ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

Link Capacity Dimensioning of Microwave Backhaul Using Real-Time Traffic In the case of the Ethio telecom Network

by: **Daniel Gonfa**

Advisor: **Dr.Yalemzewd Negash**

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

September, 2021

Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

Link Capacity Dimensioning of Microwave Backhaul Using Real-Time Traffic In the case of the Ethio telecom Network

by: **Daniel Gonfa**

Approval by Board of Examiners

	Signature
_____	_____
(Chairman,school of Graduate committe)	
Dr. Yalemzewd Negash	**Signature**
_____	_____
(Advisor)	
Dr.Ing Dereje Hailemariam	**Signature**
_____	_____
(Examiner)	
Dr. Murad Ridwan	**Signature**
_____	_____
(Examiner)	

Date: .....

Addis Ababa, Ethiopia

Abstract

As technology advances, the traffic demand has increased in tandem with the evolution of mobile service types, and prompting mobile operators to enhance their network capacity. Because optimum link capacity planning is expressed in terms of both cost and network performance, operators face a challenge in managing limited network resources such as microwave channel frequency and modulation order. Although the peak capacity of mobile services for provisioning of uplink microwave backhaul has resulted in underutilized links, it is a prevalent practice among service providers due to its ease of installation. Underutilized microwave links required higher transmission power and antenna size due to excessive modulation order, and frequency channel spacing which decreases receiver sensitivity, that affects link reliability. Thus, the fundamental purpose of this study is to optimize underutilized microwave link capacity using a real-time traffic to free up unused resources. Hence, this paper proposed an approach of link capacity dimensioning using Probabilistic approach for busy hour traffic, and SARIMA time series modeling for future traffic trend of sample sites from Addis Ababa. Therefore, the study's key findings include 40% of sites in sample space have been reduced the unused channel frequency by half, 92% of sites optimized their modulation order by one forth, and improvement of overall link percentage utilization from 31% to 60% as minimum, as well as link reliability improvement of the microwave backhaul in particular, comparatively with the existing link capacity using the proposed approach.

key Words: MW Mobile backhaul; real-time link capacity dimensioning; SARIMA Modeling.

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.

Daniel Gonfa Dechassa .....
Signature

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

Dr. Yalemzewd Negash

(Advisor) **Signature**.....

Date: .....

Acknowledgments

First and foremost, I want to praise and thank God, the Almighty, for lavishing me with numerous blessings.

I'd like to express my heartfelt gratitude to my advisor, Dr.Yalemzewed Negash,for all of the support and guidance.

My gratitude goes to Dr.Ing Dereje Hailemariam, and Dr. Murad Ridwan, my examiners, who provided valuable inputs and comments throughout the process.

Finally, I will be eternally grateful to my beloved family for their unwavering support and encouragement throughout my life.

Last but not least, all of my best friends, colleagues, and especially Heaven Girma for friendship and cooperation.

Contents

List of Figures

List of Tables

Acronyms

ACF	Auto Correllation Coefficient	29
AIC	Akaike Information Criterion	31
AR	Auto Regression	29
ARIMA	Auto Regressive Intigrated Moving Average	29
ARMA	Auto Regressive Moving Average	29
BIC	Bayesian Information Criterion	31
BSC	Base Stations Controller	2
BTS	Base Transceiver Station	2
CCDF	Complimantary Cummulative Distribution Function	15
CDF	Cumulative Density Function	10
GEV	General Extreme Value	15
GP	Generalized Pareto	15
GSM	Global System Mobile	4
HSPA	High Speed Packet Access	4
LS	Least Square	15
LTE	Long Term Evolution	4
MA	Moving Average	29

MAPE	Mean Absolute Percentage Error	32
MEA	Mean Absolute Error	32
MLE	Maximum Likelihood Estimation	15
MM	Method of Moments	15
MSE	Mean Squared Error	32
MW	MicroWave	4
PACF	Partial Auto Correllation Coefficient	29
PDF	Probability Distribution Function	10
PDH	Psychlonoues Digital Hierarchy	2
PGW	Packet Getaway	2
Q-Q	Quantile-Quantile	4
RMSE	Root Mean Squared Error	32
RNC	Radio Network Controller	2
SARIMA	Seasonal Auto Regressive Intigrated Moving Average	29
SGW	Serving Gateway	2

1 Introduction

In early 1980 up to now, mobile network evolution is transmuting from the first generation with only voice supporting capability up to fifth generation with multimedia services [4]. This evolutionary change in the sector has pushed the industry and brought an outstanding development in telecommunications technologies. During the evolution, the architectural change of access, transportation, and core of the network has evolved from simple Base Transceiver Stations(BTS) into most complicated eNodeB, from Base Stations Controller(BSC) and Radio Network Controller(RNC) into Serving Getaway(SGW) and Packet Getaway(PGW) in the core, the platform goes up to all IP from TDM and ATM, and the transmission link evolved from simple E1/T1 interface into hundred Gigabit Ethernet [5].

As it is illustrated in Figure 1.1, Backhaul comprises two major parts within the mobile transportation network. The first is often referred to as the last mile or the access, which covers the base transceiver station, that collects user traffic and transports to the near cell site gateway or optical node. The second one is an aggregation, which is used to interlink more cell site gateways for better traffic convergence and plays an intermediate role between access and core networks. It is noted that the transmission

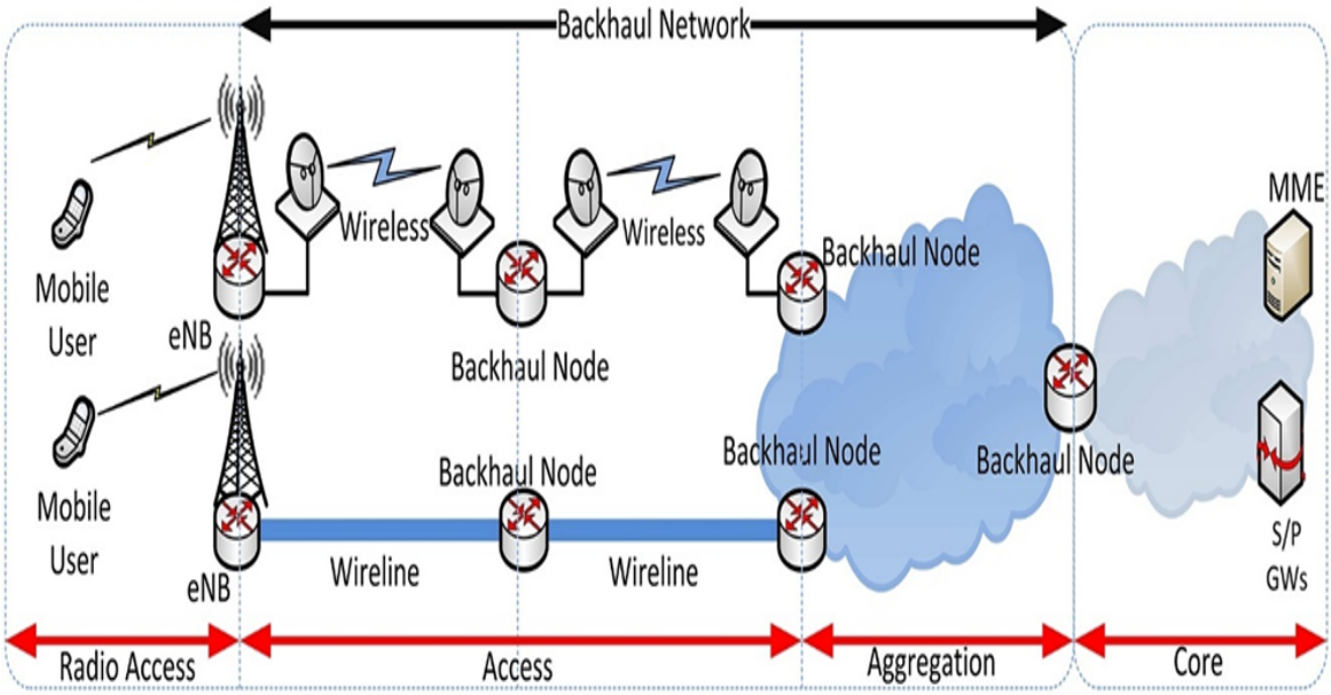


Figure 1.1: Microwave and fixed backhaul architecture [1].

technology that carries the traffic between network devices progressed as the network hierarchy changes from access to aggregations. In backhaul transmission medium selections, the choice is mainly dependent on cost, ease of deployment, and throughput requirement of the service at the access network. Thus, Service providers must consider their pros and cons for selecting which transmission technology is suitable for their network. Most providers, working in developing countries where fiber cables are not available or in dense urban areas where laying fiber is expensive, microwave access technology comes forth as a backhaul access transport solution. Therefore, it is not by chance, Ethio telecom has more than 6500 microwave access devices for supporting 7500 BTS sites across the country [6].

Link capacity dimensioning of microwave backhaul transmission is mainly dependent on channel spacing and modulation order, which are related to the cost of deployment by keeping interference and propagation loss constant. It is also considered traffic coming from cellular access network as an input, which constitutes signaling and user traffic. Once, radio network planning has been prepared based on the requirement of respective capacity, the aggregated traffic per mobile site is delivered to microwave capacity planning. Using the link capacity, the microwave backhaul planning process proceeds with channel spacing, modulations order, frequency planning, link budgeting,

interference management, and reliability to satisfy the uplink requirement of the mobile network.

1.1 Statement of the problem

Mobile capacity planning is mainly dependent on customer density, future expansion of the area, customer profile, and marketing strategy [6]. But for microwave backhaul, most providers dimension the link capacity using the aggregated peak traffic. In the process, the network contains links that are optimally dimensioned, if the traffic from the customer side is generated as it would be expected. Otherwise, congested and underutilized links will be part of the network (Figure 1.2). Congested links will be upgraded to the next level of link capacity, whereas underutilized links are considered normal by most providers, even though they are a source of resource wastage. On the other hand, during expansion, it is common practice to add link capacity based on mobile technology peak traffic capacity requirements regardless of the current utilization of the existing link, which leads to unwanted license costs.

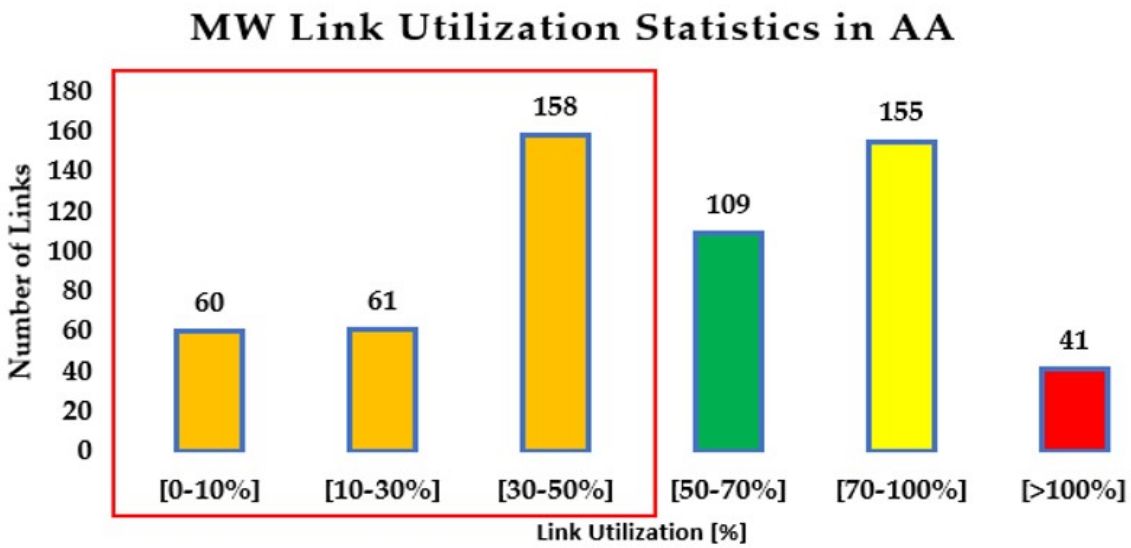


Figure 1.2: Microwave backhaul link utilization statistics[2].

Therefore, unwanted cost and wastage of resources can be minimized by optimum capacity dimensioning of the link using statistical characteristics of the traffic instead

of the rule of thumb (peak traffic dimensioning). This study attempts to show how the real-time traffic link capacity dimensioning method is useful for solving resource wastage caused by underutilized links by considering busy hour traffic (to find the peak link capacity), and studying the traffic trend (for the gradual traffic growth), to find the optimum link capacity for re-dimensioning of the link of the microwave mobile backhaul.

1.2 Objective

1.2.1 General objective

The main objective of this thesis is to dimension the link capacity of the microwave backhaul network using a probabilistic and SARIMA time-series provisioning approach based on real-time traffic to optimize the underutilized links and free up scarce resources of sample sites in Addis Ababa.

1.2.2 Specific objectives

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

1. Study the traffic behavior at the input side of microwave access side.
2. Identify the most suited traffic model for busy hour using probabilistic traffic modeling approach.
3. Extract the rate of change of the mean (Trend) of the traffic using time series analysis.
4. Dimension the microwave link capacity by combining the busy hour traffic with traffic trend and free up the unused link capacity resources.
5. Compare the proposed with the existing link capacity in terms of the capacity and related link performances.

1.3 Literature revision

Microwave link capacity is dependent on the dynamic channel characteristics of air and is limited in link capacity[7]. Hence, optimum link dimensioning is vital for improving network performance and minimizing implementation costs. Therefore, this topic aims to cover literature that are closely related to link dimensioning and optimum bandwidth provisioning of the mobile microwave backhaul.

Link dimensioning for Internet traffic to transport different services such as voice, data, and video has been studied from different perspectives and purposes. The main idea of [8] was the dimensioning of link capacity for Internet traffic using real-time data, and it demonstrates how to model Internet traffic at low aggregation time. The study took more than five different locations to collect traffic traces and justified that, the traffic at small aggregation time acquired self-similarity and traffic burstiness. The study confirmed, self-similarity and traffic burstiness manifests the heavy tail nature of the traffic, especially at the right tail of the distribution. As a result, the study's finding suggested, link dimensioning using heavy-tailed distributions such as Generalized Extreme Value (GEV) and Lognormal distributions is more appropriate for Internet traffic than Gaussian distribution at small aggregation time.

A. Pras,L.Nieuwenhuis, R.van de Meent, and M.Mandjes argued in [9] that using the rule of thumb to dimension the Internet link causes over-dimensioning, because it uses mean traffic and 30% addition as traffic margin. Hence, they suggested, an investigation of the traffic statistical characteristics such as mean, variance, aggregation time, and the percentage of tolerable loss. Based on the analysis, the formula has been devised using the above four required parameters of the measured traffic characterized by Gaussian distribution, and their result suggest that dimensioning using real-time traffic is superior to the rule of thumb.

The preceding two studies looked at how to represent Internet traffic using a probabilistic approach and traffic burstiness at low aggregation time. However, in [10] the focus was on modeling the microwave traffic using time series analysis by comparing different forecasting techniques, such as SARIMA, AR, and Multi-Layer Perceptron Neural Network to predict future traffic and dimension the link capacity accordingly. The Multi-layer Perceptron and SARIMA modeling were found to be a better models

to find the link capacity and minimized a number of underutilized microwave links. Thus, The above two approaches only considers the peak traffic modeling during the busy hour, whereas the third approach forecasts the traffic based on the traffic trend change. Therefore, taking the advantage of busy hour traffic to find the maximum link capacity, and time series modeling for future gradual traffic change to dimension the link capacity is an approach, this study follows.

1.4 Methodology

Methodologies followed to achieve the general and specific objective of this thesis:

1. Related works revision.
2. Data collection from the Network Element Server (NMS).
3. Model the traffic of busy hour and specify the link capacity for microwave using MATLAB.
4. Investigate the trend of the traffic using time series analysis using Python SARIMA Modeling.
5. Obtain the final link capacity of the microwave for re-dimensioning of the under-utilized links.

1.5 Scope and Limitations

1.5.1 Scope of the Thesis

In this research, a mathematical model based on a probabilistic approach has been developed, for the traffic measured at the busy hour of a one-second interval, to provision the link capacity by considering the blocking probability as a quality-of-service parameter. For gradual traffic change or trend, the time series SARIMA modeling has applied using five months of monitored traffic at one-hour measured granularity. As the research is a case study, Addis Ababa city's microwave backhaul of 12 sites' traffic is considered as sample space among 584 link, and the modeling is performed for aggregate traffic regardless of the service type.

1.5.2 Limitations of the Thesis

The following were the main limitations in conducting this research.

1. The dataset monitored for time series has missing traffic information for up to two days for some sites; as a result, the sample space domain is narrowed down to 12 sites and single uplink links are only considered by neglecting the hub sites.

2. Previous studies on mobile microwave backhaul link capacity dimensioning using real-time approach were either more focused on forecasting using time series alone or on link capacity dimensioning of Internet traffic during peak hours. Therefore, the lack of similar papers that are related to this research was other limitation.

1.6 Thesis contributions

In [8, 9], link dimensioning for Internet traffic was done using real-time data. However, to my knowledge, only the referenced work study focuses on campus and other Internet users. This thesis work can contribute to the service providers to re-dimension of underutilized microwave backhaul links based on the real time traffic and also helping to manage their channel frequency usage by freeing them during the process. It also assist the link capacity planner to decide the next optimum link capacity during upgrading of congested links.

1.7 Thesis organization

The remainder of this study is organized in the following manner. The second chapter explores the basic theory of probability, focusing on the types of distribution functions and curve-fitting techniques such as the quantile-quantile plot and auto-correlation coefficients. Link capacity provisioning approaches, link transparency theory, and parameter estimations for distributions are all covered in the third chapter. Time series analysis, types of models, performance evaluation metrics, and forecasting are briefly reviewed in Chapter four. Microwave link dimensioning and the link parameters that can be affected by link capacity change are revised in Chapter five. In Chapter six, the

obtained results and observations from the study are explained. Finally, Chapter seven presents the conclusion and recommendations for future work.

2

Probability theory

This chapter introduces the basic concepts and notations that are employed throughout the thesis. The chapter starts by defining the general idea of probability distributions with continuous random variables. Subsequently, it deals with heavy-tailed continuous probability distributions. This section also introduce the important characteristics of the distributions, such as the Probability Density Function (PDF) and the Cumulative Density Function (CDF). The penultimate section briefs the techniques for fitting empirical data with the best suitable probability distribution, and finally summarizes the above points.

2.1 General probability theory

The subject matter of probability deals with uncertainty, it can be a numerical measure of the likelihood that a particular event will occur (i,e ,the value varies from 0 to 1) [11]. A numerical description of an outcome of an event can be represented by a random variable, and the distribution of this variable over the domain of sample space is called

a ***Probability Distribution Function(PDF)***. Mathematically expressed as: [11]

$$P[a \leq X \leq b] = \int_{a}^{b} f(x)dx \tag{2.1}$$

Where the interval [a b] is an area of coverage of the PDF of continues function *f(x)*.
Cumulative Distributions Function (CDF) is defined as the summation of the areas of the given probability density function of the random variable X [11].

$$F(x) = P[X \leq x] = \int_{-\infty}^{x} f(z)dz \tag{2.2}$$

The first moment or Mean is an expected value of the random variable and measures the central tendency of a given data. Whereas ***Variance*** is the measure of the average distance of samples from the mean. Mathematically, the mean and variance of a continues random variable of the certain outcome expressed as:

$$\mu_X = E(X) = \int_{-\infty}^{\infty} xf(x)dx \tag{2.3}$$

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x-\mu_x)^2 f(x)dx \tag{2.4}$$

Where μ_X and σ_X^2 are mean and variance of random variable X.
Skewness is the parameter that used to measure of the non-symmetric nature of the probability distribution of a random variable about its mean. Its value can be positive (if the tail is on the right side of the mean), zero (symmetric with respect to the mean), and negative (if the tail is on the left of the mean). On the other hand, the probability distribution of a real-valued random variable's tailedness characterized by ***Kurtosis*** [11]. Like skewness, kurtosis describes the shape of a probability distribution.
NB: All the PDF, CDF, mean and variance of the random variable mentioned above assumed, the random variable X is a continuous random variable, otherwise all the integration changes to summation, and the density function changes to a probability mass function if the variable is discrete.

2.2 Heavy-tailed distributions

Heavy-tailed distributions have tails that are not exponentially bounded; or the distribution curve at the edge decays slower than exponential distributions [12]. Example of such distribution are Lognormal, Generalized Extreme Value(GEV) and Generalized Pareto(GP) distributions and mathematically characterized by:

$$P(X > x) = (1 - F(x)) > e^{(-\alpha x)} \tag{2.5}$$

Where : α is the exponential distribution's rate parameter, which is always positive, and X is a random variable. Graphically, the tail of the heavy tail distribution slowly decays

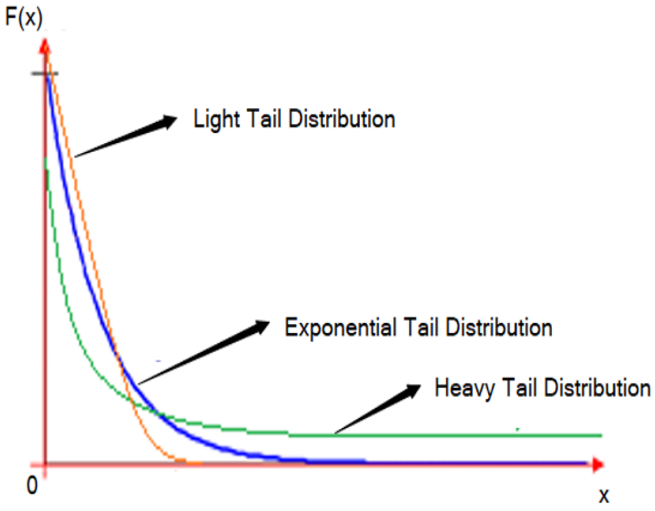


Figure 2.1: Tailed distribution comparison[11].

compared to light tailed and exponential distribution as shown in Figure 2.1.

2.2.1 Generalized pareto distribution

It is a family of continuous distributions which is often used to fit the exceedance of a certain threshold or the tail of another distribution which can be specified by three parameters (locations μ, scale σ, and shape ξ)[12]. The exponential and Pareto distributions are special cases. The kurtosis and skewness are only dependent on the shape parameter of General pareto, which implies the tailedness nature determined by the shape parameter ξ (i,e ,referred as the tail index). This distribution is applied in the field of engineering computation for reliability and is also used for the analysis of envi-

ronmental extreme events. The density function is defined as:

$$f(x) = \begin{cases} (1+\xi(\frac{x-\mu}{\sigma})^{-(\frac{\xi+1}{\xi})}; \ \xi \neq 0 \\ e^{-(\frac{x-\mu}{\sigma})}; \xi = 0 \end{cases} \tag{2.6}$$

where x represents the variable of interest or random data, μ represents the location, ξ represents the tail index, and σ represents the distribution's scale.

2.2.2 Lognormal distribution

It is a Gaussian distribution, but the random variable is quantified by the natural logarithm of x. It is defined only as positive real values and is applicable for engineering, science, and modeling of Internet traffic [8]. Its variance is the parameter that determines the kurtosis and skewness nature of the distribution. The Lognormal probability density function expressed as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp^{-\frac{1}{2}(\frac{\ln x-\mu}{\sigma})^2} \tag{2.7}$$

where x is a random variable, μ is the mean of the natural logarithm of x.(i.e.$\ln x$)

2.2.3 Generalized extreme values distributions

A distribution developed by extreme value theory which combines, the type I (Gumbel), the type II (Fréchet), and type III(Weibull) extreme value distributions. The distributions functions defined as GEV (μ ,σ ,ξ),where (μ ε R),($\sigma > 0$), and($-\infty < \xi < \infty$), defined the location,scale and shape parameter respectively. Besides the determination of heavy tail nature,the value of shape defines whether the distribution is Weibull, (for $\xi < 0$,) Fréchet (for $\xi > 0$) and Gumbel (for $\xi = 0$). GEV distribution can apply for estimation of the extreme nature of the statistics such as temperature, and risk analysis[13]. The probability density function mathematically expressed as:

$$f(x) = \frac{1}{\sigma}(A(x))^{\xi+1}e^{-A(x)} \tag{2.8}$$

Where:

$$A(x) = \left\{ \begin{array}{l} (1+\xi(\frac{x-\mu}{\sigma}))^{-(\frac{1}{\xi})}; \xi \neq 0 \\ e^{-(\frac{x-\mu}{\sigma})}; \xi = 0 \end{array} \right\} \tag{2.9}$$

2.3 Goodness of fit plots

In statistics, fitting is the process of matching empirical data to a known probability distribution. Goodness of fit is a measure of fitness and can be done either using an analytical approach such as the Kolmogorov-Smirnov test (for continuous observed data) [14] or the Chi-square test (for categorical random data), or by using graphical comparison of the observed data with a known probability distribution (Q-Q and Probability density plots). Hence, the subsequent section discuss about probability density and quantile quantile plot.

2.3.1 Probability density comparison plot

This fitting method (Figure 2.2) is done by plotting a histogram plot of the observed data against a density function of the selected distribution. It is a very informative comparison plot and helps to see where the main discrepancies have occurred between the empirical data and the nominated density function.

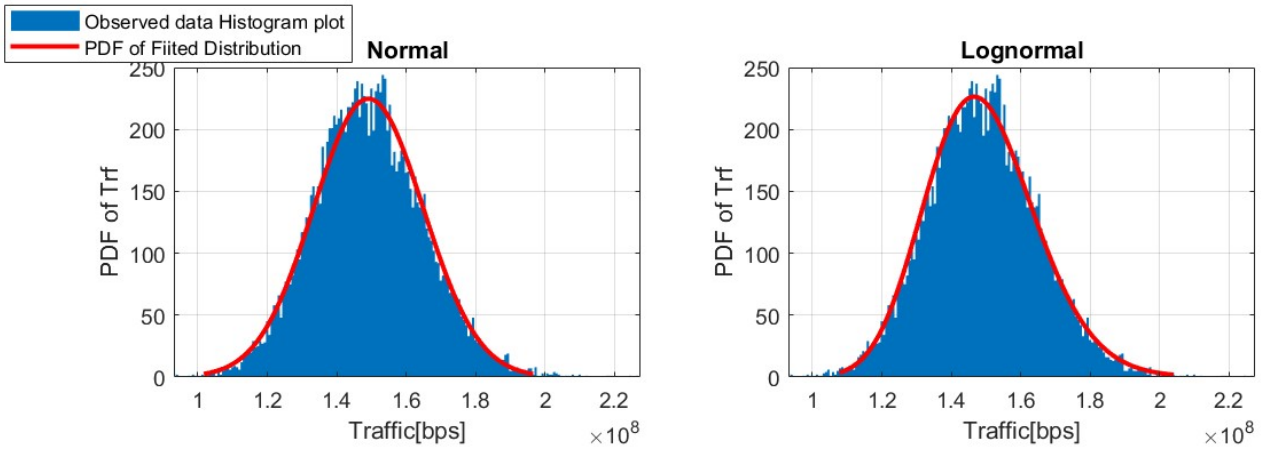


Figure 2.2: Example of probability density function plot.

2.3.2 Quantile-quantile plots

The Q-Q plot (Figure 2.3) is a powerful visualization plot, which is used to evaluate the degree of fitness between selected distributions and observed data. The graph of the plot can be done using coordinates of (x_i, s_i)[15].
where , $x_i = F^{(-1)}(\frac{i}{n+1})$ which is the Complimentary CDF of distribution F, and s_i is the value of i^{th} sample from n number of observed data. The scattered points of the two quantiles follow a straight fitting line, if and only if, every pair of(x_i, s_i) matched in the sample space.

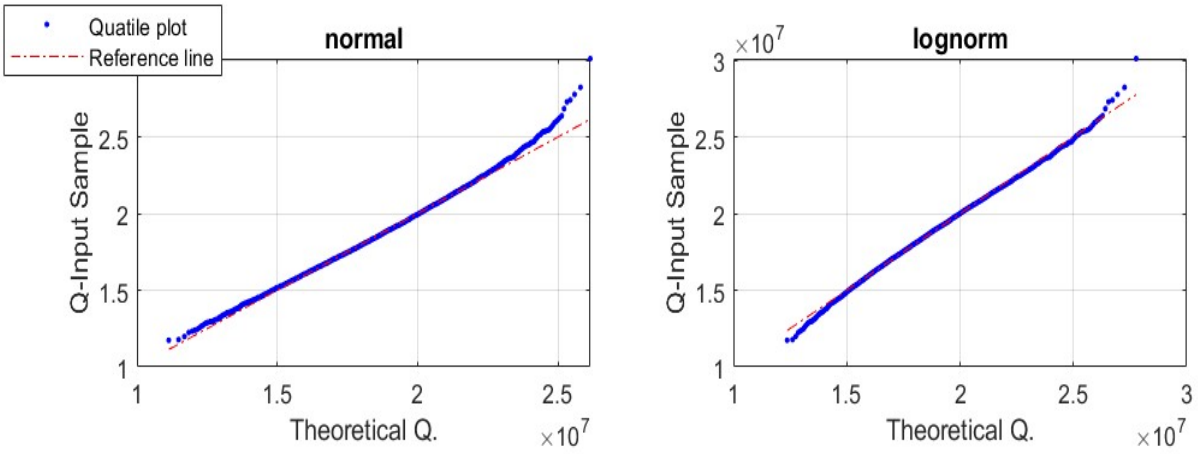


Figure 2.3: Example of quantile quantile plot.

2.3.3 Linear correlation coefficient

Apart from the visualization test, the linear correlation coefficient is used to determine how the observed data fits with a selected probability distribution. In probability theory, covariance is a measure of the joint probability of two random variables, which shows the directions of the relationship either in tandem or in opposite directions [12]. On the other hand, the magnitude of normalized linear correlations of the two random variables (ranging from -1 to 1) can express the degree of relations among the observed data and the intended distribution. Therefore, the normalized correlation of observed data with

known probability distributions is defined as:

$$\gamma = \frac{Cov(s_i, x_i)}{\sigma_s \sigma_x} = \frac{\sum_{i=1}^{n} (s_i - \mu_s)(x_i - \mu_x)}{\sqrt{\sum_{i=1}^{n} (s_i - \mu_s)^2 \sum_{i=1}^{n} (x_i - \mu_x)^2}} \tag{2.10}$$

where $\mu_s = \frac{1}{n} \sum_{i=1}^{n} s_i$, $\mu_x = \frac{1}{n} \sum_{i=1}^{n} x_i$, and n is the number of samples. The value of $\gamma \, \varepsilon (0, 1]$, signifies the two variables are positively correlated and the similarity increases as the value reaching to 1.

Among other curve-fitting approaches, the two visual techniques (Q-Q and Probability Density Comparison plot) and linear correlation coefficient ($\gamma \geq 0.95$), have been selected for this study because of their familiarity and ease of implementation.

2.4 Parameter estimations techniques

Parameter estimation is a method in statistics that uses sample data to estimate the parameters of the known distribution. Empirical data can be modeled by deterministic or known mathematical equations, unless it is random. However, most Internet traffic, by its nature, is completely random and treated by a probabilistic modeling approach [16]. Fitting the observed data with the selected probability distribution using parameter estimation techniques such as Method of Moments (MM), Least Square (LS), and Maximum Likelihood Estimation (MLE) is required before selecting the probability distribution. Because of the property of minimum variance as the sample size increases, and several popular statistical software packages (including MATLAB) provide an algorithm for maximum likelihood estimation, and therefore, MLE has been selected in this study. Thus, a brief mathematical definition of MLE is explained as follows:
Given a set of random data that is both independent and identically distributed X, the unknown parameter of the distribution, θ and the probability distribution f(x). Therefore, the likelihood equation of the random variable X for unknown parameter θ is expressed as:

$$L(\theta | x_i) = \prod_{i}^{n} f(x_i | \theta) \tag{2.11}$$

Taking the partial differentiation with respect to the parameter θ and equating it to zero will yield the value of each parameter of the probability distributions, for example, the Gaussian distribution parameters are mean and standard deviation. The mathematical work is not as simple as it appears; hence, it is better to find the parameters of the distribution by using MLE as an estimation algorithm and a MATLAB, R, or Python fitting distribution package. Once these parameters of the distribution have found, generating the random variable using CCDF of the distribution for curve fitting and link capacity provisioning is possible.

In summary, the probability distribution is a distribution that describes the behavior of random variables in the sample space. It can be classified as discrete or continuous, and some are heavy-tailed, such as Lognormal, Generalized Pareto, Weibull, and GEV. A Q-Q plot, probability density plot, and correlation coefficient comparison are the approaches to find the best fitting distribution for the observed data. MLE is used for the probability distribution parameter estimation which is built in statistical tools packages of MATLAB throughout this study.

3

Link capacity provisioning

Provisioning of the link in communication is a process of dimensioning the link capacity of the network by keeping the required quality of service [17]. In a broad sense, there are two types of link provisioning: real time traffic provisioning, which is a reactive type of dimensioning that is applied for link optimizations, and capacity upgrading. The proactive approach uses planning parameters such as customer and traffic profile, and marketing strategy as an input. The main objective of this chapter is to introduce link capacity provisioning approaches using real-time traffic (Gaussian, Meent's formula, and Heavy-Tail) provisioning approaches in detail, and provisioning terminology such as link transparency, and blocking probability.

3.1 Link capacity provisioning using real-time traffic

It is standard practice to use the peak traffic of a mobile service for dimensioning the uplink of the mobile sites in green areas where no traffic information (such as customer profile) is available [17]. But it may create underutilized links and required link optimization after installation using real-time traffic especially, for microwave backhaul

where link capacity is related to scarce resources. Link dimensioning should consider the quality of service such as variable delay or jitter, latency, packet loss. Such parameters are defined per service level. Whereas in this study the traffic data was collected in an aggregated form regardless of service type. Therefore, blocking probability has been selected as a traffic quality parameter.

3.1.1 Blocking probability

Blocking is a telecommunications term that refers to the available channel's insufficiency for the given traffic within the specified period. Alternatively, when traffic exceeds the available channel capacity and causes a packet drop, this indicates that service is blocked, and its value is measured in percent [8, 18]. In order to function effectively, every telecom service has a packet loss threshold, according to [3], packet loss of less than 0.3% is preferred for Long Term Evolution(LTE) services in microwave mobile backhaul, whereas IPTV requires less than 0.1%. Voice services, on the other hand, typically have a packet loss or blocking probability of 0.1%. As a result, a blocking probability of 0.01% (99.99% of service passage) is used in this study to dimension the maximum possible link capacity. The required maximum packet loss of respective mobile services is summarized in Table 3.1.

Table 3.1: Backhaul requirement of services from users' perspective[3]

Performance(KPI)	GSM	WCDMA	4G
Delay	Max:40ms	Max:30ms	Max:20ms
Jitter	Max:10ms	Max:10ms	Max:2ms
Packet loss	Max:10^{-3}	Max:10^{-3}	Max:10^{-3}

3.1.2 Link transparency

Link transparency is a networking term that describes a link's ability to send data through a network transparently, or it is a link property that allows users of a certain application to access remote resources as if they were local. Mathematically, link transparency can be defined as the likelihood of the given traffic A(T) exceeding the link

capacity C at the given time T is less than or equal to the quality control parameter ε.

$$P(A(T) \geq CT) \leq \varepsilon \tag{3.1}$$

Where T is the aggregation time, in which the trace of the traffic has been collected.

Until now, it has been discussing how to fit observed data with probability distributions (Chapter 2), link transparency, and blocking probability (for securing the quality of service of the link). The next step is to formulate a link capacity dimensioning equation based on the distributions that have been defined in the literature, [8, 19, 20], stating that *Internet traffic is mostly characterized by Gaussian and heavy-tailed distributions such as (GEV, Lognormal)*, and mobile traffic has already evolved from voice to full data communication, hence it is convincing to treat mobile backhaul traffic similarly, besides the curve fitting techniques' confirmation. Thus, in the following section, the provisioning of the link capacity based on these distributions has been discussed.

3.1.3 Gaussian distribution provisioning approach

One of the candidate distributions that might fit with the Internet data for provisioning is Gaussian. Therefore, under the Gaussianity behavior, the traffic A(T) is represented by A(T) $\sim N\ [\mu T,\ \sigma(T)]$, μT and $\sigma(T)$ are mean and standard deviation of the observed traffic of A(T)[bits] at the aggregation time T[second]. Graphically, shown in Figure 3.1 with link capacity representations of C[bps] at the tail of the distribution. For elaborating further, once the given data is represented by such distribution (Q-Q plot and Auto correlation's coefficient ($\gamma \geq 0.95$)) then, the traffic A(T) can be represented by the following probability density function.

$$A(T) \cong \frac{1}{\sigma(T)\sqrt{2\pi}} e^{-\frac{(A(T)-\mu T)^2}{2\sigma^2(T)}} \tag{3.2}$$

Transforming into normal standard distribution using Z parameter shown in Figure 3.1, taking the z value of Equation (3.3) and substituted in place of A(T), then link transparency equation becomes:

$$z = \frac{A(T) - \mu T}{\sigma(T)} \tag{3.3}$$

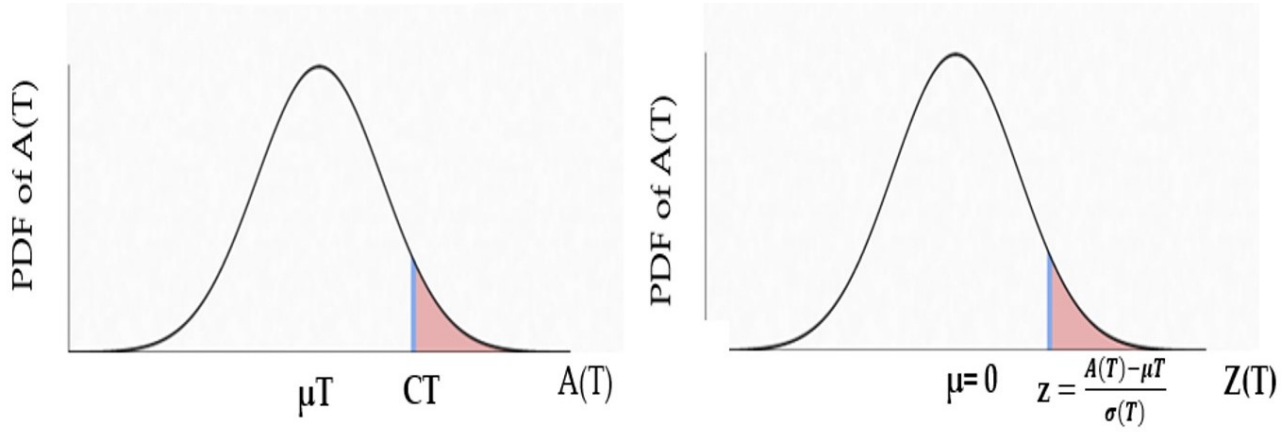


Figure 3.1: PDF of Gaussian distributions with standardization.

$$P(z\sigma(T)+\mu T \geq CT) \leq \varepsilon \tag{3.4}$$

the mathematical procedure for finding the link provisioning formulla using Guassian distribution follows from Equation (3.5) to (3.7):

$$1-P(z \leq \frac{CT-\mu T}{\sigma(T)}) \leq \varepsilon \tag{3.5}$$

$$(1-\varepsilon) \leq P(z \leq \frac{CT-\mu T}{\sigma(T)}) \tag{3.6}$$

$$\Phi^{-1}(1-\varepsilon) = \frac{CT-\mu T}{\sigma(T)} \tag{3.7}$$

The capacity of the link, that satisfy the quality of service parameter ε for offered traffic A(T) at an aggregation time T calculated as:

$$C = \mu + \frac{1}{T}\Phi^{-1}(1-\varepsilon)\sigma(T) \tag{3.8}$$

Where C is capacity of the link [bits/sec], μ is average traffic in [bits/sec], ε is quality of service parameter in decimal, $\sigma(T)$ is traffic standard deviation in [bits] , Φ^{-1} is the inverse of CDF of Gaussian distribution.

From Equations (3.8), link capacity depends on mean traffic and the additional value that accounts for a deviation from the mean (at the right tail of the distribution). As the value of blocking probability goes to zero, the link will offer the maximum traffic which the approach allows, whereas, when the blocking probability is equal to 0.5, the link capacity can handle the mean traffic only.

3.1.4 Meent's link capacity provisioning approach

This approach was mainly done by Remco van de Meent[18], His idea started by assuming the Internet traffics is fairly Gaussian. The initial concept was derived from the Markov inequality and finalized by Chernoff upper bound conditions. Based on the analysis, given any non-negative random variable X, located at the positive tail of the distribution, and the probability of X which is greater than the positive value A is bounded by the magnitude of the expectations of X divided by the value of A. This is

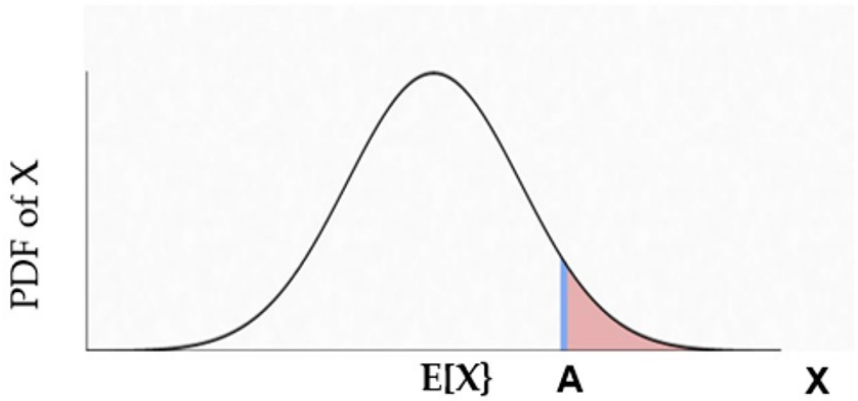


Figure 3.2: Markov inequality.

called, a Markov inequality and is mathematically expressed as:

$$P(X \geq A) \leq \frac{E(X)}{A} \tag{3.9}$$

Applying the chernoff upper bound condition, Equation (3.9) need to be exponential function by selecting the positive value θ (changing the analysis using Moment generating function instead of first moment or Mean), and combining the boundary condition with link transparency, the link provisioning equation becomes:

$$P(A(T) \geq CT) = P(e^{\theta A(T)} \geq e^{\theta CT}) \leq E(e^{\theta A(T) - \theta CT}) \tag{3.10}$$

where $E(e^{\theta A(T)})$ is moment generating function of the given traffic of A(T).
Since the value of $\theta \geq 0$ using Chernoff upper bound condition, there exist at least one θ to hold Equation (3.10) and the link transparency equation of the given traffic A(T)

becomes:

$$\exists\theta : E(e^{\theta A(T)-\theta CT}) \leq \varepsilon \tag{3.11}$$

applying the natural logarithm at both sides of Equations (3.11) and taking link capacity C to the left by rearranging, then link capacity equation becomes:

$$\exists\theta : C \geq \frac{\log E(e^{\theta A(T)}) - \log \varepsilon}{\theta T} \tag{3.12}$$

Hence, by keeping the performance requirement of a given link (blocking probability) at aggregation time T, the generic bandwidth provisioning approach that Meent's has proposed defined as:

$$C(T,\varepsilon) \geq \min_{\theta\geq 0} \frac{\log E(e^{\theta A(T)}) - \log \varepsilon}{\theta T} \tag{3.13}$$

The observed traffic A(T) is assumed to be Gaussian, with N[0, 1], using the standardized normal transformation (for mathematical simplicity). Therefore, the moment generating function is defined for zero mean and variance of one as:

$$E(e^{\theta Z}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\theta Z} e^{-\frac{1}{2}(Z^2)} dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Z-\theta)^2} e^{\frac{1}{2}\theta^2} dz = e^{\frac{1}{2}\theta^2} \tag{3.14}$$

Keep in mind that:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Z-\theta)^2} dz = 1 \tag{3.15}$$

But the traffic A(T) defined as:

$$A(T) = \mu\, T + Z\sigma(T) \tag{3.16}$$

by substituting A(T) of Equation (3.16) in place of Z in Equation (3.14) and equate it in the same way, the moment generating function of the Gaussian fitted observed traffic becomes:

$$E(e^{\theta A(T)}) = e^{\frac{1}{2}\theta^2\sigma^2 + \mu T\theta} \tag{3.17}$$

taking the logarithim at both sides

$$\log E(e^{\theta A(T)}) = \frac{1}{2}\theta^2\sigma^2 + \mu T\theta \tag{3.18}$$

By inserting Equation (3.18) into Equation (3.13), then generic link provisioning equation becomes:

$$C(T,\varepsilon) \geq \min_{\theta\geq 0} \frac{\log E(e^{\theta A(T)}) - \log\varepsilon}{\theta T} = \min_{\theta\geq 0} \frac{\frac{1}{2}\theta^2\sigma^2 + \mu T\theta - \log\varepsilon}{\theta T} \tag{3.19}$$

using the minimium value of $\theta = \sqrt{\frac{-2\log\varepsilon}{\sigma^2}}$ that satisfied the boundary condition in Equation (3.19), the final link provisioning of Meent's approach defined as:

$$C(T,\varepsilon) = \mu(T) + \frac{1}{T}\sqrt{(-2\log\varepsilon)\sigma^2(\,T)} \tag{3.20}$$

From Equations (3.20), it is clear that the link capacity is primarily determined by the mean, and the remaining expression takes into account for traffic burstiness that is determined by aggregation time T, blocking probability ε , and traffic variance.

3.1.5 Heavy-tailed distribution provisioning approach

In traffic measurement, the probability of links reaching peak value is quite low, and it only happens once in a while, otherwise the traffic is constant. These extreme link capacities are located at the positive tail of the distributions in the histogram plot. Hence, compared to exponentially bounded distributions, the addition of this extremeness makes the tail of the distribution heavier or fatter. Therefore, this type of traffic is treated by extreme value distributions such as Generalized Pareto (GP) and Generalized Extreme Value (GEV). Thus, in this study, the GEV, GP, and Lognormal link capacity provisioning approaches have been chosen to represent the heavy tail nature of the traffic for link capacity provisioning. Because the Lognormal distribution is easy to demonstrate mathematically, the provisioning formula for heavy tail distribution with link transparency equation, expressed as follows:

$$P(A(T) \geq CT) = \frac{1}{\sqrt{2\pi}\sigma(T)} \int_{CT}^{\infty} e^{-\frac{1}{2}\left(\frac{\log(A(T))-\mu T}{\sigma(T)}\right)^2} dA \tag{3.21}$$

$$P(\frac{A(T)}{T} \geq C) = \int_{C}^{\infty} \frac{1}{\sqrt{2\pi}\sigma(T)} e^{-\frac{1}{2}\left(\frac{\log\frac{A(T)}{T}-\mu T}{\sigma(T)}\right)^2} dA \tag{3.22}$$

$$P(\frac{A(T)}{T} \geq C) = 1 - \int_{-\infty}^{C} \frac{1}{\sqrt{2\pi}\sigma(T)} e^{-\frac{1}{2}\left(\frac{\log\frac{A(T)}{T}-\mu T}{\sigma(T)}\right)^2} dA \tag{3.23}$$

$$F(C) = \int_{-\infty}^{C} \frac{1}{\sqrt{2\pi}\sigma(T)} e^{-\frac{1}{2}\left(\frac{\log\frac{A(T)}{T}-\mu T}{\sigma(T)}\right)^2} dA \tag{3.24}$$

using the link transparacy equation

$$P(\frac{A(T)}{T} \geq C) \leq \varepsilon = 1 - F(C) \leq \varepsilon \tag{3.25}$$

Therefore, based on the CDF of a given traffic A(T) and the blocking probability,the link capacity C, using Lognormal provisioning approach defined as:

$$C = F^{-1}(1-\varepsilon) \tag{3.26}$$

NB: The above approach also applied for Generalized Extreme Value(GEV), and Generalized Pareto(GP). In general, the provisioning of link dimensioning using heavy-tailed probability distribution can be performed by following the basic procedure shown in Figure 3.3 below.

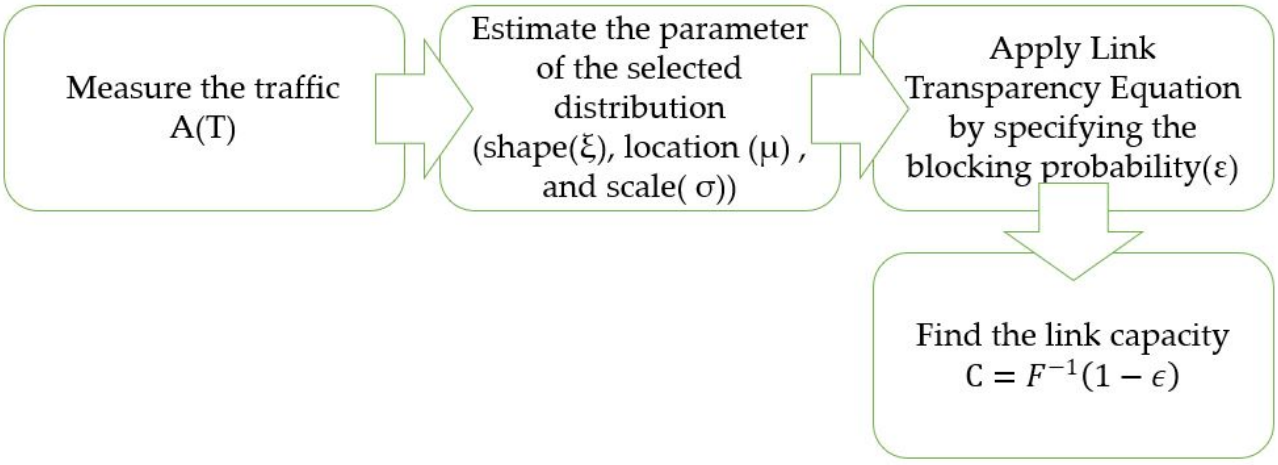


Figure 3.3: Link provisioning for heavy-tailed distributions.

3.2 Summary of link capacity provisioning

So far, it has been emphasized on link provisioning approaches based on Gaussian, Meent's formula, and heavy-tailed distributions. The goal of the link transparency equation with blocking probability is to dimension the link optimally under quality of service criteria, and the major principle behind all provisioning approaches is to represent the extreme value of the distribution, which implicitly represents the peak link capacity.

4

Time series analysis

Time series analysis is a dynamic study of field that has captivated the interest of academics in recent decades. The primary goal of time series modeling is to collect and extensively analyze past data observations in order to build an appropriate model to predict the future [21]. This chapter's main objective focuses on time series analysis, especially SARIMA modeling and evaluation metrics.

4.1 General definition of time series

A time series is a series of data points that are measured over successive periods. It is defined as a set of vectors x(t), t = 0,1,2,..., where x (t) represents a stochastic process at an elapsed of the chronological order time t. A time series is univariate if it contains only one variable; otherwise, it is multivariate and might be continuous or discrete. Observations in a continuous time series are measured at every instance of time, whereas in a discrete series, the measurement is defined as a time interval with consecutive observations of equally separated periods, such as hourly, daily, weekly, monthly, or yearly [21].

4.2 Components of time series

In general, the time series is defined by four main components, which can be determined from the observed data. These are ***Trends***, ***Cyclical, Seasonal***, and ***Random*** components. The Trend is a tendency of a time series that shows the long-run change of the series, either increasing, decreasing, or constant throughout the series, or it can be considered as the rate of change of mean over time. Detrending is the process of removing a trend from the time series.

The Seasonal component is the second component of time series decomposition, and it is a periodical fluctuation that shows the same pattern at a regular interval of time. A Cyclic component of a time series is created by circumstances that cyclically repeat themselves, and the duration might go from a single period of an hour to a yearly basis. This type of redundancy rarely happens, compared to seasonality. Last but not least, the Random component is considered to be a residual part, after the above three components have been removed from the observed data, and it is completely undetermined because of it's stochastic nature.

4.3 Decomposition of the time series

Once the time series components have been discovered, decomposing the observed data into the above mentioned four components is highly useful before detailed time series analysis. Hence, time series decomposition can be done in either a multiplicative or additive manner. When the decomposition is multiplicative, it indicates that the four series components are not independent or there is non-linearity among the components. The additive decomposition approach, on the other hand, demands non dependency among the four components and requires a linear relationship. The two decomposition approaches mathematically given by [21]:

$$Y(t) = T(t).S(t).C(t).I(t) \tag{4.1}$$

$$Y(t) = T(t) + S(t) + C(t) + I(t) \tag{4.2}$$

where Y(t) is the observation and T(t), S(t) C(t), and I(t) are the Trend, Seasonal, Cyclical, and Irregular variation of the time series respectively.

4.4 Stationarity of time series

Time series variables are commonly assumed to be independent and identically distributed random variables that can be defined by a specific probability distribution. Or, the statistical parameters (mean and variance) are independent of time. Mathematically

$$E(y(t)) = E(y(t-k)) \tag{4.3}$$

$$\text{var}(y(t)) = \text{var}(y(t-k)) \tag{4.4}$$

Where y(t) is the value of the series at time t and y(t-k) is the value of the series k lags time steps.
However, the series is not exactly independent, identically distributed; rather, it follows a regular pattern in the periodic intervals (it has seasonality). Thus, referring Equations (4.3) and (4.4) to define the time series leads to accepting strong stationary properties, which is not feasible in practice. Instead, assuming as weakly stationary series, if and only if the observed data meet the following preconditions.

- The mean $\mu(t)$ of the series is constant and does not depend on time.
- The autocovariance function only depends on the time lag k.

4.4.1 Stationarity test of time series

Though the presence of wide sense stationarity in a time series is obvious in most cases, it is better to verify its validity before proceeding with the analysis. The Augmented Dickey-Fuller (ADF) test is one of several tests available for the stationarity test. This test is based on the null hypothesis, which states that if a time series contains a unit root, it is non-stationary. Otherwise, it is stationary, as evidenced by a p-value smaller than a certain confidence interval value, such as 0.05 [22]. When the ADF test indicates that the series is not stationary, it is important to make it stationary using the mathematical approach known as time series differencing.

4.5 Linear time series models

4.5.1 SARIMA model

The most common and widely used time series model is the ARIMA modeling. The key assumption required to create this model is the linearity of time series. This model combines two well-known models: Autoregressive (AR) and Moving Average (MA). In AR, the next value in the sequence is a linear function of the preceding observations, whereas in MA, the next value in the series is determined by a linear function of the residual errors at a previous time. The combination of two models produces an Autoregressive Moving Average (ARMA), However, time series data may require a time difference or integration to make the series stationary besides seasonality. Hence, Seasonal ARIMA introduced to accommodate all requirements of ARIMA (p, d, q) and seasonal (P, D, Q). In other words, the SARIMA model is a seasonal ARIMA model. It can be defined mathematically as :

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \tag{4.5}$$

Where:

$$\phi_p(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \phi_3 B^3 ..... - \phi_p B^p \tag{4.6}$$

$$\theta_q(B) = 1 + \theta_1 B^1 + \theta_2 B^2 + \theta_3 B^3 ..... + \theta_q B^q \tag{4.7}$$

$$\Phi_P(B^s) = 1 - \Phi_P B^{1s} - \Phi_P B^{2s} - \Phi_P B^{3s} ..... - \Phi_P B^{Ps} \tag{4.8}$$

$$\Theta_Q(B^s) = 1 + \Theta_1 B^{1s} + \Theta_2 B^{2s} + \Theta_3 B^{3s} ..... + \Theta_Q B^{Qs} \tag{4.9}$$

$$\nabla^d = (1-B)^d \tag{4.10}$$

$$\nabla^D = (1-B^s)^D \tag{4.11}$$

S for seasonality, P for Seasonal AR order, Q for seasonal MA order, B is lagging operator, ϕ, Φ, θ ,and Θ are coefficients of AR(p), SAR(P), MA(q) and SMA(Q) respectively.

Hence, the SARIMA model has six parameters, with the seventh being the seasonal,

and the general SARIMA model representation defined as:

$$Y_t = SARIMA(p,d,q:P,D,Q)S \tag{4.12}$$

4.5.2 Determination of model order

The determination of p and q, as well as the corresponding seasonal values P and Q, leads to proper time series modeling. These parameters can be determined graphically using the Partial Auto Correlations (PACF) and Auto Correlations (ACF). By neglecting the correlation impact of intermediate data, Partial Auto Correlation (PACF) measures correlations between observations separated by k lags. ACF, on the other hand, will take into account all the correlation impacts between the two time series values which are k lags apart. The Covariance, ACF, and PACF are Mathematically expressed as equations 4.14 to 4.16 [21].

$$Cov(y_t, y_{t+k}) = E[(y_t - \mu_t)(y_{t+k} - \mu_{t+k})] = E(y_t, y_{t+k}) - \mu_t \mu_{t+k} \tag{4.13}$$

$$\gamma_{i,i+k} = \frac{Cov(y_i, y_{i+k})}{\sqrt{\sum_{i=1}^{n} (y_i - \mu_i)^2 \sum_{i=1}^{n} (y_{i+k} - \mu_{i+k})^2}} \tag{4.14}$$

$$\phi_k = Corr(y_t, y_{t-k} | y_{t-1,\ldots.y_{t-k+1}}) \tag{4.15}$$

where: Cov(y_t, y_{t+k}) $\gamma_{i,i+k}$ and ϕ_k are Covariance, Autocorrelations functions (ACF), and Partial Autocorrelations Functions (PACF) respectively.

Applying PACF and ACF for model order identification is cumbersome and sometimes difficult to decide if the series requires order differencing; instead, an automated grid search method can be used to find the best order. This method uses a hyperparameter tuning approach, by comparing the minimum of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). This automated method is used in this study to determine the model order of SARIMA (p, d, q) and (P, D, Q).

4.6 Forecasting metrics

Model order of time series is found using 80% of the data as training, and fitting the model is the next step to find the model coefficients. Before predicting the future, it is important to check the noise diagnosing plot to determine whether the leftover of the fitted model is purely noise (with mean zero and variance one).

For model order selection, parameters such as AIC or BIC can be used. However, after training, predicting the values and evaluating the model's performance by calculating the accuracy against test data is necessary. Model evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), (RMSE) Root Mean Squared Error, and (MAPE) Mean Absolute Percentage Error can be used for this purpose. Therefore, model performance metrics defined as follows: The absolute difference between the real (test data) and predicted value is defined as the Mean Absolute Error (MAE); minimizing MAE means forecasting the Median. The square root of the MSE is the Root Mean Square Error, and minimization leads to forecasting of traffic mean [22]. The Mean Squared Error (MSE) is calculated by averaging the square of the difference between the real and predicted values. MAPE is another type of performance measurement that compares predicted and actual data. It is unit-less, because the expression is defined as a percentage of the real value's error. Mathematically, these model measurement metrics are expressed as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \bar{Y}| \tag{4.16}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \tag{4.17}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{4.18}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - \bar{Y}|}{|Y_i|} * 100 \tag{4.19}$$

Where $\overline{Y}$ is the predicted value,Y_i is the true value and n is the number of samples.

NB: the value of the above metrics reaching zero as the model is more representing

the observed traffic.

4.7 Traffic forecasting

Forecasting is the final step in time series analysis. It is defined as a technique for anticipating events through a sequence of future times. In time series forecasting, the main assumption is that future trends will be similar to the past. The forecasting period can be short or long, depending on the size of the dataset. Five months of traffic data has collected for this study, with 20% used as a test and one month for the forecasting period to discover the trending element. After the traffic has been forecasted, the mean of the forecasted and test data in the provided range (test data size) can be determined as

$$B_T = \frac{1}{n}\sum_{i=1}^{n} B_i \tag{4.20}$$

$$A_T = \frac{1}{n}\sum_{j=1}^{n} A_j \tag{4.21}$$

The slope of the line between the Test and Forecasted points (T1,B_T)and($T2, A_T$) in Figure 4.1 represents the rate of change of traffic trends.

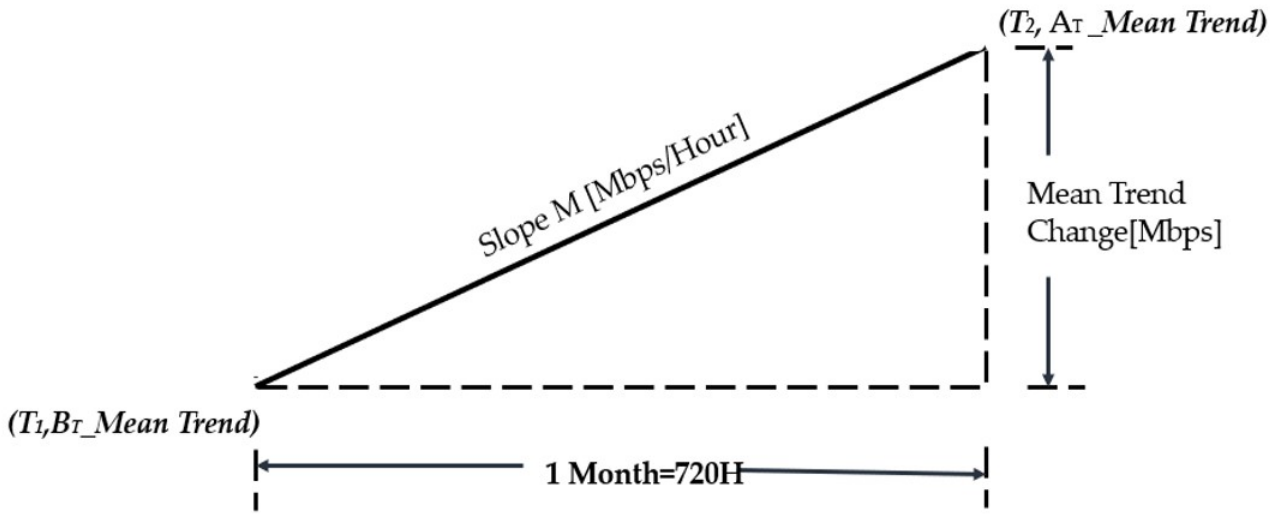


Figure 4.1: Traffic trend change calculation for 1 month.

The slope of the line is Mathematically expressed as:

$$M = \frac{A_T - B_T}{T2 - T1} \tag{4.22}$$

by assuming T1=0 and T2=T,then the forecasting Traffic Trend becomes:

$$A_F = B_T + MT \tag{4.23}$$

$$k = \frac{B_T}{A_F} \tag{4.24}$$

$$K = 2 - k = \frac{1+2z}{1+z} \tag{4.25}$$

$$z = \frac{MT}{B_T} \tag{4.26}$$

Where k is the mean trend factor that is valid in the range [0.5,1], and K is the link capacity multiplying factor. The forecasted mean is A_F, the test data mean is B_T, and the traffic trend slope is M.

5

Microwave link capacity dimensioning

5.1 Microwave communications

Microwave communication is a more cost-effective, time-saving, and simple-to-deploy transmission technology[23], particularly in remote areas where fiber transmission is impractical. The choice of technology in mobile backhaul, on the other hand, is influenced by the distance between the customer's premises and the core network, and therefore, it can be made of fiber, microwave, or satellite. In Ethiopia, the majority of mobile backhaul on the access side of the network [24] has built using microwave technology.

Microwave link design takes into account a number of technical criteria, including free space path loss, channel spacing, frequency planning, modulation scheme, and others, all of which are included in fundamental microwave dimensioning parameters. The aim of this chapter is not to introduce microwave design but rather to focus on the link capacity dimensioning and the link parameters that are directly influenced by link capacity change, such as modulation order, channel spacing, noise figure, signal to noise ratio, fade margin, and link availability. Thus, this chapter briefs these parameters

and their corresponding mathematical formulations.

5.2 Fundamental microwave dimensioning parameters

5.2.1 Line of sight and Fresnel zone

It is necessary to achieve radio visibility between sites before establishing the link for point-to-point microwave communication. Hence, the general setup necessitates the placement of microwave antennas in a higher position. When there is no obstruction between two sites, the received power is characterized by free-space path loss, which decreases with increasing distance between the transmitter and receiver [7]. On the other hand, the Fresnel zone defined as a region of ellipsoid-shaped volume around the line-of-sight that aids in defining a clear radio view between the sites. For optimum reception in a location where obtaining a clear line of sight is difficult, 60% of the first Fresnel radius is sufficient.

5.2.2 Frequency range in mobile backhaul

The frequencies used in microwave links range from 2 to 58 GHz. Based on frequency range and distance, links are classified as short, medium, or long haul; as frequency increases, the length of the link decreases. The microwave backhaul technology used in dense urban places like Addis Ababa is generally short-haul with a range of 0.2 to 9 kilometers, and the central frequency mostly varies between 26GHz, 18GHz, and 15GHz [2].

5.2.3 Link budgeting and Fade margin

Link budgeting in microwave is an accounting of all the power gains and losses that a communication signal experiences from a transmitter, through an air medium and until it reaches the receiver [25]. A fade margin, on the other hand, is a design allowance for sufficient system gain or sensitivity to handle expected fading while maintaining the required service quality. Therefore, for better link performance, the value of fade margin is significant, and it can be regulated by boosting transmitted power or increas-

ing receiver sensitivity. Thus, the fade margin must always be positive, as illustrated in Figure 5.1, and link budgeting must ensure the value as high as possible. Otherwise, the link will be unstable. Mathematically:

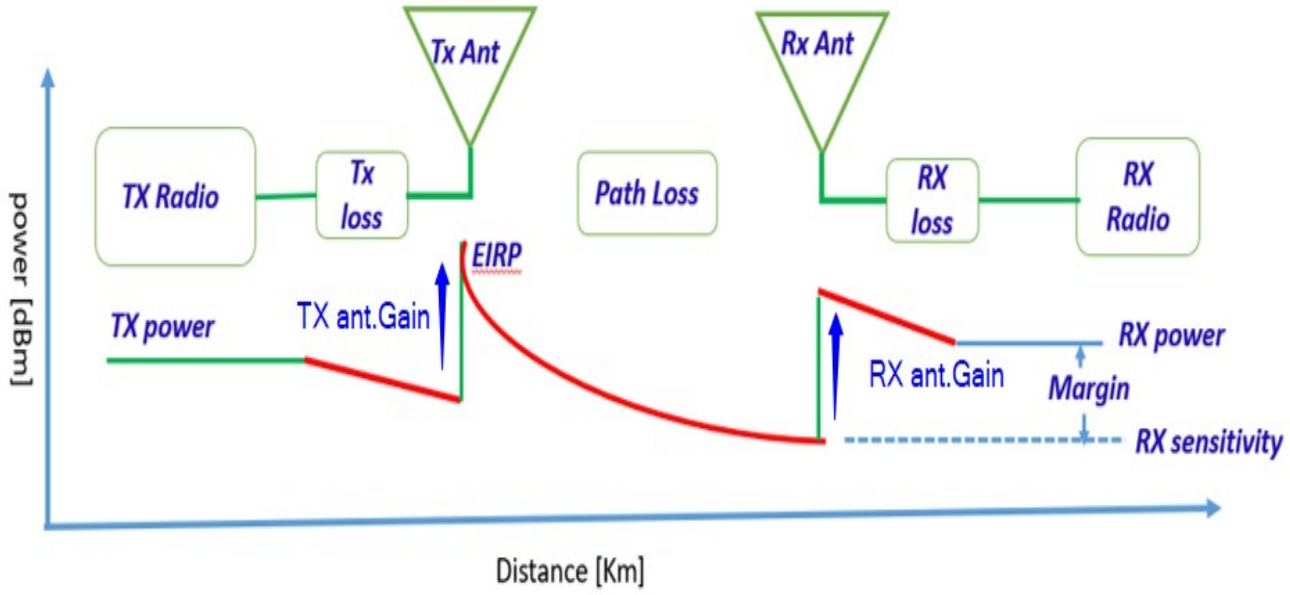


Figure 5.1: Fade margin and Link budget calculations.

$$FM(f,d)[dB] = P_{Tx}[dBm] + 2G_{T/R}[dBi] - FSL[dB] - P_{Rx(BER)}[dBm] \tag{5.1}$$

Where FM is fade margin[dB], P_{Tx}[dBm] is the transmitting power at the output of the amplifier, $2G_{T/R}$[dBi] is the antenna gain of both sides, FSL[dB] a free space path loss and P_{Rx}(BER) is receiver sensitivity at a specified bit error rate.
FSL[dB] and P_{Rx}(BER) [dBm] defined in Equations (5.2) and (5.3) respectively.

$$FSL(f,d)[dB] = 92.44 + 20\log f[GHz] + 20\log d[Km] \tag{5.2}$$

$$\mathrm{P}_{Rx(BER)}[dBm] = -114dBm + \text{Required } SNR[dB] + NF\ [dB] \tag{5.3}$$

5.2.4 Noise figure and required SNR

Noise Figure (NF): is the receiver's total thermal noise power. It is determined by Boltzmann's constant $K = 1.3806 * 10^{-23}\ Joules/^0k$, the temperature in Kelvin = 290^0k and the channel spacing in MHz. Mathematically defined as:

$$NF = K.T.B[dBm] \tag{5.4}$$

Signal-to-Noise Ratio (SNR): The SNR is primarily determined by the Bit Error Rate(P_e), modulation scheme(QAM), and order(M). The received power must be greater than the receiver's noise floor to guarantee the desired quality of the demodulated signal.Therefore, P_e, SNR, and M are related by equation as:

$$P_e = \frac{2(\sqrt{(M-1)}}{\sqrt{M}\log_2 M} erfc(\sqrt{\frac{3SNR}{(M-1)}}) \tag{5.5}$$

Where SNR is signal to noise ratio and it is expressed in decimal, not in dB.

5.2.5 Microwave link availability

It is the final calculation in the microwave link design, and primarily determined by the fade margin and the site's geolocation factor. According to ITU (ITU-R P.530), the link unavailability formula based on the worst-case scenario (worst weather conditions probability)[25] is given by

$$P_w = P_o(10^{-\frac{A}{10}}) * 100 \tag{5.6}$$

where

$$P_o = kd^{3.1}(1+|\varepsilon_p|)^{-1.29} f^{0.8} 10^{-0.0008 h_l} \tag{5.7}$$

$$k = 10^{-4.6-0.0027 dN1} \tag{5.8}$$

the reliability of the link based on the given parameters will be

$$R = (1 - \frac{P_w}{100}) \tag{5.9}$$

Where A represents the fading margin in decibels (dB), d represents the distance in kilometers (Km), f represents the frequency in gigahertz (GHz), h_l represents the altitude of the lower height site above sea level, dN1 represents the point of refractivity gradient (N units/Km), and P_o represents the fade occurrence factor. Since all the parameters, except the fade margin, remain constant for this specific study, and calculating the link

unavailability modified as shown below in equations 5.10.

$$\frac{P_w}{P_o} = (10^{-\frac{A}{10}}) \tag{5.10}$$

Which can be applied for the comparison of unavailability as a function of fade margin by keeping all the parameters unchanged, during link capacity optimization.

6

Results and Discussion

In general, capacity dimensioning of a link using real-time traffic data is performed through data preprocessing, distribution fitting, and provisioning of link capacity with link transparency as quality control. Thus, the primary purpose of this chapter is to implement the link capacity provisioning approaches outlined in the preceding chapters. It starts with a briefing of the thesis's system model, then moves to microwave link capacity provisioning based on busy hour and time series analysis to determine real time link capacity. Finally attempts to compare the results to existing link capacity in terms of link performance and utilization of scarce resources.

6.1 System model

The general outline of the proposed system model used to dimension the link capacity of the microwave backhaul network is depicted in Figure 6.1. All the activities associated with each step are explained in detail in the coming sections. The two sub-system models that are accounted for the probabilistic and time series approaches are shown below in Figure 6.2. The controlling parameters and methods are also presented besides

every step.

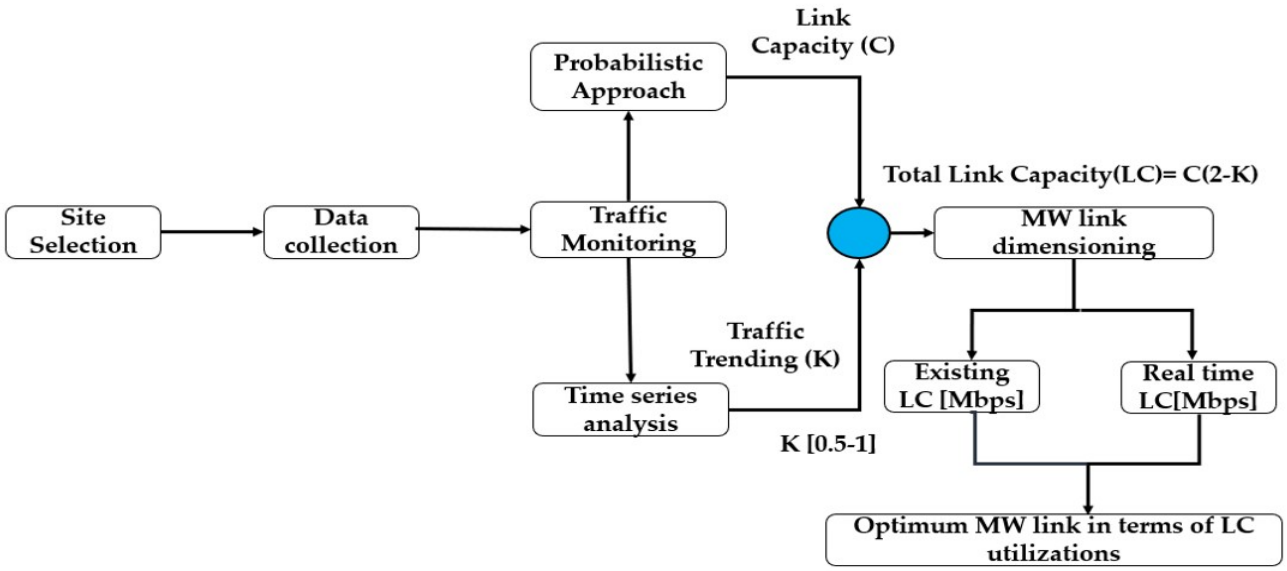


Figure 6.1: General system model.

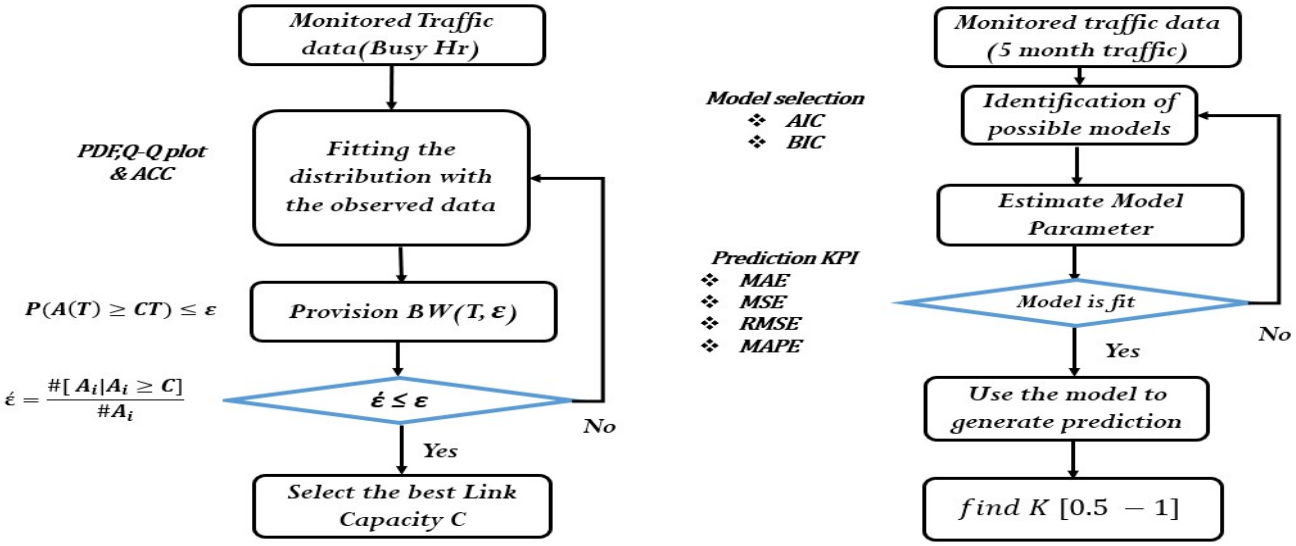


Figure 6.2: Probabilistic and Time series approach sub-models.

6.1.1 Data monitoring and collection

The data used to dimension microwave backhaul link capacity was obtained in CSV (Comma Separated Values) format from the Network Management System (NMS). It was collected from 12 mobile sites (Figure 6.3) that offer all services (voice and data) at

1-second granularity during peak hours, as well as 5 months of continuously monitored traffic data at 1-hour intervals for time series analysis. Tools such as Microsoft Excel, MATLAB, and Python were used for analysis.

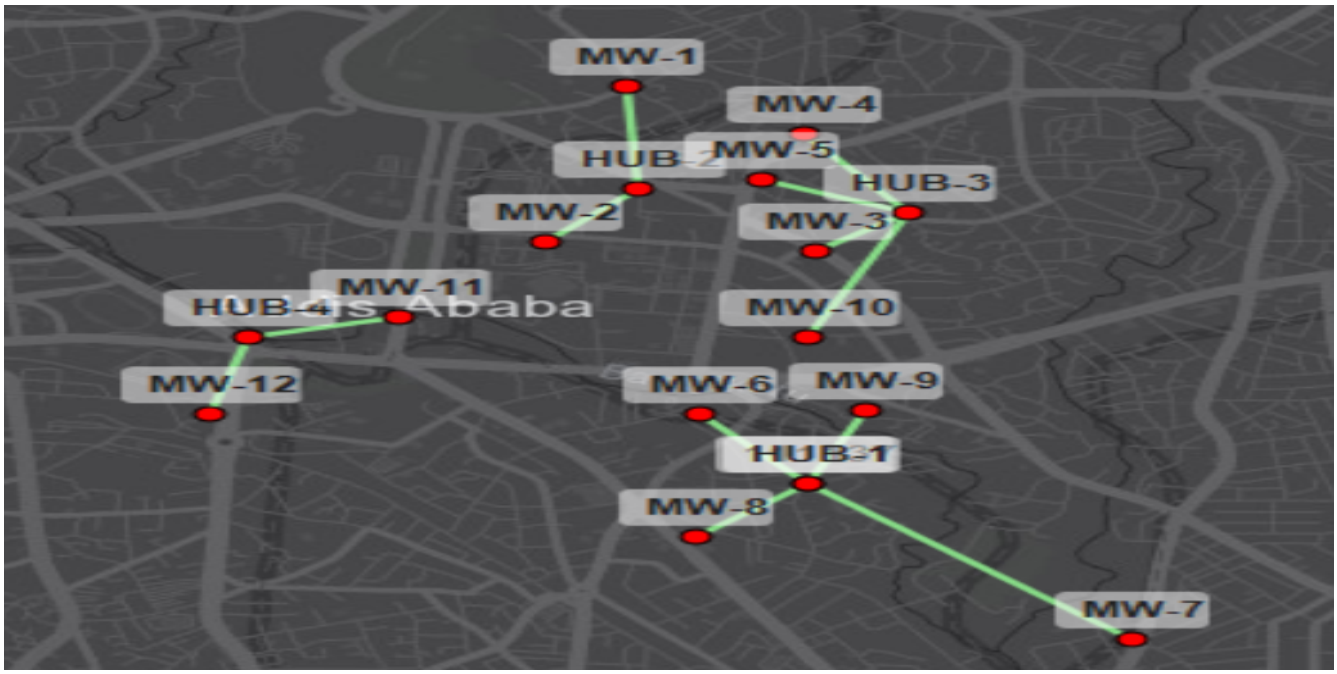


Figure 6.3: MW sample sites from AA

6.2 Probabilistic approach traffic modeling

The traffic data (1-second interval) for this approach was collected from 12 mobile sites in Addis Ababa, all of which are in the same vicinity, have single uplink, and their respective installed link capacity has been tabulated in Table.6.1 shown below.

The table lists the site's name, current uplink microwave capacity or throughput in Mbps, respective channel spacing in MHz, and Quadrature Amplitude Modulations (QAM) order.

6.2.1 Fitting probability distributions with observed data

The probabilistic approach Figure 6.2 begins with distribution selection using fitting techniques such as the probability density function plot, the Q-Q plot, and linear correlation coefficients. Maximum likelihood estimation (MLE) used implicitly during the process to estimate the parameters of the probability density function. Since MW-11 is

Table 6.1: Sample sites in Addis Ababa with actual installed capacity

Link Name	Modulations (QAM)	Channels spacing [MHz]	Throughput [Mbps]
MW-1	256	56	368
MW-2	512	56	449
MW-3	512	56	449
MW-4	256	56	368
MW-5	256	56	368
MW-6	512	56	449
MW-7	512	56	449
MW-8	256	56	368
MW-9	512	56	449
MW-10	512	56	449
MW-11	512	56	449
MW-12	256	56	368

one of the busiest site in the sample space, it has been selected for demonstrations of the study's approach. Therefore, for MW-11, Figure 6.4 of curve fitting with observed data shown and it signifies that Gaussian, GEV, and Lognormal distributions fitted well, as the area coverage of the traffic and distribution plots fit well compared to others.

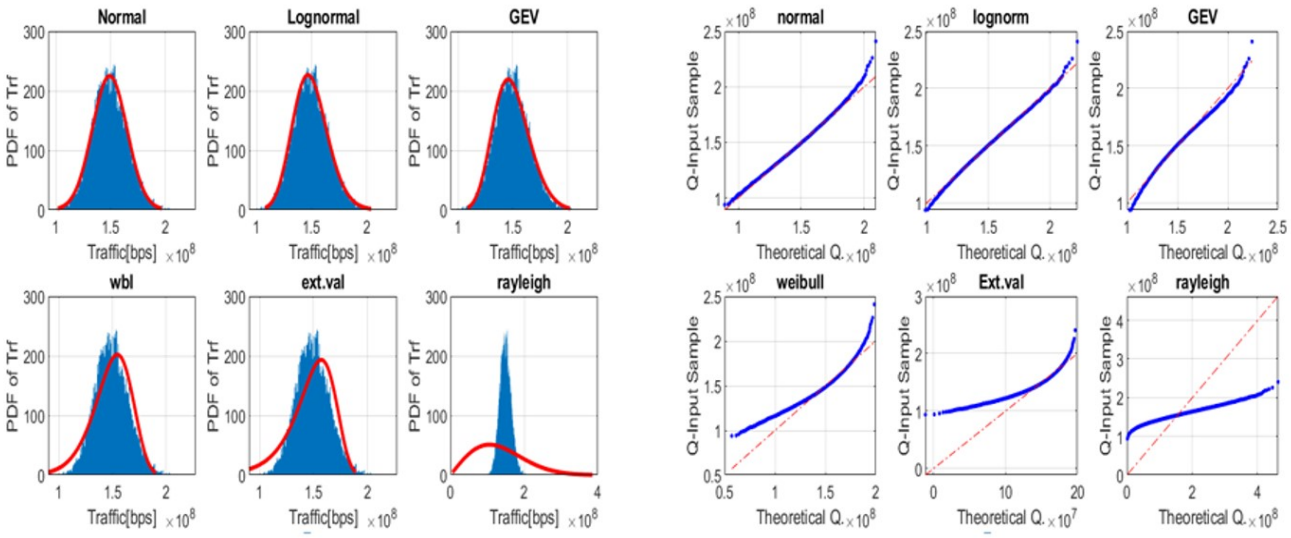


Figure 6.4: PDF fitting and q-q plot of MW-11 traffic.

The Q-Q plot, which shows how the data fits with the distribution as it fits the diagonal line, has been included along with the PDF plot. The Gaussian distribution failed to represent the traffic at the right tail of the Q-Q plot, indicating that the traffic's heavy tail nature was not bounded by an exponentially bounded distribution, whereas the Log-

normal and GEV (heavy-tailed distributions) fitted well. For further quantification, the value of the normalized correlation coefficient of the fitted distribution against the observed data has been presented in the summary of Figure 6.7. The correlation value is above 0.95 except for general Pareto, since it is selected to measure the exceedance of the traffic (peak traffic when the blocking probability reaching zero) and is accurate only at the tail of the distribution compared to the rest.

6.2.2 Link capacity vs Blocking probability

Once the probability distributions have been selected (using Q-Q plot, PDF plot, and a correlation coefficient value greater than 0.95), the next step is to calculate the link capacity using link transparency equation by fixing the value of blocking probability. Though literature [3, 8] suggests a value ranging from (0.1 to 2.5)%, 0.01% (99.99% of traffic passage) of blocking probability has been chosen for this study to cover all service requirements. In order to show the relationship between link capacity and blocking probability, the link capacity of MW-11 with the respective provisioning approach is shown in Figure 6.5. The graph has indicated that, as the value of the blocking prob-

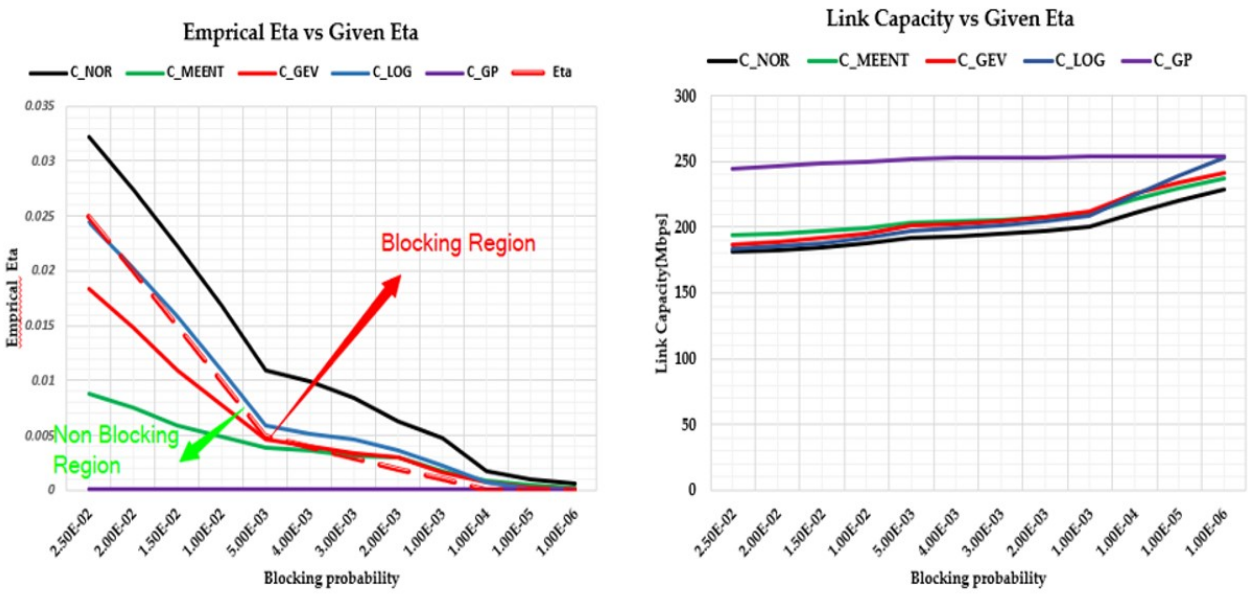


Figure 6.5: Link provisioning and blocking probability of MW-11.

ability decreases, the link capacity increases. On the other hand, from the plot of the given blocking probability of 0.01% with the calculated blocking probabilities, for the

values between (0.03 – 2.5)%, Meent's and GEV provisioning approaches are more appropriate, Lognormal is marginally fit, and the Gaussian distribution is inapplicable until the blocking probability closes to zero. General Pareto, on the contrary, can model extreme values, and is used to limit the upper link capacity when the other approaches are unable to find the optimum peak link capacity.

Link capacity provisioning of the microwave links was performed using sample traffic collected during the busiest time of day ranges from 08:00 to 23:59, as shown in Figure 6.6, by limiting the blocking probability value to 0.01%. Except General Pareto,

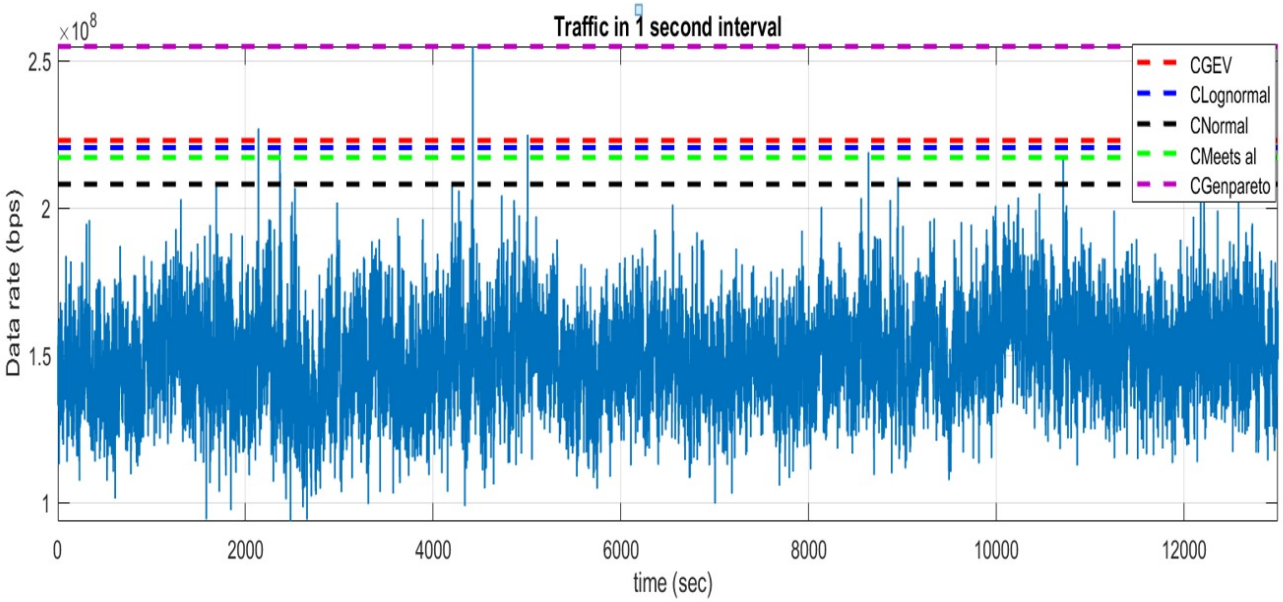


Figure 6.6: Link provisioning of MW-11

the remaining four provisioning approaches failed to dimension the link capacity with a blocking probability of 0.01%, since, the calculated blocking probability is greater than the specified value. As a result of the General Pareto provisioning approach, the link capacity of MW-11 has been determined to be 255 Mbps with empirical blocking probabilities of $7.482 * 10^{-3}$%. Figure 6.7 summarizes all results such as traffic statistics, correlation coefficients, provisioning approaches with link capacity, and empirical blocking probability using a graphic user interface designed for this study.

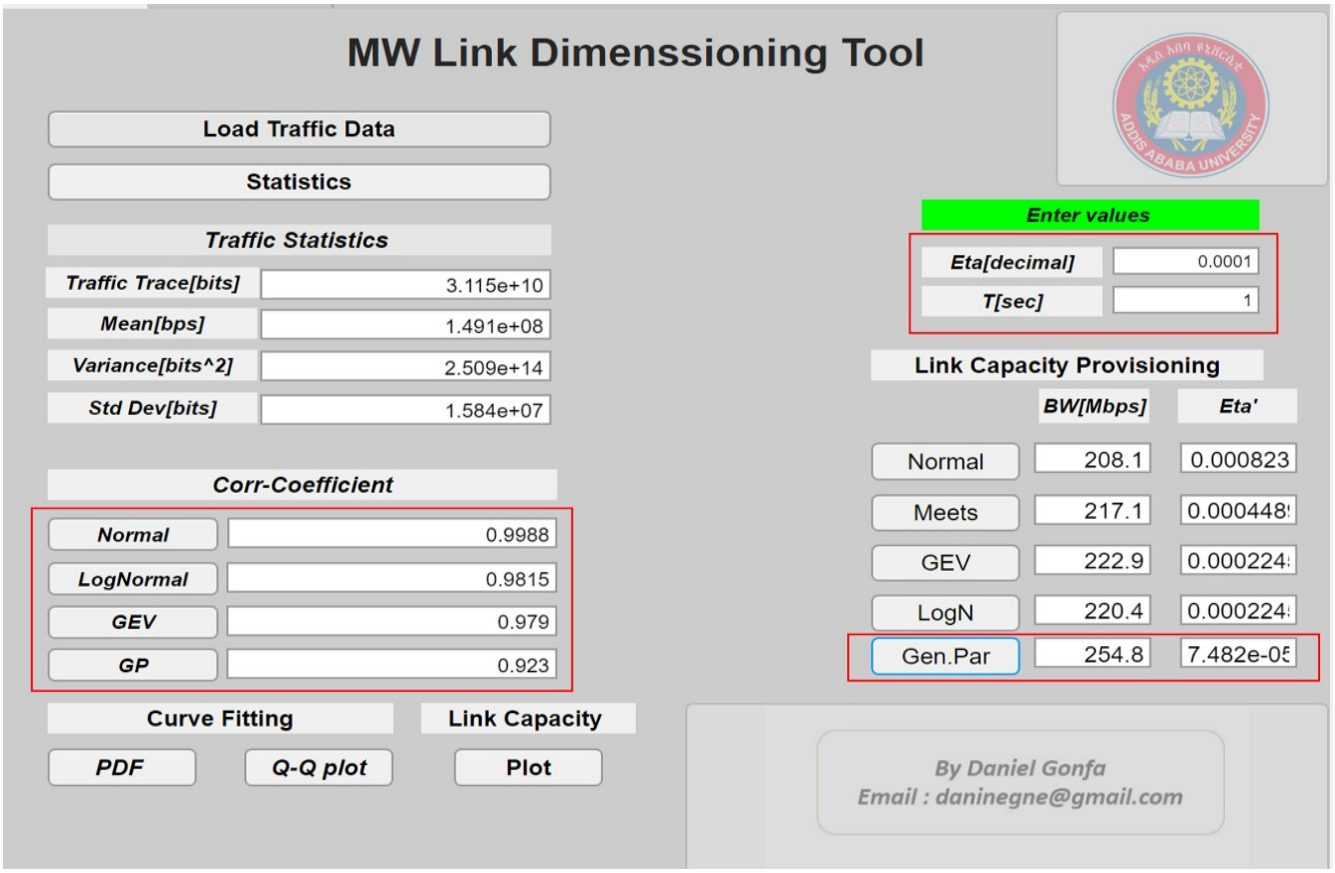


Figure 6.7: Link provisioning result for MW-11 summary

6.2.3 Link capacity provisioning of 12 sites

Following the same analysis of MW-11, all link capacity has been provisioned appropriately. Figures 6.9 and 6.10 have depicted the provisioning of all sites, as well as the summary of corresponding link capacity in Figure 6.8. Despite the fact that,the traffic was measured during peak hours, all links were not using the installed capacity, shown in Table 6.1. This could be because user traffic from the access side is not generated as planned, or peak traffic provisioning approaches are not always applicable.Thus, after installation, optimizing the link capacity with real-time traffic approach is required in timely bases. This is the primary goal of this study" *link capacity of a microwave backhaul based on actual traffic requirements*".

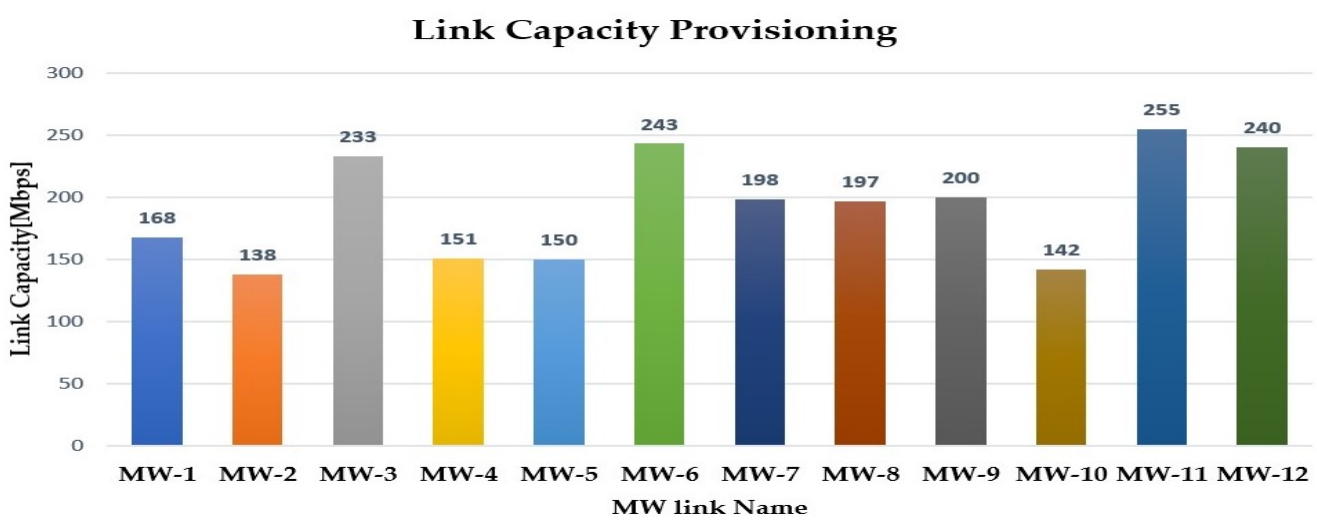


Figure 6.8: Peak link capacity result for all MW links

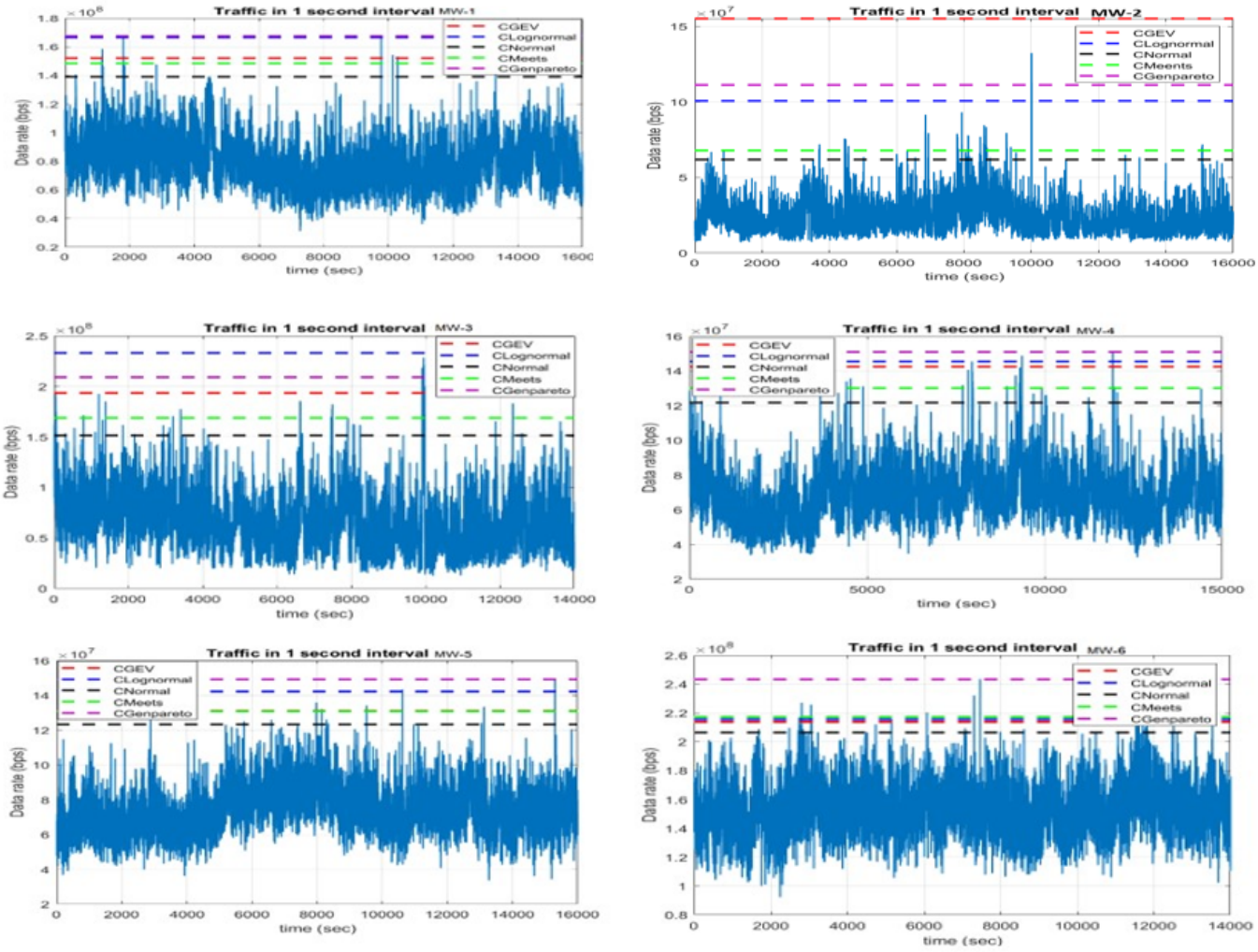


Figure 6.9: Link provisioning result for MW-1 to MW-6

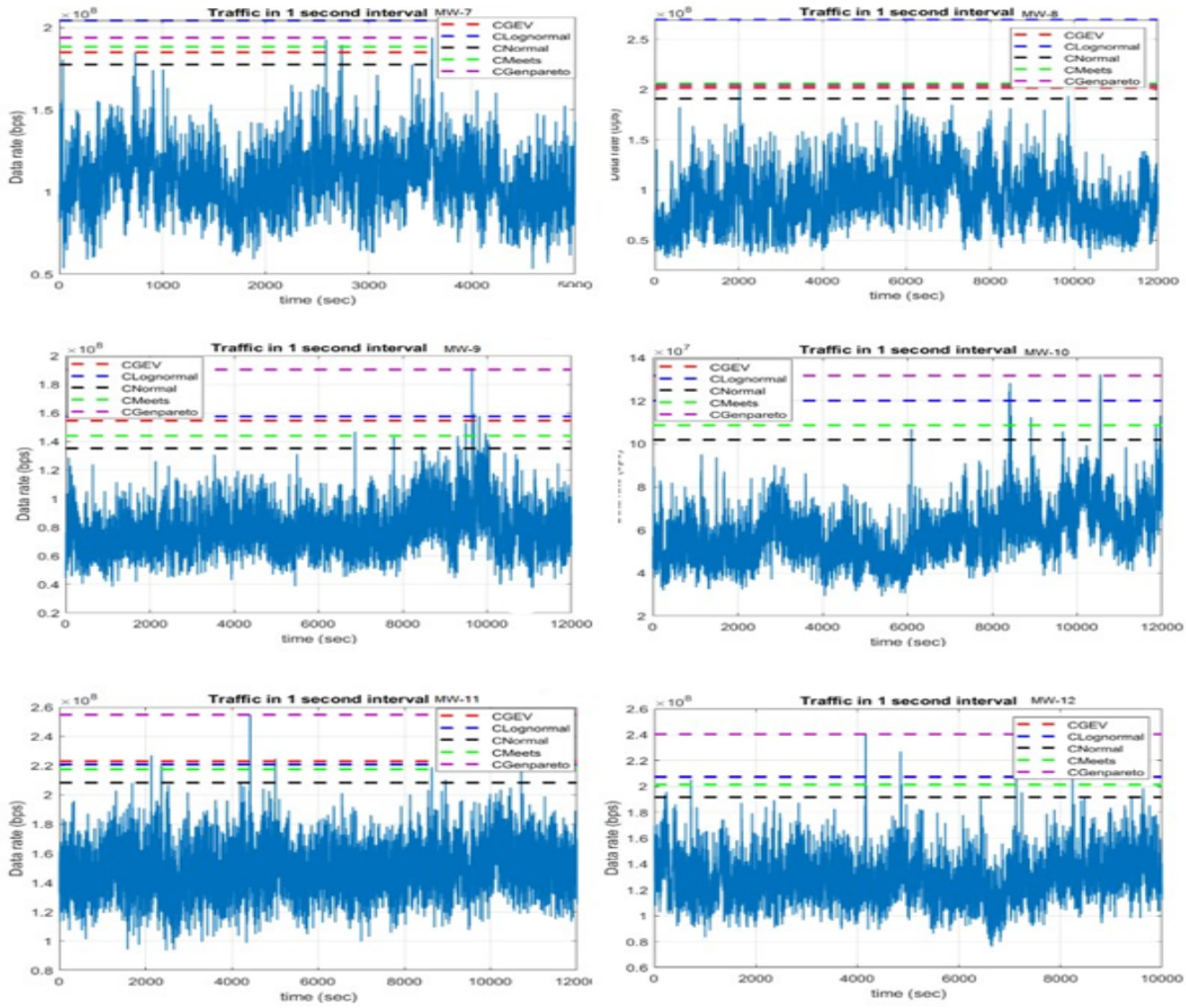


Figure 6.10: Link provisioning result for MW-7 to MW-12

After the site's link capacity has been determined using a probabilistic approach, the next step is to determine the traffic trending factor to observe, if there is any gradual traffic growth. This analysis has been carried out using time series analysis, as described in the next sections.

6.3 Time series analysis

The objective of this analysis is to determine the mean traffic variation or traffic trend. Before defining the model, it is necessary to examine the data's statistical characteristics such as mean, maximum, and minimum.

6.3.1 Traffic statistics

The general statistical description of all site's traffic (Maximum, Mean, and Minimum) is depicted in Table 6.2. According to the table, late-night traffic ranges from [2 to 35] Mbps, while busy time traffic ranges from [69 to 183] Mbps, and the mean or central tendency of traffic ranges from [35 to 130] Mbps. It should be noted that the traffic was measured in 1-hour intervals over 5 months, hence, traffic bursts were not the intended objective; rather, the mean rate of change or trend was the goal of the analysis. The

Table 6.2: Traffic statstics of MW sites

MW-NAME	Mean[Mbps]	Minimum[Mbps]	Maximum[Mbps]	Std
MW-1	71	9	123	25.3
MW-2	54.6	6	112	23.45
MW-3	60.83	6	123	22.44
MW-4	98.54	12	161	34
MW-5	48.4	4	98	26.01
MW-6	115.5	15	182	43.32
MW-7	72	7	128	29.46
MW-8	65.3	7	147	32.09
MW-9	76.3	9	128	30.84
MW-10	35	5	69	13.3
MW-11	129.5	35	183	33.39
MW-12	55.9	2	145	43.15

traffic generated from every mobile site is constituted of 2G, 3G, and 4G. The general traffic of MW-11 is shown in Figure 6.11 with average traffic plot of daily and weekly bases.

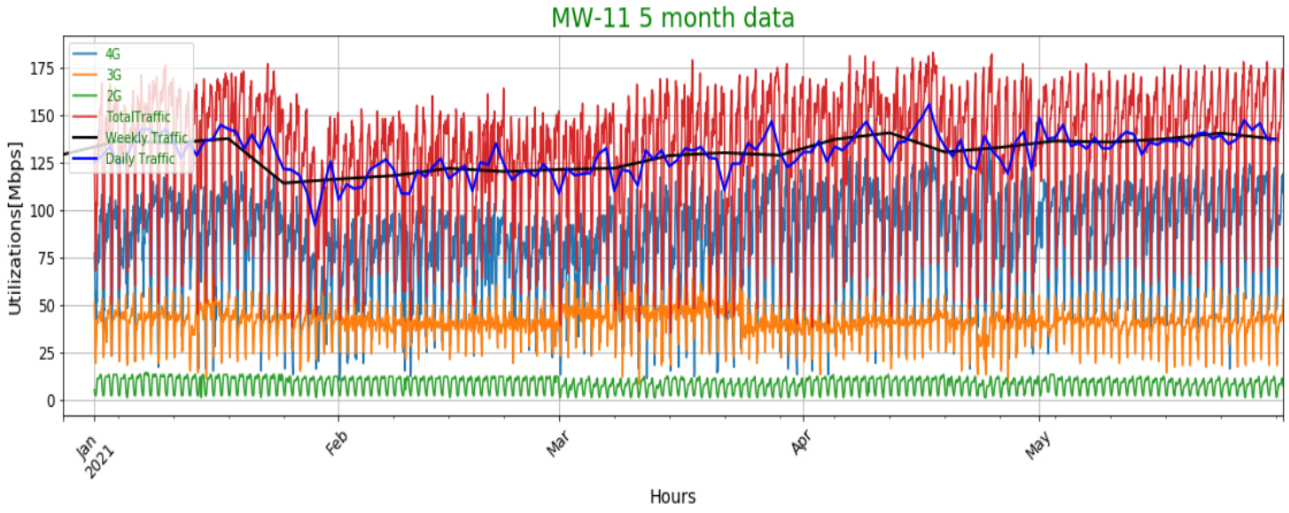


Figure 6.11: Microwave 11 general traffic

The graph has shown, the gradual traffic growth as time progressed during the monitored period. The sample traffic in Figure 6.12 has clearly showed that, the user behavior towards the services varies in the same pattern besides it repeats it self every 24-hour. From the traffic trend of each service has plotted in Figure 6.13, the trend

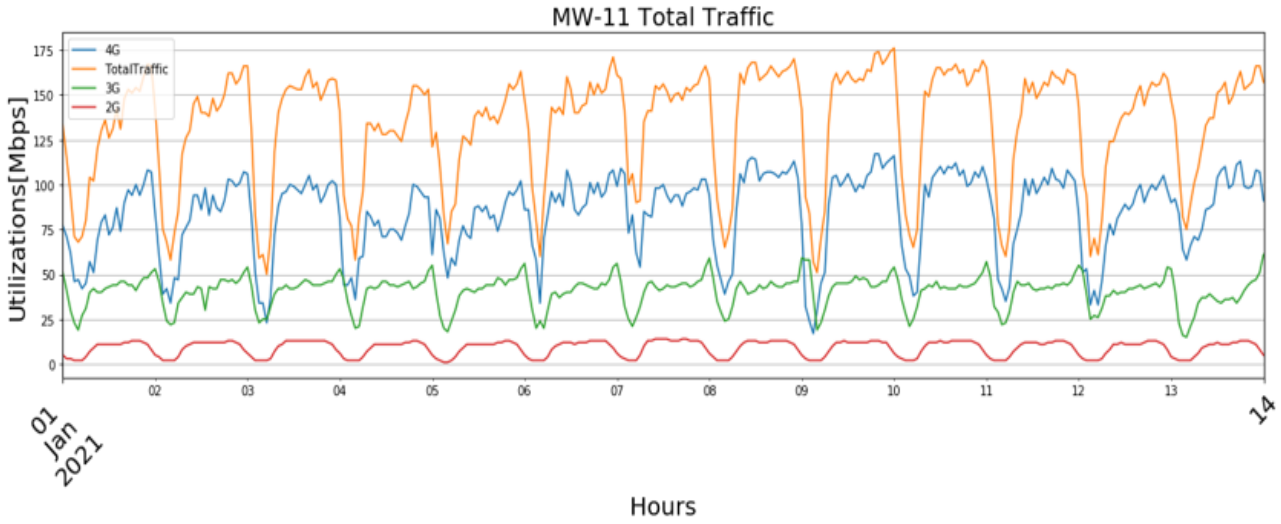


Figure 6.12: Traffic sample of 2G,3G and 4G for MW-11

came from the 4G user varies with the total traffic linearly as it is manifested in the scatter plot of each service against total traffic(Figure 6.14). The scatter plot shown in figure 6.15, the correlation between Lag-1, Lag-24 and Lag-1Week depicted that , the traffic has strongly correlated and repeats somewhat in daily and weekly bases, this means the traffic profile of the users in the site fairly unchanged or progressed slowly through time.

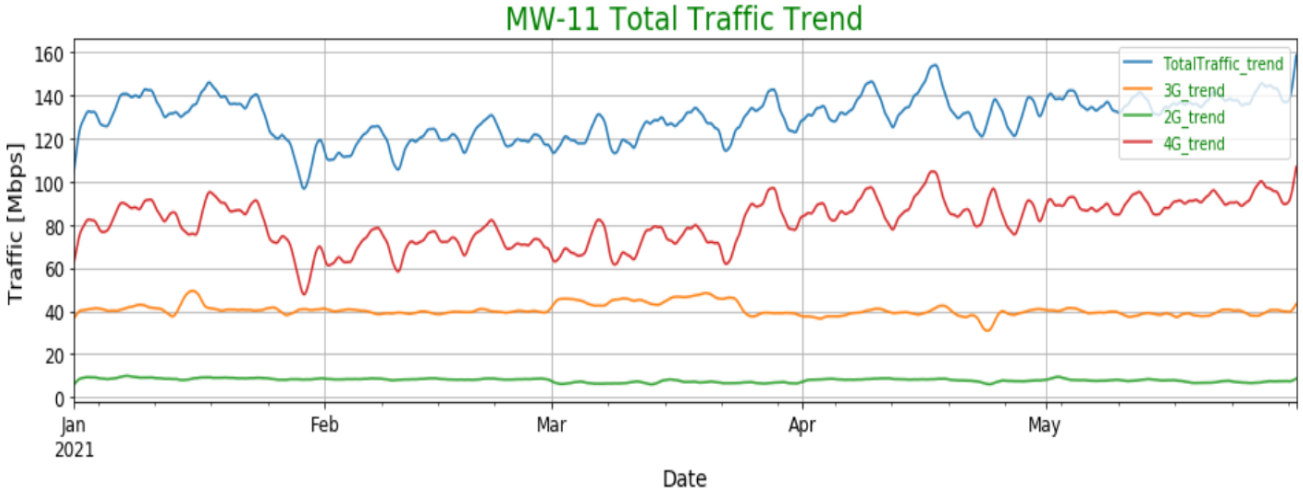


Figure 6.13: Total traffic trend of MW-11

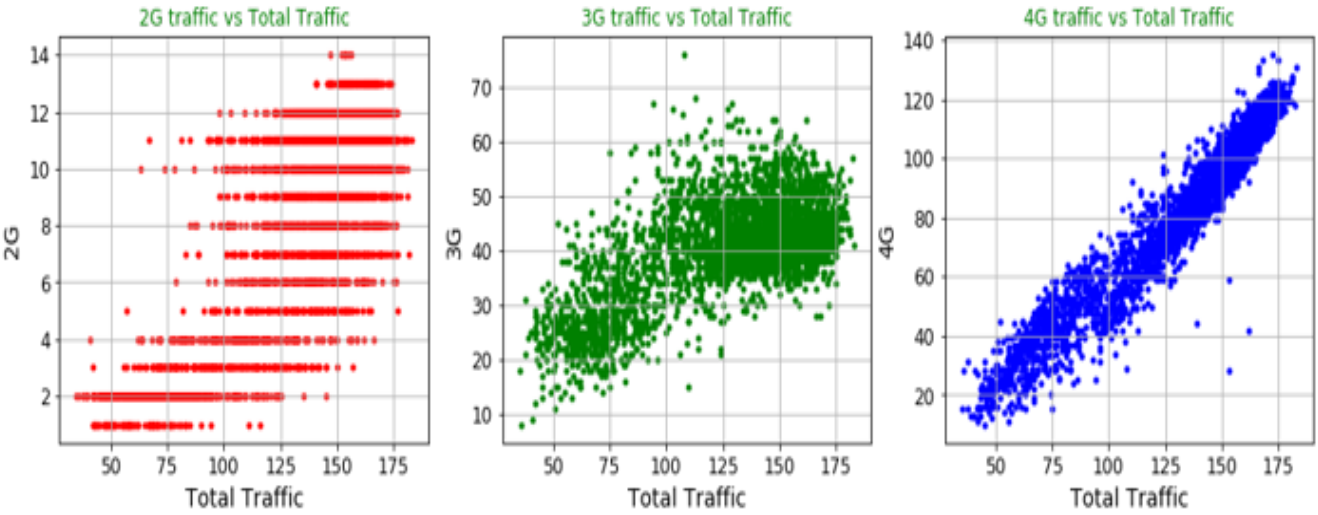


Figure 6.14: Scatter plot of 2G,3G and 4G traffic MW-11

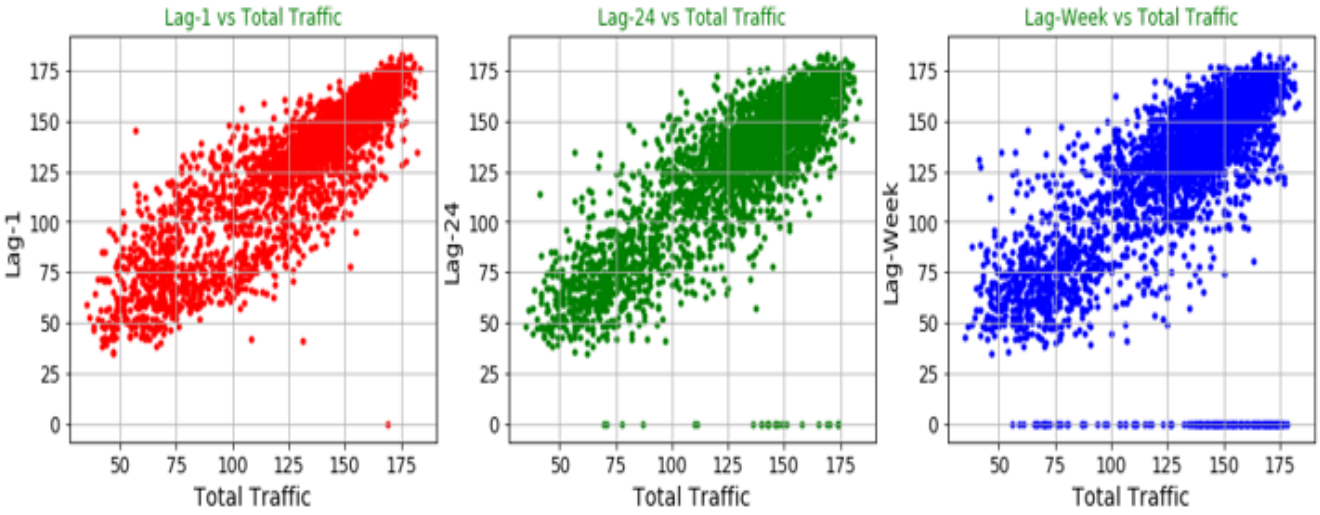


Figure 6.15: Scatter plot of Lag-1,Lag-24 and Lag-1week traffic of MW-11

6.3.2 Traffic decomposition

As mentioned in the time series discussion, in Chapter 4, time series decomposition can be done by either multiplicative or additive models. In this study, the additive model decomposition was performed, and it shows the seasonal, trend, noise characteristics of the traffic. As it is observed from the figure, the traffic has seasonality, which shows SARIMA modeling is suitable for such traffic compared to ARIMA with a period of 24 hours.

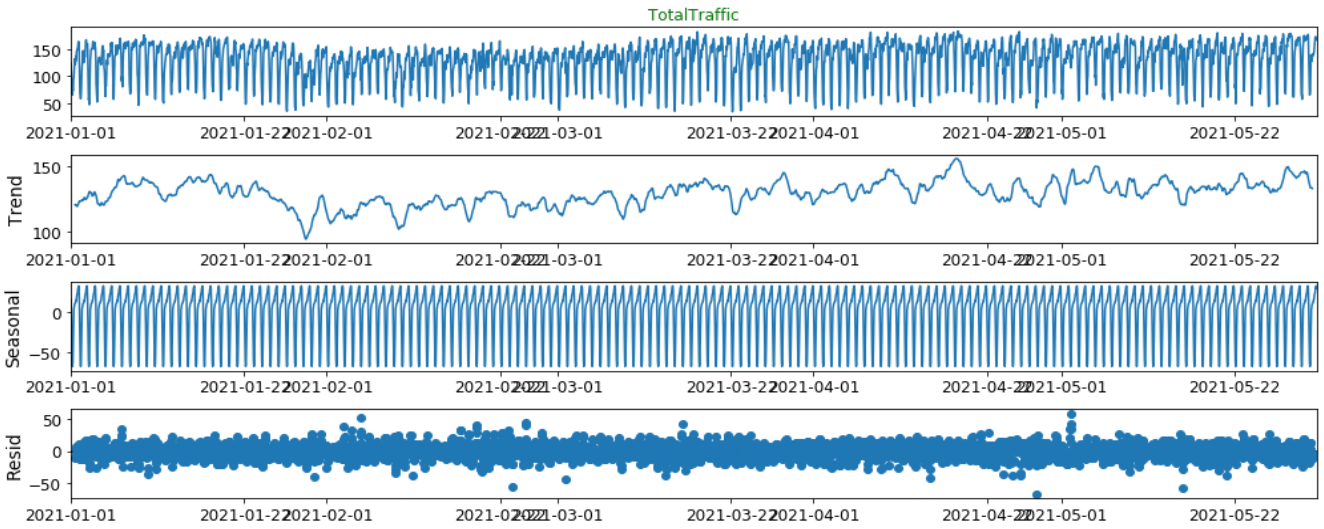


Figure 6.16: Traffic decomposition of the MW-11

6.3.3 Traffic modeling

Once, the traffic decomposition has done, the next step is model-fitting using the training data of 80% to train the SARIMA model. The model order was determined using the lowest AIC or BIC using grid search model order finding approach. As a result, the model order of the MW-11 shown in Figure 6.17 with respective coefficients. From the figure the value," $p > |z|$" are all zero, which is an indication of all the coefficients are significant in the model order. Similarly MW-1, MW-2, and MW-3 sites were discov-

SARIMAX Results

Dep. Variable:	TotalTraffic	No. Observations:	2899
Model:	SARIMAX(1, 1, 1)x(0, 1, 1, 24)	Log Likelihood	-10410.609
Date:	Fri, 24 Sep 2021	AIC	20829.217
Time:	23:56:31	BIC	20853.035
Sample:	01-01-2021 - 05-01-2021	HQIC	20837.807
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.7239	0.012	62.563	0.000	0.701	0.747
ma.L1	-0.9789	0.005	-203.854	0.000	-0.988	-0.969
ma.S.L24	-1.0786	0.008	-136.390	0.000	-1.094	-1.063
sigma2	74.4136	1.716	43.376	0.000	71.051	77.776

Ljung-Box (Q):	62.94	Jarque-Bera (JB):	1651.34
Prob(Q):	0.01	Prob(JB):	0.00
Heteroskedasticity (H):	0.91	Skew:	0.11
Prob(H) (two-sided):	0.13	Kurtosis:	6.72

Figure 6.17: Model order of the MW-11

ered to be SARIMA (111,111)24, while the remaining sites have been modeled with

SARIMA (111,011) 24. The corresponding coefficient are listed in Table 6.3.

Table 6.3: SARIMA model coefficients of respective MW sites

MW-NAME	AR(ϕ)	MA(θ)	SAR(Φ)	SMA(Θ)
MW-1	0.5601	-0.955	0.0195	-1.0364
MW-2	0.5545	-1.0158	0.018	-0.9657
MW-3	0.5649	-0.9864	0.0619	-0.9074
MW-4	0.5456	-0.9895	0	-0.8866
MW-5	0.6544	-0.9895	0	-0.09119
MW-6	0.7161	-0.9824	0	-0.95
MW-7	0.56	-1.0237	0	-0.9305
MW-8	0.7043	-1.0069	0	-1.0301
MW-9	0.5814	-0.9725	0	-0.9313
MW-10	0.5821	-0.9722	0	-0.9478
MW-11	0.7239	-0.9789	0	-1.0786
MW-12	0.7689	-1.0027	0	-0.9318

For further simplification of a given coefficients of the model, the general formula for SARIMA (p, d, q) (P, D, Q)S of the sample site of MW-11 has been shown in equations 6.1.

$$(1-\phi_1 B)(1-\Phi_1 B^{24})(1-B)(1-B^{24})y_t = (1+\theta_1 B)(1+\Theta_1 B^{24})\varepsilon_t \qquad (6.1)$$

where: (1-$\phi_1 B$) is the AR order one (AR(p=1)),(1-$\Phi_1 B^{24}$) is the Seasonal AR order of one (SAR(P=1)), (1-B) is nonseasonal difference of an order of one (d=1),(1-B^{24}) is the corresponding seasonal differencing (D=1),(1+$\theta_1 B$)is the Moving Average order one (MA(q =1)),and (1+$\Theta_1 B^{24}$) is Seasonal Moving Average (SMA(Q=1)).
In equation 6.1 of the series, by substituting $y_t - y_{t-1} = z_t$ and $y_t - y_{t-25} = z_{t-24}$ then ,the SARIMA(p,d,q)(P,D,Q)S = SARIMA(1,1,1)(0,1,1)24, simplified as:

$$z_t = \phi_1(z_{t-1} - z_{t-25}) + z_{t-24} + \Theta_1\varepsilon_{t-24} + \theta_1\Theta_1\varepsilon_{t-25} + \varepsilon_t \qquad (6.2)$$

and the corresponding MW-11 link traffic model equation expressed as:

$$z_t = 0.7234z_{t-1} + z_{t-24} - 0.7234z_{t-25} - 0.9788\varepsilon_{t-1} - 1.0788\varepsilon_{t-24} + 1.06\varepsilon_{t-25} + \varepsilon_t \qquad (6.3)$$

Where z_t, z_{t-1}, z_{t-1}, z_{t-24}, and z_{t-25} are values in their respective lags and ε_t , ε_{t-1}, ε_{t-24}, ε_{t-25} are propagated errors.

Before making the prediction using the fitted model, plotting the diagnosing plot of

the residual is very important to check the leftover nearly Gaussian noise (Figure 6.18), therefore, the graph shown the leftover is nearly Gaussian and hence there is no big spikes needs to address before prediction.

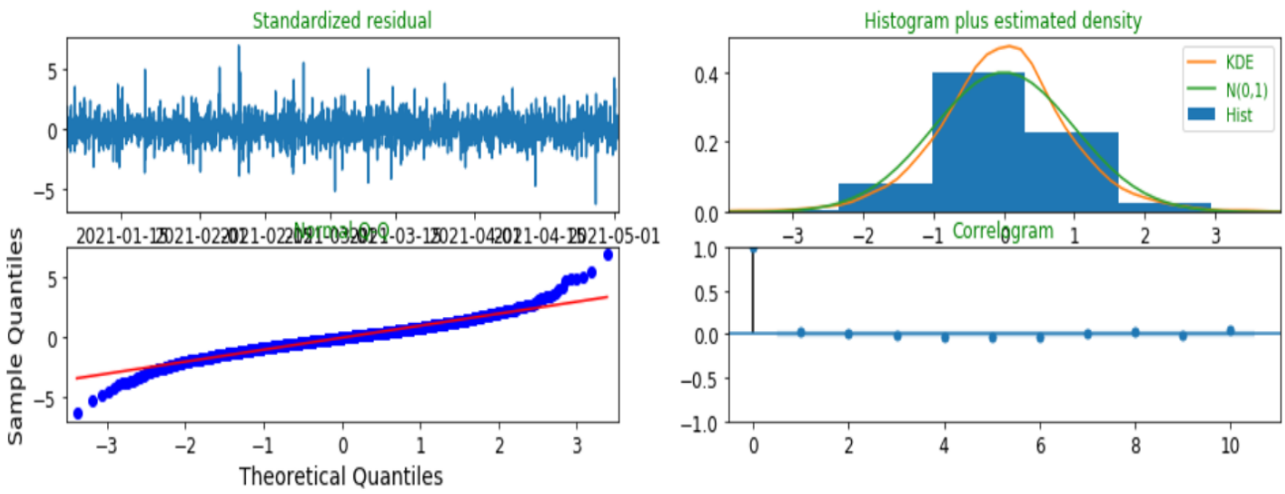


Figure 6.18: Diagonising plot of MW-11

6.3.4 Prediction with 20% test data

Test and predicted plot of MW-11 has been plotted and shown in Figure 6.19, Accordingly, all MW sites in the sample space have been fitted with their respective model for model performance metrics and tabulated in Table 6.4.

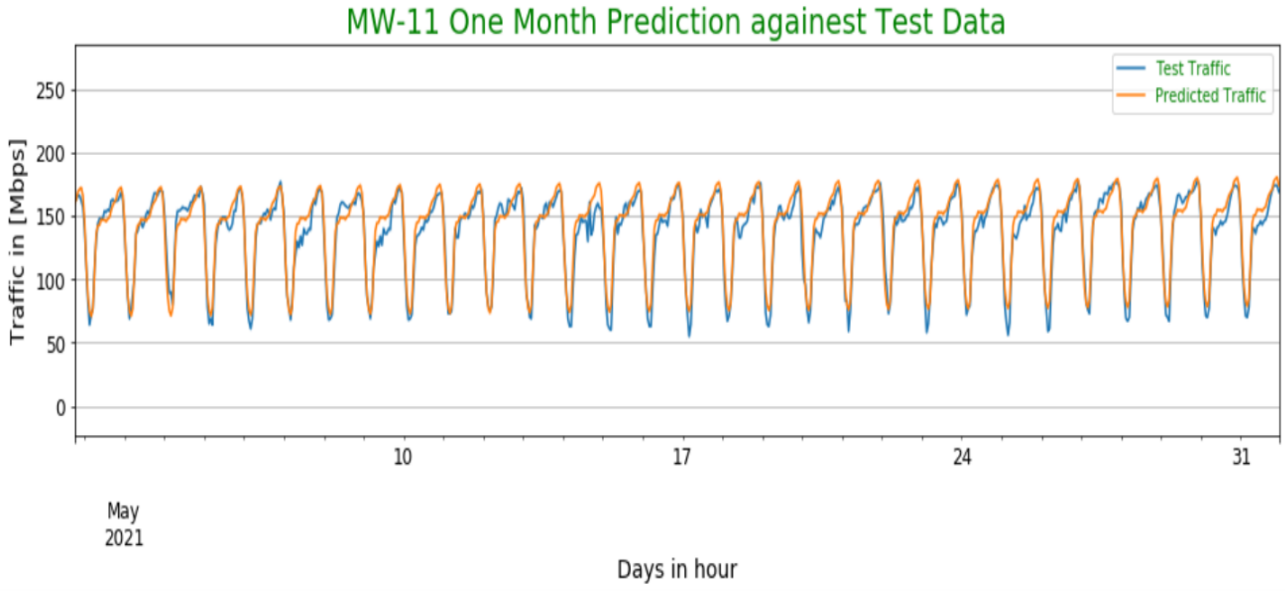


Figure 6.19: Predictions and Test plot of MW-11

Table 6.4: SARIMA model evaluations metrics

MW-Name	R^2	MAE	MSE	RMSE	MAPE
MW-1	0.92659	5.05847	43.27256	6.57819	8.851513
MW-2	0.96449	3.51713	16.77285	4.09547	6.716843
MW-3	0.92772	4.26632	23.67505	4.8657	7.069096
MW-4	0.94844	3.21414	14.0775	3.752	4.623581
MW-5	0.961622	3.749687	22.836023	4.77	11.277
MW-6	0.9648	6.03459	55.05646	7.42	5.662893
MW-7	0.93179	4.17687	19.98187	4.47011	7.690806
MW-8	0.9226	3.04855	14.79763	3.84677	4.81635
MW-9	0.93866	6.0106	49.6947	7.04945	10.748191
MW-10	0.92073	3.35265	13.57465	3.68438	9.102947
MW-11	0.93721	6.49841	66.46043	8.15233	5.494064
MW-12	0.96502	2.22122	9.300398	3.04966	6.539117

6.3.5 Forecasting of the traffic and traffic trend(K)

Forecasting of the traffic with a minimum RMSE leads to forecasting of the mean, thus with the lowest error (Creating error buffer of RMSE <10), traffic has been projected using the size of the test data, and Figure 6.20 have illustrated the forecasted traffic with test data of MW-11. From the graph, the traffic forecasting shifts the traffic mean

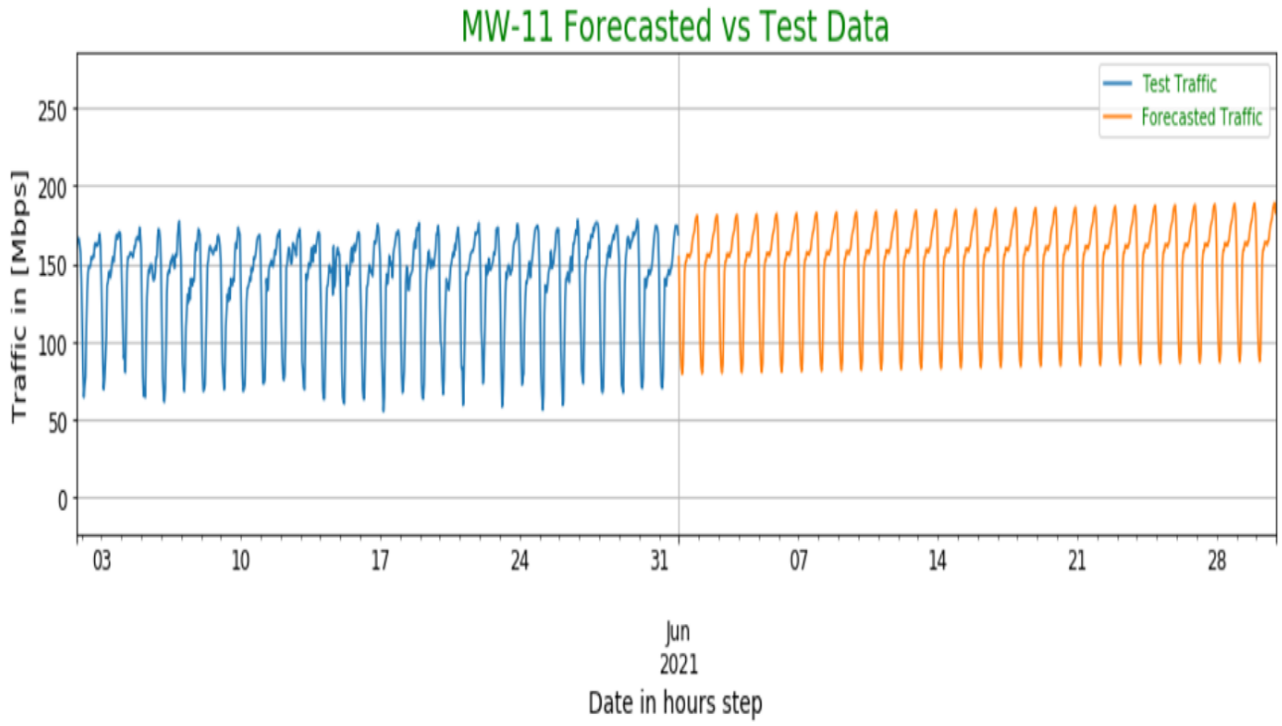
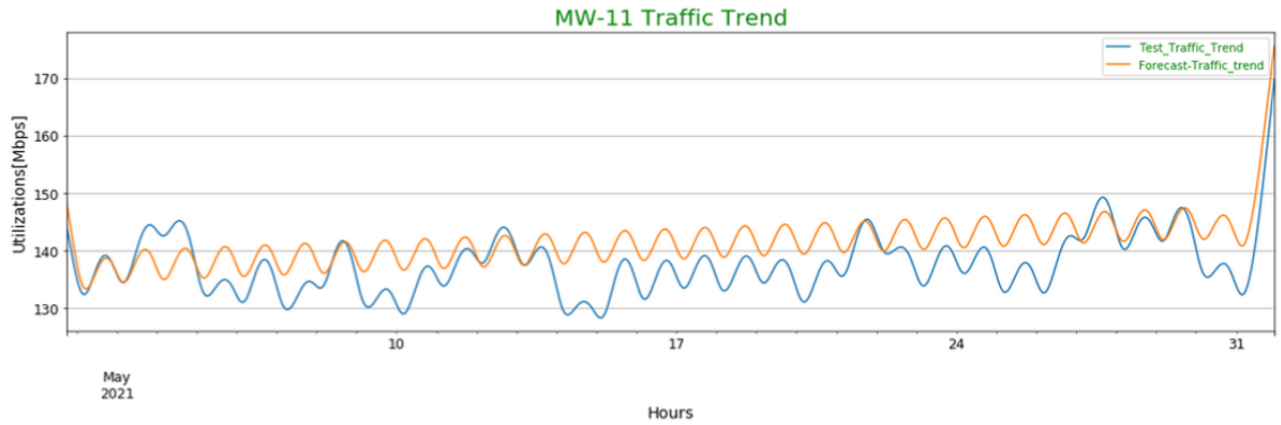


Figure 6.20: Traffic forecasting of MW-11

upward, which signifies the traffic tends to increase (Figure 6.21). The trend factor K is

found by calculating the slope M of the line that connects the two trends' mean values, the mathematical formulation is carried out as shown below.



Test and Forecast trend Mean values

```
: ytest.mean()
: 137.6198347107438
: pred.predicted_mean.mean()
: 141.47683767307697
```

Figure 6.21: Test and Predicted traffic trend comparison of MW-11

$$A_{MW-11}(T) = \frac{1}{n}\sum_{i=1}^{n} A_i = 141.47Mpbs \tag{6.4}$$

$$B_{MW-11}(T) = \frac{1}{n}\sum_{j=1}^{n} B_j = 137.6Mpbs \tag{6.5}$$

Where A(T) and B(T) are mean of Forecast and Test traffic trend [Mbps].

$$M_{11} = \frac{A_{MW-11}(T) - B_{MW-11}(T)}{744} = \frac{141.6 - 137.6}{744} = 0.00502\frac{Mbps}{Hour} \tag{6.6}$$

Therefore, the trend variation ratio K for 3 month period (assuming the traffic optimization period is carried out every 3 month) calculated as:

$$AF_{11} = B_{MW-11}(T) + M_{11}T$$

$$AF_{11} = 137.61Mbps + 0.00502\frac{Mbps}{Hour} * 2160Hours$$

$$AF_{11} = 148.4Mbps$$

$$K_{MW-11} = \frac{B_{MW-11}(T)}{AF_{11}(T)} = \frac{148.4Mbps}{137.6Mbps} = 0.927 \tag{6.7}$$

Thus, MW-11 has a positive traffic trend (slope value is positive), hence it contributes to link capacity dimensioning by a factor of 2-0.927=1.073, or by 7.3%, whereas on MW links whose traffic trend is negative, or the K value is not in the interval of [0.5,1], time series analysis has no contribution to link capacity or the link dimensioning factor will be 2-K=2-1=1. In this case, the link capacity using a probabilistic approach is enough. From probabilistic link capacity dimensioning, the link capacity of MW-11 was 255Mbps, and using the K value from the time series analysis, the total link capacity (LC) becomes:

$$LC_{MW-11} = C(2-K) = 255(1.073) = 273.6Mbps \tag{6.8}$$

Based on the time series findings, the trend factor of all microwave links is calculated and tabulated in Table 6.5. Hence, only three links did not exhibit traffic mean growth (the forecasting mean was less than or equal to the test data mean, implying a K value of one), and thus the link capacity found by the probabilistic approach was sufficient for those links and the maximum link traffic growth depicted on the MW-11 is 7.3% which has a minimum traffic of 35Mbps at the late night and 183Mbps peak traffic with mean traffic of 129.5 Mbps or the busy site in the sample space. Whereas, the minimum traffic growth or nearly constant traffic change found in MW-6. In general , the traffic change of the MW links shows a gradual traffic change of less than 10% in time of the sample space. The total and predicted traffic trend (mean value) of the sample sites depicted in Figure 6.22,6.23, and 6.24 in order to show the forecasted trend and total traffic trend travel concurrently with the same sign of slope (decreasing, increasing, or constant).

Once the total link capacity (LC) has been found using the busy hour traffic (C[Mbps]) and K[%] value, mapping of the appropriate channel spacing and modulation order would be the next step dealt with in the MW link capacity comparison.

Table 6.5: Traffic trend factor K[%] value for 3 month forecasting period

MW-Name	Forecasted Mean	Test Mean	Slope(M)	(K-1) (%)
MW-1	79.82	78.84	0.0014	3.7
MW-2	61.9	61.18	0.001	3.410167351
MW-3	65.51	66.5	-0.00141	0
MW-4	95.31	95.76	-0.000625	0
MW-5	43.33	45.73	-0.00291	0
MW-6	126.06	125.64	0.0006	0.56
MW-7	69.24	68.71	0.00074	2.26
MW-8	76.96	76.3	0.000916667	2.529381707
MW-9	71.41	70.92	0.0007	2.03
MW-10	45.62	44.99	0.000875	2.687
MW-11	141.47	137.6	0.005027778	7.3
MW-12	60.16	58.77	0.0018	6.205

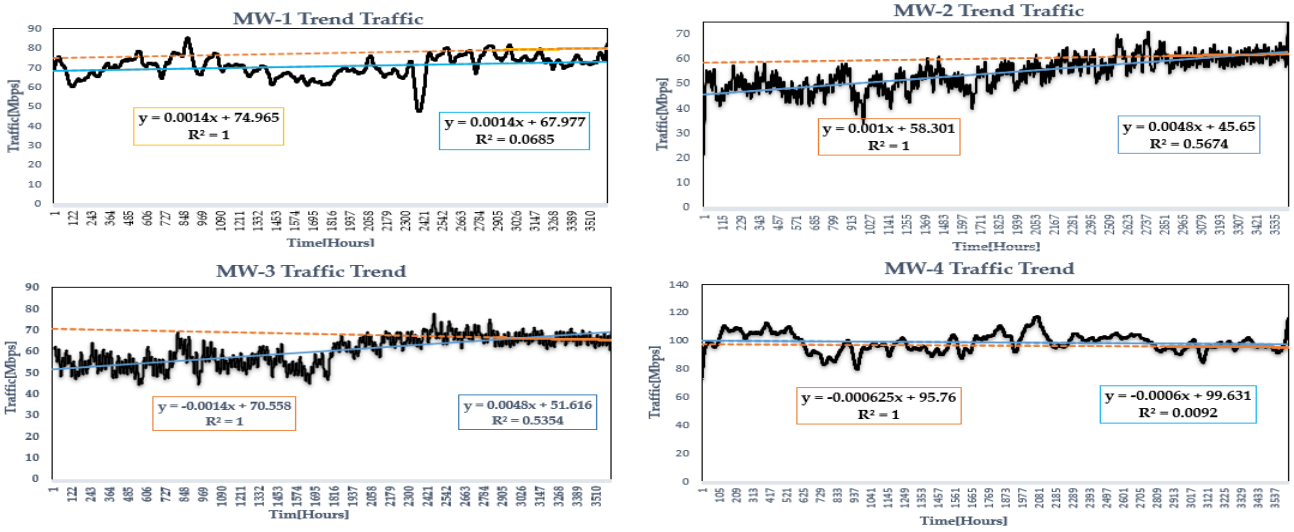


Figure 6.22: Comparison of real vs forecasted traffic trend MW-1-4

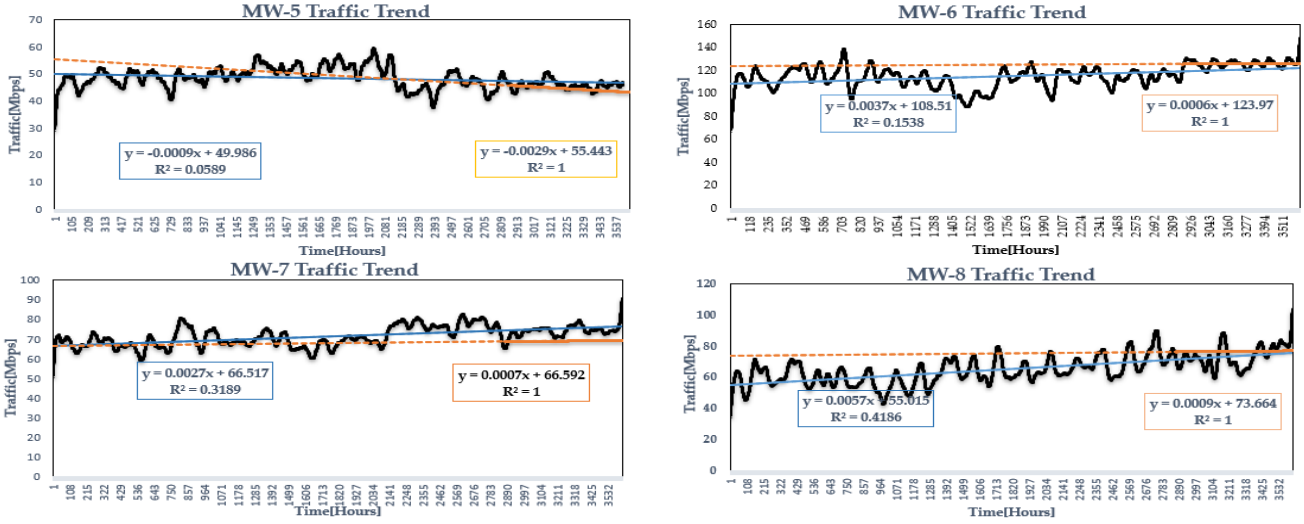


Figure 6.23: Comparison of real vs forecasted traffic trend MW-5-8

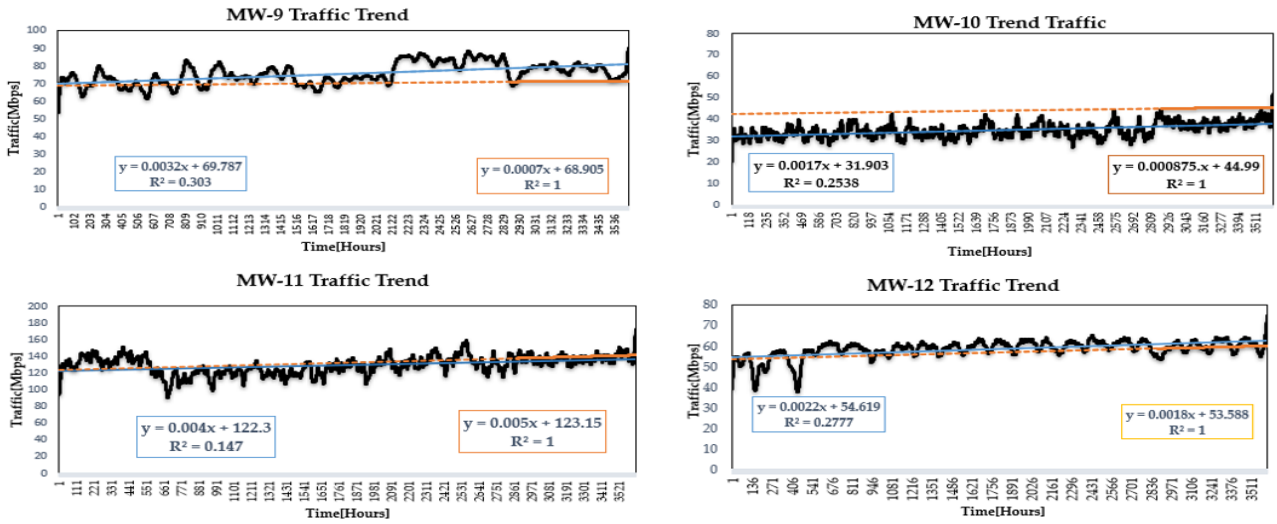


Figure 6.24: Comparison of real vs forecasted traffic trend MW-9-12

6.4 Microwave link capacity comparison

6.4.1 Throughput mapping

Throughput mapping with modulation order and channel spacing based on the data found in the manufacturing manual has been done in this section [2]. All the links were assigned to their respective channel spacing, modulation order, and range of throughput that the calculated total link capacity resided in, as shown in Table 6.6. Discrepancies between the installed and the actual capacity usage creates modulation order and channel spacing changes, which are related to the required SNR, as well as the noise figure of the MW link. Therefore, the following section deals with the SNR and Noise Figure(NF) computations based on the assigned modulation order and channel spacing.

6.4.2 SNR and Noise figure computations

Based on the modulation order and channel spacing from Table 6.6, the calculated value of the required SNR using 10^{-6} Bit Error Rate(BER) for each modulation order, and the receiver's noise figure for the corresponding channel spacing have been computed and tabulated in Table 6.7 and 6.8 respectively. For every modulation order decrement by half, resulting in a 3dB gain. For example, MW-11 was installed with a 512 QAM modulation order, the calculated link capacity resides on a modulation order of 128

Table 6.6: MW throughput mapping [32]

MW-NAME	LC[Mbps]	Chan.Spacing[MHz]	Mod.order (QAM)	Throughput[Mbps]
MW-1	174	28	256	182 to 230
MW-2	143	28	128	160 to 203
MW-3	233	56	128	323 to 409
MW-4	151	28	128	160 to 203
MW-5	149	28	128	160 to 203
MW-6	245	56	128	323 to 409
MW-7	203	56	128	323 to 409
MW-8	202	56	128	323 to 409
MW-9	205	56	128	323 to 409
MW-10	146	28	128	160 to 203
MW-11	274	56	128	323 to 409
MW-12	255	56	128	323 to 409

QAM, so the SNR link requirement changed from 32.6 dB to 26.6dB. Hence, by real traffic link capacity dimensioning, 6dB gain has been found.

Table 6.7: Required SNR for QAM modulation

MW-Name	Mod.order	Error value	SNR	New SNR[dB]	Old SNR[dB]
MW-1	256	0.9999964	455.43	29.6	29.6
MW-2	128	0.9999959	897.81	26.6	32.6
MW-3	256	0.9999964	455.43	29.6	32.6
MW-4	128	0.9999959	897.81	26.6	29.6
MW-5	128	0.9999959	897.81	26.6	29.6
MW-6	128	0.9999959	897.81	26.6	32.6
MW-7	128	0.9999959	897.81	26.6	32.6
MW-8	128	0.9999959	897.81	26.6	29.6
MW-9	128	0.9999959	897.81	26.6	32.6
MW-10	128	0.9999959	897.81	26.6	29.6
MW-11	128	0.9999959	897.81	26.6	32.6
MW-12	128	0.9999959	897.81	26.6	29.6

On the other hand, the noise figure requirement is directly related to channel spacing, and therefore, if there is a frequency change from 56MHz to 28MHz, the required noise figure decreases by 3dB or the 3dB gain of receiver sensitivity has been found besides freeing of 28MHz channel spacing.

Table 6.8: Noise figure gain vs Channel spacing

MW-Name	LC[Mbps]	Channel[MHz]New	Channel[MHz]Old	NF Gain[db]
MW-1	174	28	56	3
MW-2	143	28	56	3
MW-3	233	56	56	0
MW-4	151	28	56	3
MW-5	149	28	56	3
MW-6	245	56	56	0
MW-7	203	56	56	0
MW-8	202	56	56	0
MW-9	205	56	56	0
MW-10	146	28	56	3
MW-11	270	56	56	0
MW-12	255	56	56	0

6.4.3 Fade margin computations

Transmitter power, free space path loss, and receiver sensitivity are the parameters used to determine the fade margin. All parameters are unchanged in this study, except receiver sensitivity and modulation order. Transmitted power of 1dBm was chosen for all sites to make the calculations more comparable, and all parameters such as antenna diameter (0.3 meters with 60% radiation efficiency), frequency (26GHz), and respective link distance remained as it has built. The fade margin of each link has been calculated using Equation 5.1 and listed in Table 6.9 by specifying fade-margin-old and fade-margin-new, respectively for planned and real-time traffic values. From the table, the

Table 6.9: Sensitivity and fade margin of MW links at 26GHz

MW-Name	dis[km]	FSL[db]	Ptx + 2GTR	Sensity-Old[dbm]	Sensit-New[dbm]	FM-Old[db]	FM-New[db]
MW-1	0.68	117.4	73.2	-66.9	-69.9	22.72	25.76
MW-2	0.79	118.1	73.2	-63.9	-72.9	18.4	27.44
MW-3	0.92	120.015	73.2	-63.9	-66.9	20.1	23.11
MW-4	0.89	119.721	73.2	-63.9	-69.9	20.4	26.4
MW-5	0.65	116.99	73.2	-66.9	-69.9	23.1	29.3
MW-6	0.82	119.37	73.2	-63.9	-69.9	18.1	24.11
MW-7	1.04	121.1	73.2	-63.9	-69.9	16.02	22.02
MW-8	0.81	118.91	73.2	-66.9	-69.9	21.19	24.19
MW-9	0.77	118.46	73.2	-63.9	-69.9	18.6	24.66
MW-10	0.68	117.41	73.2	-63.9	-69.9	19.72	28.72
MW-11	0.72	117.88	73.2	-63.9	-69.9	19.2	25.24
MW-12	0.85	119.32	73.2	-63.9	-66.9	20.8	23.80

fade margin change ranges from 3 to 9 dB, which has a considerable impact on link availability. Hence, the link unavailability has been computed in the following section based on the given fade margin values.

6.4.4 Link unavailability computations

The unavailability of the link in percent has been calculated by applying equation 5.10. The fade margin of the existing and newly calculated value is already found in Table 6.9 above. Figure 6.25 and 6.26 have shown the link unavailability for MW-11 before and after optimizations respectively, that has been shown one forth decrement, as the fade margin of the link increases by 6dB. The comparison of link unavailability for all

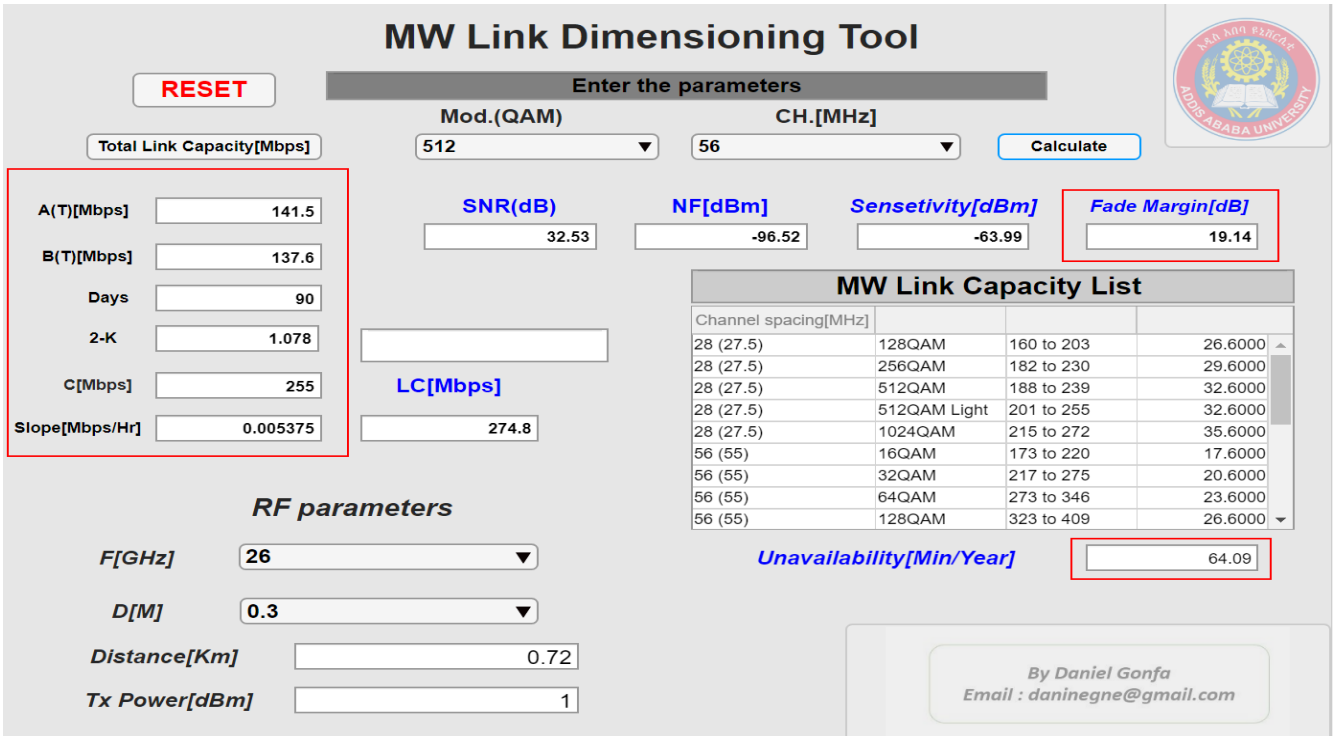


Figure 6.25: MW-11 unavailability before optimization

sites in the sample space before and after optimization has been summarized and shown in Figure 6.27. The unavailability of the link decreases as the fade margin increases, and this can be accomplished by using real-time link capacity optimization while reducing channel spacing and modulation order. As listed in Table 6.9, the fade margin gain varies from 3dB spanned to 9dB, and as the fade margin of the link increased by 3dB, the unavailability of the link decreased by half and when it increased by 6dB, the

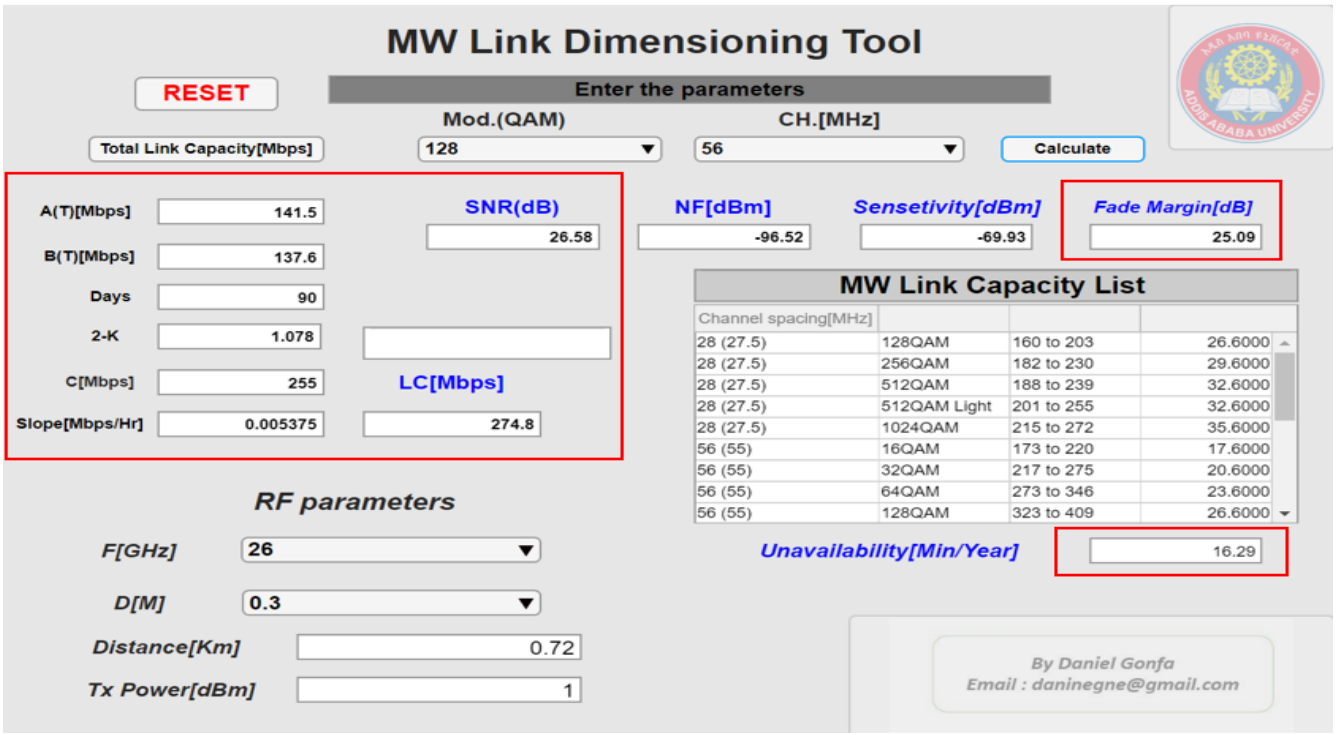


Figure 6.26: MW-11 unavailability after optimization

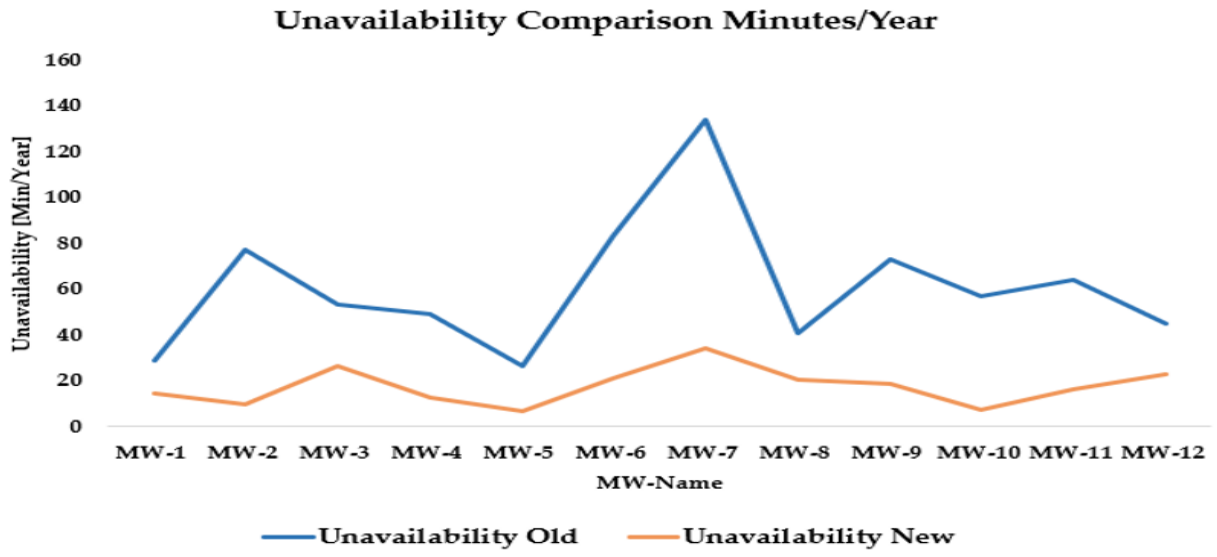


Figure 6.27: Unavailability comparison of all sites

unavailability decreased by one fourth. Therefore, this improvement makes the MW link more robust comparatively, without increasing the transmission power or antenna diameter.

6.4.5 Link utilization improvement

Low percentage of utilization (underutilized links) in microwave backhaul network was one of the thesis's motivations (Figure 1.2). Thus, Figure 6.28 shows the optimized percentage of link utilization plotted against the total installed capacity of sample sites using equation 6.9.

$$Capacity - utilization(\%) = 100 * \frac{Actual[Mbps]}{Installed[Mbps]} \tag{6.9}$$

Before re-dimensioning, the minimum and maximum link utilization were 31% and

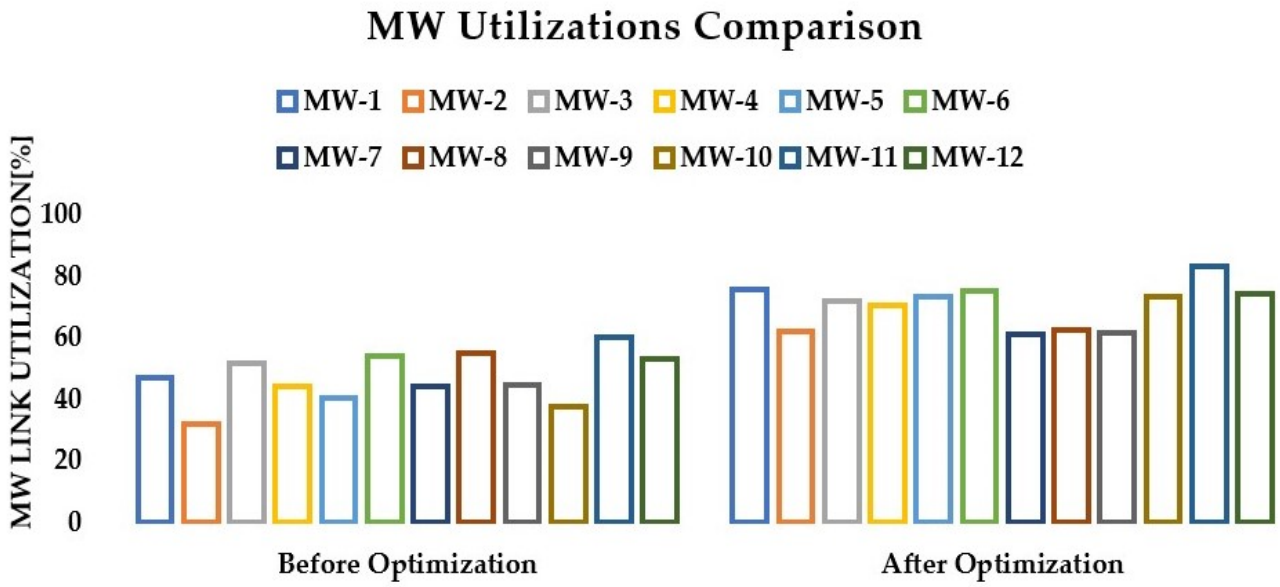


Figure 6.28: Link utilization improvement

60% respectively, whereas, after re-dimensioning, the link utilization increased to 61% as a minimum and 83% as a maximum. In general, the link capacity provisioning based on real-time traffic helps to improve the performance of underutilized links in the case of microwave backhaul, while optimizing the usage of scarce resources, such as channel spacing and modulation order.

7

Conclusion and Future work

7.1 Conclusions

The data demand of mobile customers is increasing from time to time, and service providers are continuously upgrading their microwave backhaul network. During upgrading, either installing a new or increasing the link capacity of the existing site can be a solution. While upgrading the congested link, finding the optimum link capacity for required traffic is a challenge for the planning team, and hence implementing peak traffic as a rule of thumb is considered as an option. In the process, when traffic links do not transport as much as planned, links become underutilized. Unfortunately, these underutilized links not given attention since they are considered normal even though there is a wasting of resources such as channel frequency and unwanted cost for license activation of modulation order. Thus, this thesis presented an approach to find the optimum link capacity of microwave backhaul using probabilistic and time series analysis. The study has taken 12 sites of traffic data from Addis Ababa as an input. The output of the study indicates that, link dimensioning using a real-time provisioning approach improves the overall utilization of the microwave link (minimized underutilized links)

in general and link performance in particular, besides freeing non-utilized resources such as channel spacing (frequency MHz) and modulation order. As it is shown in Figures 7.1 and 7.2 below, 40% of the total sites from the sample space freed 28MHz of frequency, and also 11 sites are changed their modulation order by one fourth.

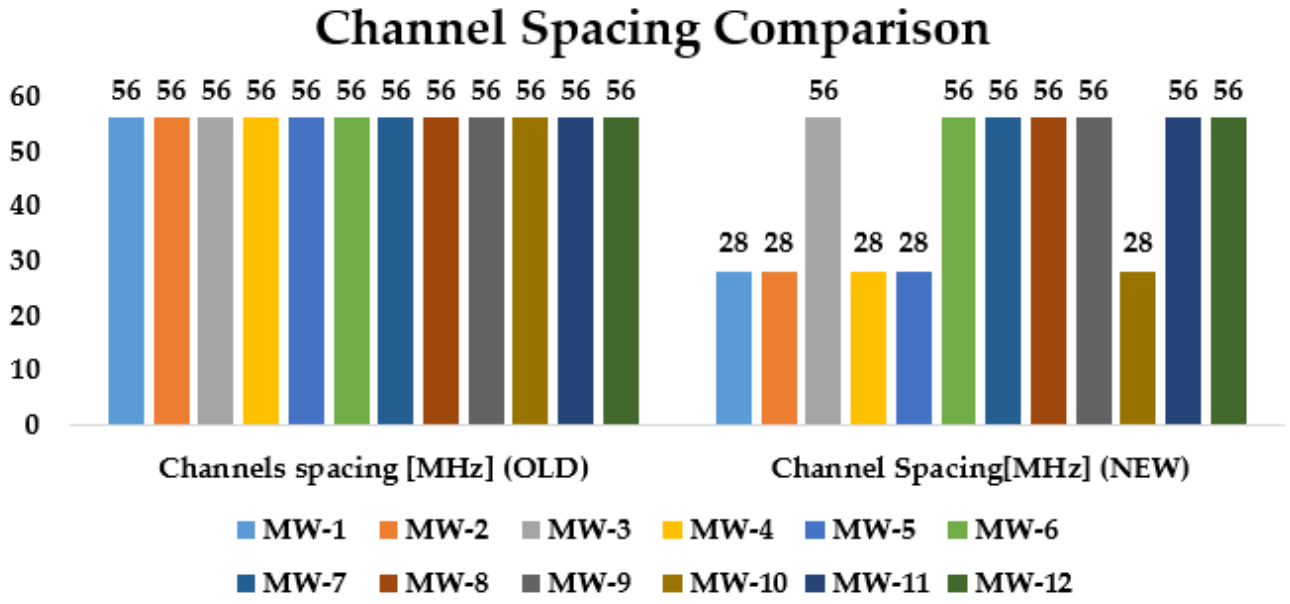


Figure 7.1: Channel spacing comparison

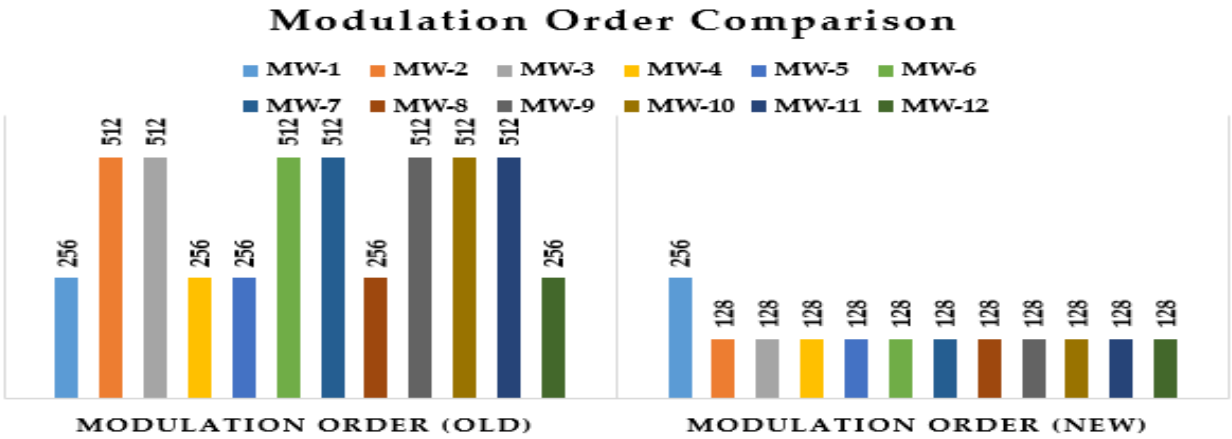


Figure 7.2: Modulation's order comparison

From busy hour link capacity provisioning approach traffic model, GEV, Lognormal and Meent's link capacity approach are more appropriate to model the traffic of site at peak hours. Whereas, from time series analysis, the traffic change in the long run or the trend varies at small rate which is less than 10% even at the busiest site of the sample

space. Therefore, dimensioning the link capacity of MW link using busy hour traffic and 10% as traffic margin can generalized the total link capacity of each links in the sample space to optimize network resources and save unwanted costs.

7.2 Future work

Based on the conclusions, it is important to point out the following considerations in future work:

1. This link capacity dimensioning approach can be adopted to the optical backhaul uplink dimensioning, since it is only based on the aggregate traffic of the link instead of traffic type.

2. The busy hour link provisioning approach can be applied for argument reconciliations between customers and service providers using blocking probability as a service level agreement parameter.

3. Network resources are primarily dependent on traffic demand generated from the user side, which is unpredictable by nature. Dimensioning the link based on a static link traffic approach is quite cumbersome as the number of links increases, and it is difficult to track traffic changes on a timely basis. Instead, upgrading this approach to a dynamic provisioning approach can be considered as one of the research areas.

4. Last but not least, Blocking probability is defined by service providers and service types. Defining the value per service is one of the challenges in the telecommunications industry. Therefore, applying this approach to find the optimum blocking probability per service can be considered as another research area.

Refrences

[1] K. A. Fidele, W. A. Syafei, *et al.*, "Denial of service (dos) attack identification and analyse using sniffing technique in the network environment," in *E3S Web of Conferences*, vol. 202, p. 15003, EDP Sciences, 2020.

[2] H. M. E. team, "Addis ababa microwave high level design in ethio telecom," 2014.

[3] A. Rusan and R. Vasiu, "Emulation of backhaul packet loss on the lte s1-u interface and impact on end user throughput," in *2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, pp. 529–536, IEEE, 2015.

[4] A. R. Mishra, *Fundamentals of cellular network planning and optimisation: 2G/2.5 G/3G... evolution to 4G*. John Wiley & Sons, 2004.

[5] K. M. S. Huq, S. Mumtaz, and J. Rodriguez, ""an overview of 4g system-level energy-efficiency performance"," *Energy Management in Wireless Cellular and Ad-hoc Networks*, pp. 45–64, 2016.

[6] R. M. Nuru and A. Abebe, *Optimal Sizing of Grid-PV Hybrid System for ethio telecom Access Layer Devices and Its Economic Feasibility*. PhD thesis, Addis Ababa University Addis Ababa, Ethiopia, 2017.

[7] G. Kizer, *Digital microwave communication: engineering point-to-point microwave systems*. John Wiley & Sons, 2013.

[8] M. Alasmar and N. Zakhleniuk, "Network link dimensioning based on statistical analysis and modeling of real internet traffic," *arXiv preprint arXiv:1710.00420*, 2017.

[9] A. Pras, L. Nieuwenhuis, R. van de Meent, and M. Mandjes, "Dimensioning network links: a new look at equivalent bandwidth," *IEEE network*, vol. 23, no. 2, pp. 5–10, 2009.

[10] A. Mahmood, L. B. M. Kiah, S. R. B. Azzuhri, and A. N. Qureshi, "Wireless backhaul network optimization using automated kpis monitoring system based on time series forecasting," in *2018 IEEE World Symposium on Communication Engineering (WSCE)*, pp. 1–6, IEEE, 2018.

[11] A. Papoulis and S. U. Pillai, "Probability, random variables, and stochastic processes," 2002.

[12] A. Luceno, ""fitting the generalized pareto distribution to data using maximum goodness-of-fit estimators"," *Computational Statistics & Data Analysis*, vol. 51, no. 2, pp. 904–917, 2006.

[13] H. B. Hasan, N. B. Salam, and S. B. Kassim, "On the use of gev distribution: A case study of temperature in malaysia," in *2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE)*, pp. 1–6, IEEE, 2012.

[14] Z. Djurovic, B. Kovacevic, and V. Barroso, "Qq-plot based probability density function estimation," in *Proceedings of the Tenth IEEE Workshop on Statistical Signal and Array Processing (Cat. No. 00TH8496)*, pp. 243–247, IEEE, 2000.

[15] J. Liu, W. Zhang, J. Yuan, D. Jin, and L. Zeng, "Monitoring the spatial-temporal effect of internet traffic based on random matrix theory," in *2008 33rd IEEE Conference on Local Computer Networks (LCN)*, pp. 258–265, IEEE, 2008.

[16] R. Norden, "A survey of maximum likelihood estimation: Part 2," *International Statistical Review/Revue Internationale de Statistique*, pp. 39–58, 1973.

[17] H. Chen and Y. Liang, "Analyzing of multi service backhaul network based on ip radio access network technology," in *2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS)*, pp. 189–192, IEEE, 2016.

[18] R. Van De Meent, *Network link dimensioning: a measurement & modeling based approach*, vol. 6. Remco van de Meent, 2006.

[19] M. Alasmar, R. Clegg, N. Zakhleniuk, and G. Parisis, "Internet traffic volumes are not gaussian–they are log-normal: An 18-year longitudinal study with implications for modelling and prediction," *IEEE/ACM Transactions on Networking*, 2021.

[20] H. Van den Berg, M. Mandjes, R. Van de Meent, A. Pras, F. Roijers, and P. Venemans, "Qos-aware bandwidth provisioning for ip network links," *Computer Networks*, vol. 50, no. 5, pp. 631–647, 2006.

[21] B. Vishwas and A. Patel, *Hands-on Time Series Analysis with Python*. Springer, 2020.

[22] F. Divina, M. Garcia Torres, F. A. Goméz Vela, and J. L. Vazquez Noguera, "A comparative study of time series forecasting methods for short term electric energy consumption prediction in smart buildings," *Energies*, vol. 12, no. 10, p. 1934, 2019.

[23] G. Dengia, T. Jemal, and S. Catolos, "Microwave link design between jimma main campus and agaro branch,"

[24] R. Kaduskar and A. D. Kavishwar, "Mobile backhaul network," in *2011 International Conference on Information and Network Technology, IPCSIT*, vol. 4, 2011.

[25] A. Hilt, "Availability and fade margin calculations for 5g microwave and millimeter-wave anyhaul links," *Applied Sciences*, vol. 9, no. 23, p. 5240, 2019.