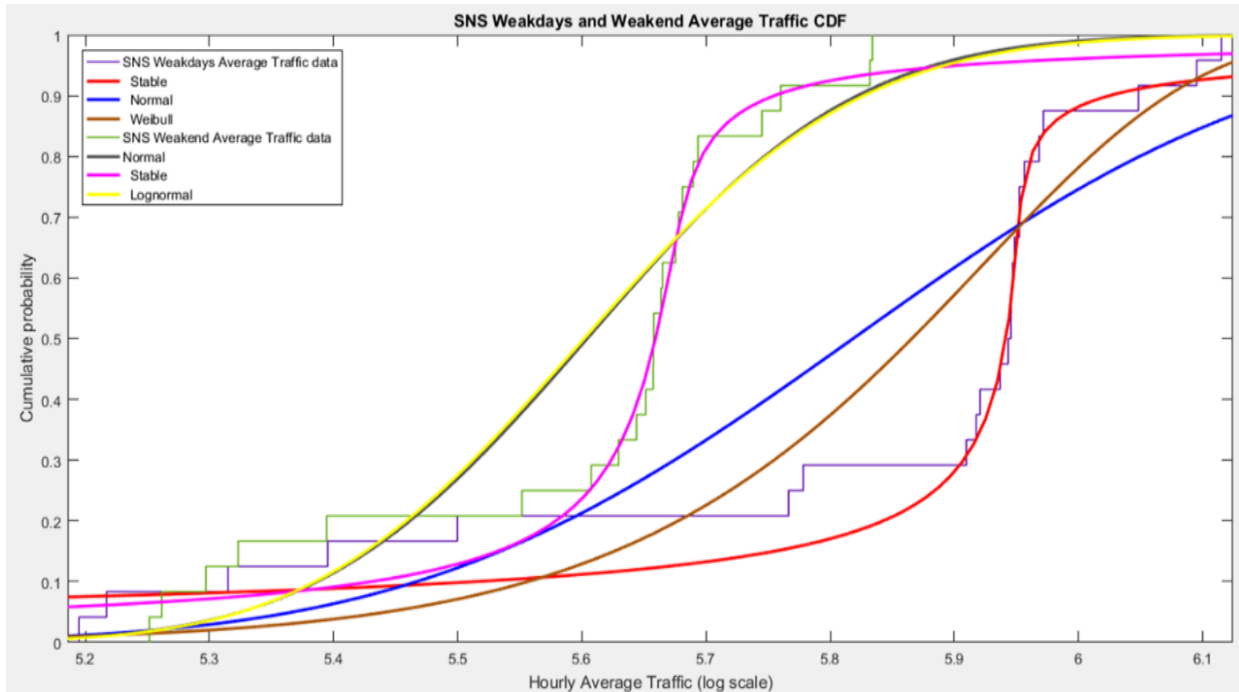


Figure 20 Social networks real traffic with candidate distribution models

Here also Maximum log-likelihood Estimation method is used for evaluating the candidate distribution in time domain. The tables below show the result for fits to the real data of each application service by the candidate distributions.

Table 6 List of candidate distributions and their fit to streaming traffic data

Distribution	Likelihood	
	Weekdays	Weekend
Normal	6.41689	16.1198
Weibull	11.4651	19.2342
Lognormal	5.89068	15.8475
Stable	19.2308	23.1952

Table 7 list of candidate distributions and their fit to SNS traffic data

Distribution	Likelihood	
	Weekdays	Weekend
Normal	-2.51207	9.01127
Weibull	2.44477	10.765
Lognormal	1.8807	8.64422
Stable	12.1988	14.37

Table 8 List of candidate distributions and their fit to Web browsing traffic data

Distribution	Likelihood	
	Weekdays	Weekend
Normal	6.41689	16.1198
Weibull	11.4651	19.2342
Lognormal	5.89068	158475
Stable	19.2308	23.1952

The same as the spatial domain, the other approach used to check candidate distribution for modeling the temporal distribution, is probability plot. Probability plot allows visual analysis of the goodness of fit of the distribution to the data as shown below.

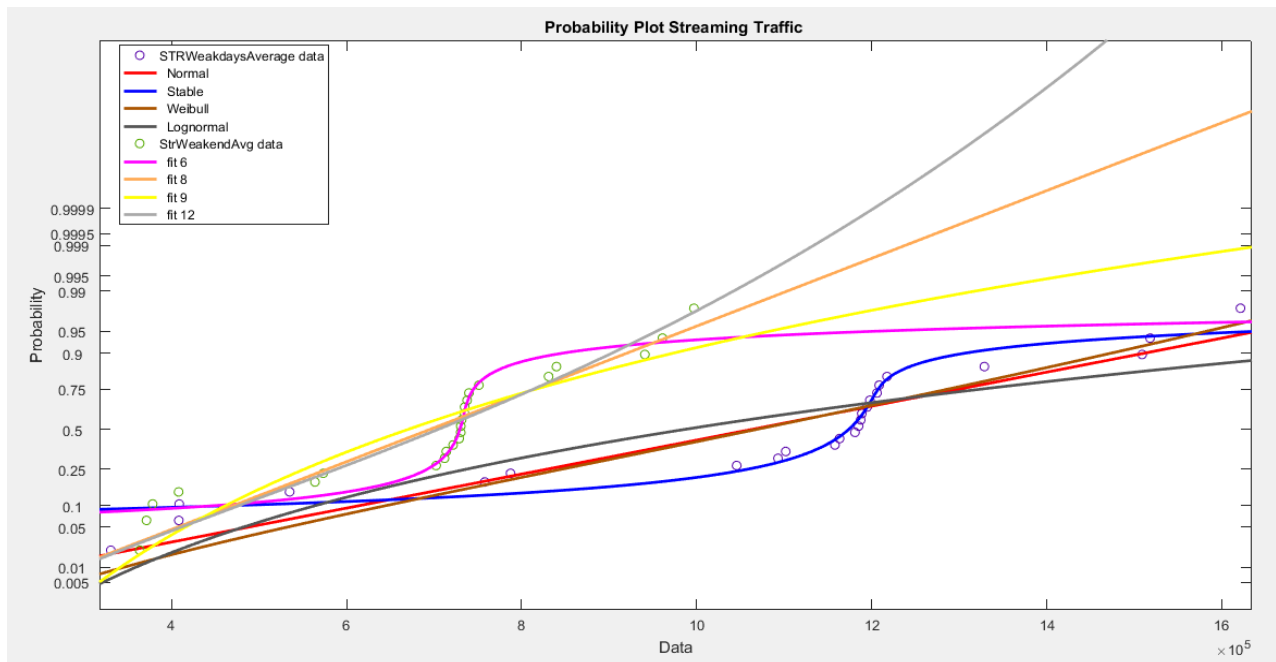


Figure 21 Streaming traffic probability plot.

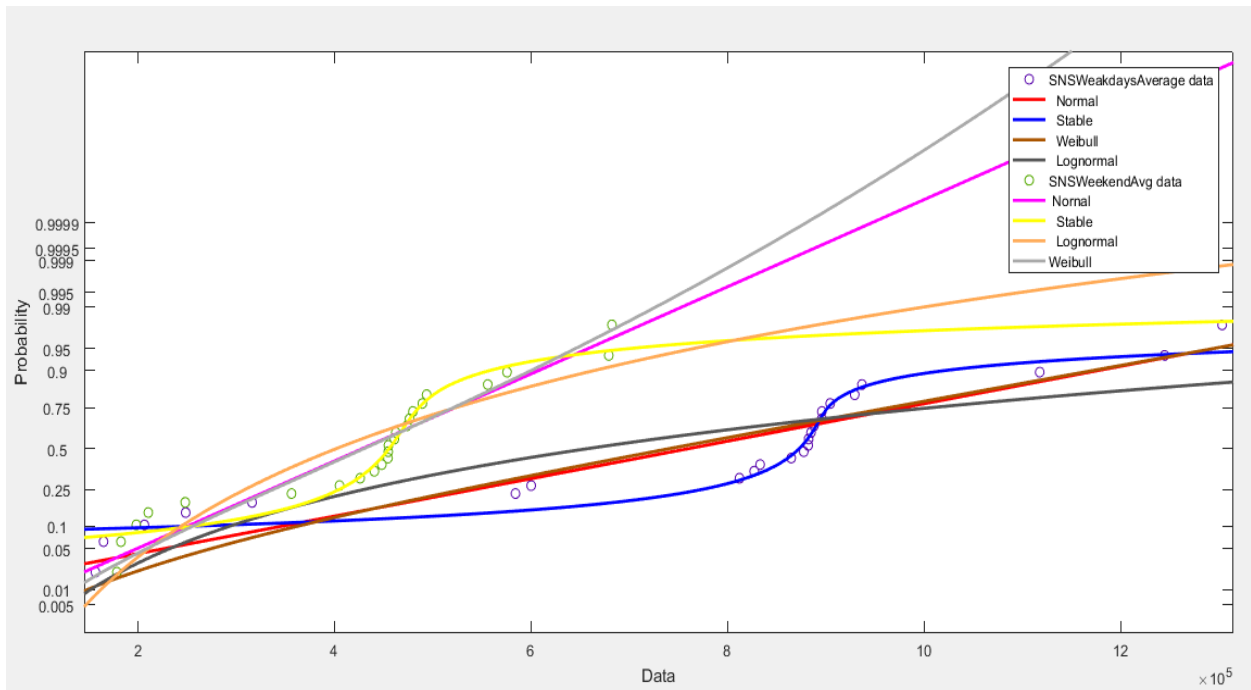


Figure 22 Social networks traffic probability plot

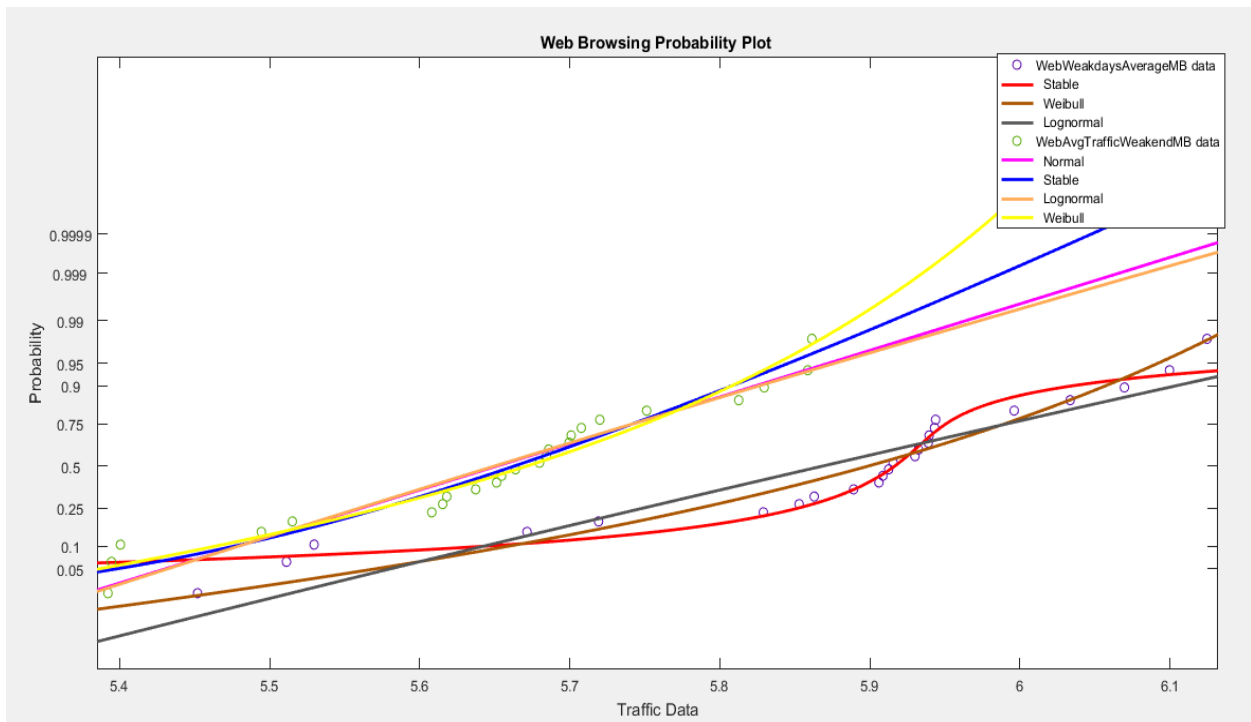


Figure 23 Web browsing traffic probability plot

4.5 Result and Discussions

This section presents the main results of this case study, where the statistical distribution modeling techniques used to model the spatial and temporal distribution of the selected three application services. Based on the traffic demand of AA area of application services, the selection has done and starts modeling the application services, namely Streaming, SNS and Web Browsing, in the time domain.

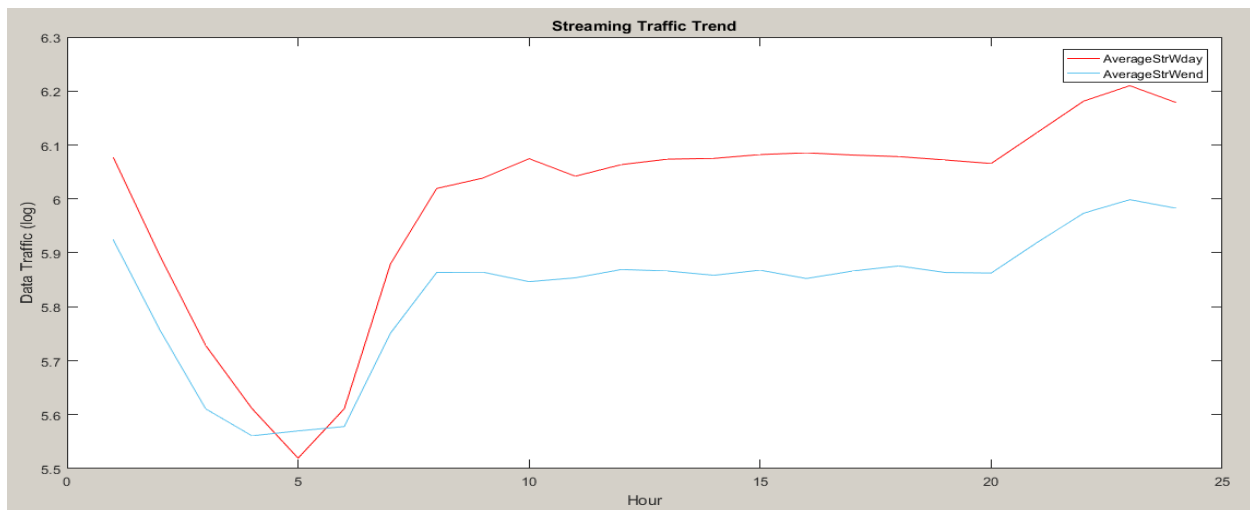


Figure 24 Streaming temporal traffic trend

In the time domain (shown in Figure 24) as a sample, the Streaming temporal traffic trend, it is observed that for each application services, for both weekdays and weekends they have nearly constant hourly traffic patterns.

Following the analysis, then first the parametric distribution model candidates identified to fit the three UMTS application services traffic, namely Weibull, Lognormal, Stable and Normal. For the selected application services Streaming, SNS and Web Browsing, for both weekday and weekend data sets, the models compared and ranked if which one has the best approximation against the observed data.

Searched for the optimal model structure based on a fit statistic of the model , the Maximum log-likelihood value and the visual aids ,the Probability Plot, proposed Stable distribution model provides a better fit than Weibull, Lognormal, Normal models of the real data for both weekday and weekend of the three application services.

The best and general model guarantee for the representation of the temporal distribution of data variations throughout the weekdays and the weekend of the three services is the Stable distribution model. The parameter estimates of Stable distribution model that fit the distribution to the selected application services real data are obtained in the below table.

Table 9 Estimated parameters and standard error for the temporal distribution

Application Service	Parameter	Estimate		Std. Err.	
		Weekdays / Weekend	Weekdays / Weekend	Weekdays / Weekend	Weekdays / Weekend
Streaming	Alpha(α)	0.578256	0.985874	0.136444	0.178943
	Beta (β)	-0.431022	-0.053393	0.223419	0.364889
	Gamma (γ)	0.01799	0.019671	0.004667	0.0053511
	Delta ()	6.07346	5.86332	0.0036193	0.0055963
Social Networks	Alpha(α)	0.542903	0.784384	0.130343	0.181299
	Beta (β)	-0.3602297	-0.294091	0.226006	0.27845
	Gamma (γ)	0.0217774	0.0347042	0.0058068	0.0091152
	Delta ()	5.94473	5.66334	0.0040266	0.00870921
Web Browsing	Alpha(α)	0.797607	1.7556	0.184097	NaN
	Beta (β)	-0.288524	-1	0.282235	NaN
	Gamma (γ)	0.0429974	0.0852454	0.010892	NaN
	Delta ()	5.9216	5.67576	0.0109008	NaN

Secondly, in space domain, in order to understand the real spatial distribution of traffic demand, the traffic density is considered instead of the cell traffic. As shown in Figure 25 as a sample, firstly the spatial distribution of streaming traffic in city of Addis Ababa is visualized, as the first step to effective modelling has been done.

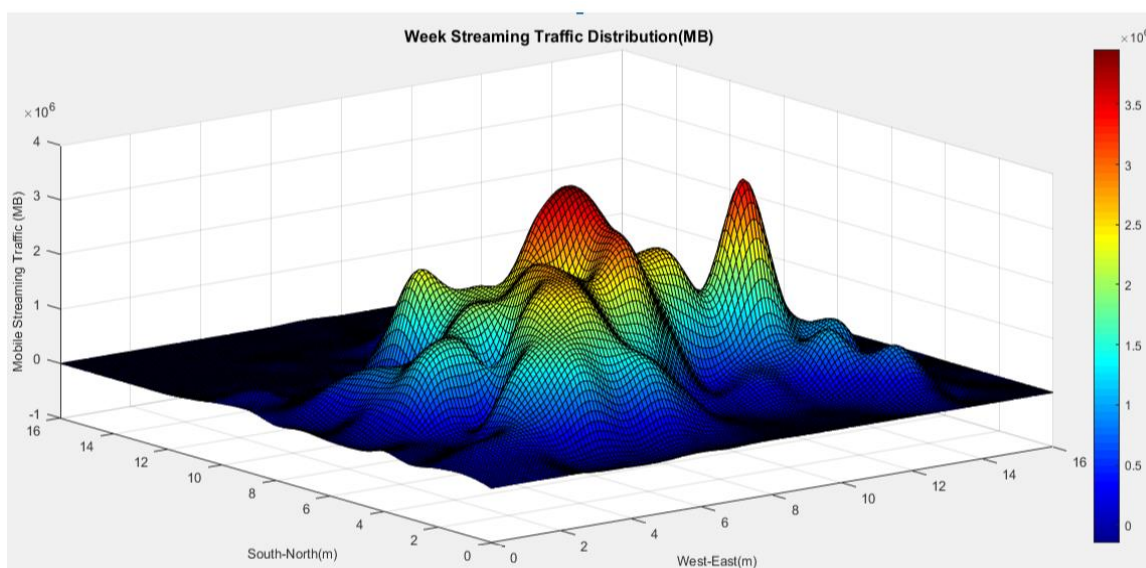


Figure 25 Streaming spatial traffic distribution

The continuous distribution model candidates identified to fit the three UMTS application services traffic that are Weibull, Lognormal, Gamma and Normal. Using a Maximum Log-likelihood value and probability plot, the candidate distribution models are compared and ranked that approximated the observed data. Weibull distribution has been selected for modeling the spatial distribution of the selected application services.

The parameter estimates and standard errors of the Weibull Distribution model that fit the distribution to the selected application services real data are obtained in the below table.

Table 10 Estimated parameters and standard error for the spatial distribution

Application Service	Parameter	Estimate	Std. Err.
Streaming	Mu (μ)	5.6579	0.0608318
	Sigma (σ)	5.6368	0.485173
Social Networks	Mu (μ)	5.48798	0.0561779
	Sigma (σ)	7.94481	0.05115
Web Browsing	Mu (μ)	5.5232	0.0572929
	Sigma (σ)	7.83553	0.504311

Chapter 5: Conclusion and Future Work

5.1 Conclusion

With the increasing usage of diversified application services, data traffic generated by these networks is increasing. Moreover, the data consumption pattern is changing in time and space. Understanding and characterizing traffic demands and dynamics associated with different mobile application services are instrumental for operators to improve application services usability, optimizing network service quality and helps as input for technical and business strategies.

The thesis presents, using Statistical distribution modeling techniques, modeling the spatial and temporal distribution of the application services traffic based on the observed data collected from AA UMTS network.

For the selected application services that are Streaming, SNS and Web Browsing the candidate models compared and evaluated if which one has the best approximation against the observed data. It has been proven that the Stable distribution model guarantees the representation of the temporal distribution of the data for both weekdays and weekends and the Weibull distribution model for the spatial distribution of above mentioned application service traffic data.

5.2 Future work

Concerning these conclusions, it is importance to point out the following considerations in future work:

- Applicability of the proposed models of the application services needs confirmation for a next-generation network since the analysis performed herein investigates UMTS networks.
- The obtained models target only to model separately the temporal and spatial distributions for the selected application services but it would be more practical to use the spatio-temporal traffic model which can produce spatial inhomogeneity, temporal variations.
- Application service traffic is dynamic as more and more application is available every day so continuous learning to model application service is needed since it is crucial to the design and optimization of networks to assure QoS for all application service.

References

- [1] Y. Abera, *“Spatiotemporal Mobile Data Traffic Modeling : The Case of UMTS Network in Addis Ababa , Ethiopia ADDIS ABABA INSTITUTE OF TECHN,”* 2018.
- [2] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, *“Characterizing Geospatial Dynamics of Application Usage In A 3G Cellular Data Network,” Proc. - IEEE INFOCOM*, no. 1, pp. 1341–1349, 2012.
- [3] M. Alasmar, *“On the Distribution of Traffic Volumes in the Internet and its Implications,”* 2017.
- [4] C. Lagerstedt, A. Aurelius, and A. Ab, *“Understanding Internet User Behavior : Towards a Unified Methodology,”* vol. 5, no. 3, pp. 153–162, 2012.
- [5] G. Gómez and R. Sánchez, *“End-To-End Quality of Service Over Cellular Networks ,”* 2005.
- [6] J. Perez-Romero O. Sallent R. Agustí, *“Radio Resource Management Strategies in UMTS,”* 2005.
- [7] R. Li *et al.*, *“The Learning and Prediction of Application-level Traffic Data in Cellular Networks,”* pp. 1–14.
- [8] S. Wang, X. Zhang, and J. Zhang, *“An Approach for Spatial-temporal Traffic Modeling in Mobile Cellular Networks,”* pp. 203–209, 2015.
- [9] C. Marquez, M. Gramaglia, M. Fiore, A. Banchs, C. Ziemlicki, and Z. Smoreda, *“Not All Apps Are Created Equal,”* pp. 180–186, 2017.

-
- [10] L. Guan, X. Zhang, Z. Liu, Y. Huang, R. Lan, and W. Wang, "Spatial Modeling and Analysis of Traffic Distribution Based on Real Data from Current Mobile Cellular Networks," pp. 135–138.
- [11] M. Zeng and J. Wu, "Characterizing and Predicting Individual Traffic Usage of Mobile Application in Cellular Network," pp. 852–861, 2018.
- [12] C. From, "Spatial Modeling of the Traffic Density in Cellular Networks," pp. 80–88, 2014.
- [13] C. Networks, "Dissertation Radio Access Network Dimensioning for 3G UMTS " 2009.
- [14] J. Salonen, A. Toskala, and H. Holma, "UMTS Services and Applications," 2002.
- [15] H. Kaaranen, A. Ahtiainen, L Laitinen "UMTS Networks Architecture, Mobility and Services."2001
- [16] C. Kappler, *UMTS NETWORKS AND BEYOND*. .
- [17] D. C. Montgomery and G. C. Runger, "Applied Statistics and Probability for Engineers" Third Edition.
- [18] W. K. Härdle, *Basic Elements of Computational Statistics*.
- [19] Z. Honggang Z. Zhifeng, Li. Rongpeng, Z. Yifan "Stable Distribution Modeling Based on Big Data & User- Behavior"2016:
- [20] "Statistical Models and Methods for Lifetime Data" Second Edition.

-
- [21] V. Pappas, K. Adamidis, S. Loukas, "*A Family of Lifetime Distributions*," vol. 2, 2012.
- [22] N. Matloff, "*From Algorithms to Z-Scores : Probabilistic and Statistical Modeling in Computer Science.*"
- [23] cowan.Glen , "*Statistical Data Analysis*," 1998 .
- [24] A. Mahdavi and D. Kundu, "*A New Method for Generating Distributions with an Application to Exponential Distribution*," pp. 1–25.
- [25] J. Bolding *et al.*,"*Statistical Models and Methods for Life Time Data*," Second Edition.
- [26] R. Rigby, M. Stasinopoulos, G. Heller, and F. De Bastiani, "*Distributions for Modelling Location , Scale and Shape : Using GAMLSS in R*," 2017.
- [27] A. Chikondi and P. Mkolesia, "*Parameter Estimation Methods for Different Probability*," no. March, 2017.
- [28] J. P. Nolan, "*Models for Heavy Tailed Data*," 2018.
- [29] R. Analysis, "*Distribution (Weibull) Fitting*," pp. 1–40.
- [30] *SpringerBriefs in Regional Science.*
- [31] S. A. Ahmed, "*Comparative study of four methods for estimating Weibull parameters for Halabja , Iraq*," vol. 8, no. 5, pp. 186–192, 2013.
- [32] L. Preserve, "*comparing graphical method and a modified method to fit weibull distribution*," 2012.