CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

3GPP	3rd Generation Partnership Project
3GPP2	3rd Generation Partnership Project 2
AICC	Akaike Information Criterion with Correction
AMBR	Aggregated Maximum Bit Rate
AR	Autoregressive Model
MA	Moving Average Model
ARMA	Autoregressive Moving-Average Model
ARIMA	Autoregressive Integrated Moving-Average Model
AF	Application Function
CS	Circuit Switched
CSCF	Call Session Control Function
DL	Downlink
UL	Uplink
E-UTRAN	Evolved-UTRAN
EMM	EPS Mobility Management
eNodeB	evolved NodeB
EPC	Evolved Packet Core
EPS	Evolved Packet System
GERAN	GPRS EDGE Radio Access Network

GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile communications
GTP	GPRS Tunnelling Protocol
HLR	Home Location Register
HSDPA	High Speed Downlink Packet Access
HSPA	High Speed Packet Access
HSS	Home Subscriber Server
HSUPA	High Speed Uplink Packet Access
IETF	Internet Engineering Task Force
IEEE	Institute of Electrical and Electronics Engineers
IMS	IP Multimedia Subsystem
IMSI	International Mobile Subscriber Identity
ITU	International Telecommunication Union
LTE	Long Term Evolution
MBMS	Multimedia Broadcast and Multicast Service
MME	Mobility Management Entity
NGN	Next Generation Network
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access

PGW	PDN Gateway
PCEF	Policy and Charging Enforcement Function
PCRF	Policy and Charging Rules Function
PDP	Context Packet Data Protocol Context
PDN	Packet Data Network
PLMN	Public Land Mobile Network
PS	Packet Switched domain
PSTN	Public Switched Telephone Network
RNC	Radio Network Control
SGW	Serving Gateway
SAE	System Architecture Evolution
SDF	Service Data Flow
SGSN	Serving GPRS Support Node
SIP	Session Initiation Protocol
SM	Session Management
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
UTRAN	Universal Terrestrial Radio Access Network

WCDMA	Wide band Code Division Multiple Access
GBR	Guaranteed Bit Rate
Non-GBR	non-Guaranteed Bit Rate
AGBR	Adaptive GBR
PGBR	Pure GBR
BW	Bandwidth
Gn	SGSN to GGSN Interface
MBR	Maximum Bit Rate
QUIC	Quick UDP Internet Connections

CHAPTER 1

INTRODUCTION

The Long Term Evolution (LTE) telecommunication technology has been introduced to provide more capabilities and functionalities to support innovative mobile services. The LTE represents a revolution in telecom technology to provide faster communication and higher data transmission with improved coverage and spectrum efficiency as well as optimized radio access network.

The LTE technology introduces new architectural changes which indicate that the EPC core network is more centralized with more responsibilities to be considered while the radio access network is more distributed. Furthermore, the LTE architecture simplifies the radio access network side by considering only one system which is the "eNodeB" and introduces complicated functionalities and intelligence towards the EPC network. Therefore, the EPC systems are supposed to provide new/evolved responsibilities in LTE technology compared to the 3G & 2G technologies (3GPP TS 36.300 (2015)). Hence, the EPC network has more challenges to be resolved and more intelligence to be provided to meet the expectations of the evolved technology (LTE) and the Next Generations Network services.

The real-time and conversational LTE services require guaranteed resources to be strictly allocated for the whole lifetime of the service call. The LTE mobile core network (EPC) resource allocation approach is inadequate with regards to the guaranteed resources used by those services. More precisely, the EPC mobile gateway system is not capable of properly utilizing the unused bandwidth of the guaranteed resources when the mobile service is not fully using the reserved bandwidth. In this thesis, we will focus on optimizing the guaranteed resource utilization for the LTE mobile services and present an adaptive approach which enhances the resource reservation for the LTE mobile guaranteed services. Our approach provides techniques to: analyze the ongoing mobile guaranteed traffic usage, provide time-series models that mathematically represent the conducted data, forecast the mobile service guaranteed resource consumption, identify the wasted/unused resources, and utilize these resources by other services. Our approach introduces a novel type/method of resource allocation in 3GPP standards. Our experiments will be conducted on datasets captured on an emulated LTE environment. The goal of our experiments is to show that our approach is feasible and beneficial in enhancing the resource allocation for the LTE mobile services and increasing the overall throughput of the LTE/EPC networks.

1.1 Overview: LTE/EPC Networks

The telecommunication evolution 4G/LTE technology came in to provide higher data rate and lower delay with improved coverage and spectrum efficiency. The LTE systems provide capable signaling as well as optimized radio transmission in the radio access network.

Through comparing the telecom wireless technologies such as 2G, 3G and WiMAX with the 4G/LTE radio access system, the LTE system provides evolution to the telecommunication as it offers higher data rate, bigger capacity, lower delay, and more improvement on coverage and spectrum efficiency (Shin *et al.* (2008b) and Verizon (2010)). The radio access network part of the LTE system has only one node which is the Evolved Node B (eNodeB) while the 3G technology in the Universal Mobile Telecommunications System (UMTS) has two nodes: NodeB and Radio Network Controller (RNC). 2G technology also has two nodes: Radio Base Station (RBS) and Base Station Controller (BSC). This architectural change helps to offer less transmission delay in the radio access network and moves more intelligence towards the core network. Also, 3GPP has proposed key system features such as the default PDN session for the User Equipment (UE), and also proposed new architecture of the policy control and charging to have more control on the mobile broadband services.

The evolved packet system (EPS) defines a single core network -all IP based- for multiple heterogeneous accesses that provide triple play services in the next generation networks. In the 2G and 3G technologies, there were two separate core networks: the circuit-switched core network which, delivers the telephony services over circuit switching methodology, and the packet-switched core network, which delivers the wireless mobile broadband data services over the packet switching methodology using all-IP network. In the LTE technology, the evolved packet core network (EPC) is considered as a single mobile core network to run all services required by the wireless user equipments (3GPP TS 36.300 (2015)).

The Packet Data Network (PDN) session is established in the LTE/EPC network for UE connectivity to the internet or to any other network. As part of the PDN session, the UE can have a default bearer and dedicated bearer(s). The default bearer is used for session connectivity and best-efforts traffic. It is established once the UE attaches to the network and an IP address will be assigned to it. It is the responsibility of EPC gateway system to assign the IP address and maintain the UE PDN session(s) and their bearers. On the other hand, the dedicated bearers can be activated to run specific services that require special QoS requirements. Based on the guarantee criteria of the resource reservation, the dedicated bearer can be classified as guaranteed or non-guaranteed bearer (Ekstrom (2009)).

The dedicated guaranteed bearer (GBR) will be used to run special services that require bandwidth to be reserved for the whole lifetime of the service/call. Usually, the GBR bearer is established once the UE demands a service that is provisioned to trigger the guaranteed bearer creation. Conversational voice & video and real-time gaming are examples of the services that would use GBR bearer. On the other hand, the dedicated non-guaranteed bearer (non-GBR) can be used to run services that require special priorities, but it does not require bandwidth to be reserved for the whole lifetime of the service. The non-GBR can remain established for long time, as it does not require bandwidth to be strictly reserved.

1.2 Problem Statement

As part of our work, we will concentrate on the challenges that the EPC network encounters and highlight the related weaknesses which could reduce the systems capabilities and efficiency in the EPC network. By studying the current EPC network design, we can observe that many limitations and weaknesses can be identified. As indicated in the 3GPP standards (3gpp 36.300), the LTE/EPC dedicated guaranteed bearers require guaranteed resources/bandwidth to be strictly allocated for the whole lifetime of the running service call. No other services can share the guaranteed resources, even if they are not fully used.

The EPC network internal design is inadequate with regards to the resource reservation techniques used to carry out the guaranteed dedicated services. So when an LTE mobile is running any service that requires guaranteed bandwidth reservation and this bandwidth is not fully used, the unused guaranteed bandwidth is considered as wasted resources in the EPC gateway and consequently the whole EPC network gets affected. Additionally, the EPC gateway does not have the capabilities to utilize properly the unused guaranteed bandwidth when it is not fully used by the mobile service.

Considering that guaranteed bandwidth is not fully utilized, the operational EPC network throughput and capacity is less than the actual capabilities. According to the 3GPP standards in (3GPP TS 36.300 (2015)), the EPC network indicates that resource allocation for bearers is only based on guaranteed and non-guaranteed rules. The EPC network only provides strict and static implementation of the bearer resource allocation and does not introduce any capabilities to have dynamic or adaptive way for resource allocation. Furthermore, the EPC systems, according to 3GPP standards, do not have any intelligence to provide traffic forecasting to predict the future situation of the running traffic in order to perform more enhanced bandwidth utilization and resource allocation. All of these issues may lead to scalability problem at the EPC gateway system, which would require extra cost for operator's network expansion.

1.3 Objectives

In this thesis, we will focus on enhancing resource utilization for LTE mobile services. Our main objective is to design and model an *adaptive* technique which improves the resource reservation for the LTE mobile guaranteed services and minimizes the wasted resources of the guaranteed bandwidth allocation in the LTE/EPC network. To achieve our objective, we introduce a novel technique that provides smart, efficient and adaptive approach for the LTE bearers resource allocation. The new concept of adaptive guaranteed bearer will provide an intermediate class between the strict guaranteed and the very open non-guaranteed resource allocation types.

This concept of adaptiveness would provide flexible implementation which allows us to apply some forecasting methods to estimate the wasted bandwidth in the guaranteed resources and utilize them accordingly to reduce the waste and increase the system throughput. Furthermore, the approach indicates that several mobile sessions of the adaptive guaranteed services can be conducted together in order to forecast the overall usage of all those sessions, estimate a pool of unused resources and utilize some of those resources for other non-guaranteed running services. Whenever the adaptive guaranteed service(s) request back the contributed resources, they will get it back from a standby resources pool which will be designated for this purpose.

As part of our work, we need to study the traffic profile and characteristics of the LTE guaranteed services according to the 3GPP standards. This would be required for dataset formulation, analysis and examination. Datasets will be the input for our proposed approach. Furthermore, we aim to provide an algorithm for mathematically representing the LTE mobile guaranteed traffic dataset through conducting time-series models. This includes using and investigating the time-series models; AR(p), MA(q), ARMA(p,q) and ARIMA(p,d,q). The modeling will be validated for the applied time-series to find suitable time-series models that can give better results.

In addition, our research provides the data forecasting method based on the designed time-series model where the predictor function will be used to estimate the potential unused bandwidth of the guaranteed bearers in the EPC network. The forecast error will be calculated to ensure higher efficiency. The idea is to release part of the forecasted guaranteed unused bit-rate to be utilized by other running services. This represents the benefit of our proposed approach in increasing the LTE/EPC gateway capacity and throughput.

Finally, a safety model will be proposed to ensure avoiding any disturbance for the contributing guaranteed bearers. A safety technique will be established based on the forecast error provided by the time-series forecast model. The safety model ensures the resource availability when the guaranteed bearers require their resources back.

1.4 Methodology Overview

In this research, our approach proposes a new concept of having *adaptiveness* in the guaranteed bearers implementation to provide flexibility which allows us to apply some forecast methods in order to estimate the wasted bandwidth in the guaranteed resources and utilize them accordingly to reduce the waste and increase the system throughput.

To achieve this, our approach indicates that the guaranteed bearers can be further classified - based on the criteria of adaptiveness- into two types: (i) *adaptive-guaranteed bearers* that allow adaptive resource allocation and contribute the unused guaranteed resources. All contributed resources will be added up in a pool. These bearers can be considered as *contributing-bearers*, and (ii) *pure-guaranteed bearers* that cannot accept any adaptive resource allocation even if they are not fully used. The later represents the current behavior of the guaranteed bearer as defined by the 3GPP standards. Our approach would manage unused resources into two pools. One can be considered as guaranteed standby resources to be used any time by the guaranteed bearers once the resources are requested back, and the other pool of the unused resources can be utilized by the non-guaranteed bearers. Our approach considers these bearers to be *acquiring-bearers* based on their willingness to utilize more resources to reach the MBR limit.

Considering the concept of having adaptive resource reservation in guaranteed mobile services, we should have robust techniques that can forecast and predict the usage of the guaranteed resources for short-term period for the ongoing bearers in the EPC gateway system, by forecasting the usage of the resources, the unused bandwidth will be measured and estimated.

Time-series will be used in this research work to model the data of the bit-rate usage of the mobile LTE guaranteed services. Time-series analysis includes methods for analyzing series of data in order to get meaningful characteristics of that data. One of the most important data characteristics is the *Stationarity*. The ACF & PACF functions will be used to examine the data stationarity. Based on the characteristics of the dataset examined, several time-series models will be investigated to provide an appropriate model to represent the dataset. For stationary dataset, the models Autoregressive AR(p), Moving Average MA(q) and Autoregressive Moving Average ARMA(p,q) can be used. For non-stationary dataset, the Autoregressive Integrated Moving Average ARIMA(p,d,q) model can be considered as it uses the "integrated" property to transform the given dataset to stationary series through applying the Differencing technique (Robert H. Shumway (2010)). Along with the Differencing technique, Box-Cox Transformation will be helpful in converting the data to become stationary. The resulted time-series model will be utilized in the prediction function that will provide the forecasted unused resources in the LTE/EPC network.

To evaluate the forecast results, the Mean Squared Error (MSE) will used to show the accuracy of both. The chosen time-series model and the prediction parameters selection. The forecasting error behavior, which is driven from the MSE, will be used in calculating the standby resources pool that will be reserved for the guaranteed bearers to be used any time once the resources are requested back.

1.5 Thesis Contribution

In this thesis, we propose and design an adaptive technique which enhances the resource reservation for the LTE mobile guaranteed services. In the proposed approach, we introduce a novel algorithm for resource allocation in 3GPP standards. The algorithm aims to include *adaptiveness* in the guaranteed bearer resource allocation. The algorithm also consolidates the guaranteed traffic usage in one pool to estimate the expected waste of the resources and utilize them properly. Our technique ensures and guarantees the resource availability through designing the "Safety Model" which complements our algorithm. We also propose a framework which analyzes the mobile data traffic and determines the mathematical characteristics. As part of the framework, we utilize some methods that can help in converting the mobile data series into stationary data.

In this thesis, we also propose another algorithm which mathematically represents the mobile data series into time-series model. Different models and techniques were used to design the time-series, validate the model and perform comparison in order to get better data representations. The proposed algorithm also helps to forecast the consolidated mobile data guaranteed resource consumption. The forecast process will help to identify the unused resources in the LTE/EPC system.

Our technique would help to maximize the use of resources by other mobile services. This would increase LTE/EPC system throughput and capacity. It would also help to avoid network expansion at telecom operators that could be caused by scalability problem. We believe that our technique will help network operators especially because our technique concentrates on improving usage of telecom network resources in particular for LTE mobile networks.

1.6 Thesis Outline

The rest of the thesis is organized as follows: in Chapter 2, we present some background about LTE/EPC network architecture, LTE bearer resource reservation, QoS mechanisms and some challenges of the LTE/EPC network. We also provide a detailed description about the related work which has been done in the QoS and Resource Allocation in LTE/EPC network. In addition, we provide a review of the related research work which has been done in the time-series modeling and forecast fields. In Chapter 3, we provide a full description of our proposed approach, explain the mathematical modeling of our approach and provide details of the methodologies and algorithms that would be used to achieve our objectives.

In Chapter 4, we apply our approach to a single guaranteed service that carries video conversational call where we analyze the dataset, prepare it for modeling, find the time-series model that fits the data, validate the the model, perform the data forecast and estimate the wasted resources. In Chapter 5, we study a bigger dataset which contains several guaranteed bearers that carry video conversational calls. We apply our approach on this dataset in order to validate the feasibility. In the time-series modeling, several experiments were conducted to find the model that better fits the data. Data forecasting was performed and our approach was able to provide and estimate the resources gain that can be used by other services.

In Chapter 6, a real-life dataset was analyzed and studied. The dataset consists of several guaranteed bearers that carry different kinds of guaranteed services which would reflect a real-life scenario. Our approach was applied to find the time-series model. The data forecast was performed to provide the unused resources. Data simulation experiments were executed to show the benefit that our approach would provide. Finally, in Chapter 7, we summarize and conclude the thesis and we present some future directions.

CHAPTER 2

LITERATURE REVIEW AND BACKGROUND

2.1 Introduction

In this Chapter, we present some background about LTE/EPC network architecture, LTE bearer resource reservation, QoS mechanisms and some challenges of the LTE/EPC network. We also provide detailed description about the related work which has been done in the QoS and Resource Allocation in LTE/EPC network. In addition, we provide a review about the related research work which has been done in the time-series modeling and forecast fields.

2.2 Background

2.2.1 LTE Network Architecture

Overview

The telecommunication evolution LTE (Long Term Evolution) technology came about to provide higher data rate and lower delay with improved coverage and spectrum efficiency. The LTE system provides capable signaling as well as optimized radio transmission and radio access network.

Through comparing the telecom wireless technologies such as 2G, 3G and WiMAX with the LTE radio access system, the LTE system provides evolution to the telecommunication as it offers higher data rate, bigger capacity, lower delay and more improvement on coverage and spectrum efficiency (Shin *et al.* (2008b) and Verizon (2010)). The radio access network part of the LTE system has only one node which is the Evolved Node B (eNodeB) while the 3G technology in UMTS has two nodes, NodeB and Radio Network Controller (RNC). 2G technology also has two nodes, Radio Base Station (RBS) and Base Station Controller (BSC). This architectural change helps to offer less transmission delay in the radio access network and moves more intelligence towards the core network. Also, 3GPP has proposed key system features such as the default PDN session for the User Equipment (UE), and also proposed the new architecture of the policy control and charging to have more control on the mobile broadband services.

LTE and Evolved Packet Core Network Overview

The radio access network in the LTE system is composed of only one kind of node, Evolved NodeB (eNodeB). Access stratum (AS) protocols such as Medium Access Control (MAC), Radio Link Control (RLC) and Radio Resource Control (RRC) are located in the eNodeB. LTE provides higher capacity by using the Orthogonal Frequency Division Multiplexing (OFDM) as the radio access technology. LTE uses OFDM for the data carried in the downlink direction (from the radio base station to the UE) (Ericsson (2013) and Ericsson (2008)).

The evolved packet system (EPS) defines a single core network, all IP based, for multiple heterogeneous accesses that provide triple play services in the next generation networks. In the 2G and 3G technologies, there were two separate core networks: the circuit-switched core network which delivers the telephony services over circuit switching methodology, and packet-switched core network which delivers the wireless mobile broadband data services over the packet switching methodology using all-IP network. In the LTE technology, the evolved packet core network (EPC) is considered as a single mobile core network to run all services required by the wireless user equipments (3GPP TS 36.300 (2015)).

Figure 2.1 adapted from (3GPP TS 36.300 (2015)) demonstrates the LTE/EPC architecture and highlights how the User Equipment (UE) accesses the internet and other networks through running the data traffic via the eNodeB, SGW & PGW systems. The Evolved Packet Core network (EPC) consists of the following elements:

- Mobility Management Entity (MME);
- Serving Gateway (SGW);

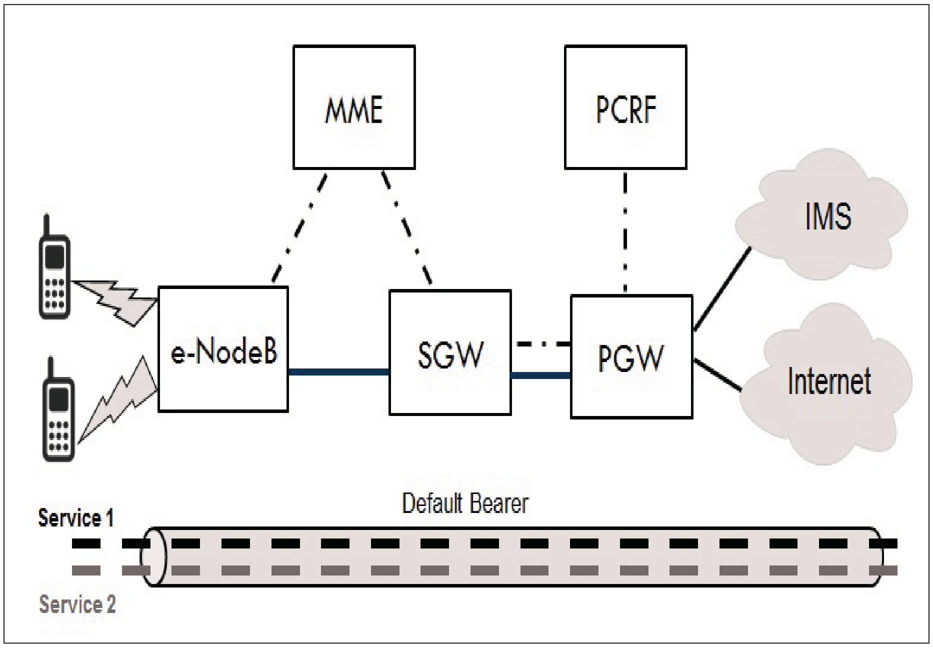

Figure 2.1 LTE Network Design
Adapted from 3GPP 36.300 (2015, p. 23)

- Packet Data Network Gateway (PGW);
- Policy and Charging Rule Function (PCRF).

The MME is responsible for the control plane function. MME controls the mobility management with the eNodeB in the LTE radio access network. MME also controls the session management with the SGW and PGW to activate and maintain the Packet Data Network sessions (PDN). The user plane is handled by the SGW node. It transports the user plane to/from the Evolved NodeB (eNodeB) in the LTE radio access network. PDN-GW or PGW provides the gateway functionality to connect the telecom world to the data communication world, e.g. IMS or packet data network (PDN). The PGW gateway is the key node for maintaining and controlling the PDN session management with the help of the SGW towards the MME; the PGW gateway allocates the IP address to the UE to be able to access the IMS, internet and corporate services. The PGW gateway has some capabilities and intelligence to perform packet inspection and service classification on the UE user plane going through the gateway. The traffic classification capabilities allow the PGW to perform credit control charging functionalities toward the Online Charging System (OCS) and to perform the charging differently based on running services. PGW also plays another rule as Policy Control Enforcement Function (PCEF), towards the Policy and Charging Rule Function (PCRF), to perform policy and QoS control enforcement on the running user plane. The Policy and Charging Rule Function (PCRF) controls QoS policy and charging for users and services and communicates that towards the PCEF (PGW) to enforce and apply the rules (3GPP TS 23.882 (2008)).

Policy and Charging Control (PCC)

Policy and charging control (PCC) provides some capabilities for service-aware QoS, policy and charging control. PCC is used in the evolved packet core network (EPC), which is defined as part of the 3GPP Release 8 specifications, and has evolved significantly to support policy and charging control for multiple-access technologies, roaming and mobility (3GPP TS 29.212 (2014) and 3GPP TS 29.213 (2014)). PCC was designed to be independent from the radio access technology. For this reason PCC can be easily adapted to be used in EPS system. In PCC 3GPP Release 8, more features and capabilities have been added; all of these new capabilities make PCC more suitable to meet EPS system requirements. These requirements include: support for mobile IP-based protocols in the EPS, roaming and mobility between heterogeneous radio access networks (Balbas *et al.* (2009)).

2.2.2 EPS Bearer and QoS Concepts

The Evolved Packet System Bearer

An EPS bearer uniquely identifies packet flows that run between the UE and the gateway. These flows receive a common QoS treatment and policy enforcement. A packet flow is defined and filtered by five-tuple parameters: the source and destination IP address, source and destination port number and protocol ID.

All packet flows which construct the bearer, mapped to the same service, receive the same packet-forwarding treatment (e.g., policy control and authorization, scheduling policy, queue

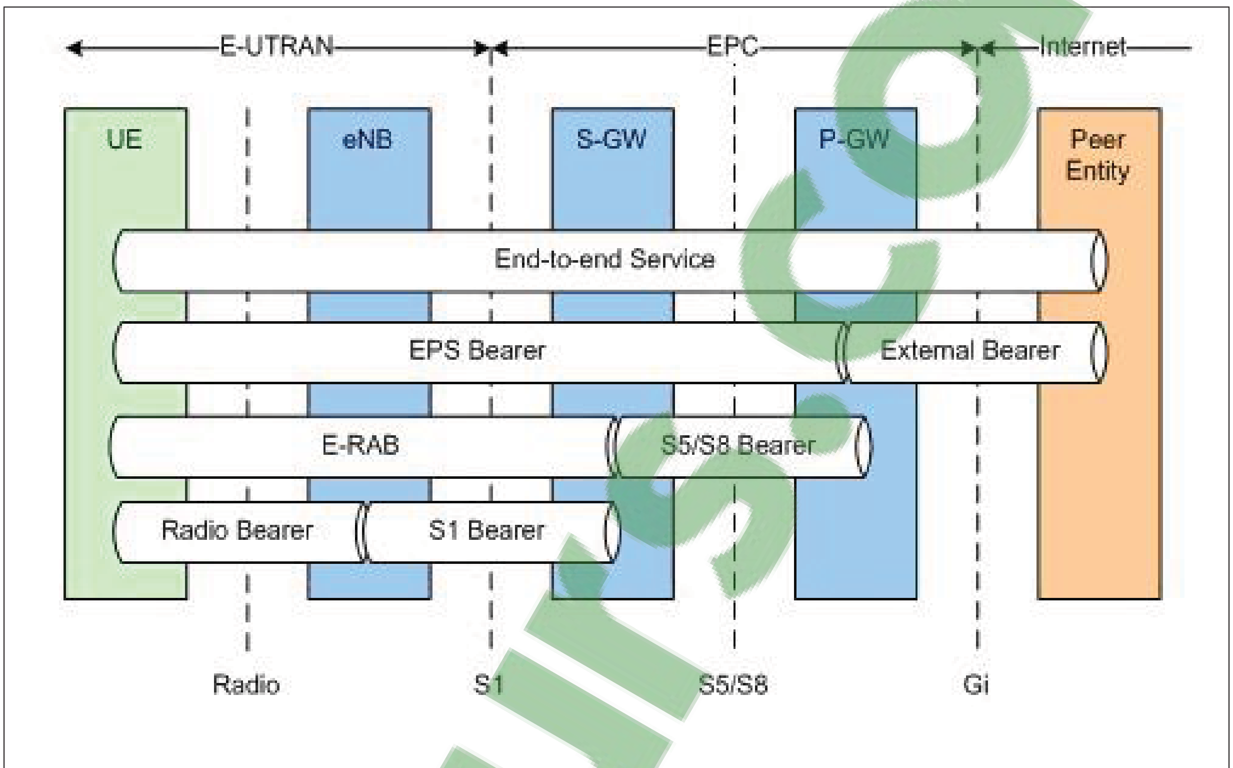

Figure 2.2 EPS Bearer Architecture
Taken from 3GPP 36.300 (2015, p. 122)

management policy, traffic-shaping policy, link-layer configuration, etc.). The bearer enables traffic separation and provides different treatment for the traffic flows, belonging to the corresponding bearer, based on the QoS requirements required (Ekstrom (2009)).

Figure 2.2, taken from 3GPP TS 36.300 (2015), shows the EPS bearer architecture and identifies the EPS bearer which runs between the UE and the PGW gateway, the EPS bearer consist of the radio bearer (between the UE and the eNB), S1 bearer (between the eNB and the SGW) and the S5/S8 bearer (between the SGW and the PGW).

EPS QoS Concepts & Parameters

In EPS system, a UE can have a default bearer which is used for basic connectivity. The default bearer is set up when the UE attaches to the network where one default bearer exists for each UE IP address, and it is kept for that UE as long as it is connected to the network.

Because the default bearer can remain established for long periods, the 3GPP specifications mandate that the default bearer is a non-guaranteed-bit-rate bearer (non-GBR). A UE also can have dedicated bearers which can be used for different services that require different QoS requirements. The dedicated bearer can be either a non-guaranteed or a guaranteed bearer (3GPP TS 36.300 (2015)).

A GBR bearer typically is established "on demand", because it reserves bandwidth resources by having special QoS attributes. On the other hand, a non-GBR bearer can remain established for long periods of time because it does not reserve any bandwidth resources. The QoS concept in the evolved packet system is based on classes/services.

According to the the 3GPP standards, each dedicated bearer will have QoS class identifier (QCI) which indicates the QoS information that the bearer would require while running the service/call through the UE in the LTE/EPC network. The QCI is given by the network based on the class/type of the bearer. The QCI specifies QoS requirements that the bearer's payload (user-plane) will receive.

The allocation and retention priority (ARP) specifies the control-plane treatment that the bearers receive. Recently, a new optional Information Element (IE) called evolved ARP is added to the GRPS Tunneling Protocol (GTP) messages that are received and sent over the Gn interface in applicable procedures between the Transport Node (SGSN) and the core network gateway (GGSN) (3GPP TS 23.401 (2015)). Previously, the ARP values were 1 to 3, where the evolved ARP values are 1 to 15 which gives more possibilities to treat the traffic differently. This functionality ensures that high-priority subscribers get prioritized access when congestion occurs and improves the service level granularity.

Resource Reservation for Guaranteed Dedicated Bearer

The guaranteed dedicated bearer is established based on prior signaling. When the user tries to access/start a specific service (e.g., a SIP data call), the application server (e.g., P-CSCF) detects the ongoing signaling, needed to establish data call, and instructs the PCRF to establish an EPS dedicated bearer for that data call. The PCRF server initiates that by sending a special diameter message to the PCEF server (PGW) to request the PCEF (PGW) to establish the EPS dedicated bearer. The PCEF (PGW) establishes the dedicated bearer with guaranteed resources. The diameter message which triggers the dedicated bearer establishment contains the IMSI, TFT and the QCI information elements. The PCEF (PGW) uses the diameter message elements to establish the dedicated bearer towards the transport and radio networks.

Table 2.1 QCI Information
Adapted from 3GPP 23.203 (2015, p. 46)

QCI	**Resource Type**	**Resource Reservation**	**Priority**	**Services**
1	GBR	GBR/MBR	2	Conversational voice
2	GBR	GBR/MBR	4	Conversational video (video calling & Live Streaming)
3	GBR	GBR/MBR	3	Real time gaming
4	GBR	GBR/MBR	5	Non-conversational video (buffered streaming)
5	non-GBR	MBR	1	IMS Signaling
6	non-GBR	MBR	6	Video (Buffered Streaming)
7	non-GBR	MBR	7	Video(Live Streaming), Interactive Gaming
8	non-GBR	MBR	8	Vidoe (Buffered Streaming)
9	non-GBR	MBR	9	Vidoe (Buffered Streaming)

According to the 3GPP standards (3GPP TS 23.203 (2015)), some services require dedicated bearer with guaranteed resources to ensure a better quality of service. Table 2.1 shows some examples of the services that require dedicated and guaranteed bearers. These services are conversational voice, conversational video (Live Streaming), real-time gaming and non-conversational video (buffered streaming). The handling of EPC bearers resource reservation in LTE/EPC network is illustrated by Figure 2.3 (3GPP TS 23.203 (2015)) where the PDN connection and its bearers connect the UE applications to the services via the radio and core networks.

QoS Mechanisms

The QoS control in EPS can be controlled through different mechanisms; these mechanisms are divided into control-plane signaling and user-plane functionalities. The Policy Control Rule Function system (PCRF) in the network determines how each packet flow for each subscriber must be handled in terms of the QoS parameters. These QoS parameters will be associated with corresponding packet flows which run through the bearer. The policy control engine (PCRF) communicates the policy and charging control (PCC) rules with the EPC gateway (PGW) which is responsible to establish a new bearer or modify an existing bearer with respect to the authorization rules and QoS parameters received from the PCRF. This is considered as applying the QoS mechanisms at the user-plane level (3GPP TS 23.401 (2015)).

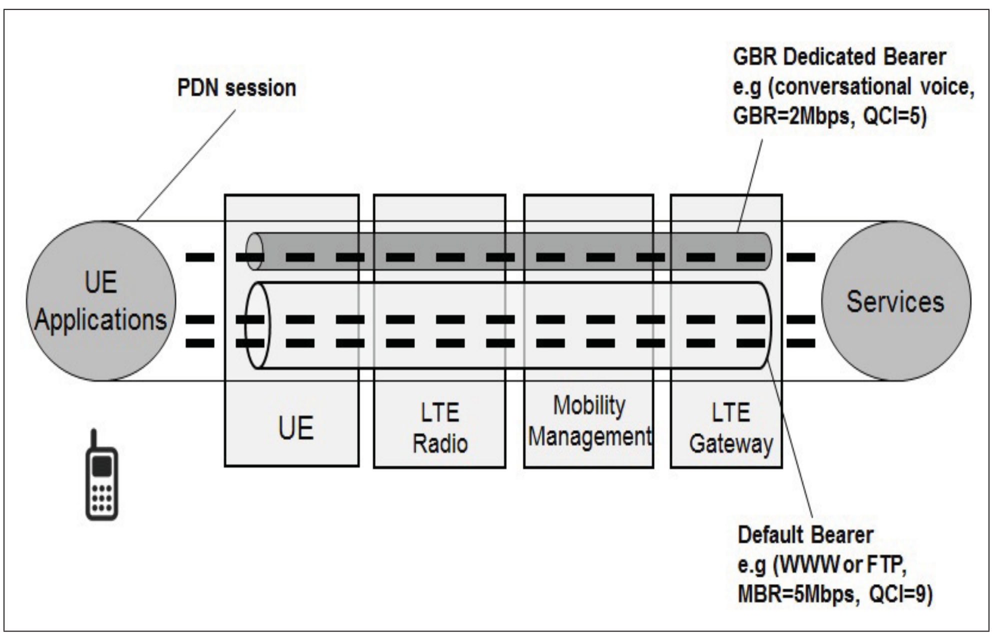

Figure 2.3 EPC Bearers Resource Reservation

Using the deep packet inspection techniques at the EPC gateway helps to identify the packet flows for the bearers. Specifying these bearers helps the EPC gateway to apply the policy control rules and the QoS mechanisms on the corresponding data bearers. Using bandwidth policing, EPC gateway can identify certain packet flows and throttle the bit-rate experienced by that particular packet flow without modifying the bearer-level QoS parameters. This is called "traffic policing" in 3GPP standards.

For bearer-level functions on non-GBR bearers, the PGW performs bandwidth policing based on the MBR value(s) for both uplink and downlink traffic; whereas the LTE RAN performs bandwidth policing based on the terminal related MBR value for both uplink and downlink traffic in the air interface (Ekstrom (2009) and 3GPP TS 23.203 (2015)). Table 2.1 shows the mapping of the standardized QCI values to the standardized guaranteed and non-guaranteed services with their priorities . To allow traffic separation in the transport network, the EPC gateway and the LTE RAN implement a QCI to DSCP mapping function. The purpose of this function is to make a translation from bearer-level QoS (QCI) to transport level QoS (DSCP).

2.2.3 Complexity of the EPC Network

The LTE architecture simplifies the radio access side by considering only one node (eNodeB) and pushes more complex functionalities and design towards the core network (EPC). The EPC network is supposed to provide more responsibilities in LTE compared to the 3G & 2G technologies. With the new architectural design in LTE, EPC network consist of four nodes which are: MME, SGW, PGW & PCRF. This will add more complexity to the coordination between these nodes and make the EPC network responsible for more work to accomplish. Apparently, there are several important factors to be aware of, these factors are: (LTE radio capacity increase, 50B devices expected by 2020, more advanced services introduced and new responsibilities for the EPC network) (Ericsson (2008)). With all of those factors, the EPC network has more challenges to be resolved and more intelligence should be provided to meet the expectations of the evolved technology (LTE) and the Next Generations Network services. Some of these challenges can be summarized as follows:

- Coordination between the radio access (eNodeB) and the EPC network nodes for centralized and distributed network architecture;
- Mobile core network capacity and resource optimization;

- End-to-end QoS control and coordination with the underlying transport;
- Traffic classification and deep packet inspection;
- Policy and Charging Control coordination;
- Scalability and network optimization.

The PDN gateway in EPC (PGW) plays an important role in handling these challenges. PGW is in charge of activating the PDN session and run the IP traffic of the UE's. PGW is responsible of the control and user planes; it is also responsible of performing the deep packet inspection through analyzing the running user plane.

In addition, PGW plays an important role of the end-to-end QoS control as it enforces the QoS for the downlink traffic in the EPC network. Furthermore, PGW interacts with the PCRF server to get the policy control rules and apply them on the running user-plane based on packet inspection results; PGW also communicates with the OCS for the online charging support. It is obvious that PGW carries out several functionalities and responsible of handling many challenges in the EPC network. With all of those mentioned, PGW would have more complex operations, scalability problems and be more sensitive from the changes in the carried traffic profile.

One EPC network could cover a metropolitan area which could have several radio access networks that are all connected to the same EPC network. This centralized design would indicate that any weakness or deficiency in the EPC network would consequently degrade the throughput for all the radio access networks and mobile users connected in that area.

2.3 Literature Review

Many research work activities have been done in the areas of the 4G/LTE technology and the Evolved Packet System (EPS) especially about the topics of Bearer QoS control, policy and charging control, EPS bearer design, EPS signaling in LTE network. Also, many researchers have paid attention to the time-series methods and how can be used for forecast and prediction. In this section we will review/discuss some techniques and research work done for those topics.

2.3.1 QoS and Resource Allocation in EPS system

In (3GPP TS 23.401 (2015), Ekstrom (2009), 3GPP TS 23.882 (2008), 3GPP TS 29.213 (2014) and 3GPP TS 36.300 (2015)) researchers describe the QoS concept of the EPC network and explain the "bearer" terminology which was introduced in LTE technology. A UE also can have dedicated bearers which can be used for different services that require different QoS requirements. The dedicated bearer can either be guaranteed or non-guaranteed bearer. Through checking the technical or research work on LTE technology, there was nothing proposed or discussed about having semi or adaptive guaranteed resource reservation.

Many research activities in (Molazem Tabrizi *et al.* (2011), Youjun *et al.* (2006), Wang *et al.* (2006) and Chandra and Helenprabha (2014)) have been done for bandwidth allocation optimization for the LTE/LTE-A radio channel but nothing has been done for optimizing bandwidth allocation in the EPC core network. According to (Mitra and Agrawal (2016) and Rodriguez (2015)), it is indicated that 5G technology will continue focusing on intelligent resource allocation scheme for cognitive radio links which somehow would positively affect the core network but no focus is proposed on optimizing the resource allocation setup at the core network especially that the typical dedicated guaranteed bearer resource allocation is still present.

The authors in (Molazem Tabrizi *et al.* (2011)) conducted some topics on bandwidth utilization on the radio channel between the base station and the mobile system. The algorithm transmits multiple variable-bit-rate (VBR) video streams from a base station to mobile in wireless networks. The algorithm transmits video streams in bursts to save the energy of mobile devices. The algorithm is adaptive to the changes in the bit-rates of video streams and allow the base station to transmit more video data on time to mobile receivers. This approach could only handle the buffered streaming video; conversational video and live streaming might not benefit from this algorithm. This algorithm does not have enough capability to provide bandwidth utilization for all different types of traffic running on the base station. Some protocols e.g. SPDY (Speedy) & QUIC (Quick UDP Internet Connections) have been introduced to focus on the data bandwidth and to make the internet faster through compressing the bandwidth as in SPDY (Belshe and Peon (2012)) or by estimating the bandwidth in each direction to avoid congestion as in QUIC (Roskind (2013)), both protocols are running at the client/server level.

In (Ekstrom (2009)), the author describes the QoS concept of the EPS system and explains the new terminology (bearer) introduced in LTE technology. This paper explains the difference between the guaranteed and non-guaranteed bearers by showing their characteristics. Each data session will have several pipes of data (bearers) where each one will have different treatment and QoS attributes which are defined according to the running service in the bearer. This paper gives an end to end use case where the EPS system inspects the signaling of a call and triggers the initiation of a dedicated bearer to run the requested call to ensure the quality of the running service. Author in (Balbas *et al.* (2009)) explains the architecture of the Policy and Charging Control (PCC) and its elements. The paper shows the role provided by the PCC about policy and charging control and how it contributes to the EPS system by adding valuable functionalities which add more intelligence to the EPS system.

The work in (Luo *et al.* (2010)) proposes a new method for EPS QoS control using the IPv6 Flow Label field to represent QoS information about the corresponding flow. The authors highlight a problem that, in the EPC network, each node needs to have an inspection rule to identify the bearer to apply the QoS policies corresponding of the running bearer. All of these tasks require more processing and storage resources at the EPC network nodes. To reduce the waste of resources, the authors provide a technique to pass the information between the nodes using a customized IPv6 flow label and apply the QoS attributes without inspecting the traffic based on the TFT fields. The flow is expected to be uniquely identified with its source and destination IP addresses as well as flow label.

The research work in (Shin *et al.* (2008a)) proposes a model for the LTE radio interface signaling for session and bearer control, this model utilizes the LTE features agreed in 3GPP such as: network-controlled service, QoS aggregation and default IP access service. The authors of this paper claim that the provided architectural changes help to offer less transmission delay in the radio access network. The authors in (Li and Shen (2011)) present high level system architecture for monitoring and troubleshooting harmonization of CoS/QoS based on DSCP (Diffserv Code Point) mappings that optimize end-to-end network performances over multiple LTE network elements. The research work in (Anas *et al.* (2008)) presents combined admission control and scheduling for QoS differentiation in LTE uplink direction.

In (Alcatel-Lucent (2009)), the authors explained the challenges that are facing the EPS system. Some of these challenges are the end-to-end QoS control and deep packet inspection overhead in the LTE network. Research work in (Armitage (2003) and Bell (2003)) discuss the QoS in the IP networks which can be also applied to the LTE from some aspects with respect to the end to end QoS control.

2.3.2 Time-Series and Forecast Models

The main role of conducting the time-series methods is to forecast the future and extrapolate the events in near or far future. Time-series are utilized in several fields or applications to get some future knowledge based on the current and past values of the forecasted series; e.g traffic flow, weather, economy, internet traffic usage and power load. Identifying the time-series depends on the conducted field/application, time period and characteristics of the forecasted time-series. The time-series model considered as statistical method of dynamic relation between the past/current observations and the forecasted variables (**?**).

The constant-level forecasting model includes very simple methods such as: the last-value forecasting method which uses the last observed value as the next forecast. The Averaging Forecasting Method determines the forecast based on the average of all the previously observed values, Moving-Average Forecasting Method does not include all the previous values; instead it provides the average of the most recent values observed (Rivett (1968)). Exponential smoothing depends on the last observation and the preceding forecast based on some weighted constant which is called smoothing constant. It implies a recursive relation between all the observations with decreasing weights for the earliest observations. There are also other alternative models for the exponential smoothing technique which are slightly different from the general one with more attention to the most recent forecasting error (Rivett (1968) and Hyndman *et al.* (2008)).

The author in (He *et al.* (2001)) uses the Exponential Smoothing model with some attention to the periodicity of the patterns. The proposed model was designed and implemented to perform network traffic forecast. This paper indicates that the network traffic is characterized with strong periodicity through applying some analysis of the network transmission rate. The proposed forecasting model dynamically selects the smoothing parameters to get near optimal results for the forecast. The proposed model demonstrates the effectiveness of the results through applying some error calculation and evaluation. The Stochastic time-series is considered to be one of the most popular approaches that has been applied for forecasting in several applications. Based on the research work conducted in (Basu *et al.* (1996), You and Chandra (1999), Yin and Lin (2005) and Svoboda *et al.* (2008)), the following mentioned time-series models are the most popular and mainly considered for traffic forecast in most of the wireless networks, telecommunications and internet applications. Some of these models are: the (i) Autoregressive (AR) model, (ii) the Moving-Average (MA) model, (iii) the Autoregressive Moving-Average (ARMA) model and (iv) the Autoregressive Integrated Moving-Avenge (ARIMA) model. The Seasonality is a time-series component that indicates periodic behavior in the time-series models (Moghram and Rahman (1989)).

In the Autoregressive model, the current value of the time-series is represented in terms of the time-series previous values and some random noise. The Moving-Average model indicates that the current value of the time-series is represented using the current and previous values of a white noise series. This noise series is constructed from the forecast errors when the observations become available. The Autoregressive Moving-Average (ARMA) model considers both previous model AR and MA. The current value of the time-series is represented Based on the time-series previous values and the white noise current and previous values. The time series explained previously; AR, MA and ARMA models are utilized for the stationary model. It indicates that the mean and variance for those time-series do not change with time. If the data in the series is not stationary, an initial step is applied to remove the non-stationary property of the data, this is known as "integrated". The Autoregressive Integrated Moving-Average model is considered as a generalization model for the ARMA model. The generalized ARIMA(p,d,q) indicates whether the autoregressive, integrated, or moving-average are conducted in the model (Moghram and Rahman (1989) and De Gooijer and Hyndman (2006)).

Many research works (Davis *et al.* (2000), Liu and Mao (2005) and Dai and Li (2009)) have utilized the time-series to perform forecast on the variable video traffic flow. The research work in (Davis *et al.* (2000)) analyzes the VBR video traffic into the queuing system to represent the time series to forecast the buffer size of the VBR video in away to better allocate the resources and to ensure the QoS requirements. This model uses the Markov chain to model the system in correspondence with video frames sent during the session, the work involves nonlinear autoregressive time series models. The author in (Dai and Li (2009)) uses the ARMA model to provide forecast for the dynamic variable bit-rate MPEG video traffic. This model concentrates on short-term period prediction.

The work in (Kalle *et al.* (2012)) applies the time-series and specifically the ARIMA model to forecast the usage of real-time applications for video streaming over the RTP protocol. The forecast results are utilized to better control and reduce the power consumption in the battery for the mobile station browsing the video media. The forecast model is proposed to be run in the eNodeB system which represents the radio station in the LTE technology. It is not an easy consideration to have dynamic power control over the air interface especially that other application could be running on the mobile and using same channel. Other consideration should be applied for any other traffic.

The Autoregressive Moving Average (ARMA) and the Autoregressive (AR) models are proposed by the authors in (Nomura *et al.* (1989) and Xu and Qureshi (1999)) to determine the statistical characteristics of the video traffic discussed. These methods are utilized for offline video applications without conducting any application for online video traffic.

The research work in (Basu *et al.* (1996), Sivakumar *et al.* (2011), You and Chandra (1999) and Yin and Lin (2005)) have applied the time-series models to provide forecast on the traffic usage of the IP networks. The author in (Basu *et al.* (1996)) investigates parametric time-series models that can be used to forecast aggregated data traffic in the internet. The author investigated some datasets collected from different universities networks to provide traffic forecast. The research work considers the traffic as stationary process. Considering this and based on non-random periodicity shown by the author. The Auto-Regressive Moving-Average (ARMA) has been used as a time-series for many datasets. The author indicates that in some datasets where more TCP applications are present (FTP-control, WWW, TELNET, SMPT). The ARIMA process should be used considering the stationarity property. The research work indicates that the forecasting model has potential application in dynamic resource allocation.

The research work in (Sivakumar *et al.* (2011)) conducts the Hidden Markov model and the Neural Network model to predict and forecast number of wireless devices connected to the Access Point in the wireless network that varies with time. The Author highlights that the forecasted number of devices could give some indications about the ongoing traffic which would help in early allocation of the network resources; that would help to avoid congestion. Based on some simulation results, the author concluded that the Neural Network will give better forecast for the varying number of wireless devices. It is important to mention that wireless devices can differ from the capability perspectives; also user applications can be different with respect to traffic volume required. Based on that, number of devices will not always give correct indication about the ongoing traffic in the wireless network.

The work in (Akinaga *et al.* (2005)) proposes a method of forecasting the radio traffic to give some prediction about the overall usage of the radio channels in the mobile communication system. The forecast results would help the radio access systems to have better admission control to avoid any congestion in the resources. The author discusses the mobile traffic characteristics based on individual behavior, regular pattern, or temporary incidents. This research work proposes a new forecasting method based on the discussed mobile traffic characteristics.

The research work in (Svoboda *et al.* (2008)) proposes a method of forecasting traffic load in the 3G Packet Core Network (PCN). The authors considered some simple and sophisticated time-series modeling approaches; e.g. exponential regression and the ARMA approaches. The research work provides long-term forecast of the mobile traffic running in the PCN network to have better planning of the packet core network resources to handle any forecasted boast in the mobile traffic during the year. The dataset of the forecast approach was a mobile traffic trace collected from real 3G mobile network. The authors studied the trace to have more understanding of the dataset characteristics. To represent the forecasting error, two representations were used: the (i) Mean Absolute Error (MAE) which is the mean of the absolute differences between the real observations and the forecasted data, and the (ii) Mean Squared Error (MSE) which is the mean of the absolute squared differences between the real observations and the forecasted data. The MSE error representation gives more weight to larger differences than smaller ones. The authors obtained that the simple and sophisticated approaches give similar results of forecasts for less than 100 days. For longer forecasts, the ARMA approach will deliver better performance results.

To summarize, through checking technical and research work done on LTE technology, there was nothing proposed or discussed about having semi or adaptive guaranteed resource reservation. Furthermore, research activities lack of providing adaptive resource reservation for LTE mobile services at the network or system level. It is important to mention also that lots of research work have been conducted for bandwidth allocation optimization for the LTE/LTE-A radio channel but very minimal work has been done to optimize bandwidth allocation in the EPC core network.

CHAPTER 3

PROPOSED APPROACH: ENHANCED RESOURCE UTILIZATION FOR LTE MOBILE SERVICES

3.1 Introduction

The ultimate goal of our research is to advance the state of the art in the LTE and mobile broadband technologies. Particularly, our research will serve to provide more capable and adaptive resource reservation techniques for the LTE mobile services to reduce wasted resources and increase the overall throughput of the EPS system. The approach helps to improve the resource reservation for LTE mobile services through forecasting the actual bandwidth consumption of the guaranteed service using time-series modeling, identifying the wasted/unused guaranteed bandwidth and helping to utilize the unused bandwidth by other non-guaranteed bearers. The approach would help to increase the overall throughput of the LTE/EPC network. In this chapter, we provide a full description of our proposed approach, explain the mathematical modeling of our approach and provide in detail the methodologies and algorithms that are used to achieve our objectives.

3.2 Approach Overview

By conducting a literature review, we noticed that EPS system design lacks optimization techniques for the resources reservation of the EPS bearers. By conducting the 4G technology design details of the bandwidth reservation for the guaranteed and non-guaranteed bearers, we concluded that by utilizing the unused bandwidth of some guaranteed bearers, that hold reserved bandwidth through the guaranteed QoS parameter, we can perform some optimization and utilize the unused bandwidth for other bearers (i.e., these bearers are willing to consume more resources). This shall increase the capability of the telecom network and improve the service availability.

In this research, our approach proposes a new concept of having *adaptiveness* in the guaranteed bearer implementation which will provide an intermediate class between the strict (guaranteed) and the very open (non-guaranteed) resource allocation types. This concept of adaptiveness would provide flexible implementation which allows us to apply some forecast methods to estimate the wasted bandwidth in the guaranteed resources and utilize them accordingly to reduce the waste and increase the system throughput. Furthermore, the approach indicates that several mobile sessions of the *adaptive* guaranteed services can be conducted together to forecast the overall usage of all those sessions, estimate a pool of the unused resources and utilize some of those resources for other non-guaranteed running services. Also, this unused resources could be utilized by the system to establish new non-guaranteed bearers. Whenever the *adaptive* guaranteed service(s) request the contributed resources, they will get them from a standby resources pool which will be designated for this purpose.

The approach indicates that the guaranteed bearers can be further classified, based on the criteria of adaptiveness, into two types: (i) *adaptive-guaranteed bearers* that allow adaptive resource allocation and contribute the unused guaranteed resources; all contributed resources will be added up in one pool; these bearers can be considered as *contributing-bearers*. (ii) *pure-guaranteed bearers* that cannot accept any adaptive resource allocation even if they are not fully used. This represents the current behavior of the guaranteed bearer as defined by the 3GPP standards. This could be considered for public safety, emergency and alarm/alert services that should maintain pure guaranteed resources. It also can be considered for any other services defined by the operator to be pure guaranteed.

In the LTE/EPC networks, the non-guaranteed bearers are given Maximum bit-rate (MBR) which is provided based on availability. Hence the non-guaranteed bearers may not reach the MBR bit-rate. That being said, our approach considers the non-guaranteed bearers to be *acquiring-bearers* based on their willingness to utilize more resources to reach the MBR limit. The *acquiring-bearers* will consume some resources of the unused guaranteed resources pool contributed by the *contributing-bearers*. It is important to mention that non-guaranteed bearer will never be contributing as they do not have any guaranteed resources. Table 3.1 highlights the new terminology proposed by our approach.

Table 3.1 New Concept of the LTE Guaranteed Bearers

Bearer Type	**QCI**	**Resource Reservation**	**Guaranty Level**	**Contributing or Acquiring**
GBR1	1	Guaranteed GBR/MBR	Adaptive GBR	Might be Contributing
GBR2	2	Guaranteed GBR/MBR		
GBR3	3	Guaranteed GBR/MBR	OR	
GBR4	4	Guaranteed GBR/MBR	Pure GBR	no Contribution
non-GBR5	5	non-Guaranteed MBR	NA	Might be Acquiring
non-GBR6	6	non-Guaranteed MBR	NA	Might be Acquiring
non-GBR7	7	non-Guaranteed MBR	NA	Might be Acquiring
non-GBR8	8	non-Guaranteed MBR	NA	Might be Acquiring
default-BR	9	non-Guaranteed MBR	NA	Might be Acquiring

Furthermore, our approach indicates that the contributed resources can be utilized by the *acquiring-bearers*; but as soon as the *contributing-bearers* request the bandwidth back, the resources should be available to ensure the quality of the guaranteed services as expected. To safely perform the resources transition for the guaranteed resources once it is requested back, we introduce the concept of "Safety Model" which depends on the forecasting error to avoid any disturbance for the guaranteed services. The Safety Model indicates to have several categories or pools of the unused bandwidth: (i) one resources pool can be considered as guaranteed standby resources to be used any time by the GBR bearers once the resources are requested back immediately; the resources reserved in this standby pool can be considered as the *Safety Threshold*. This threshold shall be defined and calculated based on the forecasting error resulted from the forecast model. (ii) The other resources pool or category will be utilized by the *acquiring-bearers* to increase the EPC gateway system capacity. The *Safety Threshold* shall adapt dynamically based on the forecast process results. The forecasting error and the demand of requesting the resources back by the *contributing-bearers* (Albasheir and Kadoch (2015)).

3.3 Solution Modeling

The bandwidth usages of the *adaptive* GBR bearers will be all summed together in one pool (*GBR_used*); the *GBR_used* represents the summation of all used guaranteed resources for adaptive GBR1, GBR2, GBR3, GBR4 instances according to Table 3.1. The bandwidth reserved of those guaranteed bearers will be all added together in another pool (*GBR_reserved*); the *GBR_reserved* represents the summation of all reserved guaranteed resources for adaptive GBR1, GBR2, GBR3, GBR4 instances. The data series of *GBR_used* will represent the dataset on which the forecast process will be performed. The Adaptive GBR instance will be considered as contributing. If the guaranteed reserved resources are higher that the guaranteed used resources, the *pure-guaranteed bearers* will be excluded from this calculation. The *GBR_used* & *GBR_reserved* are represented by the equations 3.1 & 3.2 which are only applicable when the *GBRij* is Adaptive, where n represents the total number of all adaptive GBR bearers and k represents the QCI number of the GBR bearers:

$$GBR\_used = \sum_{i=1}^{k}\sum_{j=1}^{n} GBRij\_usage, \text{where GBRij is Adaptive;} \tag{3.1}$$

$$GBR\_reserved = \sum_{i=1}^{k}\sum_{j=1}^{n} GBRij\_reserved, \text{where GBRij is Adaptive;} \tag{3.2}$$

The forecast process will be applied on the *GBR_used* data and the prediction outcome will be represented by *GBR_used_F* data series. Based on the forecast results, we can calculate the difference between the *GBR_used_F* and *GBR_reserved* to provide the forecasted unused

GBR bit-rate which is referred to as *GBR_unused_F*, as indicated in Equation 3.3. The *GBR_-unused_F* represents the overall forecasted unused/wasted guaranteed resources in the EPC gateway system that will be considered for further utilization.

$$GBR\_unused\_F = GBR\_reserved - GBR\_used\_F \tag{3.3}$$

As part of the Safety Model, we include the *Safety Threshold* which is a reserved standby resources pool of the contributed resources to be used by the *contributing-bearers* in case they request back the guaranteed resources. The *Safety Threshold* can be determined based on the Root Mean Square Error (RMSE) (Armstrong and Collopy (1992)), the *RMSE* value only gives an overall forecast error of the whole forecast experiment but it does not reflect the forecast error behavior with time. The *Safety Threshold* should be driven based on *RMSE* error behavior which can be a series of *RMSE* error observations with time. The forecast *RMSE* error behavior can be formulated by calculating a series of *RMSE* errors where the *mean*, used to calculate *RMSE*, is the value h which is the corresponding forecasting steps.

$$RMSE\_Behavior_{t=n} = \sqrt{1/h \sum_{i=n-h+1}^{n} (\hat{X}_i - X_i)^2} \tag{3.4}$$

Equation 3.4 shows the *RMSE* error behavior calculation at $t = n$ and the forecasting step value of h. Since the *Safety Threshold* will always rely on the forecast error at specific moment, we consider the RMSE error behavior at moment (t = n) to be used for *Safety Threshold* calculation at moment (t = n + 1) as stated in Equation 3.5. The Equation indicates that *Safety Threshold* will maintain a minimum value of "RMSE" in case the **"RMSE Behavior"** at (t=n) is lower than the "RMSE" value.

$$safety\_threshold_{t=n+1} = \begin{cases} RMSE\_Behavior_{t=n}, & RMSE\_Behavior_{t=n} \geq RMSE, \\ RMSE, & RMSE\_Behavior_{t=n} < RMSE \end{cases} \tag{3.5}$$

The Equation 3.6 shows the *bitrate_tobe_utilized* which is the guaranteed unused bit-rate that can be released and utilized for other running services (*acquiring-bearers*), the *bitrate_tobe_utilized* represents the gain/benefit of our proposed approach. This can be calculated by deducting the *Safety_Threshold* limit from the forecasted unused GBR bit-rate (*GBR_unused_F*).

$$bitrate\_tobe\_utilized = GBR\_unused\_F - Safety\_Threshold \tag{3.6}$$

3.4 Solution Methodology

3.4.1 Time-Series and Stationary Data

Considering the concept of having adaptive resource reservation for the mobile LTE guaranteed resources to minimize the waste, we should have robust techniques that can forecast and predict the usage of the guaranteed resources for short-term period for the ongoing GBR bearers in the EPC gateway system. By forecasting the usage of the resources, the unused bandwidth will be measured and estimated. Time-series will be used in this research work to model the data of the bit-rate usage of the mobile LTE guaranteed service. The resulted time-series model will be utilized in the prediction function and data forecast.

Time-Series is a sequence of observations that are measured at consecutive points in time spread out at identical time intervals. Time-series analysis includes methods for analyzing series of data to get meaningful characteristics of that data. Furthermore, time-series are utilized in several fields or applications to get some future knowledge based on the current and past values of the data observation series (Karapanagiotidis (2012)). Furthermore, the time-series models are considered as statistical methods of dynamic relation between the past/current observations and the forecasted variables (Karapanagiotidis (2012)). In this research, the time-series models will be utilized to forecast and predict the usage of the guaranteed resources in the mobile services. The traffic usage or bit-rate used by EPC gateway system to run the guaranteed service represents the dataset.

Based on the characteristics of the dataset examined, several time-series models will be discussed to provide an appropriate model to represent the dataset. Based on the research work conducted in (Basu *et al.* (1996), You and Chandra (1999), Yin and Lin (2005) and Svoboda *et al.* (2008)), the following mentioned time-series models are the most popular and mainly considered for traffic forecast in most of the wireless networks, telecommunications and internet applications. For stationary dataset, the models: Autoregressive AR(p), Moving Average MA(q), Autoregressive Moving Average ARMA(p,q) can be used. For non-stationary dataset, the Autoregressive Integrated Moving Average ARIMA(p,d,q) model can be considered as it uses the "integrated" property to transform the given dataset to stationary series through applying the differencing technique. Those models are further explained below:

- Autoregressive Model (AR): The autoregressive model attempts to predict an output of a system based on the previous outputs. In this model, the current value of the time-series is represented in terms of its previous values and some random noise;
- Moving Average Model (MA): The Moving-Average model indicates that the current value of the time-series is represented using the current and previous values of a white noise series. This noise series is constructed from the forecast errors when the observations become available;
- ARMA: The Autoregressive Moving-Average (ARMA) model considers both previous model AR and MA. The current value of the time-series is represented using the time-series previous values and the current and previous values of the white noise;
- ARIMA: The Autoregressive Integrated Moving-Average model is considered as a generalization model for the ARMA model. The time series explained previously; AR, MA and ARMA models are utilized for the stationary model which indicates that the mean and variance for those time-series do not change with time. If the data in the series is not stationary, an initial step is applied to remove the non-stationary property of the data; this is known as "integrated" and this is where the ARIMA model got introduced. The generalized ARIMA(p,d,q) indicates if the autoregressive, integrated, or moving-average are conducted

in the model according to (Moghram and Rahman (1989) and De Gooijer and Hyndman (2006)).

To compare the time-series model forecast results and evaluate the performance for each model, two representations for forecasting error can be used: the (i) Mean Absolute Error (MAE) which is the mean of the absolute differences between the real observations and the forecasted data, and the (ii) Mean Squared Error (MSE) which is the mean of the absolute squared differences between the real observations and the forecasted data. This one gives more weight to larger differences than to smaller ones, (Robert H. Shumway (2010)). The resulted forecasting error from the time-series model will be used in calculating the *Safety Threshold* explained previously in Section 3.2.

To understand the data characteristics and determine the time-series model to use, data *Stationarity* is a key thing to study and investigate. Stationarity indicates that time-series statistical characteristics do not change with time; instead they depend on the difference between the dataset observations. This *difference* can be referred as *lag*. In other words, a time-series $\{X_t, \quad t = \pm 0, 1, ...\}$ is considered to be stationary if it has statistical characteristics similar to those of the *shifted* time-series $\{X_{t+h}, \quad t = \pm 0, 1, ...\}$, for each lag h. The Autocorrelation Function (ACF) and Autocovariance Function (ACVF) are important indicators to measure the stationarity status of the time-series. The mathematical representation of the Autocovariance Function of $\{X_t\}$ time-series with *mean* $\mu_x(t)$ at $lag = h$ is represented by Equation 3.7:

$$\begin{aligned} \gamma x(h) &= Cov(X_{t+h}, X_t) \\ &= E[(X_{t+h} - \mu_x(t+h))(X_t - \mu_x(t))] \end{aligned} \tag{3.7}$$

Taken from Brockwell and Davis (2006)

And, the mathematical representation for the Autocorrelation Function of $\{X_t\}$ at $lag = h$ is is represented by Equation 3.8:

$$\rho x(h) = \frac{\gamma x(h)}{\gamma x(0)} = Cor(X_{t+h}, X_t) \tag{3.8}$$

Taken from Brockwell and Davis (2006)

3.4.2 Overall Structure

In this research, our overall solution structure is summarized by Figure 3.1 which demonstrates our solution methods and the flow-chart states.

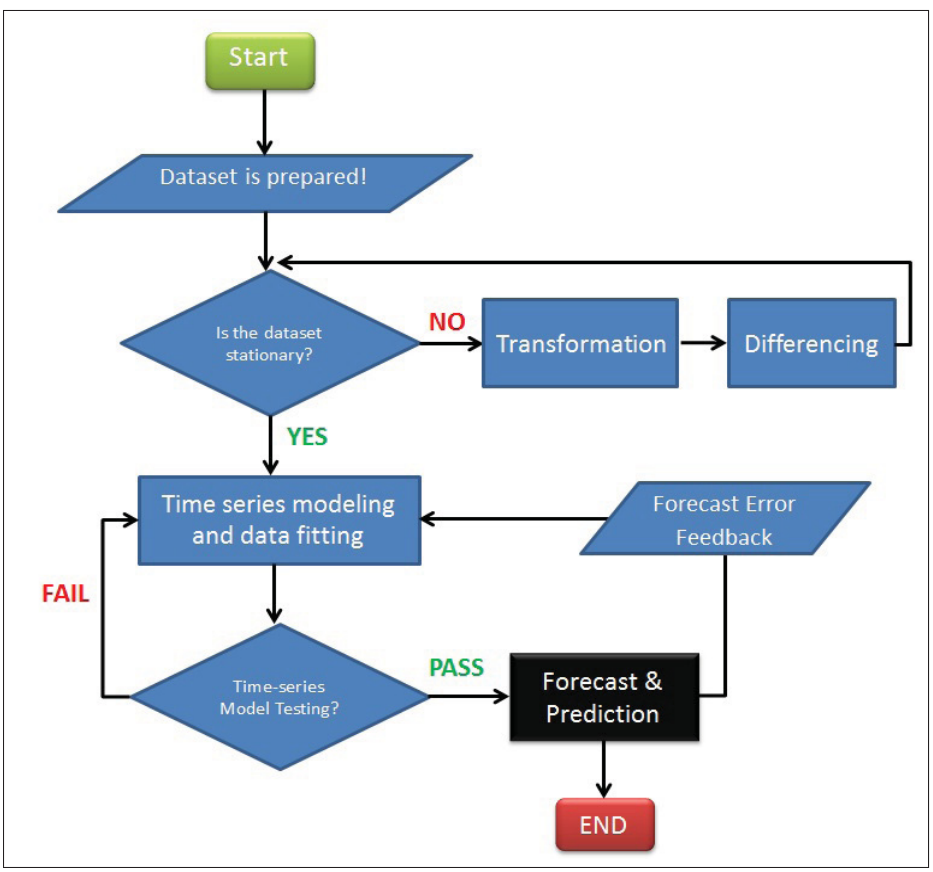

Figure 3.1 Time-Series Modeling Structure

Algorithm 3.1 Time-Series Modeling & Forecast for LTE Service Bandwidth Usage

1: Examine the stationarity status of the LTE service dataset Y_n by calculating ACF & PACF.
2: **if** the dataset Y_n is not stationary **then**
3: Use the Box-Cox Transformation to stabilize the data variability by the choosing value of λ.
4: Use the differancing method to eliminate any possible trend or seasonal affect by finding suitable value of d.
5: **else**
6: Fit the time-series X_n with AR(p), MA(q) or ARMA(p,q) model based on the ACF & PACF values and calculate the AICC values.
7: **end if**
8: **for** $i = 1$ `to` n **do**
9: `Fit` X_n `with AR(`i`) using Innovations algorithm preliminarily estimation and calculate AICC. Find AR(`p`) with minimum AICC-AR`
10: **end for**
11: **for** $i = 1$ `to` n **do**
12: `Fit` X_n `with MA(`i`) using Innovations algorithm preliminarily estimation and calculate AICC. Find MA(`q`) with minimum AICC-MA`
13: **end for**
14: **for** $i = 1$ `to` n, $j = 1$ `to` n **do**
15: `Fit` X_n `with ARMA(`i,j`) using Innovations algorithm preliminarily estimation and calculate AICC. Find ARMA(`p,q`) with minimum AICC-ARMA`
16: **end for**
17: Find the time-series model with minimum AICC values in AICC-AR, AICC-MA & AICC-ARMA.
18: Specify the p & q values from the chosen time-series model, specify the value d from step 4.
19: Based on Steps 17 & 18, initialize the Maximum Likelihood Estimation (MLE).
20: **return** the ARIMA model found from the MLE estimation.
21: Find the predictor function $(P_n X_{n+h})$ based on the definitive ARIMA model found in Step 20.
22: Run the prediction function $(P_n X_{n+h})$ on the dataset.
23: **return** the unused LTE service bandwidth forecast.

Based on the demonstration in Figure 3.1, once the LTE bit-rate dataset is prepared, some methods are required to determine if the provided data is stationary or not. This can be achieved through calculating the ACF & PACF functions. Based on the outcome, Transformation and/or

Differencing techniques could be applied on the dataset to convert it into stationary state. Once the dataset is confirmed to be in stationary state, time-series models will be used to mathematically fit and represent the dataset. The chosen time-series model of the dataset will be tested and validated to ensure the accuracy of representing the dataset. If the time-series model fails in the validation state, more adjustment would be needed on the model before proceeding to the next state. After the dataset is properly modeled and represented through time-series, the traffic forecast and prediction state would predict and provide the future value(s). The *Forecast Error* will be calculated once the future time is reached.

Algorithm 3.1 summarizes the main phases and steps that will be performed in our experiments. Firstly, our algorithm examines the dataset to check the data stationarity. The algorithm will determine if the transformation is needed to get stationary data. Once the data is confirmed to be stationary through the ACF/PACF functions, several known time-series models will be used to find a model which fits our data. The preliminarily estimation and the Maximum Likelihood Estimation (MLE) will be utilized to find our time-series model that represents the data. Finally, the predictor function will be determined based on our time-series model and the forecast process will be performed on the data. In this thesis, we will be using MATLAB and ITSM (Lee and Strazicich (2002)) tool for all the experimental work and simulations.

CHAPTER 4

EVOLVED APPROACH FOR LTE VIDEO SERVICE RESOURCE RESERVATION

4.1 Introduction

In this Chapter, we will apply our approach on a single guaranteed service that carries video conversational call where we analyze the dataset, prepare it for modeling, find the time-series model that fits the data, validate the the model, perform the data forecast and estimate the wasted resources.

4.2 Dataset Analysis

In this section, we will present our dataset, provide full analysis of the studied data, check the stationarity status of the dataset and perform the required methods to convert it to stationary if needed.

4.2.1 Dataset: LTE Guaranteed Service

As shown in Table 2.1, it is obvious that guaranteed resources are used for conversational voice & video (Live Streaming) calls, real-time gaming and non-conversational video (buffered streaming). It is obvious that LTE networks allocate guaranteed resources for several types of video traffic which is costly on the LTE systems end to end. According to the report in (Ericsson (2015)), video traffic represents the largest segment of mobile data traffic in LTE networks and it is continuously increasing. In this chapter, we concentrate on the conversational video traffic and its guaranteed resource reservation in LTE networks. So our examined dataset for this experiment will be a conversational video traffic that uses LTE guaranteed resources. Also, we focus on the used bit-rate of the running video call through the LTE network considering the resource reservation at the EPC gateway system.

The examined dataset is a conversational video service represented by a series of used bit-rates during the guaranteed bearer call. The call is captured through a simulated LTE/EPC environment called *nwEPC* (Chawre (2010)); a guaranteed bearer session was used with guaranteed/reserved bit-rate (GBR)=2.2 Mbps. We refer to the number of bit-rate observations as (n) and we also consider $\{Yn\}$ as our time-series of the video bit-rate observations. Figure 4.1 shows the dataset of LTE conversational video call with $n = 300$ seconds. This dataset will be studied and conducted through our model phases and referred to as LC-100.

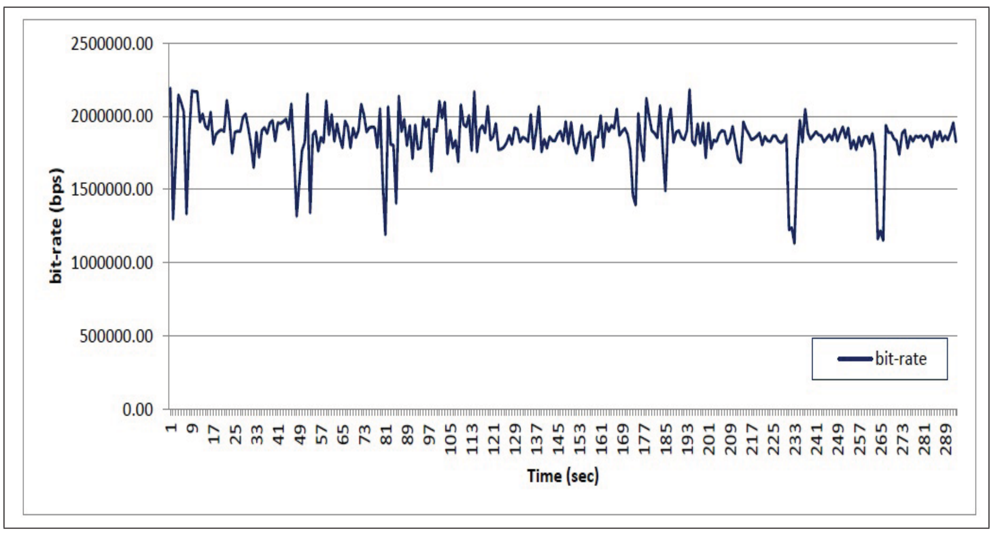

Figure 4.1 LC-100 Conversational Video Dataset

4.2.2 Data Characteristics Analysis

According to the typical representation of time-series, the LTE video bit-rates can be represented using the *Classical Decomposition Model* of typical time-series (Brockwell and Davis (2006)), this can be shown in *Equation* 4.1. Where Y_t is the observation at time t, m_t is a "trend component", s_t is a "seasonal component" and X_t is a residual or "random noise component" which is stationary with mean equal to zero.

$$Y_t = m_t + s_t + X_t, \quad t = 1, \ldots, n \tag{4.1}$$

Taken from Brockwell and Davis (2006)

Considering the *Classical Decomposition Model* of time-series shown in *Equation* 4.1, it is possible to have data trend, seasonality or any data variability with time which indicates that the examined dataset is non-stationary since the statistical characteristics depend on time. Transformation & Differencing are the procedures used to eliminate data trend and seasonality and also can be used to stabilize the data variability to generate new time-series with stationary properties; the new time-series will be the residual or the noise component $\{X_t\}$.

Once the data is transformed and the trend & seasonal affects are eliminated -in case they exist-, the ACVF & ACF functions can be calculated on resulted the noise component. As mentioned earlier, ACVF & ACF provide useful information to understand the data characteristics and the stationarity status of the time-series. It also helps to measure the dependency between the data observations at different time which plays an important role in the forecast and prediction of the future observations. Also, the Partial Autocorrelation Function (PACF) provides more understanding of the data characteristics especially that it shows the autocorrelation between the data observations after removing any linear dependency if it exists between the given observations.

Furthermore, the ACF & PACF functions can be calculated to examine the residual and discover if it represents Independent and Identically Distributed random data (IID), in case it is, no dependency will exist between the observed values and no forecast will be possible. To examine the IID status of the residual, 95% of the calculated sample ACF/PACF values should reside within the confidence bounds $\pm 1.96n^{-1/2}$. This would indicate that the data residual is IID random data. Otherwise, IID hypothesis will be rejected and the residual values will have dependency among each others.

4.2.3 Transformation

The objective of data transformation is to produce data with no apparent deviations from stationarity. If the magnitudes of seasonal and noise fluctuations increase linearly with time, then the resulted transformed data will have fluctuations of more constant magnitude. Box-Cox transformation is useful when the variability of the dataset increases or decreases with time. The variability can often be made nearly constant by choosing a suitable value of λ. Box-Cox transformation can be executed through specifying the value of the parameter λ and applying the *Equation* in 4.2. If the original observations are $y1, y2, ..., yn$, the Box-Cox transformation function $f_\lambda(y)$ converts them to $f_\lambda(y1)$, $f_\lambda(y2)$, . . . , $f_\lambda(yn)$ (Box *et al.* (2015)).

$$f_\lambda(y) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(y) & \lambda = 0. \end{cases} \tag{4.2}$$

Taken from Box *et al.* (2015)

In particular, for positive data whose standard deviation increases linearly with time, the variability can be stabilized by choosing λ closer to zero. We found that the variability of the dataset LC-100 can be made more constant at λ=0.

4.2.4 Differencing

Differencing can be used to eliminate any possible trend and seasonal effect in the data aiming to produce stationary data. Furthermore, the data ACVF & ACF functions provide useful information if the differencing is needed to be applied; e.g. having slowly decreasing ACF function would indicate that the data is not stationary and differencing shall be applied. Differencing procedure can be applied to replace the original series $\{A_t\}$ by $\{B_t\}$ for some positive integer d where $B_t = A_t - A_{t-d}$, Differencing can be applied at different lags d (Box *et al.* (2015)).

The ACF function was calculated for 40 different lags for the data before and after the differencing procedure is applied. Figure 4.2 shows the ACF function calculated for the dataset

LC-100 before differencing, it is obvious that the ACF values somehow decrease slowly which indicates that the data is not stationary and Differencing would be needed.

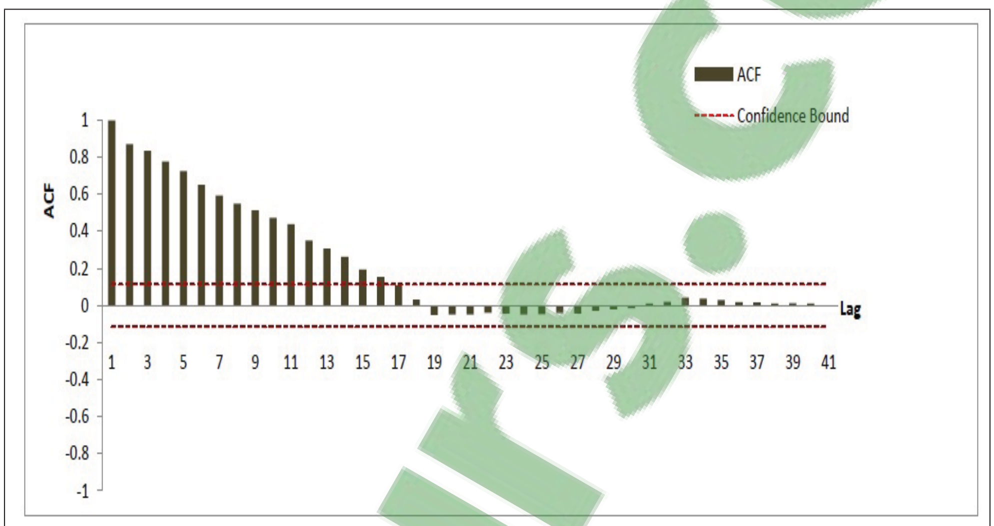

Figure 4.2 Sample ACF of the Conversational Video Dataset

The differencing procedure was applied on the transformed data (which was done at λ=0 according Box-Cox procedure in *Equation* 4.2) at different values of d. We found that differencing at d=*1* is good enough to convert the data into stationary. Figure 4.3 demonstrates the data after applying the Box-Cox transformation, the differencing procedure and *mean* being subtracted. According to *Classical Decomposition Model* highlighted previously in *Equation* 4.1, the resulted data represents the residual or the random noise component $\{X_t\}$.

Figures 4.4 & 4.5 show the ACF & PACF functions calculated for the dataset LC-100 after differencing was performed at d=*1*. According to these figures, we can see that dependency does exist between the data observations in the very early values of the *lag*.

Considering that trend and seasonal effect will be eliminated after the Differencing procedure is applied and data has no deviation from stationarity, the outcome data would represent the "residual" or the "noise component" X_t according to *Equation* 4.1. As mentioned previously, to ensure that dependency exists between the residual observations and IID hypothesis does not apply, more than 5% of the ACF/PACF calculated values should fall outside the confidence

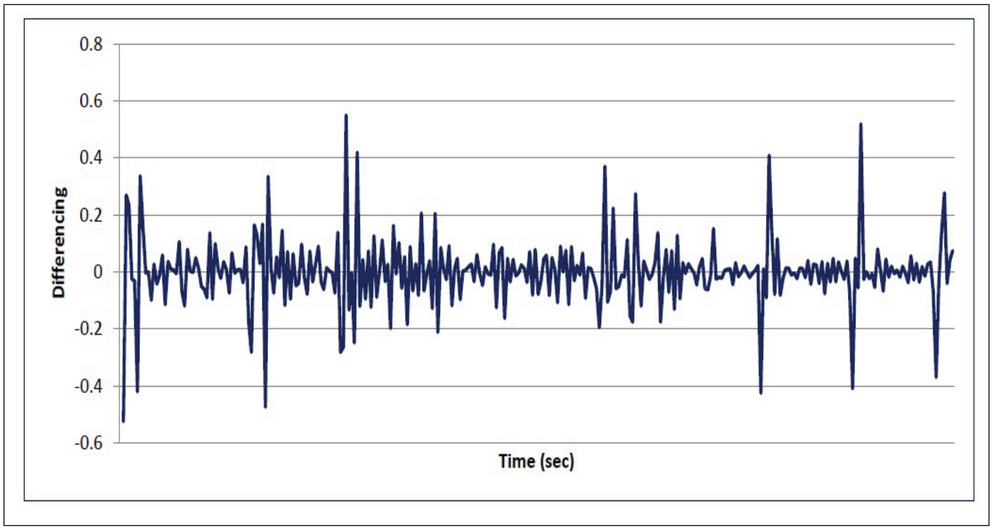

Figure 4.3 LC-100 Dataset after Transformation & Differencing

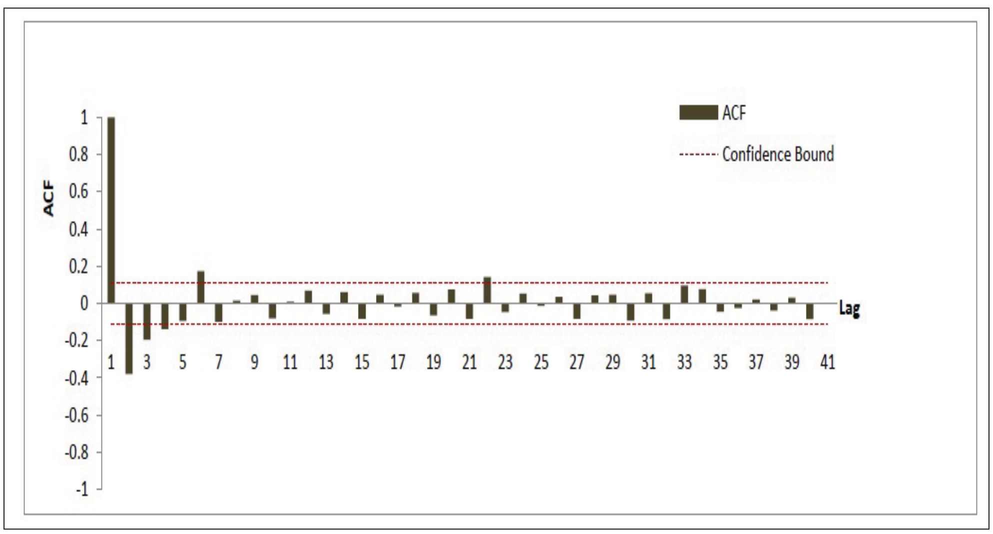

Figure 4.4 ACF Function of the Differenced Data at $d = 1$

bounds $\pm 1.96 n^{-1/2}$ which equal to ± 0.1131 for the LC-100 dataset. The confidence bound is represented by the dashed line in Figures 4.4 & 4.5. The ACF/PACF functions were calculated based on 40 different lags; based on that, there should be more than 2 values of ACF or PACF ($0.05 \times 40 = 2$) residing outside the confidence bounds to consider that the data/residual does not represent IID data. The confidence bounds is ± 0.1131 (which is calculated at $n = 300$ and represented by the dashed line). Figures 4.4 & 4.5 show that the ACF & PACF functions have

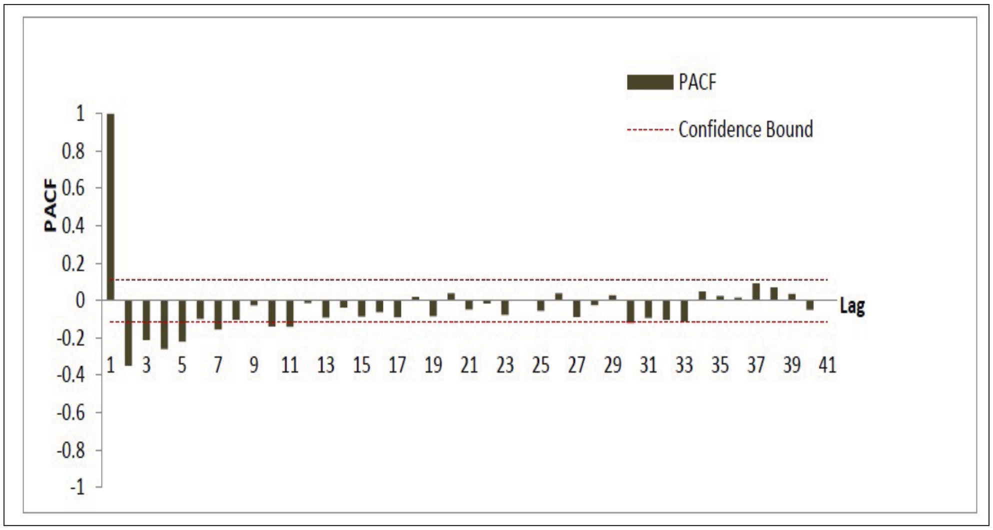

Figure 4.5 PACF Function of the Differenced Data at $d = 1$

more than 5% of the calculated values residing outside the confidence bounds, this rejects the IID hypotheses of the LC-100 dataset and demonstrates that the data is stationary. Interested reader can refer to (Albasheir and Kadoch (2014)) for more details on how to transform LTE video traffic into stationary. It is important to mention that the ARIMA time-series model includes the Differencing technique to convert the non-stationary dataset to stationary before applying the time-series modeling.

4.3 Time-Series Modeling and Data Forecast

In this Section we will present the studied time-series models, design the time-series model that fits the experimented data, validate the modeled time-series and perform the data forecast and prediction.

4.3.1 Time-Series Models

The time-series models are considered as statistical methods of dynamic relation between the past/current observations and the forecasted variables (Karapanagiotidis (2012)). Based on the characteristics of the dataset examined, several time-series models will be discussed to provide an appropriate model to represent the dataset. For stationary dataset, the following models can be used: $AR(p): X_t = Z_t + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p}$, $MA(q): X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$ and $ARMA(p,q): X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$. In all these models, Z_t is the white noise function with σ^2 variance and zero *mean*, $\quad Z_t \sim WN(0,\sigma^2)$.

For non-stationary dataset, the Autoregressive Integrated Moving Average ARIMA(*p,d,q*) model can be considered, ARIMA model uses the "integrated" property to transform the given dataset to stationary series through applying the differencing technique (Brockwell and Davis (2006)). Once the differencing is performed and the value of *d* is determined, the ARIMA time-series will follow the process in AR, MA or ARMA time-series, (for further discussions about time-series, interested readers can go to (Brockwell and Davis (2006)). The ACF & PACF are important functions that help in modeling and fitting the data into time-series. Our examined dataset (LC-100) was found to be non-stationary and differencing was used to convert it to stationary data. Based on that, the ARIMA time-series would be the one chosen to model this dataset. To test time-series models, the AICC criterion (Hurvich and Tsai (1989)) is a major technique that can be used to test the time-series model and choose the most appropriate model that can better fit the dataset in our experiments. This can be done through checking ϕ_p, θ_q, p, q in the candidate models that minimize the AICC value, the AICC is shown in *Equation* 4.3.

$$AICC = -2ln\ \ L(\phi_p, \theta_q, S(\phi_p, \theta_q)/n) + 2(p+q+1)n/(n-p-q-2) \tag{4.3}$$

Taken from Hurvich and Tsai (1989)

4.3.2 Time-Series Modeling Experiments

Finding an appropriate AR(*p*), MA(*q*) or ARMA(*p,q*) model to represent the observed stationary data includes a number of problems to be solved. These involve the choice of *p* and *q* (order selection) and the estimation of the mean, the noise component variance and the coefficients $\{\phi_i, i = 1, \ldots, p\}$, $\{\theta_i, i = 1, \ldots, q\}$. The ACF & PACF functions provide good indicators about the range where the *p* and *q* orders can be estimated. The model, used to find the final selection, depends on a variety of goodness of fit tests that mainly use minimization of the AICC statistic criteria (Hurvich and Tsai (1989)). Considering that p and q orders can be determined from an estimated range, the ϕ and θ coefficients can be found by utilizing the data observations of the stationary time-series through using the Maximum Likelihood Estimation (MLE) to estimate ϕ and θ. The Likelihood Estimation of the ARMA Process is highlighted in *Equation* 4.4 (Brockwell and Davis (2006)).

$$L(\phi,\theta,\sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n r_0 \dots r_{n-1}}} exp\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\frac{(X_j-\hat{X}_j)^2}{r_{j-1}}\} \tag{4.4}$$

Taken from Brockwell and Davis (2006)

By differentiating $\ln L(\phi,\theta,\sigma^2)$ partially with respect to σ^2 and noting that $\hat{X}_j$ and r_j are partially independent of σ^2, it was found that the maximum likelihood estimators of $\hat{\phi}$, $\hat{\theta}$ and $\hat{\sigma}^2$ satisfy the following equations (Brockwell and Davis (2006)):

$$\hat{\sigma}^2 = n^{-1}S(\hat{\phi},\hat{\theta}) \tag{4.5}$$

$$\text{where;} \quad S(\hat{\phi},\hat{\theta}) = \sum_{j=1}^{n}(X_j-\hat{X}_j)^2/r_{j-1} \tag{4.6}$$

Taken from Brockwell and Davis (2006)

and $\hat{\phi}$, $\hat{\theta}$ are the values of ϕ, θ that minimize

$$l(\phi,\theta) = \ln(n^{-1}S(\phi,\theta)) + n^{-1}\sum_{j=1}^{n}\ln r_{j-1} \tag{4.7}$$

Taken from Brockwell and Davis (2006)

The MLE algorithm is very beneficial in finding accurate estimation of the coefficients but it is complicated and time consuming in solving the system equations. To make MLE feasible, fast and realistic, we do specify initial parameter values with which MLE begins the search. This can be achieved through Preliminary Estimation. The closer the preliminary estimates are to the maximum likelihood estimates, the faster the search will generally be. In our work, we consider the Innovations algorithm (Brockwell and Davis (1988)) which provides initialization parameters for the MLE Estimation and indicates some estimation of the order selection. AICC will be used to measure which model is better fitting the data, calculation of the Maximum likelihood Estimation for an AR(*p*), MA(*q*) or ARMA(*p,q*) models will be greatly simplified by using the Innovations algorithm. Our experiments use the Innovations algorithm preliminarily estimation via choosing different estimation limits of *p* & *q* in AR(*p*), MA(*q*) or ARMA(*p,q*) models:

Experiment 4.1: $AR(p), p \in [1-10], MA(q), q = 0, d = 1$

Table 4.1 Preliminarily Estimation for Experiment 4.1

p	d	q	AICC
1	1	0	-408.390
2	1	0	-418.382
3	1	0	-432.815
4	1	0	-442.226
5	1	0	-439.103
6	1	0	-440.558
7	1	0	-439.679
8	1	0	-438.115
9	1	0	-437.265
10	1	0	-435.659

In Experiment 4.1 and as indicated by Table 4.1, the Preliminarily Estimation was executed using the Innovations algorithm within the given limits of p, d, q for only the Autoregressive model $AR(p) \quad p \in [1-10], MA(q) \quad q = 0, \quad d = 1$. Table 4.1 shows the calculated $AICC$ values for each examined order of p, d, q, the table indicates that Autoregressive model would provide better fitting of the examined dataset at $p = 4, d = 1, q = 0$ equivalent to $ARIMA(4, 1, 0)$ with minimum $AICC$ value of -442.226.

Experiment 4.2: $AR(p), p = 0, MA(q), q \in [1-10], d = 1$

Table 4.2 Preliminarily Estimation for Experiment 4.2

p	d	q	AICC
0	1	1	-434.685
0	1	2	-441.445
0	1	3	-435.738
0	1	4	-439.097
0	1	5	-443.159
0	1	6	-440.918
0	1	7	-432.018
0	1	8	-439.664
0	1	9	-432.045
0	1	10	-433.981

Table 4.2 shows the calculated $AICC$ values for each examined order of p,d,q for Experiment 4.2. The Preliminarily Estimation was executed using the Innovations algorithm within the given limits of p,d,q for only the Moving-Average model $MA(q) \quad q \in [1-10], AR(p) \quad p = 0, \quad d = 1$, the table indicates that Moving-Average model would provide better fitting of the examined dataset at $p = 0, d = 1, q = 5$ equivalent to $ARIMA(0,1,5)$ with minimum $AICC$ value of -443.159.

Experiment 4.3: $ARMA(p,q), p \in [1-10], q \in [1-10], d = 1$

In experiment 4.3, the Preliminarily Estimation was executed using the Innovations algorithm within the given limits of p,d,q for the Autoregressive Moving-Average model $ARMA(p,q) \quad p \in [1-10], \quad q \in [1-10], \quad d = 1$. Table 4.3 shows the calculated $AICC$ values for some examined orders of p,d,q, the table indicates that Autoregressive Moving-Average model would provide better fitting of the examined dataset at $p = 5, d = 1, q = 5$ equivalent to $ARIMA(5,1,5)$ with minimum $AICC$ value of -459.498. In this Table, we only show 10 combinations with the lowest AICC values.

Table 4.3 Preliminarily Estimation for Experiment 4.3

p	d	q	AICC
3	1	3	-445.119
4	1	3	-449.214
4	1	4	-450.809
5	1	3	-451.677
5	1	4	-456.735
5	1	5	-459.498
5	1	6	-451.232
6	1	2	-443.699
7	1	3	-442.860
7	1	5	-436.933

The parameters found in the $ARIMA(5,1,5)$ were used to initialize the MLE estimation; the MLE estimation was executed on different combinations from $p \in [1-5]$, $q \in [1-5]$, $d = 1$. According to our experiment, the MLE estimation provided better fitting of the examined dataset at $p = 4, d = 1, q = 5$ or $ARIMA(4,1,5)$ with minimum $AICC$ value of -474.516. It is important to mention that $ARIMA(4,1,5)$ order was examined previously in the Preliminarily Estimation without giving the minimum value of $AICC$, this is because the Preliminarily Estimation could not find the right values of the parameters (ϕ_p & θ_q) that give the minimum value of $AICC$. The resulted time-series model that fits our data is indicated in Equation 4.8. This Equation will be used in the prediction operation to find the data forecast.

$$
\begin{aligned}
X(t) &= -0.387X(t-1) + 0.271X(t-2) + 0.635X(t-3) \\
&+ 0.481X(t-4) + Z(t) - 0.340Z(t-1) - 0.787Z(t-2) \\
&- 0.684Z(t-3) + 0.104Z(t-4) + 0.705Z(t-5)
\end{aligned} \tag{4.8}
$$

4.3.3 Time-Series Model Testing

Once the data is fitted into a time-series model, it is important to test the model and ensure that the model represents the examined data. If the model is appropriate then the forecast errors should represent a white noise $\{Z_t\}$. This can be achieved through calculating the residuals which are the one-step prediction errors (Box *et al.* (2015)). The residuals $\{W_t\}$ is shown in Equation 4.9:

$$\hat{W}_t = (X_t - \hat{X}_t)/\sqrt{r_{t-1}} \tag{4.9}$$

where; $r_{t-1} = E(X_t - \hat{X}_t)^2/\sigma^2$

Taken from Box *et al.* (2015)

Where the predictor of X_t, based on data observations up to $t-1$, is represented by $\hat{X}_t$, and white noise variance of the fitted model is represented by σ^2. Assuming that the fitted time-series model is capable of generating our data, then the calculated residuals would represents observations with properties similar to those of a white noise process. This can be checked through calculating the ACF & PACF of the residuals. Figures 4.6 & 4.7 show the ACF & PACF functions of our fitted model residuals. We can see that there is no data exceeding the confidence bounds and this indicates that the calculated residuals have similar properties of the white noise process and indicates an IID random data. With that being said, we conclude that the fitted time-series model is an appropriate representation of the examined data.

4.3.4 Data Forecast & Prediction Results

Once the dataset is properly represented by a suitable and tested model, the data forecasting technique will use this model to predict the future observation(s). The future value(s) $X_{n+h}, h > 0$ of the stationary time-series can be predicted based on the previous values of $Xn, \ldots, X1$, with minimum Root Mean Square Error (RMSE) which has the same unit as the quantity

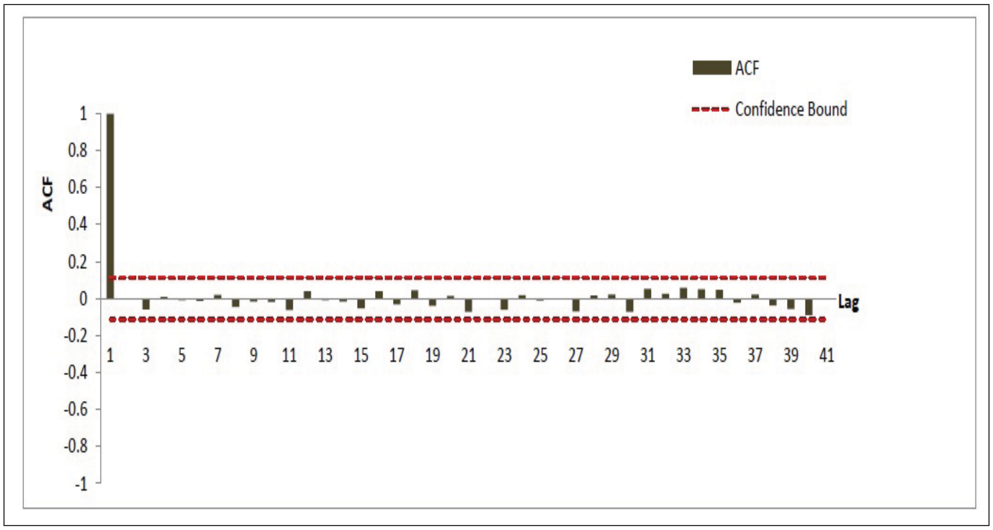

Figure 4.6 ACF of the fitted model residual

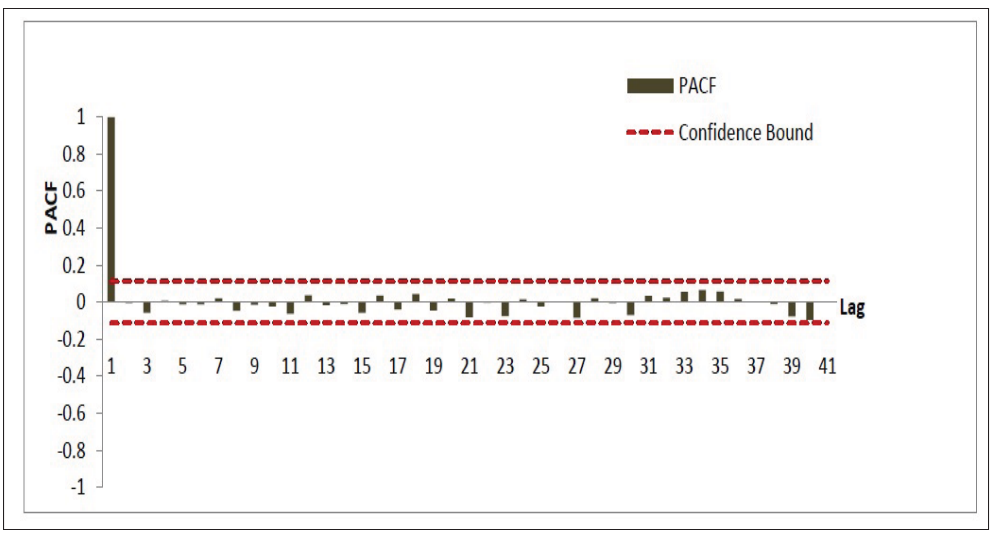

Figure 4.7 PACF of the fitted model residual

being estimated (Armstrong and Collopy (1992)). The time-series predictor can be denoted by P_nX_{n+h} and represented by Equation 4.10 where $m = max(p,q)$ (Brockwell and Davis (2006)):

$$P_n X_{n+h} = \begin{cases} \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \widehat{X}_{n+h-j}), \quad 1 \le h \le m-n, \\ \sum_{i=1}^{p} \phi_i P_n X_{n+h-i} + \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \widehat{X}_{n+h-j}), \quad h > m-n, \end{cases} \tag{4.10}$$

Taken from Brockwell and Davis (2006)

The data prediction was applied on a random subset of 50 seconds of the dataset LC-100. The forecast was executed on different values of ”h” which indicates the forecast steps measured by seconds. In fact it was conducted on 3 different values of ”h”: $h = 5$, $h = 10$, $h = 15$. Figure 4.8 shows the forecast results with $h = 5$ where the root mean squared error is $RMSE = 72.511$ kbps. The forecast results for $h = 10$ is shown in Figure 4.9 with $RMSE = 104.429$ kbps. The third forecast experiment was executed for $h = 15$ and the results are indicated by Figure 4.10 with $RMSE = 112.654$ kbps. By comparing the $RMSE$ error of the 3 experiments, we can see that the forecast experiment with $h = 5$ gives the lowest $RMSE$ error and the forecast experiment with $h = 15$ gives the highest $RMSE$ error value. Apparently, we can see that forecast with $h = 5$ provides more accurate values than the forecast for $h = 10$ & $h = 15$. Furthermore, forecast with $h = 5$ would require more processing power considering the frequency of the forecast execution, while the forecast with $h = 10$ or $h = 15$ would require less processing power.

4.4 Resource Reservation Enhancement

Based on the data forecast outcome of the dataset LC-100, we can estimate the potential unused guaranteed resources and utilize them accordingly to increase system throughput and capacity. The potential unused GBR bit-rate forecast (*GBR_used_F*) is shown in Equation 3.3, the *GBR_reserved* was assumed to be 2.2 Mbps for the LC-100 dataset. Figure 4.11 shows the resulted *GBR_used_F* based on the random subset of 50 seconds studied earlier. As stated before in Section 3.2, to safely perform the resources transition for the GBR resources once it

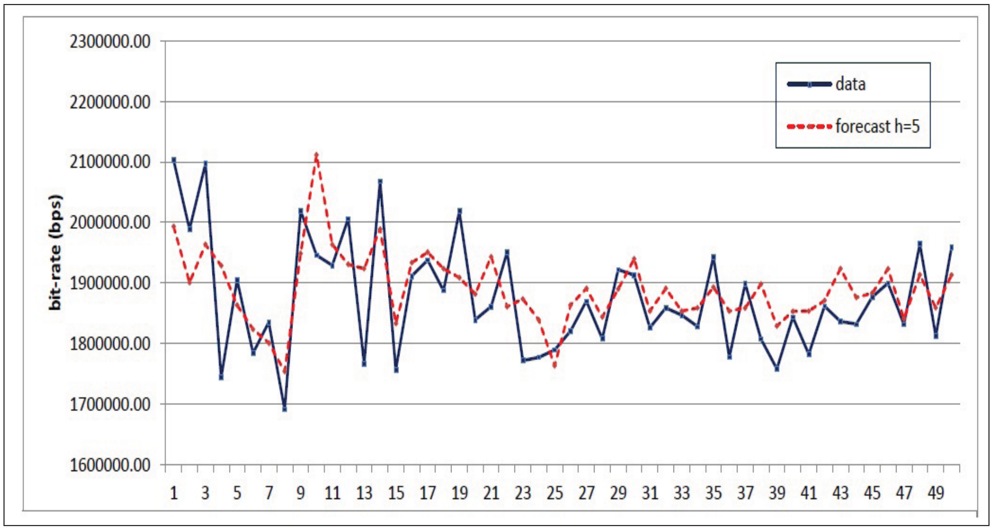

Figure 4.8 Forecast results with $h = 5$

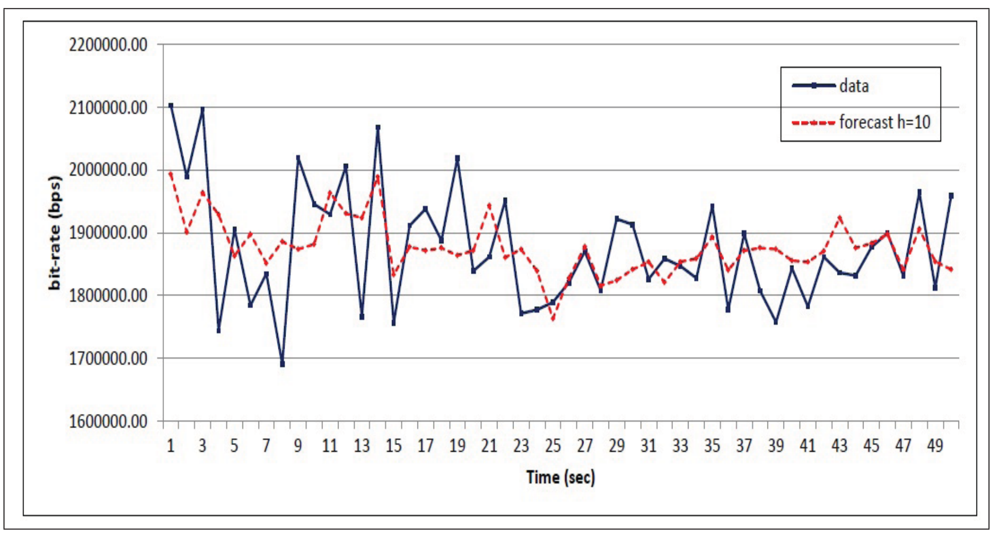

Figure 4.9 Forecast results with $h = 10$

is requested back by the "*contributing-bearer*", we introduce the *Safety Threshold* to avoid any disturbance for the LTE guaranteed services.

The *Safety Threshold* can be determined based on the *RMSE* forecast error. The previously calculated *RMSE* error value gives the overall error of each experiments but it does not reflect the forecast error behavior with time. To fully understand the forecast *RMSE* error behavior with time, we calculate a series of *RMSE* errors for each experiment based on the value

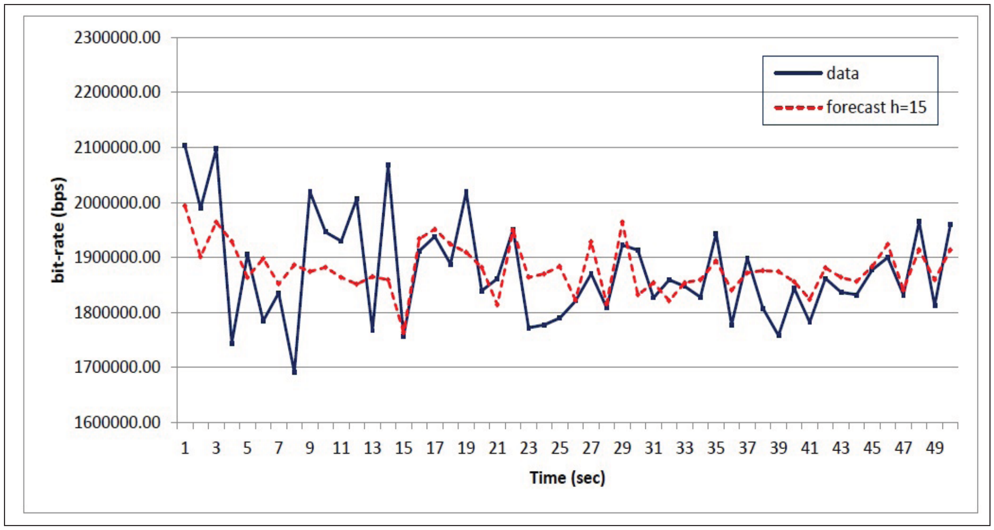

Figure 4.10 Forecast results with $h = 15$

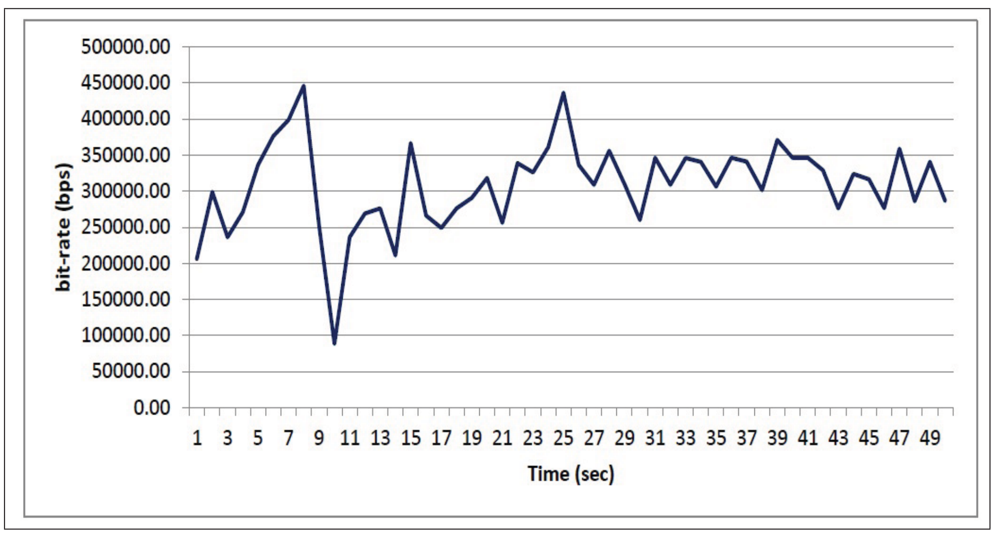

Figure 4.11 Potential Unused GBR Bit-rate Forecast

of h where the value h is considered for the *RMSE mean* calculation. The *RMSE* error behavior ($RMSE\_Behavior_{t=n}$) is highlighted in Equation 3.4 Section 3.3. The forecast *RMSE* error behavior will be continuously changing and consequently it will provide more accurate information for re-calculating the *Safety Threshold*.

Figures 4.12, 4.13 & 4.14 show the *RMSE* error behavior for $h = 5$, $h = 10$, $h = 15$ respectively. We notice that the *RMSE* error behavior for the lowest value of $h = 5$ provides more details about the error fluctuation than the other ones for $h = 10$ & $h = 15$. It actually shows the highest and lowest values of *RMSE* error with time, while *RMSE* error behavior for $h = 10$ & $h = 15$ hide those errors considering the difference in the *mean* value for *RMSE* error calculation. The value of h can be selected based on the *RMSE* error and *RMSE* error behavior. Furthermore, it is important to take in considerations that the small value of h would consume more CPU resources because of the frequency of the forecast execution but it will also provide more accurate results, while the large value of h would consume less CPU resources; but there would be a risk of having less accurate forecast results. For this experimental work, we choose to have the *Safety_Threshold* calculation based on *RMSE* error behavior with $h = 5$ which would ensure more safety and protection of the guaranteed resources availability when it is compared to the *RMSE* error behavior calculation with $h = 10$ & $h = 15$. Since the *Safety Threshold* will always rely on the forecast error at specific moment. We consider the RMSE error behavior at moment $(t = n)$ to be used for *Safety_Threshold* calculation at moment $(t = n+1)$. The $Safety\_Threshold_{t=n+1}$ is stated in Equation 3.5 in Section 3.3.

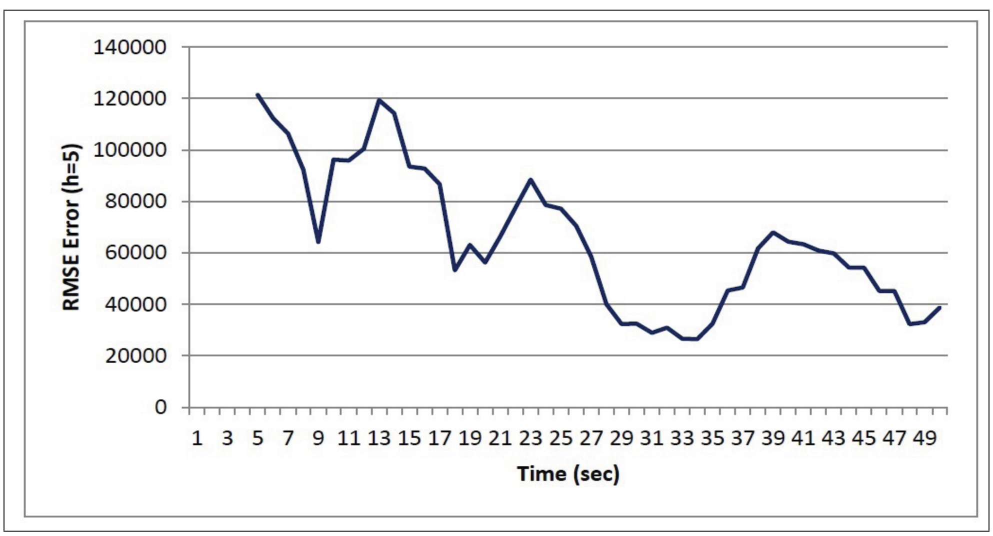

Figure 4.12 *RMSE* Error Behavior for $h = 5$

The Equation 3.5 indicates that *Safety_Threshold* will maintain a minimum value of *RMSE* in case the $RMSE\_Behavior_{t=n}$ is lower than the *RMSE* value which was calculated at 72.511 kbps for $h = 5$.

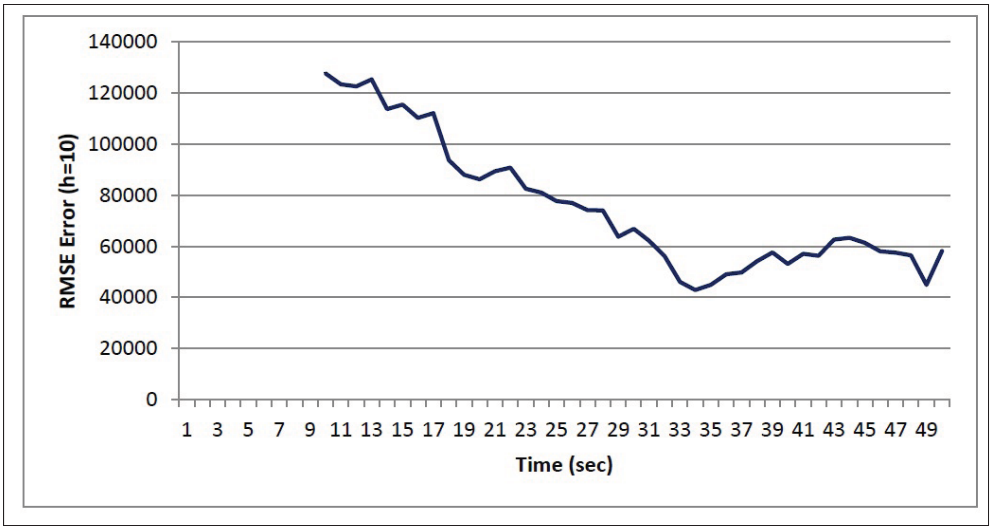

Figure 4.13 *RMSE* Error Behavior for $h = 10$

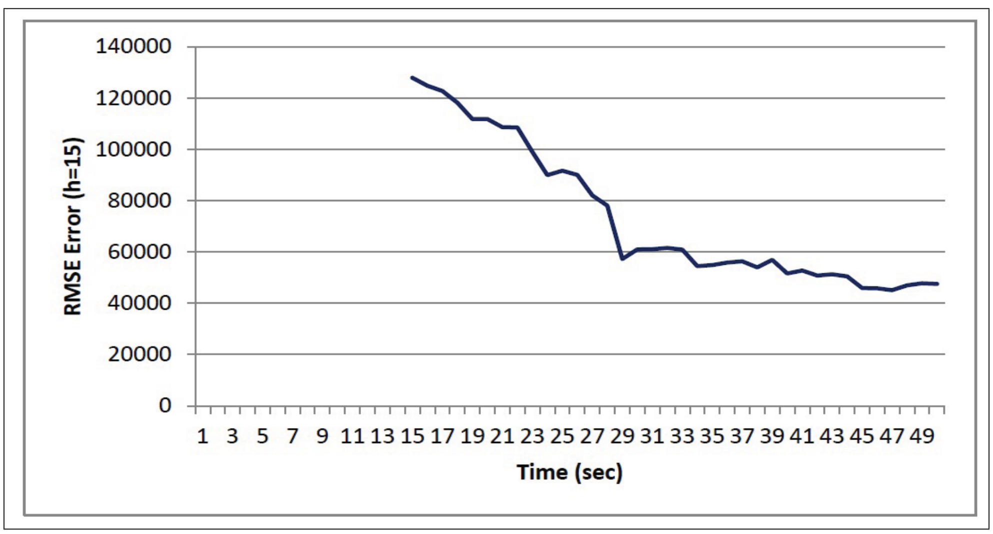

Figure 4.14 *RMSE* Error Behavior for $h = 15$

The $bitrate\_tobe\_utilized$ is shown in Equation 3.6 and indicates the guaranteed unused bit-rate that can be released and utilized for other running services. This can be calculated by deducting the *Safety Threshold* limit from the potential GBR unused bit-rate forecast. Figure 4.15 shows the guaranteed unused bit-rate that can be utilized for this experimental work ($bitrate\_tobe\_utilized$).

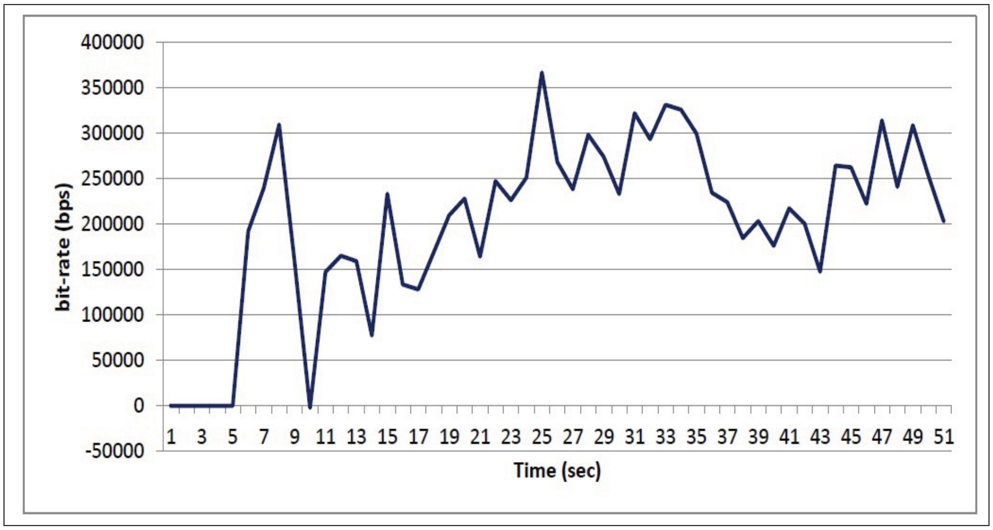

Figure 4.15 Guaranteed Unused Bit-rate that can be Utilized

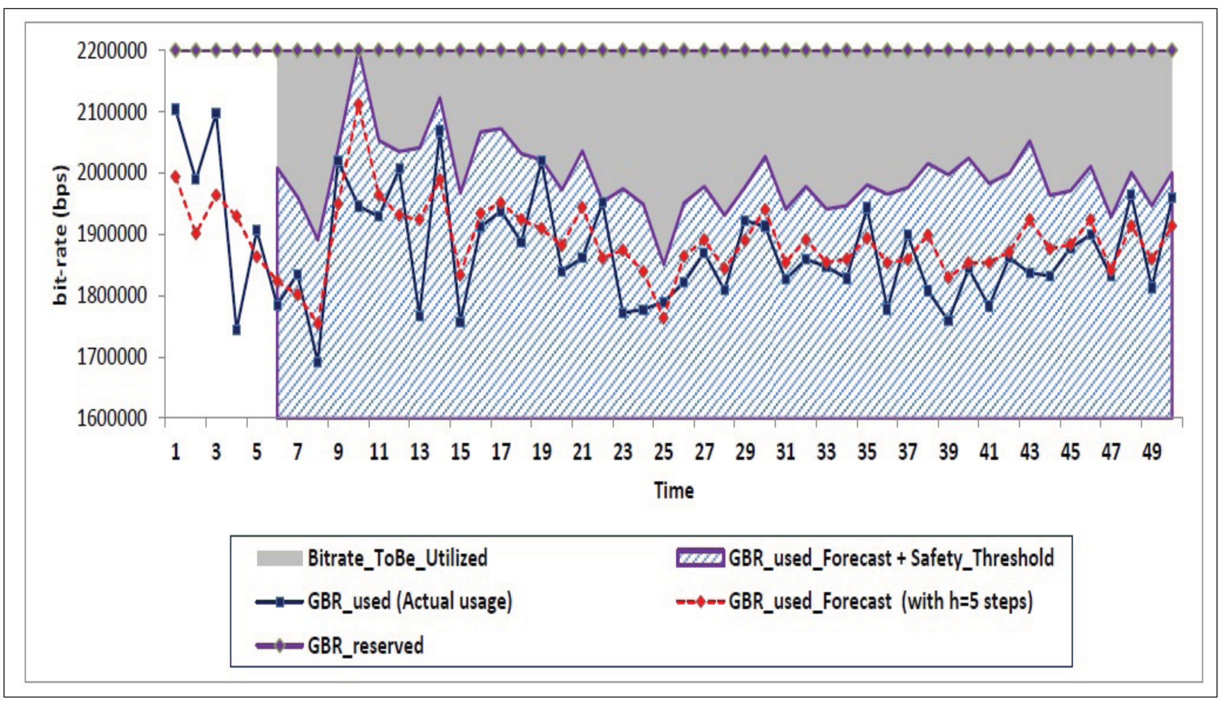

Figure 4.16 Overall view of bandwidth allocation

According to our proposed method in Section 3.2, the *acquiring-bearers* will be able to utilize the *bitrate_tobe_utilized*. This represents the released bit-rate or the gain/benefit of our proposed method. The average gain of the *bitrate_tobe_utilized* based on the studied 50 seconds subset is "185.411 kbps". This represents about 8.43% of gained resources. Figure 4.16 shows the overall view of the reserved bit-rate, actual used bit-rate, forecasted bit-rate and the *Safety Threshold* limit which is added on top of the forecast values. The grey highlighted area in Figure 4.16 represents the bit-rate bandwidth that can be released and utilized by other services (*bitrate_tobe_utilized*), which also represents the gain of our proposed model. In case the *GBR contributing-bearer* requests some bandwidth back, the *Safety Threshold* will be available to be used to ensure the quality of the guaranteed service.

To summarize, in this chapter, we applied our evolved approach for LTE video service resource reservation where we analyzed the dataset, transformed it to stationary and prepared it for modeling. We also found the time-series model that fits the data, performed the data forecast and estimated the wasted resources that can be utilized by other services.

CHAPTER 5

ENHANCED RESOURCE RESERVATION FOR GUARANTEED MULTI VIDEO LTE SESSIONS

5.1 Introduction

In this chapter, we will study a bigger dataset which contains several guaranteed bearers that carry mutli video conversational calls. We apply our approach on this dataset to validate the feasibility. Also, we will apply the time-series modeling through conducting several experiments to find the model that better fits the data, data forecast will be performed and our approach is expected to provide and estimate the resources gain that can be used by other services.

5.2 Dataset Formulation and Analysis

Based on the 3GPP Standardized QCI Characteristics highlighted in Table 2.1 (3GPP TS 23.203 (2015)), the table indicates that guaranteed resources are used for conversational voice, conversational video, real-time gaming and non-conversational video (buffered streaming). According to a study in (Ericsson (2015)), video traffic represents the largest portion of mobile data usage in LTE/EPC networks and this is why we focus on LTE video traffic in this experiment.

To formulate our dataset, we study several voice & video conversational calls and their guaranteed resource reservation in EPC gateway. According to our approach described in Chapter 3, in Equation 3.1, our examined dataset includes all those GBR calls by summing their used bit-rates in one pool (*GBR_used*) which represents our dataset for this work. The *GBR_reserved* represents the summation of all the reserved GBR bit-rates for those calls.

Table 5.1 shows the 10 voice & video conversational calls conducted in this study with the bit-rate information, those calls were captured in simulated LTE/EPC environment called *nwEPC*

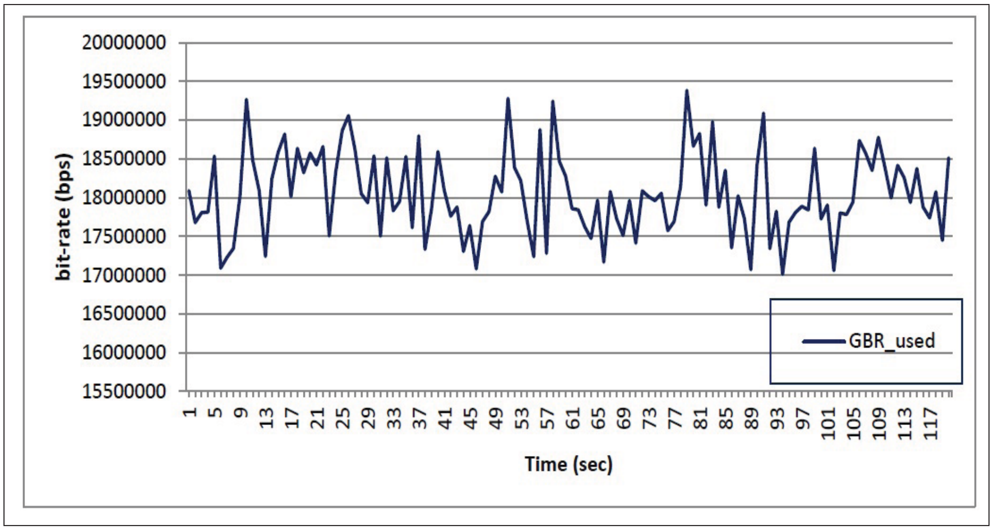

Figure 5.1 Dataset: Total GBR_used

(Chawre (2010)). Figure 5.1 shows the *GBR_used* dataset according to Equation 3.1, the *GBR_reserved* was calculated based on those 10 calls and found to be (22 Mbps).

Table 5.1 Dataset Formulation: voice & video conversational calls

Application	GBR bit-rate	call duration (sec)	number of calls
Facetime	1.3 Mbps	120	2
Skype	1.6 Mbps	120	2
Hangouts	4.3 Mbps	120	2
Viber	2.0 Mbps	120	2
Tango	1.8 Mbps	120	2
Total	22 Mbps	-	10

The Autocorrelation Function (ACF) has been calculated for the *GBR_used* dataset and it was found that the ACF values somehow decrease slowly which indicates that the data is not stationary, this indicates that the data may contain some data trend, seasonality and/or data variability. Hence, Box-Cox transformation (Box and Cox (1964)) & differencing technique (Brockwell and Davis (2006)) should be applied on the dataset *GBR_used* to generate new data with stationary properties.

By conducting Box-Cox & differencing techniques, it was found that Box-Cox transformation is needed to be executed at $\lambda = 0$ to stabilize any data variability. The differencing technique is also needed to be to eliminate any data trend and seasonality. The differencing procedure was applied on the transformed data at different values of d. We found that differencing at d=*5* is good enough to convert the data into stationary. Figure 5.2 shows the *GBR_used* dataset after Box-Cox & differencing techniques were applied at the previously indicated parameters.

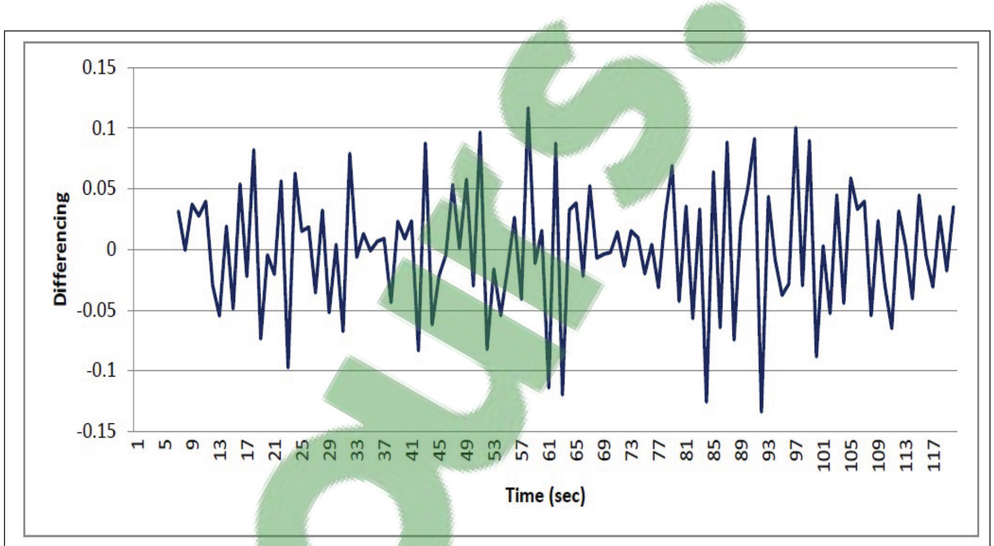

Figure 5.2 Dataset after Transformation & Differencing

By calculating the ACF & PACF functions -shown in Figures 5.3 & 5.4-, we can conclude that the transformed data represents stationary time-series. We refer to the resulted data by $\{X_n\}$ which represents the random noise component or the residual of the original dataset.

To ensure that dependency exists between the residual observations, the Independent and Identically Distributed (IID) hypothesis should be rejected. Based on the ACF/PACF functions, the IID hypothesis is rejected as there are more than 5% of the ACF/PACF calculated values falling outside the confidence bounds $\pm 1.96n^{-1/2}$ which is equal to ± 0.1789 where $n = 120$. The confidence bound is represented by the dashed line in the figure. In addition, the ACF/-PACF output indicates that the examined data is correlated with itself at several lags. This will be important in the time-series modeling discussed in the next section.

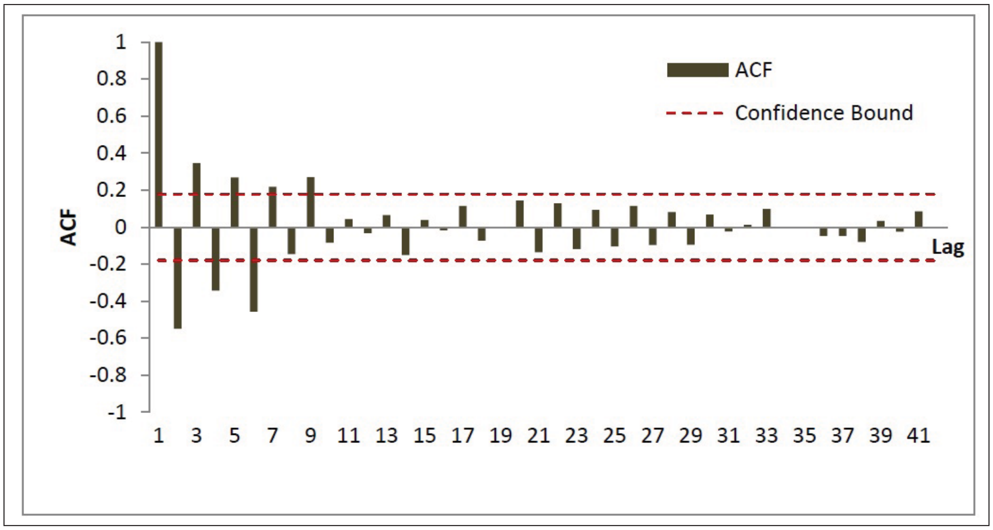

Figure 5.3 ACF after Transformation and Differencing

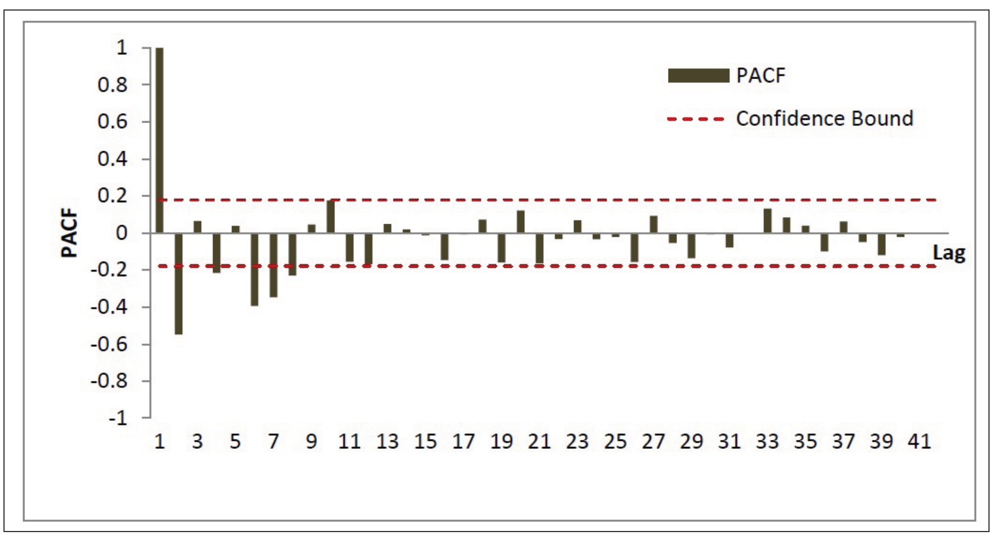

Figure 5.4 PACF after Transformation and Differencing

5.3 Modeling of Time-Series

Considering that the original dataset was not stationary, the Autoregressive Integrated Moving Average ARIMA(p,d,q) model (Brockwell and Davis (2006) and Box *et al.* (2015)) should be considered to model our time-series $\{X_n\}$. The ARIMA model is usually used when the data is not stationary. Through the differencing technique performed in Section 5.2, we found that $d = 5$ which leads to have ARIMA(*p,5,q*).

To complete the ARIMA time-series modeling, we have several problems to solve which includes: the determination of p and q (order selection) and the estimation of the mean, the noise component variance and the coefficients $\{\phi_i, i = 1, \ldots, p\}$, $\{\theta_i, i = 1, \ldots, q\}$. These coefficients formulates the ARIMA model that represents $\{X_n\}$. Both ACF & PACF in Figures 5.3 & 5.4 indicate that the data is correlated with itself at different lags. This indicates that the Autoregressive and Moving Average modeling components will be present in the ARIMA model. The ϕ and θ coefficients can be found by utilizing the data observations of the stationary time-series through using the Maximum Likelihood Estimation (MLE) (Brockwell and Davis (2006)).

Since the MLE is complicated and time consuming in solving the system equations, we do specify initial parameter values for the MLE to begin the search. This can be achieved through Preliminary Estimation and more specifically we use the Innovations algorithm (Brockwell and Davis (1988)) which provides initialization parameters for the MLE Estimation and indicates some estimation of the order selection.

Through the Innovations algorithm, we performed several iterations to find a suitable model that can be used to initialize the MLE operation. Based on the lags that show data correlation in the ACF & PACF, the Innovations algorithm was executed for $p \in [1-5], \quad q \in [1-8], \quad d = 5$, Table 5.2 indicates the top 10 combinations found within the specified ranges for p & q according to the AICC criterion (Hurvich and Tsai (1989)). According to experiment results indicated in Table 5.2, we found that $ARIMA(2,5,6)$ provides the lowest $AICC$ value of -438.591 which provides an appropriate model that can used for MLE initialization. The AICC criterion (Hurvich and Tsai (1989)) has been used in our experiments to test the time-series model and choose the most appropriate model that can better fit the data.

By using the MLE, we perform several experiments to find a suitable time-series model that better represents our data. We found again that $ARIMA(2,5,6)$ provides the lowest $AICC$ value

Table 5.2 Preliminarily Estimation using Innovations algorithm for: $ARMA(p,q) \quad p \in [1-5], \quad q \in [1-8], \quad d = 5$

p	d	q	AICC
2	5	3	-429.016
3	5	3	-424.754
2	5	4	-430.098
3	5	4	-425.429
2	5	5	-432.431
3	5	5	-428.990
2	5	6	-438.591
3	5	6	-432.532
5	5	6	-419.114
4	5	8	-420.009

of -442.708. The resulted time-series model that fits our data is indicated in Equation 5.1. This Equation will be used in the data forecast section.

$$
\begin{aligned}
X(t) &= -0.2870X(t-1) + 0.04703X(t-2) + Z(t) \\
&- 0.4030Z(t-1) + 0.01992Z(t-2) - 0.1021Z(t-3) \\
&- 0.04738Z(t-4) - 0.8829Z(t-5) + 0.5425Z(t-6)
\end{aligned}
\tag{5.1}
$$

Once the data is fitted into a time-series model, it is important to test the model and ensure that the model represents the examined data. As explained in Section 4.3.3, if the model is appropriate, then the forecast errors should represent a white noise $\{Z_t\}$. This can be achieved through calculating the residuals which are the one-step prediction errors (Box *et al.* (2015)). The residuals calculation is highlighted in Section 4.3.3.

Figures 5.5 & 5.6 show the ACF & PACF functions of our fitted model residuals. We can see that there is no data exceeding confidence bounds and this indicates that the calculated residuals have similar properties of white noise process and indicate an IID random data. With that being said, we conclude that the fitted time-series model is an appropriate representation of the examined data.

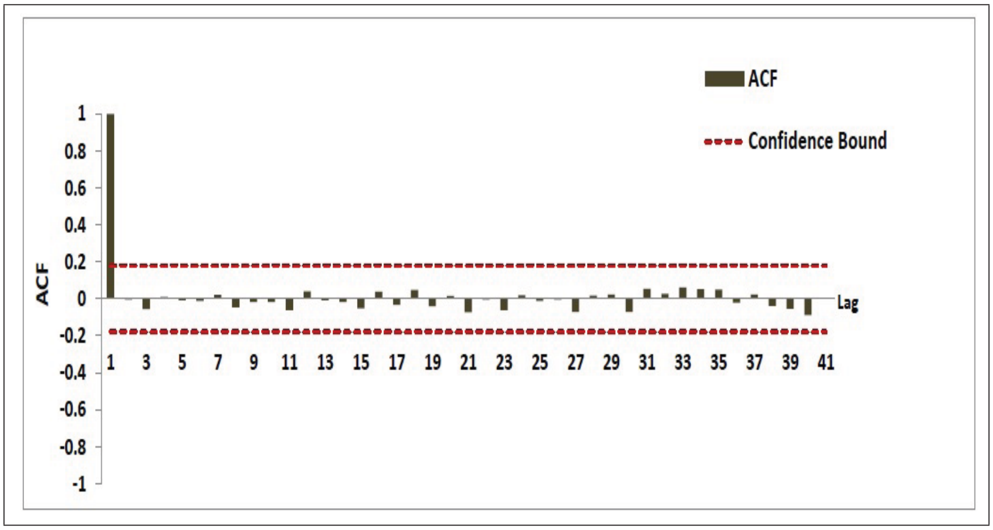

Figure 5.5 ACF of the fitted model residual

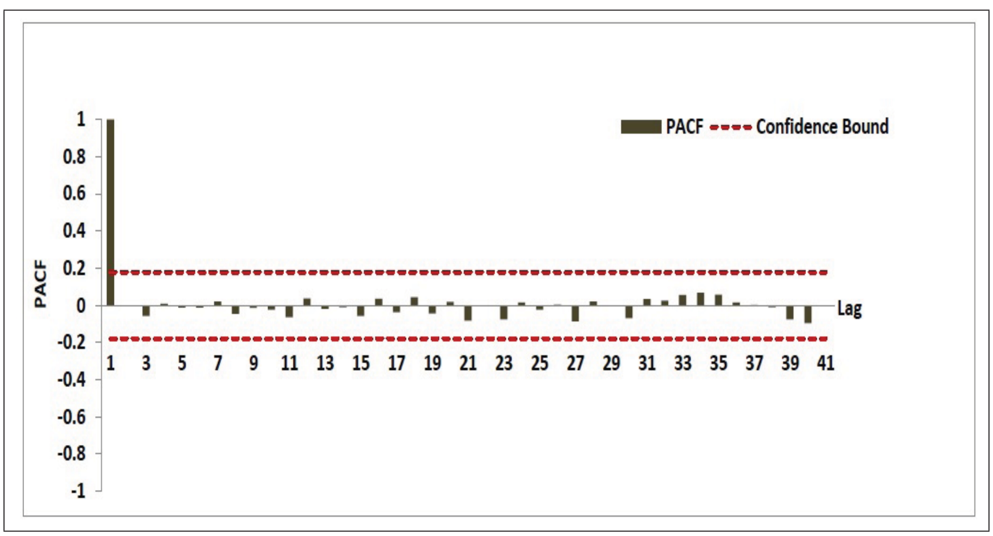

Figure 5.6 PACF of the fitted model residual

5.4 Data Prediction and Forecast

The future value(s) $X_{n+h}, h > 0$ of the stationary time-series can be predicted based on the previous values of $Xn, \ldots, X1$ (Karapanagiotidis (2012)), with minimum Root Mean Square Error (RMSE) (Armstrong and Collopy (1992)). The time-series predictor can be denoted by $P_n X_{n+h}$ and represented by Equation 5.2 where $m = max(p,q)$ (Brockwell and Davis (2006)):

$$P_n X_{n+h} = \begin{cases} \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \widehat{X}_{n+h-j}), & 1 \leq h \leq m-n, \\ \sum_{i=1}^{p} \phi_i P_n X_{n+h-i} + \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \widehat{X}_{n+h-j}), & h > m-n, \end{cases} \tag{5.2}$$

Taken from Brockwell and Davis (2006)

With the help of the ARIMA time-series model found in the Section 5.3 and the Prediction Equation 5.2, the data prediction was executed on a random subset of 60 seconds of our time-series $\{X_n\}$. The forecast was executed on different values of "h" at: $h = 5$, $h = 10$ & $h = 15$. The value "h" indicates the forecast steps measured by seconds. Once the forecast is executed on the time-series, the differencing and the box-cox transformation are being reverted on the time-series forecast results to produce the original dataset forecast.

Figures 5.7, 5.8 & 5.9 show the forecast results of the values $h = 5$, $h = 10$ & $h = 15$ respectively. By calculating the $RMSE$ error of the 3 experiments, we can see that the forecast experiment with $h = 5$ gives the lowest $RMSE$ error at $RMSE = 669.739$ kbps and the forecast experiment with $h = 15$ gives the highest $RMSE$ error value at $RMSE = 817.606$ kbps. The forecast experiment with $h = 10$ gives $RMSE = 729.923$ kbps. Clearly, the forecast experiment with $h = 5$ provides more accurate values than the forecast for $h = 10$ & $h = 15$; but it would require more processing power considering the frequency of the forecast execution, while the forecast with $h = 10$ or $h = 15$ would require less processing power. In this experiment, we will consider the forecast with the value of $h = 5$.

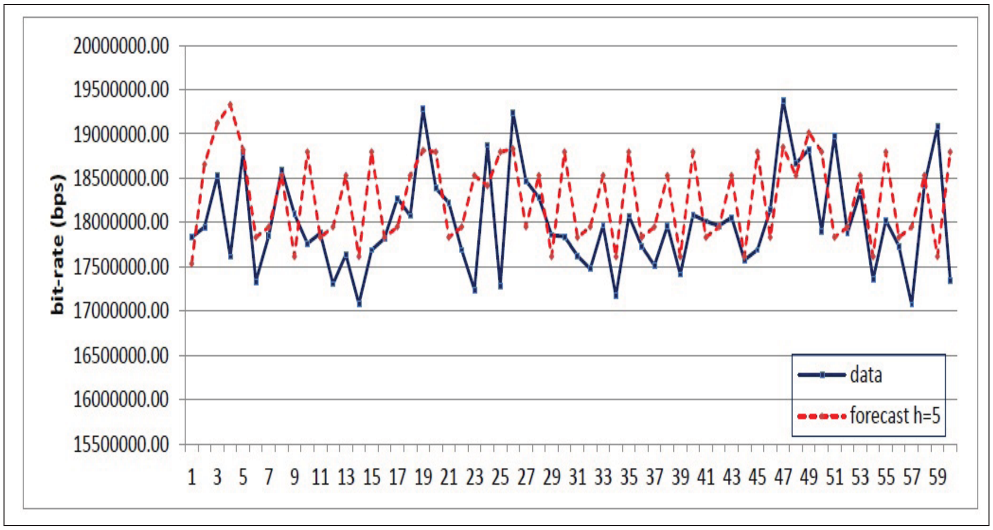

Figure 5.7 Forecast results with $h = 5$

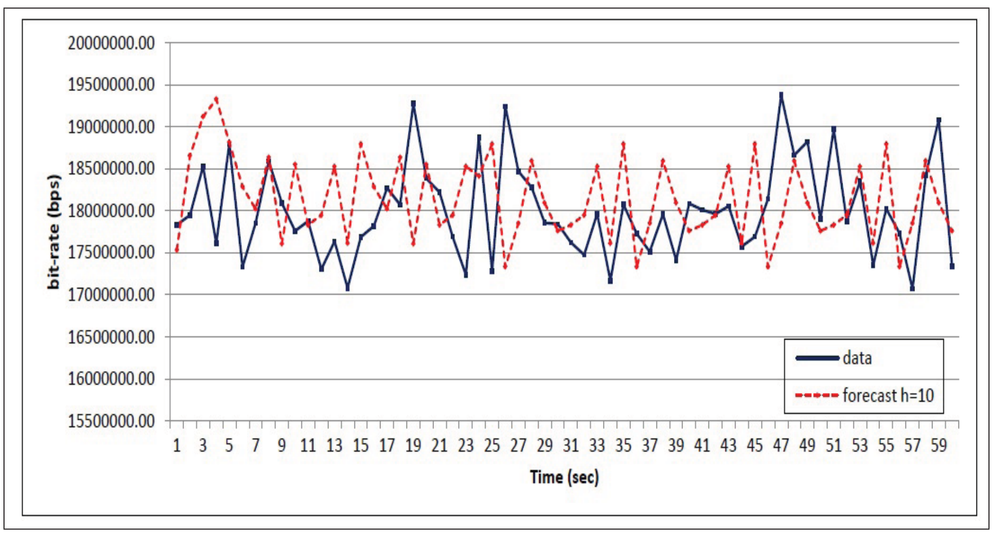

Figure 5.8 Forecast results with $h = 10$

5.5 Enhanced Resource Reservation

Based on the prediction results from the forecast experiment with $h = 5$ -referred to as *GBR_used_F*-, we can calculate the *GBR_unused_F* series based on Equation 3.3 stated in Section 3.3. The *GBR_unused_F* can be used to calculate the *bitrate_tobe_utilized* as shown in Equation 3.6 by deducting the *Safety_Threshold* limit from the forecasted unused GBR bit-rate (*GBR_unused_F*).

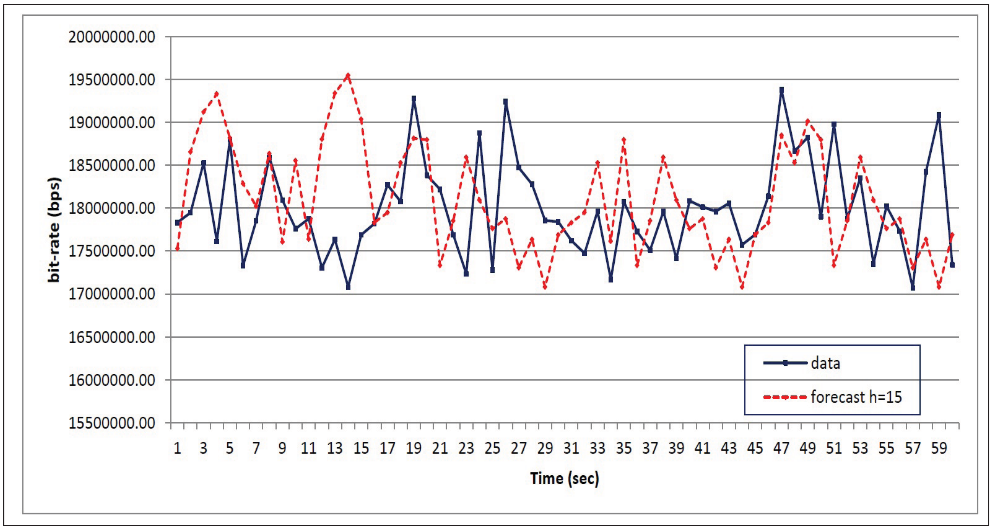

Figure 5.9 Forecast results with $h = 15$

The *Safety_Threshold* is calculated based on Equation 3.5 and according to the *RMSE_Behavior* that is calculated according to the forecast experiment with $h = 5$. Figure 5.10 shows the *RMSE* error behavior which is used to formulate the *Safety_Threshold*. The *Safety_Threshold* represents a reserved standby resources pool of the contributed resources to be used by the *contributing-bearers* in case they request back their guaranteed resources. The *bitrate_tobe_utilized* of our experiment is calculated and shown in Figure 5.11. The *bitrate_tobe_utilized* represents the guaranteed unused bit-rate that can be released and utilized for other running services (*acquiring-bearers*); it also represents the gain/benefit of our proposed approach.

The average gain of the *bitrate_tobe_utilized*, based on the studied 60 seconds subset, is "3022.954 kbps". This represents about 13.74% of gained resources. Figure 5.12 shows the overall view of the *GBR_reserved* bit-rate, the *GBR_used* bit-rate, the *GBR_used_F* bit-rate and the *Safety_Threshold* limit which is added on top of the *GBR_used_F* forecasted values. The grey highlighted area in Figure 5.12 represents the bit-rate bandwidth that can be released and utilized by other services. It also represent the gain/benefit of our proposed model. In case the *GBR contributing-bearer* requests some bandwidth back, the *Safety_Threshold* will be available to be used to ensure the quality of the guaranteed services.

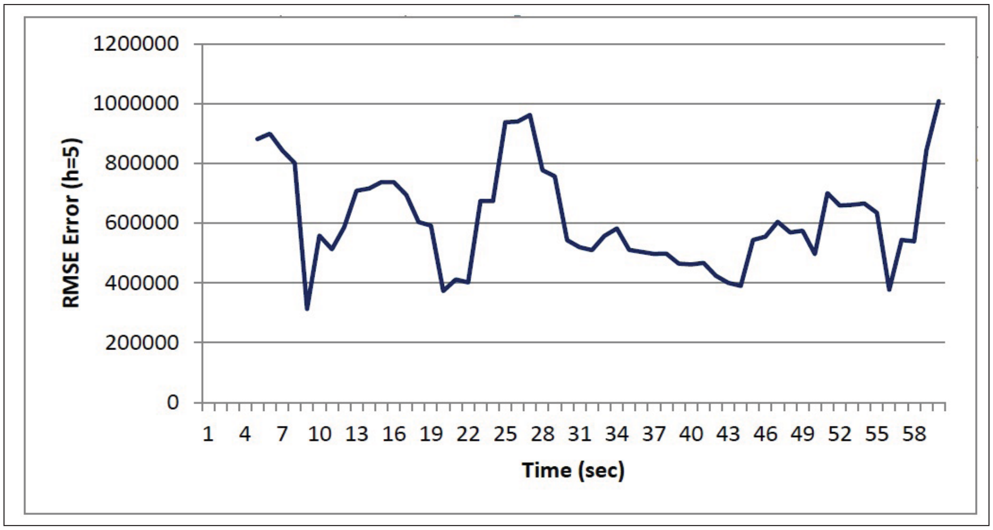

Figure 5.10 *RMSE* Error Behavior for $h = 5$

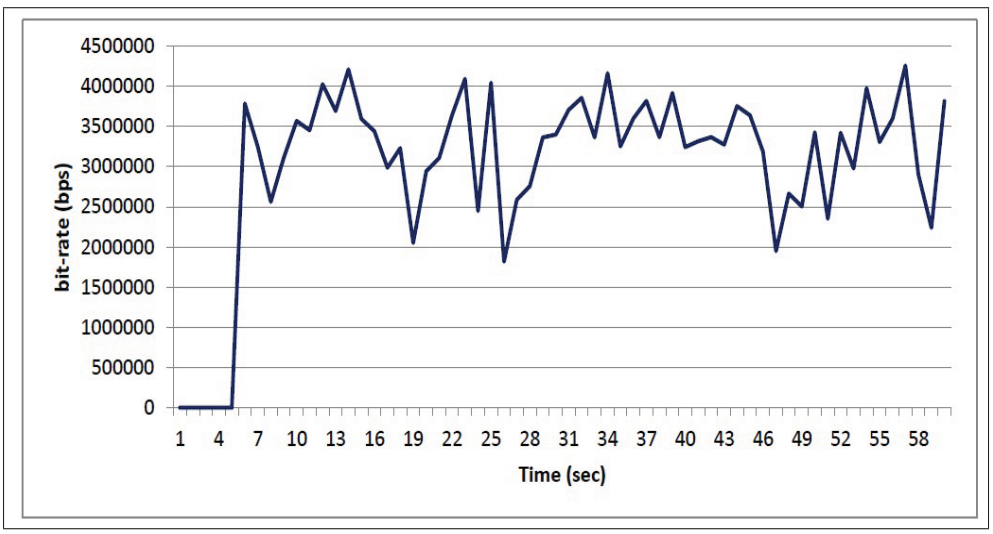

Figure 5.11 Guaranteed Unused Bit-rate that can be Utilized

In this chapter, we studied bigger dataset which contains several guaranteed bearers that carry mutli video conversational calls. We applied our enhanced approach of resource reservation on the dataset to validate the feasibility of our approach. Also, we applied the time-series modeling through conducting several experiments to find the model that better fits our data. Data forecast was performed and our approach was able provide and estimate the wasted resources that can

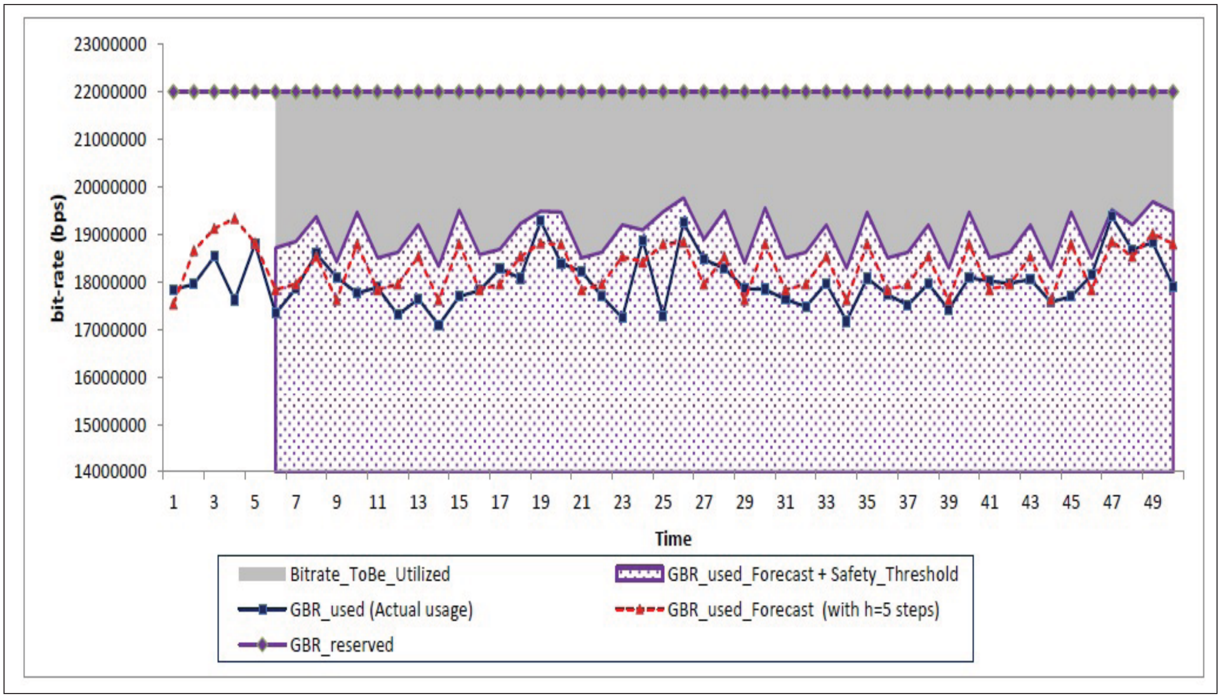

Figure 5.12 Overall view of bandwidth allocation

be reserved differently and used by other services to reduce the wasted resources and increase the system capacity.

CHAPTER 6

RESOURCE UTILIZATION ENHANCEMENT FOR GUARANTEED MULTI-MEDIA LTE SESSIONS

6.1 Introduction

In this chapter, a more complex real-life dataset will be analyzed and studied. The dataset consists of several guaranteed bearers that carry different kinds of guaranteed services which would reflect a real-life scenario. Our approach will be applied to find the time-series model. The data forecast will be performed to provide the unused resources. Data simulation experiment will be executed to show the benefit/gain that our approach would provide.

6.2 Multi-Media Dataset Formulation

As mentioned earlier in Table 2.1 from 3GPP standards (3GPP TS 23.203 (2015)), guaranteed resources can be allocated for several services, namely conversational voice, conversational video, real-time gaming and non-conversational video. The guaranteed resources are used to allocate dedicated network resources and bitrate to ensure the quality of the provided service. Hence, real-life guaranteed LTE traffic would be a mix of all those services but with different ratios.

To formulate a dataset that represents real-life guaranteed LTE traffic, we study several Multi-media services that require guaranteed resources and can be categorized under the guaranteed services according to Table 2.1.

Table 6.1 shows a set of 10 conversational video calls, the application, bit-rate information, call duration in seconds and number of calls for each application. This kind of mobile service is classified under QCI "2" according to the QCI Information & Characteristics indicated in Table 2.1. In Table 6.2. We highlight another set of conversational video, this data represents 5 live video streaming sessions from different known live streaming sources. The bit-rate information and session duration in seconds are highlighted in the Table. This kind of mobile service could be classified under QCI "2" according to the QCI Information & Characteristics indicated in Table 2.1.

Table 6.1 Dataset: voice & video conversational calls

Application	**GBR bit-rate**	**call duration (sec)**	**number of calls**
Facetime	1.3 Mbps	120	2
Skype	1.6 Mbps	120	2
Hangouts	4.3 Mbps	120	2
Viber	2.0 Mbps	120	2
Tango	1.8 Mbps	120	2
Total	22.0 Mbps	-	10

Table 6.2 Dataset: Conversational video (Live Streaming) calls

Application	**GBR bit-rate**	**call duration (sec)**	**number of calls**
Ustream	5.0 Mbps	120	1
LiveStream	6.0 Mbps	120	1
LiveStream	5.0 Mbps	120	1
NBC news	4.0 Mbps	120	1
ABC	4.5 Mbps	120	1
Total	24.5 Mbps	-	5

Table 6.3 shows 5 buffered video streaming sessions from known video providers. Full information about bit-rate and session duration are highlighted in the Table. This kind of mobile service is classified as non-conversational video under QCI "4" according to the QCI Information & Characteristics indicated in Table 2.1. Table 6.4 shows 2 real-time gaming sessions along with the bit-rate information and session duration. This kind of mobile service is classified under QCI "3".

Those calls were captured in LTE/EPC simulated environment called *nwEPC* (Chawre (2010)) using guaranteed bit-rates. Table 6.5 shows a summary of all the data sessions provided in Tables 6.1, 6.2, 6.3 & 6.4 and formulates the overall dataset that we will conduct in this chapter.

Table 6.3 Dataset: non-Conversational video (buffered streaming) calls

Application	**GBR bit-rate**	**call duration (sec)**	**number of calls**
YouTube 720p	4.5 Mbps	120	1
YouTube 1080p	6.0 Mbps	120	1
Shomi Streaming	2.5 Mbps	120	1
Netflix 720p	3.5 Mbps	120	1
Netflix 1080p	5.0 Mbps	120	1
Total	21.5 Mbps	-	5

Table 6.4 Dataset: Real-time Gaming calls

Application	**GBR bit-rate**	**call duration (sec)**	**number of calls**
StreamMyGame	4.0 Mbps	120	1
StreamMyGame	6.0 Mbps	120	1
Total	10.0 Mbps	-	2

According to Table 6.5, our dataset consists of several guaranteed services from different QCI. The dataset contains 22 sessions with total of 78 Mbps for the GBR bit-rate.

Table 6.5 Total Dataset Formulation

Service Category	**Total GBR bit-rate**	**number of calls**
Conversational-video (video calling)	22.0 Mbps	10
Conversational-video (Live Streaming)	24.5 Mbps	5
non-Conversational-video (buffered streaming)	21.5 Mbps	5
Real-time Gaming	10.0 Mbps	2
Total	78.0 Mbps	22

According to our approach described in Section 3.3 and in Equation 3.1, our examined dataset includes all those GBR calls by summing their used bit-rates in one pool (*GBR_used*) which represents our dataset for this experiment. The *GBR_reserved* represents the summation of all the reserved GBR bit-rates for those calls. Figure 6.1 shows the *GBR_used* dataset according to Equation 3.1. The *GBR_reserved* was calculated based on those 22 calls and found to be (78 Mbps).

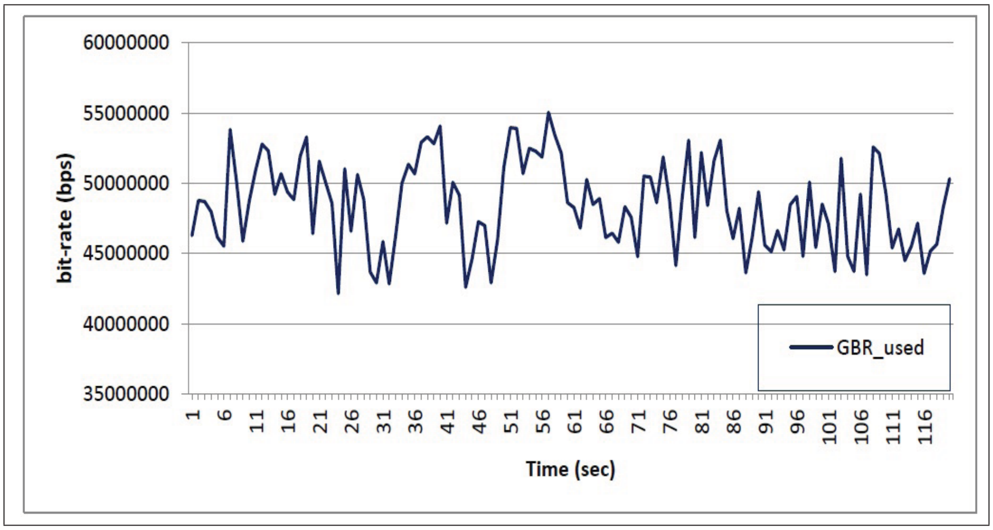

Figure 6.1 Dataset: Total GBR_used

6.3 Dataset Analysis & Preparation

To perform dataset analysis, we need to examine if dataset is stationary which would be important for the data modeling. To achieve that, the ACF function has been calculated for the *GBR_used* dataset and it is demonstrated in Figure 6.2. Based on the ACF demonstration, it is obvious that the ACF values somehow does not decrease quickly which indicates that the data is not stationary. It indicates also that the data may contain some data trend, seasonality and/or data variability. To make the data stationary, Box-Cox transformation (Box and Cox (1964)) & differencing technique (Brockwell and Davis (2006)) should be applied on the dataset *GBR_used* to generate new time-series with stationary properties.

The Box-Cox & differencing techniques were applied on the *GBR_used* dateset. The Box-Cox transformation was needed to be executed at $\lambda = 0$ to stabilize any data variability. The differencing technique was applied to eliminate any data trend and seasonality. The differencing procedure was applied on the transformed data at different values of d. We found that

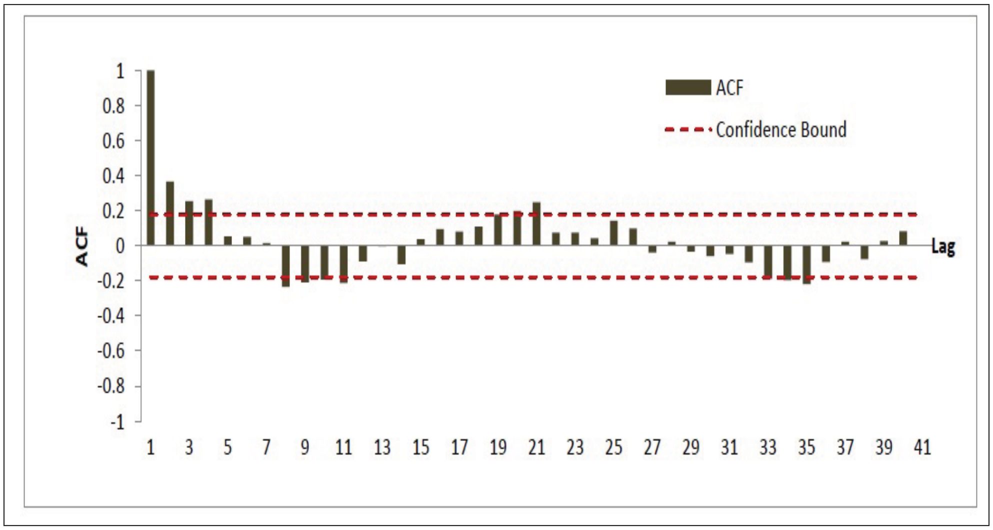

Figure 6.2   ACF of the Multi-Media Dataset

differencing at $d=1$ is good enough to convert the data into stationary. Figure 6.3 shows the *GBR_used* dataset after Box-Cox & differencing techniques were applied at the previously indicated parameters.

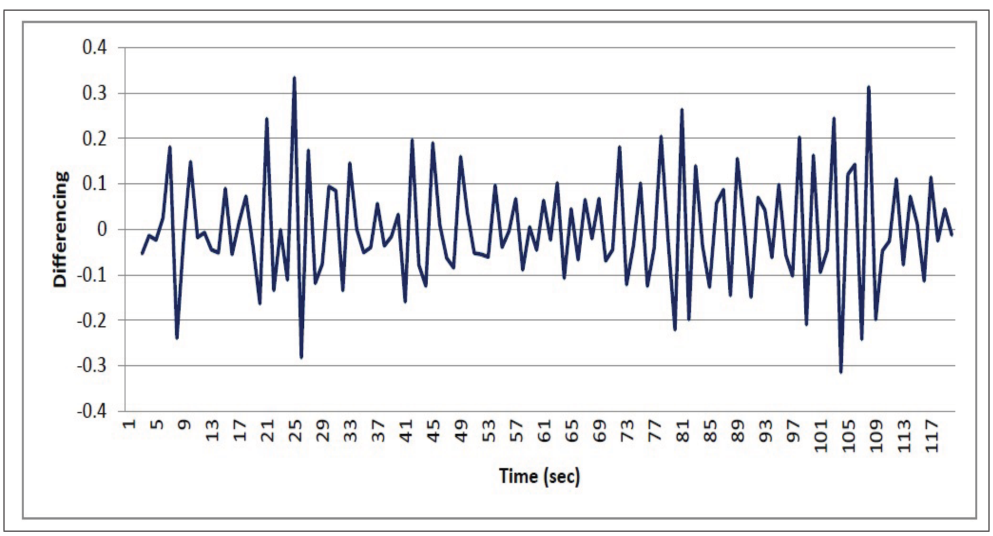

Figure 6.3   Dataset after Transformation & Differencing

After the dataset was transformed and differenced, we calculated the ACF & PACF functions on the resulted data as shown in Figures 6.4 & 6.5. The figures show that the resulted data represents stationary time-series as the ACF & PACF values decrease quickly. We refer to the resulted data by $\{W_n\}$ which represents the random noise component or the residual of the original dataset.

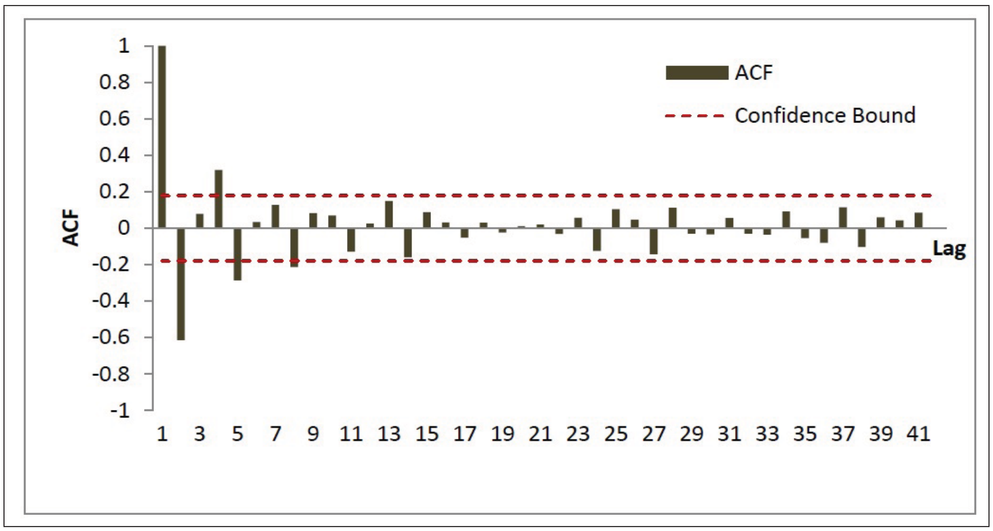

Figure 6.4 ACF after Transformation and Differencing

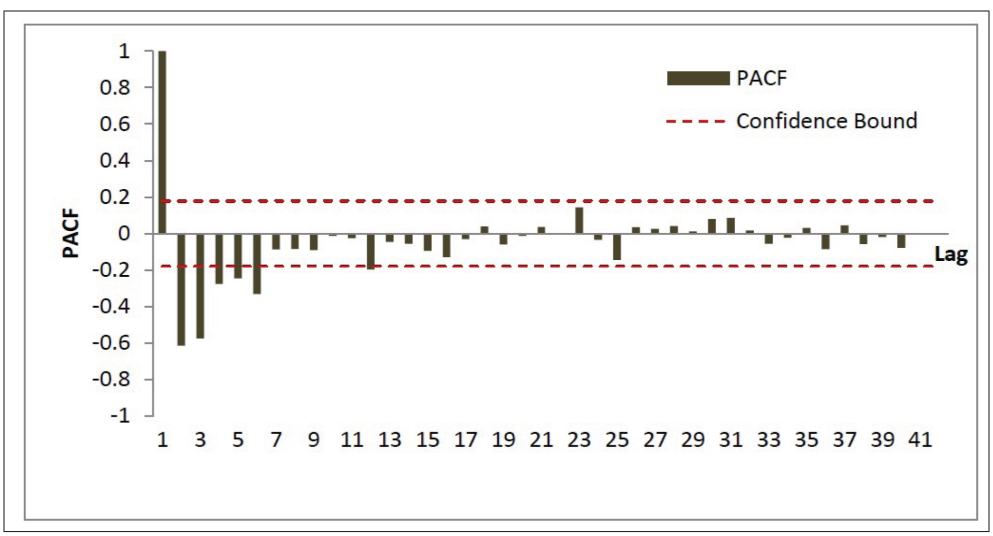

Figure 6.5 PACF after Transformation and Differencing

To ensure that dependency exists between the residual observations in $\{W_n\}$, the Independent and Identically Distributed (IID) hypothesis should be rejected. Based on the ACF & PACF functions, the IID hypothesis is rejected as there are more than 5% of the ACF & PACF values ($5\% * 40 = 2$) falling outside the confidence bounds $\pm 1.96n^{-1/2}$ which is equal to ± 0.1789 where $n = 120$, we have "4" values for ACF and "6" values for PACF that reside outside the confidence bounds. The confidence bound is represented by the dashed line in the ACF & PACF figures. In addition, the ACF & PACF output indicates that the examined data is correlated with itself at those lags that reside outside the confidence bounds, this conclusion is important in modeling the time-series $\{W_n\}$.

6.4 Time-Series Modeling

As the original dataset was found not stationary, the Autoregressive Integrated Moving Average ARIMA(*p,d,q*) model (Brockwell and Davis (2006)) should be the one used to model our time-series $\{W_n\}$. The ARIMA model is usually used when the data is not stationary. As part of the dataset preparation, the differencing technique was performed in Section 6.3 and we found that $d = 1$ would help to make the data stationary, differencing at $d = 1$ leads so far to have ARIMA(*p,1,q*) model.

To complectly model the ARIMA time-series, we need to solve several problems which are: the determination of p and q (order selection) and the estimation of the mean, the noise component variance and the coefficients $\{\phi_i, i = 1, \ldots, p\}$, $\{\theta_i, i = 1, \ldots, q\}$, these coefficients formulates the ARIMA model that represents $\{W_n\}$. As stated before, the ACF & PACF output in Figures 6.4 & 6.5 indicates that the examined data is correlated with itself at several lags. This indicates that the Autoregressive and Moving Average modeling components will be present in the ARIMA model.

As indicated previously in Section 4.3.2, finding the ϕ and θ coefficients can be done by utilizing the data observations of the stationary time-series through using the Maximum Likelihood Estimation (MLE) (Brockwell and Davis (2006)). The MLE is considered complicated and time consuming in solving the system equations. To simplify the MLE, we specify initial parameter values for the MLE to begin the search with. This can be done through the Preliminary Estimation and more specifically the Innovations algorithm (Brockwell and Davis (1988)) that is used to provide initialization parameters for the MLE Estimation and also to indicate some estimation of the order selection. Calculation of the Maximum likelihood Estimation for ARIMA(*p,1,q*) model will be greatly simplified by using the Innovations algorithm. The AICC (Hurvich and Tsai (1989)) criterion will be used in the our experiment to measure which model is better fitting the data. In our experiment, the Innovations algorithm preliminarily estimation is used to choose different estimation limits of p & q in ARIMA(*p,1,q*) model.

Through the Innovations algorithm, we performed several iterations to find a suitable time-series model that can be used to initialize the MLE operation. Based on the lags that show data correlation in the ACF & PACF, the Innovations algorithm was executed for $p \in [1-6], \quad q \in [1-8], \quad d = 1$. Table 6.6 indicates the top 10 combinations found within the specified ranges for p & q according to the AICC criterion (Hurvich and Tsai (1989)).

Table 6.6 Experiment: Preliminarily Estimation using Innovations algorithm for: $ARIMA(p,d,q) \quad p \in [1-6], \quad q \in [1-8], \quad d = 1$

p	d	q	AICC
2	1	3	-299.939
2	1	4	-300.631
3	1	4	-313.639
3	1	6	-271.808
3	1	7	-254.898
4	1	3	-298.141
4	1	4	-295.969
5	1	4	-294.000
5	1	6	-269.488
5	1	7	-271.571

According to experiment results indicated in Table 6.6, we found that $ARIMA(3,1,4)$ provides the lowest $AICC$ value of -313.639 which provide the most appropriate model via using the Innovations algorithm, Equation 6.1 shows the resulted mathematical model.

$$
\begin{aligned}
W(t) = &-0.8231W(t-1) - 0.6392W(t-2) + 0.08174W(t-3) \\
&+ Z(t) - 0.5577Z(t-1) - 0.05903Z(t-2) - 0.4778Z(t-3) \\
&+ 0.1236Z(t-4)
\end{aligned} \tag{6.1}
$$

The resulted model from the Innovations algorithm will be used to initialize the MLE operation to find more suitable model to represent our time-series. The used MLE estimation is highlighted in *Equation* 4.4 (Brockwell and Davis (2006)). Based on the MLE execution, we found better model that can better fit our time-series with lower AICC than the one found via the Innovations algorithm. Equation 6.2 indicates the resulted model ($ARIMA(3,1,4)$) via MLE, the calculated AICC for this model is -351.472. This would be the model that will be used in the data forecast experiments.

$$
\begin{aligned}
W(t) = &-0.3854W(t-1) - 0.3434W(t-2) + 0.4943W(t-3) \\
&+ Z(t) - 1.214Z(t-1) - 0.2312Z(t-2) - 0.7965Z(t-3) \\
&+ 0.7899Z(t-4)
\end{aligned} \tag{6.2}
$$

After we find the time-series model that fits our data, it is important to validate the model and ensure that it represents the examined data. As explained in Section 4.3.3, if the model is appropriate, then the forecast errors should represent a white noise $\{Z_t\}$, this can be examined through calculating the residuals which are the one-step prediction errors (Box *et al.* (2015)). The residuals calculation Equation is highlighted in Section 4.3.3.

The calculated ACF & PACF functions of our fitted model residuals are shown in Figures 6.6 & 6.7. It is clear that all ACF & PACF values are below the confidence bounds. This indicates that the calculated residuals have similar properties of the white noise process and indicates an

IID random data. Based on that, we conclude that the fitted time-series model is an appropriate representation of the examined data in this experiment.

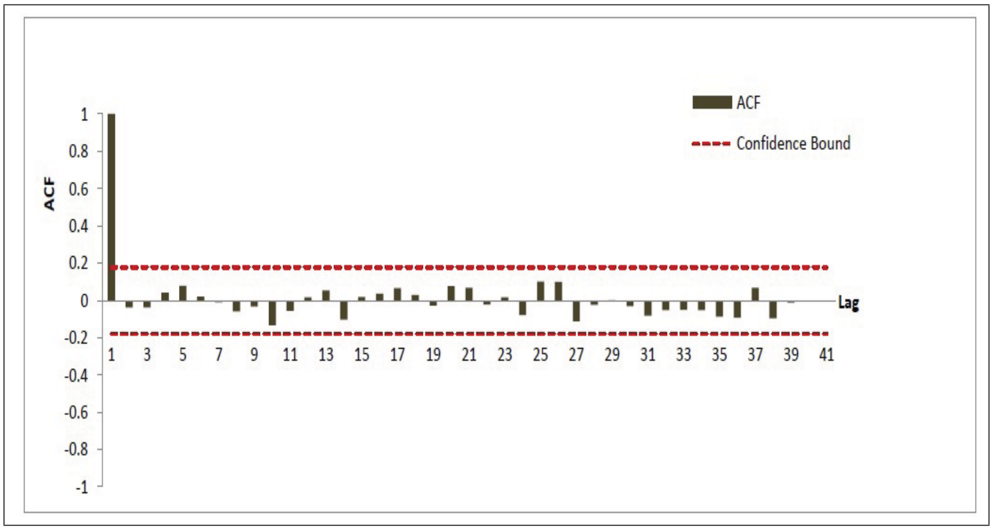

Figure 6.6 ACF of the fitted model residual

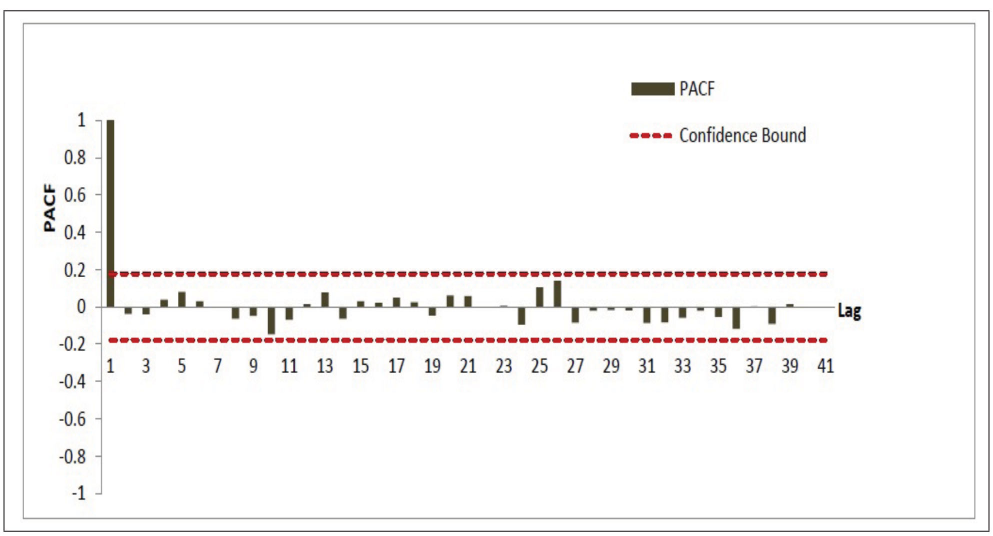

Figure 6.7 PACF of the fitted model residual

6.5 Data Forecast

Based on the determined time-series model in Section (6.4), the time-series predictor can be represented by Equation 6.3 where $m = max(p,q)$ (Brockwell and Davis (2006)); time-series predictor can be denoted by $P_n W_{n+h}$. The data forecast of the future value(s) $W_{n+h}, h > 0$ can be predicted based on the previous values of $Wn, \ldots, W1$ (Karapanagiotidis (2012)), with minimum Root Mean Square Error (RMSE) (Armstrong and Collopy (1992)) which will be used for forecast evaluation.

$$P_n X_{n+h} = \begin{cases} \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \widehat{X}_{n+h-j}), & 1 \leq h \leq m-n, \\ \sum_{i=1}^{p} \phi_i P_n X_{n+h-i} + \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \widehat{X}_{n+h-j}), & h > m-n, \end{cases} \tag{6.3}$$

Taken from Brockwell and Davis (2006)

In this experiment, the data prediction was executed on a random subset of 60 seconds of our time-series $\{W_n\}$ based on our time-series model and the Prediction Equation. The forecast was executed on different values of "h" at: $h = 5$, $h = 10$ & $h = 15$. The value "h" indicates the forecast steps measured by seconds. Once the forecast is executed on the time-series, the differencing and the box-cox transformation are being reverted on the time-series forecast results to get the original dataset forecast. Please note that the box-cox transformation and the differencing operations were applied on the original data earlier in Section 6.2 to transform the data into stationary time-series.

We demonstrate forecast results of the values $h = 5$, $h = 10$ & $h = 15$ in Figures 6.8, 6.9 & 6.10 respectively. After the forecast was performed with different forecast steps, we calculated the $RMSE$ error of the 3 experiments and the results were as follows: $h = 5$ gives the lowest $RMSE$ error at $RMSE = 2.669$ Mbps and the forecast experiment with $h = 10$ gives $RMSE$ error value at $RMSE = 3.173$ Mbps. The forecast experiment with $h = 15$ gives $RMSE = 3.184$ Mbps. From the $RMSE$ error calculations, we can see that the forecast experiment with $h = 5$ provides more accurate values than the other experiments but it would require more processing power considering the frequency of the forecast execution. The forecast with $h = 10$ or $h = 15$ would require less processing power. We see also that the $RMSE$ error calculations for the forecast experiments $h = 10$ & $h = 15$ are very close, so the forecast experiment with $h = 10$ does not really give much better results when it is compared to the forecast experiment with $h = 15$.

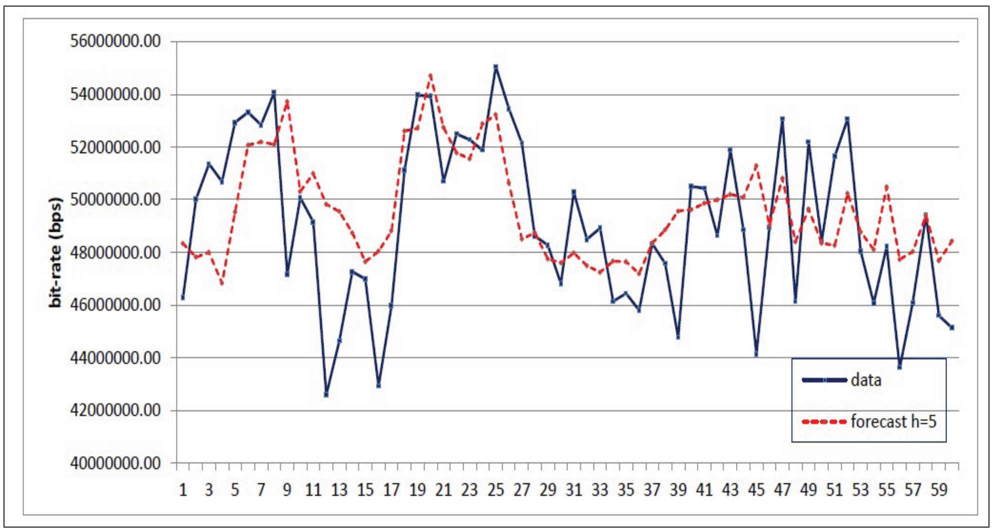

Figure 6.8 Forecast results with $h = 5$

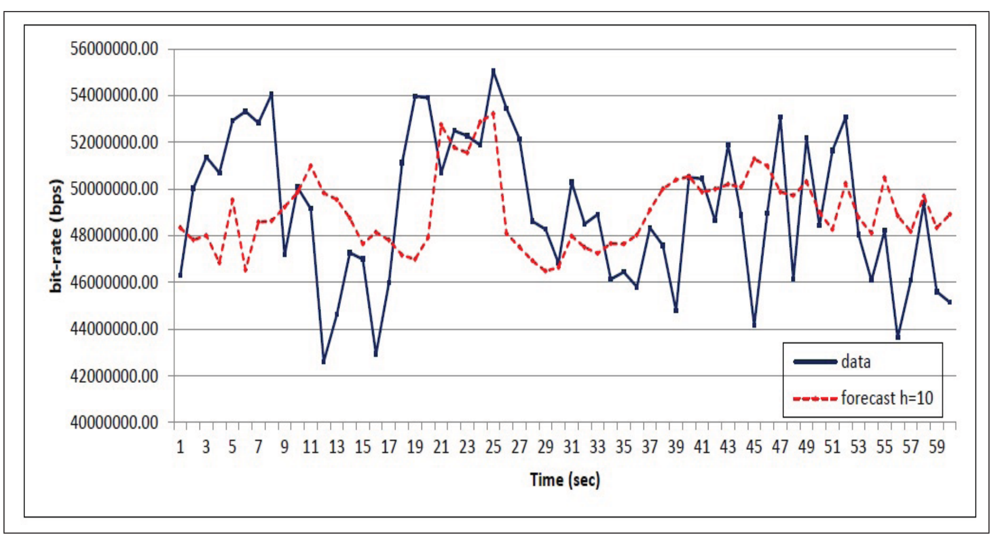

Figure 6.9 Forecast results with $h = 10$

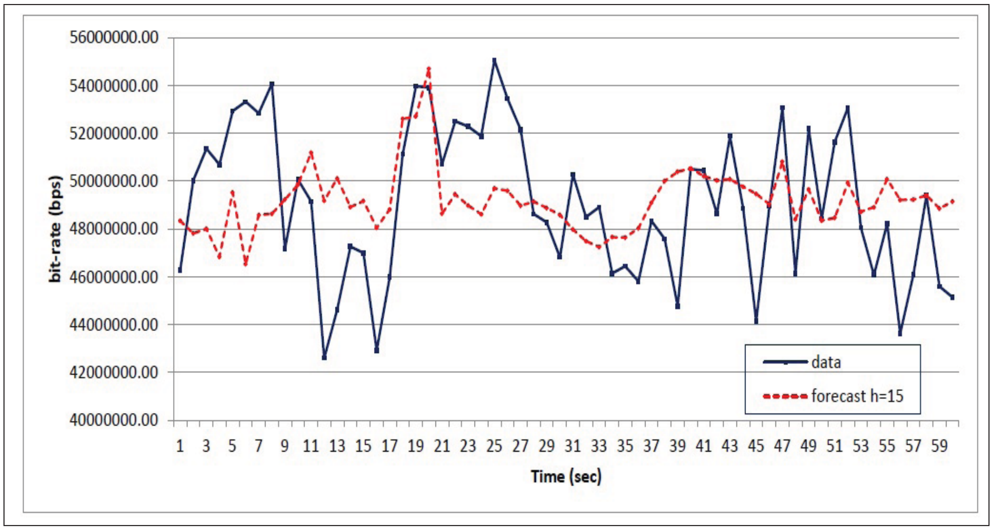

Figure 6.10   Forecast results with $h = 15$

6.6 Resource Utilization Enhancement & Simulation

6.6.1 Resource Utilization

In this experiment, we will consider the forecast with the value of $h = 5$. The resource utilization relies mainly on the forecasted unused GBR bit-rate (*GBR_unused_F*) series which can be calculated based on Equation 3.3 stated in Section 3.3, the *GBR_used_F* is the result from the data forecast experiment with $h = 5$ performed in the previous Section (6.5). Once the *GBR_unused_F* is calculated, it will be used to calculate the *bitrate_tobe_utilized* as shown in Equation 3.6 by deducting the *Safety_Threshold* limit from the forecasted *GBR_unused_F*, as stated in Section 3.3.

As explained previously in Equation 3.5, in Section 3.3, the *Safety_Threshold* is formulated according to the *RMSE_Behavior* which is calculated based on forecast experiment with $h = 5$. Figure 6.11 shows the *RMSE* error behavior which is used to formulate the *Safety_Threshold*. The *Safety_Threshold* is essential standby resources that can be used by the *contributing-bearers* in case they request back their guaranteed resources. The *bitrate_tobe_utilized* of our experiment is calculated according to Equation 3.6 and shown in Figure 6.12. The *bitrate_-*

tobe_utilized represents the guaranteed unused bit-rate that can be released and utilized for other running services (*acquiring-bearers*).

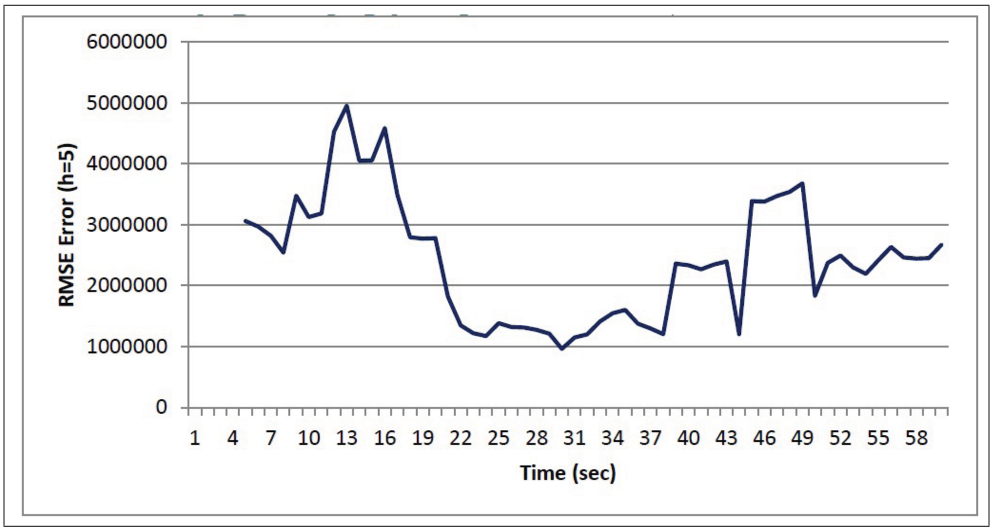

Figure 6.11 *RMSE* Error Behavior for $h = 5$

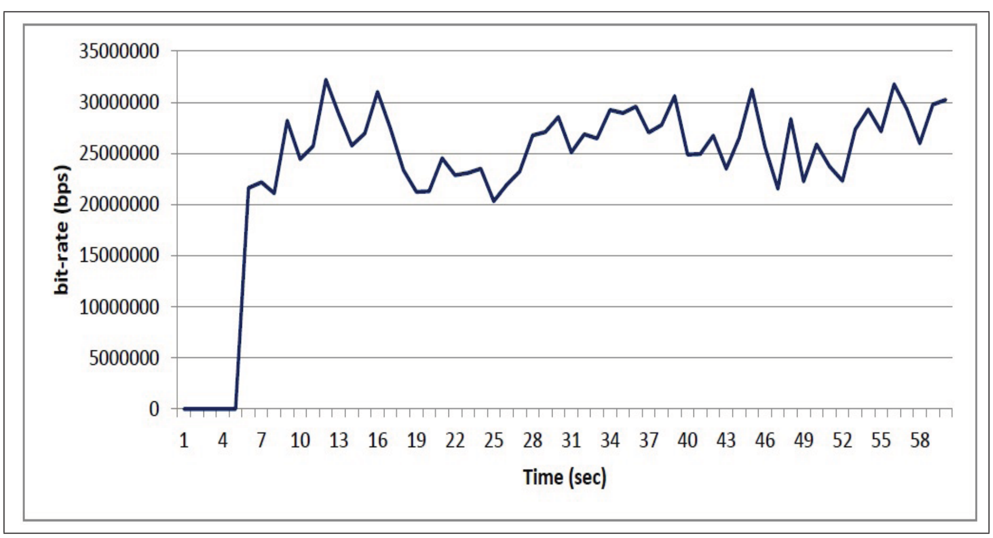

Figure 6.12 Guaranteed Unused Bit-rate that can be Utilized

The average bitrate (*bitrate_tobe_utilized*) that we can gain is "23.896 Mbps", based on the studied 60 seconds subset of the formulated dataset in this experiment. This represents about 30.63% of gained resources. Figure 6.13 shows the overall view of the *GBR_reserved* bit-rate, the *GBR_used* bit-rate, the *GBR_used_F* bit-rate and the *Safety_Threshold* limit which is added on top of the *GBR_used_F* forecasted values. The grey highlighted area in Figure 6.13 represents the bit-rate bandwidth that can be released and utilized by other services (*bitrate_-tobe_utilized*).

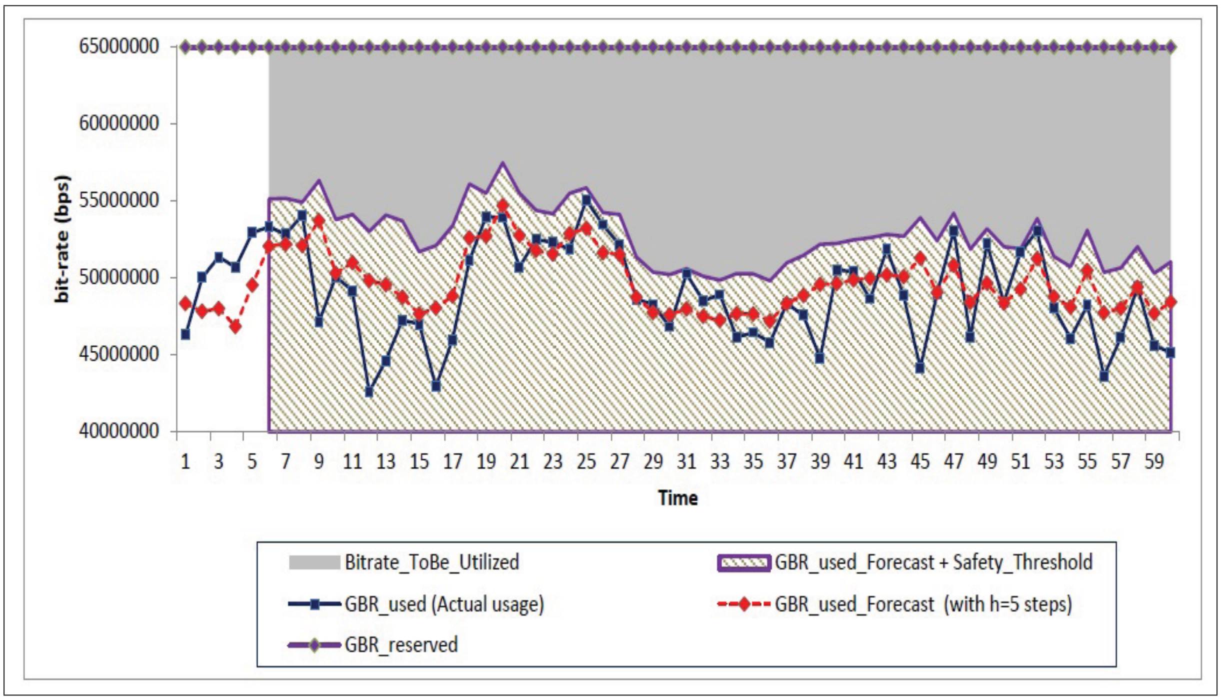

Figure 6.13 Overall view of bandwidth allocation

6.6.2 Simulation: Enhanced Resource Utilization

In the previous section, we considered the forecast experiment $h = 5$ to determine the *bitrate_-tobe_utilized* that represents the gain of our work indicated in Figure 6.12. We also determined the *Safety_Threshold* which would represent standby resources that can be used by the *contributing-bearers* in case they request back their guaranteed resources.

In this section, we will perform a simulation experiment using MATLAB to distribute the *bitrate_tobe_utilized* resources among a set of non-guaranteed bearers *acquiring-bearers*. Table 6.7 shows 40 non-guaranteed bearers which will be acquiring the *bitrate_tobe_utilized*. Those bearers are classified into 4 different sets according to the MBR which is assigned by the network; each set contains 10 non-guaranteed bearers. The "Avg used bit-rate" represents the average used bit-rate for each set. The "Potential to Consume" represents the potential for each set of bearers to consume the *bitrate_tobe_utilized* in this experiment, the "Potential to Consume" has been calculated according to the ratio of the difference between the *MBR* and the "Avg used bitrate" to the difference of the overall total between the *MBR* and the "Avg used bitrate" as explained in Equation 6.4. The "Potential to Consume" calculations are shown in Table 6.7 for each set.

$$potential\_to\_consume = (MBR - avg\_used\_bitrate) \div (total\_MBR - total\_avg\_used\_bitrate) \quad (6.4)$$

Table 6.7 Utilizing the unused bandwidth by non-GBR sessions

Set ID	**Number of non-GBR**	**MBR**	**Avg used bit-rate**	**Potential to Consume**
1	10	4.5 Mbps	3 Mbps	21.5%
2	10	3 Mbps	2 Mbps	14%
3	10	5 Mbps	3.5 Mbps	21.5%
4	10	7 Mbps	4 Mbps	43%
Total	40	19.5 Mbps	12.5 Mbps	100%

For the simulation environment, we assume that all the non-guaranteed bearers are running TCP-based application internet calls (www, e-mail, ftp, p2p file, etc.) and using QCI 9 according to (3GPP TS 23.203 (2015)). The simulation experiment was performed for 60 seconds based on the *bitrate_tobe_utilized* indicated in Figure 6.12. Each non-GBR set utilized part of the *bitrate_tobe_utilized* based on the calculated "Potential to Consume" percentage, it is important to mention that the *bitrate_tobe_utilized* was changing with time based on the forecast results.

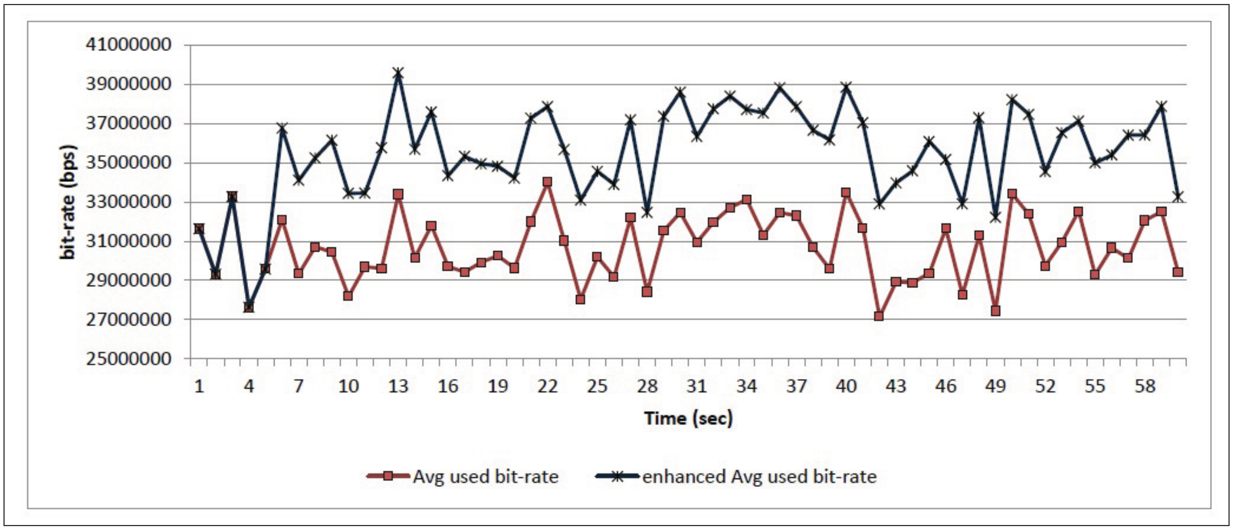

Figure 6.14 Resource Utilization for non-GBR bearers (set 1)

Figure 6.14 shows the total of the average used bit-rate -for the 10 non-GBR bearers combined (set 1)- without the enhanced resource utilization and also the average used bit-rate with the enhanced resource utilization being enabled. It is clear that we have an increase in the used bit-rate after enabling the enhanced resource utilization. The improvement in the average used bit-rate reached ”4.77” Mbps for all the bearers combined. Figure 6.15 also shows the the total of average used bit-rates for set 2 with and without the enhanced resource utilization. The improvement in the average used bit-rate reached ”3.24” Mbps for all the bearers combined in this set.

Figure 6.16 shows the results for set 3 which somehow has similar "Potential to Consume" as in set 1, the improvement in the total average used bit-rate with the enhanced resource utilization reached ”4.8” Mbps. By checking Figure 6.17, which represents the results for set 4, we can see that the improvement in the total average used bit-rate with the enhanced resource utilization reached ”9.93” Mbps. This represents the highest improvement with respect to all other sets. This can be justified by the high "Potential to Consume" percentage which was calculated at 43% for this set.

Based on this simulation experiment for enhanced resource utilization, we can see that our approach improved the average used bit-rate for those sets of non-GBR bearers. By adding up

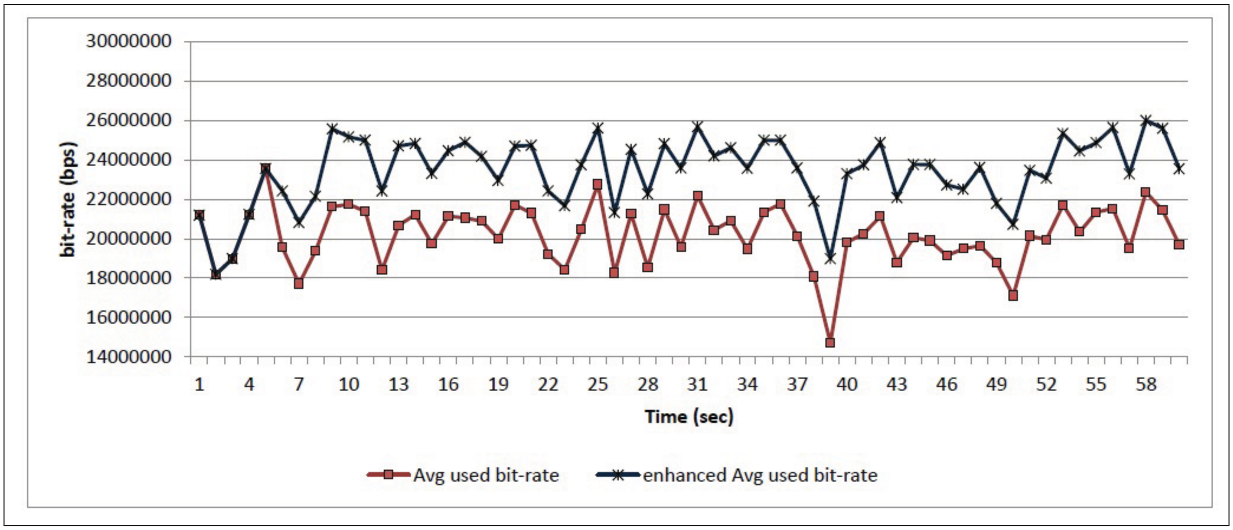

Figure 6.15   Resource Utilization for non-GBR bearers (set 2)

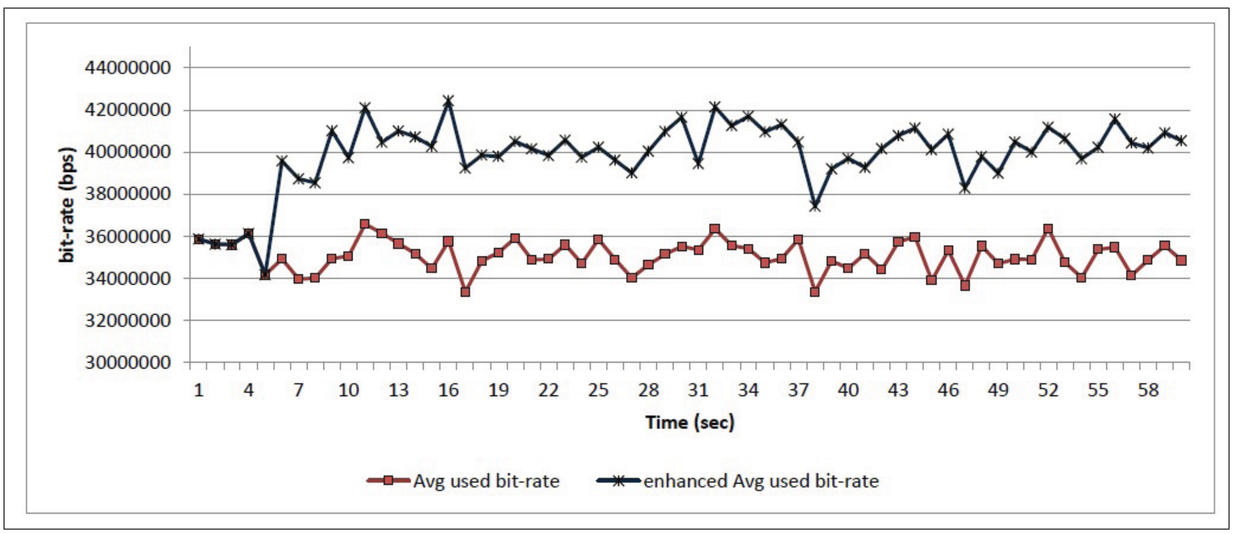

Figure 6.16   Resource Utilization for non-GBR bearers (set 3)

all the improvements of the average used bit-rates for all the 4 sets combined, the total will give ”22.74” Mbps. This value is very close to the *bitrate_tobe_utilized* value of "23.896" Mbps found previously in Section 6.6.1. The difference between those values can be justified by the willingness of the non-GBR bearers to consume resources.

In this chapter, our approach was applied to enhance the resource utilization for guaranteed Multi-Media LTE sessions. A real-life dataset was studied and analyzed, the dataset consists

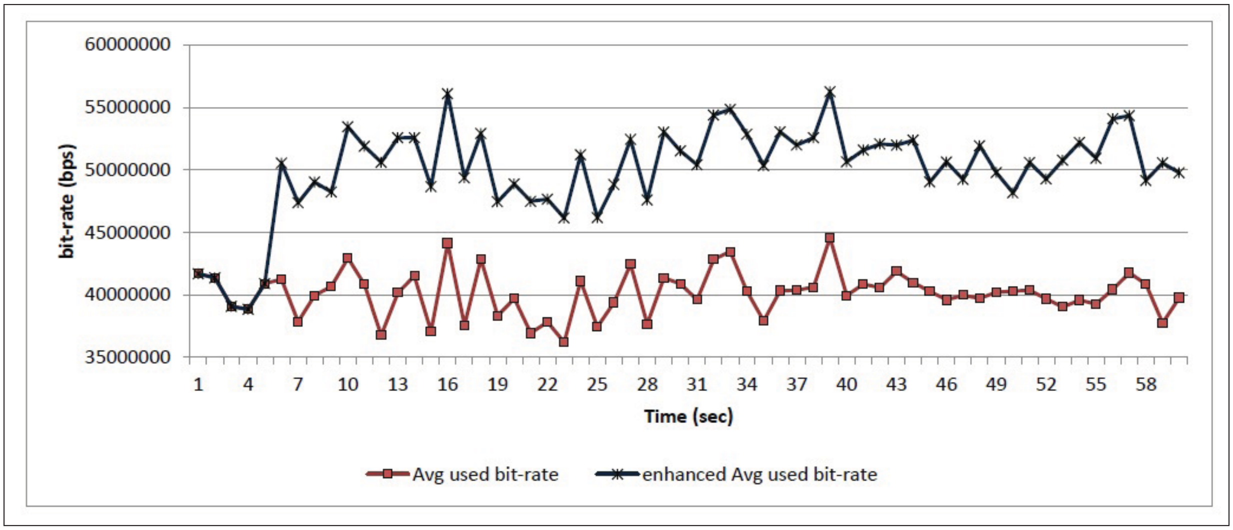

Figure 6.17 Resource Utilization for non-GBR bearers (set 4)

of several guaranteed bearers that carry different kinds of guaranteed services which reflect a real-life scenario. In addition, we were able to find the time-series model and perform the data forecast which was able to provide the unused resources. Resource utilization simulation experiment was executed where the unused/wasted resources were utilized by non-guaranteed bearers to improve the used bit-rate. This experiment demonstrated the benefit/gain that our approach would provide.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Conclusion

The LTE technology has been introduced to ensure that the needed infrastructure/technology is present to meet any requirements of the mobile services. To ensure high performance, throughput and scalability, efficient resource allocation and network optimization are very crucial in all 4G/LTE communication systems to avoid wasting the resources and ensure service availability and system efficiency.

In this thesis, we have focused on enhancing the resource utilization for LTE mobile services. Mainly, we designed and modeled an adaptive technique which improves the resource reservation for the LTE mobile guaranteed services and minimizes the wasted resources of the guaranteed bandwidth allocation in the LTE/EPC network. To achieve that, we introduced a novel technique that provides a smart, efficient and adaptive approach for the LTE bearers resource allocation. The new concept of adaptive guaranteed bearer provides an intermediate class between the strict guaranteed and the very open non-guaranteed resource allocation types as defined in 3GPP standards. Our approach was designed to consolidate the guaranteed traffic usage in one pool to estimate the waste of the resources and utilize them properly. Our technique ensures and guarantees the resource availability through designing the "Safety Model" which complements our technique.

To validate our contribution in this thesis, we proposed a framework which analyzes the mobile data traffic and determines the mathematical characteristics. As part of the framework, we utilized some methods that can help in converting the mobile data series into stationary data. In addition, we proposed an algorithm to mathematically represent the mobile data series into a time-series model. Different models and techniques were used to design the time-series, validate the model and perform comparison to get better data representations. Also, the proposed algorithm helped to forecast the consolidated mobile guaranteed resource consumption.

The forecast process helped in identifying the unused resources in the LTE/EPC system. The "Safety Model" was proposed as part of our approach to avoid any disturbance for the contributing guaranteed bearers. The model was established based on the forecast error provided by the time-series forecasting model. The "Safety Model" ensures the resource availability when the guaranteed bearers require their resources back.

Our experiments were conducted on several datasets that were captured on simulated LTE/EPC environment. In the first experiment, we applied our approach on a single guaranteed service that carries video conversational call where we analyzed the dataset, prepared it for modeling, found the time-series model that fits the data, validated the the model, performed the data forecast and estimated the wasted resources. Based on this experiment, we validated the feasibility of our approach through conducting a single video call. We showed that some improvements can be achieved in the guaranteed resource reservation mechanism. After that, we studied a bigger dataset which contains several guaranteed bearers that carry video conversational calls, we applied our approach on this dataset to validate the feasibility at bigger scale. In the time-series modeling, several experiments were conducted to find the model that better fits the data, data forecast was performed and our approach was able to provide and estimate the resources gain that can be used by other services.

In the last experiment, a more complex real-life dataset was analyzed and studied. The dataset consists of several guaranteed bearers that carry different kinds of guaranteed services which include video calling, live streaming, buffered streaming and real-time gaming sessions. The dataset reflects a more complex real-life scenario. Our approach was applied to find the time-series model. The data forecast was performed to predict the unused resources. Data simulation experiments were executed to show the benefit/gain that our approach would provide to improve system capabilities.

Finally, the experimental results showed that our approach is feasible and beneficial as it enhances the resource allocation for the LTE mobile services and increases the overall throughput of the LTE/EPC networks. It would also help to avoid network expansion at telecom operators that could be caused by scalability problem. Furthermore, our approach will affect telecom operators especially because our technique concentrates on practical improvement of telecom network resources utilization in particular the LTE mobile networks.

7.2 Future Work

Large scale LTE networks: we have evaluated our approach based on small to medium datasets. A possible future work is to evaluate our approach with larger datasets that could have a bigger variation of guaranteed sessions with different call durations. This will help to demonstrate the scalability of our approach on large-scale LTE networks.

Using Machine Learning: In this research, we have focused on time-series modeling to mathematically represent the data to perform the data forecast and prediction. A possible future work is to use Machine Learning techniques that can study the data and explore the construction of algorithms that can provide learning knowledge to make predictions on the data. These algorithms operate by building a model from example inputs to make data-driven prediction and forecasts. A comparison can be conducted to show the differences, outcome and capabilities of Machine Learning vs. time-series based on our approach for enhanced resource utilization for LTE networks.

APPENDIX I

LIST OF PUBLICATIONS

Journals:

Albasheir, S. and M. Kadoch. 2015. "Enhanced Control for Adaptive Resource Reservation of Guaranteed Services in LTE Networks". *Internet of Things Journal, IEEE,* vol. 99.

Albasheir, S. and M. Kadoch. 2016. "Evolved Bandwidth Allocation for LTE Multi-Media Mobile Services". *IEEE Journal of Communications and Networks*, IEEE, Submitted.

Conferences:

Albasheir , S. and M. Kadoch. 2014. "Stationary Transformation of Video Traffic in LTE Networks". In *Future Internet of Things and Cloud (FiCloud), 2014 International Conference on*. p. 351–355. IEEE.

BIBLIOGRAPHY

3GPP TS 23.203, G. T. S., 2015. *Policy and Charging Control Architecture*.

3GPP TS 23.401, G. T. S., 2015. *Evolved Universal Terrestrial Radio Access Network Access, GPRS Enhancements for EUTRAN Access*.

3GPP TS 23.882, G. T. S., 2008. *3GPP System Architecture Evolution: Report on Technical Options and Conclusions*.

3GPP TS 29.212, G. T. S., 2014. *Policy and Charging Control over Gx Reference Point, Provisioning and Policy Enforcement of Authorized QoS*.

3GPP TS 29.213, G. T. S., 2014. *Policy and Charging Control signalling flows and Quality of Service (QoS) parameter mapping*.

3GPP TS 36.300, G. T. S., 2015. *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN)*.

Akinaga, Y., S. Kaneda, N. Shinagawa, and A. Miura. 2005. "A proposal for a mobile communication traffic forecasting method using time-series analysis for multi-variate data". In *Global Telecommunications Conference, 2005. GLOBECOM'05. IEEE*. p. 6–pp. IEEE.

Albasheir, S. and M. Kadoch. 2014. "Stationary Transformation of Video Traffic in LTE Networks". In *Future Internet of Things and Cloud (FiCloud), 2014 International Conference on*. p. 351–355. IEEE.

Albasheir, S. and M. Kadoch. 2015. "Enhanced Control for Adaptive Resource Reservation of Guaranteed Services in LTE Networks". *Internet of Things Journal, IEEE*, vol. 99.

Alcatel-Lucent. 2009. "Introduction to Evolved Packet Core". *white paper. <https://www.alcatel-lucent.com/solutions/ip-mobile-core>*.

Alejandre, E. Z., C. G. Hernàndez, G. C. Dìaz, J. P. Díaz, and P. S. Terán. 2008. "Quality of service management efficient scheme for the universal mobile telecommunications system". *Ingeniería. Investigación y Tecnología*, vol. 9, n° 2, p. 113–129.

Anas, M., C. Rosa, F. Calabrese, K. Pedersen, and P. Mogensen. sept. 2008. "Combined Admission Control and Scheduling for QoS Differentiation in LTE Uplink". In *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*. p. 1 -5.

Armitage, G. J. 2003. "Revisiting IP QoS: why do we care, what have we learned? ACM SIGCOMM 2003 RIPQOS workshop report". *SIGCOMM Comput. Commun. Rev.*, vol. 33, n° 5, p. 81–88.

Armstrong, J. S. and F. Collopy. 1992. "Error measures for generalizing about forecasting methods: Empirical comparisons". *International journal of forecasting*, vol. 8, n° 1, p. 69–80.

Balbas, J.-J., S. Rommer, and J. Stenfelt. february 2009. "Policy and charging control in the evolved packet system". *Communications Magazine, IEEE*, vol. 47, n° 2, p. 68 -74.

Basu, S., A. Mukherjee, and S. Klivansky. 1996. "Time series models for internet traffic". In *INFOCOM'96. Fifteenth Annual Joint Conference of the IEEE Computer Societies. Networking the Next Generation. Proceedings IEEE.* p. 611–620. IEEE.

Bell, G. 2003. "Failure to thrive: QoS and the culture of operational networking". In *Proceedings of the ACM SIGCOMM workshop on Revisiting IP QoS: What have we learned, why do we care?* (New York, NY, USA 2003), p. 115–120. ACM.

Belshe, M. and R. Peon. 2012. "SPDY protocol". *Network Working Group.* *<https://tools.ietf.org/html/draft-mbelshe-httpbis-spdy-00>*.

Blum, C. and A. Roli. 2008. "Hybrid metaheuristics: an introduction". *Hybrid Metaheuristics*, p. 1–30.

Blum, C., J. Puchinger, G. Raidl, A. Roli, et al. 2010. "A brief survey on hybrid metaheuristics". *Proceedings of BIOMA*, p. 3–18.

Box, G. E. and D. R. Cox. 1964. "An analysis of transformations". *Journal of the Royal Statistical Society. Series B (Methodological)*, p. 211–252.

Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, 2015. *Time series analysis: forecasting and control*. John Wiley & Sons, ed. Fifth Edition.

Brockwell, P. J. and R. A. Davis, 2006. *Introduction to time series and forecasting*. Springer Science & Business Media, ed. Second Edition.

Brockwell, P. and R. Davis. 1988. "Applications of innovation representations in time series analysis". *Probability and Statistics, Essays in Honor of Franklin A. Graybill, J.N.*, p. 61–84.

Chandra, I. and K. Helenprabha. 2014. "Efficient Resource Allocation with QoS Guarantee on LTE-A Downlink Network Systems". *Journal of Networks*, vol. 9, n° 9, p. 2365–2370.

Chawre, A. 2010. *nwEPC - EPC SAE Gateway.* <http://sourceforge.net/projects/nwepc/>. Logiciel.

Dai, J. and J. Li. 2009. "VBR MPEG Video Traffic Dynamic Prediction Based on the Modeling and Forecast of Time Series". In *INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on.* p. 1752–1757. IEEE.

Davis, J. L., K. Chandra, and C. Thompson. 2000. "Nonlinear time-series model for VBR video traffic". In *Proceedings of the 25th Annual IEEE Conference on Local Computer Networks*. (Washington, DC, USA 2000), p. 678–. IEEE Computer Society.

De Gooijer, J. G. and R. J. Hyndman. 2006. "25 years of time series forecasting". *International Journal of Forecasting*, vol. 22, n° 3, p. 443–473.

Ekstrom, H. february 2009. "QoS control in the 3GPP evolved packet system". *Communications Magazine, IEEE*, vol. 47, n° 2, p. 76 -83.

Ericsson. 2008. "Long term evolution (LTE): an introduction". *White paper.* <*http://www.ericsson.com/news/lte-introduction-254740099*>.

Ericsson. 2013. "Ericsson Media Report: ON THE PULSE OF THE NETWORKED SOCIETY". <*https://www.ericsson.com/res/docs/2013/ericsson-mobility-report-june-2013.pdf*>.

Ericsson. 2015. "Ericsson Mobility Report: ON THE PULSE OF THE NETWORKED SOCIETY". <*http://www.ericsson.com/res/docs/2015/mobility-report/ericsson-mobility-report-nov-2015.pdf*>.

He, F., J. Li, and Y. You. 2001. "Design and implement of network traffic forecast system". In *Info-tech and Info-net, 2001. Proceedings. ICII 2001 - Beijing. 2001 International Conferences on*. p. 164 -169.

Hurvich, C. and C. Tsai. 1989. "Regression and time series model selection in small samples". *Biometrika Journal*, vol. 76, p. 297–307.

Hyndman, R., A. Koehler, J. Ord, and R. Snyder, 2008. *Forecasting with Exponential Smoothing: The State Space Approach*. Springer Series in Statistics.

Kalle, R., A. Nandan, and D. Das. 2012. "La VoLTE: Novel Cross Layer Optimized Mechanism of Video Transmission over LTE for DRX". In *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*. p. 1–5. IEEE.

Karapanagiotidis, P. 2012. "Literature review of modern time series forecasting methods". <*http://individual.utoronto.ca/paulkara/S2012%20Literature%20Review%20-%20linear%20survey.July31.pdf*>.

Lee, J. and M. C. Strazicich. 2002. "ITSM 2000 Professional Version 6.0, developed by Peter J. Brockwell and Richard A. Davis, B&D Enterprises, Inc., Copyright 1999. The Student Version is included in Introduction to Time Series and Forecasting, 1996, Springer-Verlag New York Inc.(ISBN: 0387947191). The Professional Version is obtainable from pbrockwell@ compuserve. com. Web Page of the author: http://www. stat. colostate. edu/ pjbrock". *International Journal of Forecasting*, vol. 18, n° 3, p. 455–460.

Li, L. and S. Shen. sept. 2011. "End-to-End QoS performance management across LTE networks". In *Network Operations and Management Symposium (APNOMS), 2011 13th Asia-Pacific*. p. 1 -4.

Liu, H. and G. Mao. 2005. "Prediction algorithms for real-time variable-bit-rate video". In *Communications, 2005 Asia-Pacific Conference on*. p. 664–668. IEEE.

Luo, S., X. Huang, L. Hu, and Y. Ma. oct. 2010. "A secured flow label based QoS scheme for the Evolved Packet Core in the Evolved Packet System". In *Broadband Network and Multimedia Technology (IC-BNMT), 2010 3rd IEEE International Conference on*. p. 289 -293.

Mitra, R. N. and D. P. Agrawal. 2016. "5G mobile technology: A survey". *ICT Express*.

Moghram, I. and S. Rahman. 1989. "Analysis and evaluation of five short-term load forecasting techniques". *IEEE Transactions on Power Systems*, vol. 4, n° 4, p. 1484–1491.

Molazem Tabrizi, F., J. Peters, and M. Hefeeda. 2011. "Adaptive transmission of variable-bit-rate video streams to mobile devices". *NETWORKING 2011*, p. 213–224.

Nomura, M., T. Fujii, and N. Ohta. 1989. "Basic characteristics of variable rate video coding in ATM environment". *Selected Areas in Communications, IEEE Journal on*, vol. 7, n° 5, p. 752–760.

Poh, E. and H. Ewe. july 2005. "IPv6 packet classification based on flow label, source and destination addresses". In *Information Technology and Applications, 2005. ICITA 2005. Third International Conference on*. p. 659 - 664 vol.2.

Ren, Z., X. Qi, G. Zhou, H. Wang, and D. Nguyen. "Throughput Assurance for Multiple Body Sensor Networks".

Ren, Z., X. Qi, G. Zhou, and H. Wang. 2014. "Exploiting the Data Sensitivity of Neurometric Fidelity for Optimizing EEG Sensing".

Rivett, P., 1968. *An introduction to operations research*. Science & discovery.

Robert H. Shumway, D. S. S., 2010. *Time Series Analysis and Its Applications: With R Examples*.

Rodriguez, J., 2015. *Fundamentals of 5G Mobile Networks*. John Wiley & Sons.

Roskind, J. 2013. "QUIC: Multiplexed stream transport over UDP". *Google working design document*.

Seeling, P., M. Reisslein, and B. Kulapala. 2004. "Network performance evaluation using frame size and quality traces of single-layer and two-layer video: A tutorial". *Communications Surveys & Tutorials, IEEE*, vol. 6, n° 3, p. 58–78.

Shin, J., K. Jung, and A. Park. sept. 2008a. "Design of Session and Bearer Control Signaling in 3GPP LTE System". In *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*. p. 1 -5.

Shin, J., K. Jung, and A. Park. 2008b. "Design of Session and Bearer Control Signaling in 3GPP LTE System". In *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*. p. 1–5. IEEE.

Sivakumar, R., E. Ashok Kumar, and G. Sivaradje. 2011. "Prediction of Traffic Load in Wireless Network Using Time Series Model". In *Process Automation, Control and Computing (PACC), 2011 International Conference on*. p. 1–6. IEEE.

Svoboda, P., M. Buerger, and M. Rupp. 2008. "Forecasting of traffic load in a live 3G packet switched core network". In *Communication Systems, Networks and Digital Signal Processing, 2008. CNSDSP 2008. 6th International Symposium on*. p. 433–437. IEEE.

Verizon. 2010. "LTE: The Future of Mobile Broadband Technology". *White paper.* <*http://innovation.verizon.com*>.

Wang, H., C. Huang, and J. Yan. 2006. "Efficient multiple-link adaptive bandwidth provisioning for end-to-end quality of service guarantee". In *Electrical and Computer Engineering, 2006. CCECE'06. Canadian Conference on*. p. 127–131. IEEE.

Xu, W. and A. Qureshi. 1999. "Adaptive linear prediction of MPEG video traffic". In *Signal Processing and Its Applications, 1999. ISSPA'99. Proceedings of the Fifth International Symposium on*. p. 67–70. IEEE.

Yin, S. and X. Lin. 2005. "Adaptive load balancing in mobile ad hoc networks". In *Wireless Communications and Networking Conference, 2005 IEEE*. p. 1982–1987. IEEE.

You, C. and K. Chandra. 1999. "Time series models for internet data traffic". In *Local Computer Networks, 1999. LCN'99. Conference on*. p. 164–171. IEEE.

Youjun, G., T. Hui, Z. Ping, et al. 2006. "A QoS-Guaranteed Adaptive Resource Allocation Algorithm with Low Complexity in OFDMA System". In *Wireless Communications, Networking and Mobile Computing, 2006. WiCOM 2006. International Conference on*. p. 1–4. IEEE.