University of Mississippi

eGrove

Electronic Theses and Dissertations | Graduate School

2014

Cross-Layer Capacity Optimization In Ofdma Systems: Wimax And Lte

Bimal Paudel
University of Mississippi

Follow this and additional works at: https://egrove.olemiss.edu/etd

 Part of the Electrical and Electronics Commons

Recommended Citation

Paudel, Bimal, "Cross-Layer Capacity Optimization In Ofdma Systems: Wimax And Lte" (2014). *Electronic Theses and Dissertations*. 971.
https://egrove.olemiss.edu/etd/971

This Dissertation is brought to you for free and open access by the Graduate School at eGrove. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of eGrove. For more information, please contact egrove@olemiss.edu.

Cross-Layer Capacity Optimization in OFDMA Systems: WiMAX and LTE

A Thesis
presented in partial fulfillment of requirements
for the degree of Master of Science in Electrical Eningeering
in the Department of Electrical Engineering
The University of Mississippi

Bimal Paudel

February 28, 2014

Copyright © 2014 by Bimal Paudel
All rights reserved

ABSTRACT

Given the broad range of applications supported, high data rate required and low latency promised; dynamic radio resource management is becoming vital for newly emerging air interface technologies such as Wireless interoperability for Microwave Access (WiMAX) and Long Term Evolution (LTE) adopted by international standards. This thesis considers Orthogonal Frequency Division Multiple Access (OFDMA) system, which has been implemented in both WiMAX and LTE technologies as their air interface multiple access mechanism. A framework for optimized resource allocation with Quality of Service (QoS) support that aims to balance between service provider's revenue and subscriber's satisfaction is proposed. A cross-layer optimization design for subchannel, for WiMAX, and Physical Resource Block (PRB), for LTE, and power allocations with the objective of maximizing the capacity (in bits/symbol/Hz) subject to fairness parameters and QoS requirements as constraints is presented. An Adaptive Modulation and Coding (AMC)-based cross-layer scheme has also been proposed in this thesis by adopting an AMC scheme together with the cross-layer scheme to realize a more practical and viable resource allocation. The optimization does not only consider users channel conditions but also queue status of each user as well as different QoS requirements. In the proposed framework, the problem of power allocation is solved analytically while the subchannel/PRB allocation is solved using integer programming exhaustive search. The simulation and numerical results obtained in this thesis have shown improved system performance as compared to other optimization schemes known in literature.

ACKNOWLEDGEMENTS

I express my utmost and sincere gratitude to my supervisor, Prof. Dr. Mustafa M. Matalgah, for his constant motivation, valuable guidance and support throughout this dissertation work. He always has been a source of inspiration and I value his dynamic concepts and a profound insight into the research topic without which this research would not have been a success.

I take immense pleasure in extending my thankfulness to Dr. Lei Cao for his insightful lectures and concepts. His continuous support and guidance throughout my graduate studies were of outmost value. My earnest and sincere thanks goes to Dr. Ramanarayanan "Vish" Viswanathan for his ardent support and trust in me as a department chair throughout my stay in Olemiss.

I am very much indebted to my parents and my beloved wife Jigyasha for their constant encouragement, love, inspiration and moral support throughout this work.

University, Mississippi Bimal Paudel
March 2014

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MF Maximum Fairness

3GPP Third-Generation Partnership Project

AMC Adaptive Modulation and Coding

ARP Allocation and Retention Priority

ARQ Automatic Repeat reQuest

ASN-GW ASN gateway

AWGN Additive White Gaussian Noise

BAMC Band Adaptive Modulation and Coding

BE Best Effort

BER Bit Error Rate

BLER Block Error Rate

BS Base Station

BWA Broadband Wireless Access

CLWRC Cross-layer Weighted Rate Constraint

CP Cyclic Prefix

CSI Channel State Information

DC Direct Current

eNodeB enhanced NodeB

EPS Evolved Packet System

ErtPS Extended real-time Polling Service

FDD Frequency Division Duplexing

FEC Forward Error Correction

FFT fast Fourier transform

FIFO First-in First-out

FTP File Transfer Protocol

GBR Guaranteed Bit Rate

HTTP Hyper-Text Transfer Protocol

ISI inter-symbol interference

LTE Long Term Evolution

LTE-A LTE-Advanced

MAC Media Access Control

MF Maximum Fairness

MPEG Motion Picture Experts Group

NGBR Non GBR

nrtPS nonreal-time packet service

OFDM Orthogonal Frequency Division Multiplexing

OFDMA Orthogonal Frequency Division Multiple Access

OSI Open Systems Interconnect

P2P Peer to Peer

PAPR Peak-to-Average-Power Ratio

PDF probability density function

PDN-GW Packet Data Network - Gateway

PHY Physical

PRB Physical Resource Block

PRC Proportional Rate Constraint

PSS Primary Synchronization Sequence

PUSC Partial Usage of Subcarriers

QCI QoS Class Identifier

QoE Quality of Experience

QoS Quality of Service

QAM Quadrature Amplitude Modulation

rtPS real-time Polling Service

SDF Service Data Flow

SER Symbol Error Rate

SF Service Flow

SNR signal-to-noise ratio

S/P Serial to Parallel

TDD Time Division Duplexing

TTI Transmission Time Interval

UE User Equipment

UGS Unsolicited Grant Services

VBR Variable Bit Rate

VoIP Voice over Internet Protocol

WiMAX Worldwide interoperability for Microwave Access

CHAPTER 1

INTRODUCTION

Air interface standards such as Worldwide Interoperability for Microwave Access (WiMAX) and Long Term Evolution (LTE) were introduced for the realization of a successful broadband wireless access (BWA) solution. They aim to support mobility for BWA with the capability of delivering high data rates over long ranges. For the proper support of real-time, multimedia, and bandwidth demanding applications, the standards like IEEE 802.16e for WiMAX technology and 3rd Generation Partnership Project (3GPP) for LTE technology, provides quality of service (QoS) support with scheduling services at the media access control (MAC) layer. Standards in general suggests the main principles in designing the QoS architecture and signaling framework. Five different service flows (SFs) including unsolicited grant services (UGS), real-time packet service(rtPS), extended real-time packet service (ErtPS), non real-time packet service (nrtPS), and best effort (BE) are defined in IEEE 802.16e standard. While nine different QoS class identifier that associates with them a specific QoS parameters and that belongs to the two broad resource types namely: Guaranteed Bit Rate (GBR) and Non GBR (NGBR) are defined in the 3GPP standard. Similarly, standards also suggest the use of multiple access mechanism such that the system is accessible to a number of users to share the available system resources. Orthogonal Frequency Division Multiple Access (OFDMA) scheme has been adopted in both 4G WiMAX and LTE air interface technologies as their multiple-access mechanism. Multiple access is achieved in

OFDMA by assigning subsets of subcarriers and time slots to individual users. The subcarriers allow simultaneous low data rate transmission from several users. The subsets of subcarriers considered in frequency domain are referred to as subchannels in WiMAX and physical resource blocks (PRBs) in LTE. Based on feedback information about the channel conditions, adaptive user-to-subcarrier assignment can be achieved.

Resource allocation is concerned with the proportional allocation of resources such that a system utility is optimized. Allocating resources to users based on the channel conditions of each user is one of the approaches of resource allocation or system capacity optimization. There are many non-cross-layer resource allocation algorithms found in literature and they operate only on the physical (PHY) layer. In these algorithms, basically two variables namely: total transmit power and overall system capacity are concerned. So, these resource allocation algorithms either maximize the overall system capacity while having constraint on the maximum total transmit power or minimize the total transmit power while having constraint on the minimum overall system capacity. The former approach is suitable for the bursty applications supported by the standards like WiMAX and LTE. Some of the major techniques with such an approach of maximizing the capacity while having constraints on total transmit power are, maximum throughput approach, maximum fairness approach, proportional rate approach and proportional fairness approach. In maximum throughput approach, the objective is to maximize the summation of rates from all users, while having a constraint on the total transmit-power [2]. Although this approach maximizes the system throughput, it does not establish fairness among users. The maximum fairness approach [3] focuses on achieving fairness among users while maximizing the minimum data rate among them. Therefore, the maximum fairness is called a max-min problem - maximizing the minimum data rate. Users with weak channels are deprived of resources in the maximum system throughput approach while they consume most of the resources in the maximum fairness approach. Moreover, the maximum fairness approach is not flexible enough to support services with varying QoS requirements. So the proportional rate constraints approach comes as a

modification to maximum fairness approach, where the total throughput is maximized while maintaining some sort of proportionality among users data rates [4], [5]. When the users in motion are considered, the resource allocation objective of maximizing throughput while maintaining fairness can be achieved on the long term. Being near to the base station (BS), a user will have strong channel while other users being faraway will have weak channels. But these locations keep changing for a moving user so the proportional fairness approach introduces a third dimension, latency, for optimization besides total transmit-power and total system throughput. Simply put, this approach aims to maximize the total throughput and minimize the total latency while maintaining fairness among users [6], [7].

In addition, the other approach of capacity optimization is the cross-layer approach where various cross-layer information are considered along with the channel conditions. Cross-layer design has been extensively used to achieve multiuser diversity gain. This gain is achieved due to channel-state-dependent scheduling where channel state information at the PHY layer are passed on to the packet scheduler at the MAC layer [6], [8]. A simple illustration on the multiuser diversity gain can be found in [9] and a detailed study on the packet scheduling for QoS support in IEEE 802.16 broadband wireless access system is presented in [10].

1.1 Motivation and Contribution

Radio resource management involves mechanisms by which the system controls operations such as packet scheduling, admission control, subcarrier allocation, subchannel/PRB assignment, power allocation, modulation order, and rate control. The ultimate goal is to efficiently utilize the network resources and the scarcely available radio spectrum while keeping a good grade of services. A significant improvement in the performance of the wireless network can be realized by wisely adopting the cross-layer design approach for optimizing resource allocations in order to maximize a given system utility [9]. Cross-layer design refers to protocol design done by actively exploiting the dependence between protocol layers to obtain

certain performance gains. In other words, protocols can be designed by violating the open systems interconnect (OSI) architecture, one way of achieving such design is by allowing direct communication between protocols at non-adjacent layers or sharing variables between layers. Furthermore, resource allocation for a practical system is achieved by adapting adaptive modulation and coding (AMC). The idea of taking advantage of channel fluctuation to improve the system performance was what instigated the concept of AMC, which is based on the fact that transmitting high data rate during good channel conditions and low data rate during poor channel conditions achieves a significant performance gain. The use of small constellation (modulation type) and low rate error correcting codes (coding) results in a low data rate while using higher constellation and higher rate results in a high data rate. This adaptive selection of a set of modulation and coding set depending on the channel condition can hence achieve the highest data rate while maintaining the bit error rate (BER) requirement of the system.

This thesis presents an extension of the work presented in [11], where a novel cross-layer resource allocation scheme for WiMAX has been proposed. The cross-layer fairness parameters, service urgency and service satisfaction based on the queue status and QoS requirements of users, respectively that added a new dimension to the fairness concept introduced in [11] are considered in this thesis. The thesis then presents the resource allocation scheme for LTE system based on these fairness parameters modified accordingly as applied to LTE systems. The thesis also presents an AMC-based cross-layer resource allocation based on the similar concept as in [11] implementing the AMC. Depending on the diverse QoS requirements of different users, resources can be allocated wisely; users that are well served and have no critical QoS requirements to schedule for service immediately can lag for some time allowing underserved users to access the channel. An optimization of the system performance subject to the constraints on power and cross-layer fairness parameters are studied in this work as well. The significant improvement in the performance of the system in terms of maximization of system capacity achieved with the implementation of the proposed scheme is confirmed

by the extensive simulation results.

1.2 Organization of the thesis

The remainder of the thesis is organized as follows. Chapter 2 presents a detailed review on the related works that form the foundation of this thesis and also presents an extension to the one of the work. In Chapter 3, a cross-layer resource allocation scheme as applied to LTE system, LTE-CLWRC, is proposed. In addition, Chapter 4 presents an AMC-based cross-layer resource allocation scheme, AMC-CLWRC is proposed. Finally some conclusions are drawn in Chapter 5

CHAPTER 2

TECHNOLOGIES OVERVIEW AND STATE-OF-THE-ART

2.1 OFDMA Overview

Orthogonal Frequency Division Multiplexing (OFDM) is one of the common multicarrier modulation technique. Efficient and flexible management of intersymbol interference (ISI) is the main reason for the popularity of OFDM for high data rate applications [12]. A wideband signal of bandwidth B is broken down into N narrowband signals (subcarriers) each of bandwidth B/N. An OFDM with multiple access is known as OFDMA. In general, there are three types of subcarriers in an OFDMA symbol as shown in Fig. 2.1 and are listed as follows.

- Data subcarriers: carry data symbols
- Pilot subcarriers: carry pilot symbols which can be used for channel estimation and channel tracking
- Null subcarriers: have no power allocated to them and includes the DC subcarriers and the guard subcarriers towards the edges. The DC subcarrier is used by the user equipment (UE) to track the center of the OFDMA frequency band and appears only once in the spectrum.

An OFDMA downlink transmitter is depicted in Fig. 2.2 [12]. Since a downlink transmitter forward the signal from multiple user using a single transmitter to a receiver, we

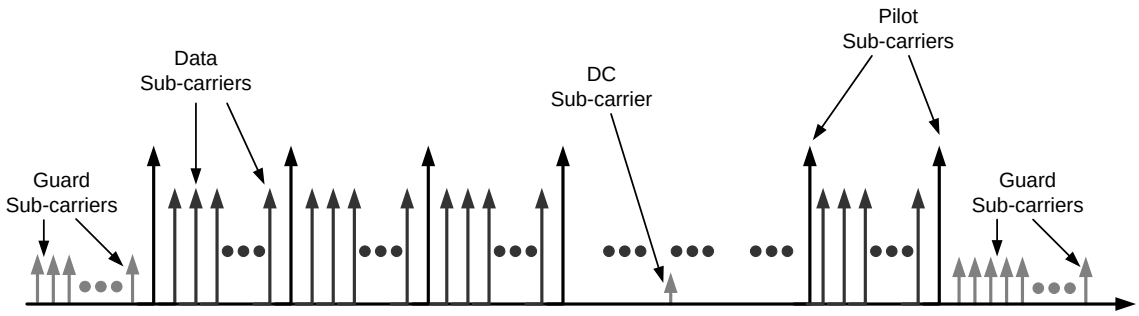

Figure 2.1. OFDMA Symbol structure

have K number of symbol mapping, serial-to-parallel (S/P) converter and subcarrier mapping blocks. Once we have subcarrier mapping for individual users then N point IFFT is performed. The IFFT operation at the transmitter allows all the subcarriers to be created in the digital domain, such that only a single radio can be used rather than multiple radios corresponding to each subcarriers. In order for the IFFT/FFT to create an ISI-free channel, the channel must appear to provide a circular convolution, hence cyclic prefix (CP) are append to the output of IFFT block. The output of the CP block is then parallel to serial converted to form a single multiuser OFDMA symbol. This multiuser OFDMA symbol is then transmitted over channel on to the receiver. An OFDMA downlink receiver is depicted in Fig. 2.3 shows . The multiuser OFDMA symbol is received by the receiver corresponding to each individual mobile device. Once the multiuser OFDMA symbol is received, it is S/P converted and then CP are removed. N point fast fourier transform (FFT) is then performed followed by frequency equalization and subcarrier demapping, such that the user receives only the subcarrier signal that belong to it. Symbol demapping is then performed finally retrieving the original k^{th} users data bits.

The remaining of the chapter is organized as follows. Section 2.2 gives the basics on WiMAX technology while Section 2.3 gives the basics on LTE technology. Similarly, some overview on AMC is presented in Section 2.4. Section 2.5 presents a detail on the state-of-the-art researches on WiMAX, LTE and AMC. This section also includes a detail on the CLWRC algorithm. Results corresponding to CLWRC algorithm is presented in Section 2.6.

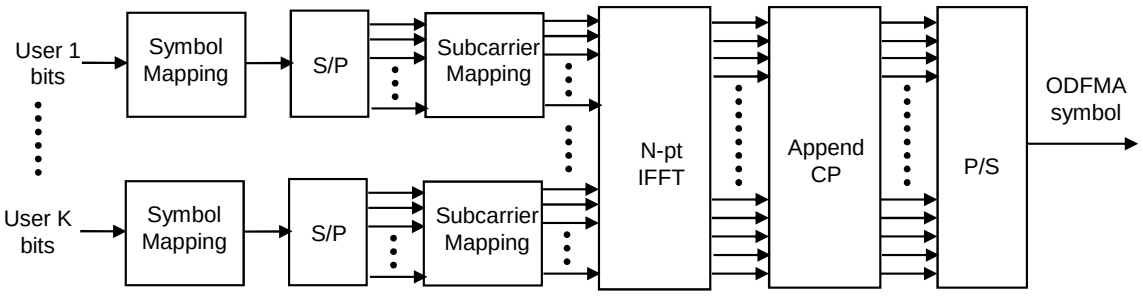

Figure 2.2. OFDMA downlink transmitter

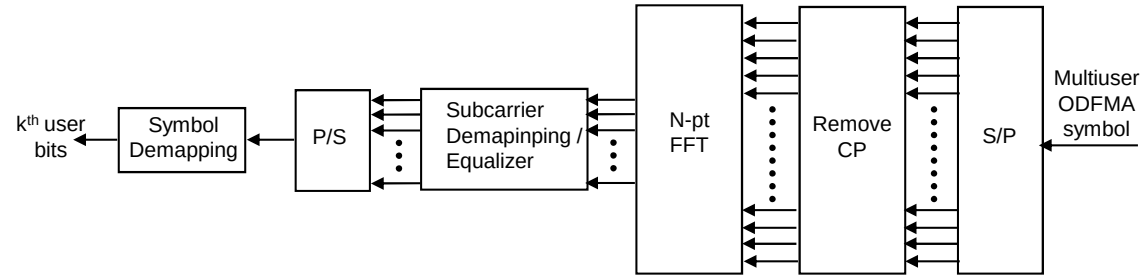

Figure 2.3. OFDMA downlink receiver

Extension to the CLWRC algorithm is presented in Section 2.7 and finally some conclusions are drawn in Section 2.8.

2.2 WiMAX Technology

WiMAX is an air interface technology based on IEEE 802.16e-2005 standard and provides a solution for delivering broadband wireless services. WiMAX uses OFDMA as its multiple access mechanism. Subcarriers are the smallest granular units in the frequency domain and OFDM symbol duration is the smallest granular unit in the time domain in OFDMA system. Since the number of subcarriers are too large in the system, subsets of subcarriers are considered together in an OFDM symbol [13]. Mapping subcarriers to a particular subset can be referred to as resource mapping. In WiMAX subsets of subcarriers is known as subchannels and there exists two subchannelization methodologies: partial usage of subcarriers (PUSCs) based on distributed subcarrier grouping and band adaptive modulation and coding (BAMC) based on adjacent subcarrier grouping [14, pp. 43]. The exact number of data

and pilot subcarriers in a subchannel depend on the subchannelization method implemented [14, pp. 273].

In WiMAX, frame duration is 5ms with a common deployment bandwidth of 5 and 10 MHz. The frame is divided into number of OFDM symbols. Since, time division duplexing (TDD) is implemented by WiMAX some of the OFDM symbols are allocated for downlink and the rest are used for uplink. Commonly the downlink uplink ratio is $1:1$ or $3:1$ [14]. The first symbol in the frame is used for preamble transmission. Preamble is used for physical layer procedures such as time and frequency synchronization and initial channel estimation. Preamble is followed by control messages that provides frame configuration information, such as MAP message length, the modulation and coding scheme and the usable subcarrier. The control message and data transmissions are then sent using subchannels. Fig. 2.4 shows a typical WiMAX frame structure. The minimum time-frequency resource that can be allocated by a WiMAX system to a given link is called a slot. Each slot consists of one subchannel over one, two or three OFDM symbols depending on the subchannelization scheme used. It is important to note that scheduling is performed by base station (BS) every frame period.

A subcarrier bandwidth of 10.94 kHz is considered in WiMAX. The WiMAX OFDMA system is implemented using IFFT/FFT. Since, IFFT/FFT can only take the values equal to 2^n, zero padding is performed such that dummy subcarriers are padded to the left and right of the useful subcarriers. In PUSCs subchannelization method, 6 subcarriers distributed pseudo-randomly across the frequency spectrum constitute a subchannel [14, pp. 42].

The WiMAX QoS framework is as shown in Fig. 2.8 [15]. QoS support is fundamental part of MAC layer design and WiMAX QoS framework is based on the Service flows (SFs). SF is a unidirectional flow of packets with a particular set of QoS attributes and the flow is between access service network gateway (ASN-GW) and user equipment (UE). The traffic mapping to appropriate SF is done at the ASN-GW for downlink and at UE for uplink. The scheduler at the MAC layer determines how radio resources are assigned among multiple SFs

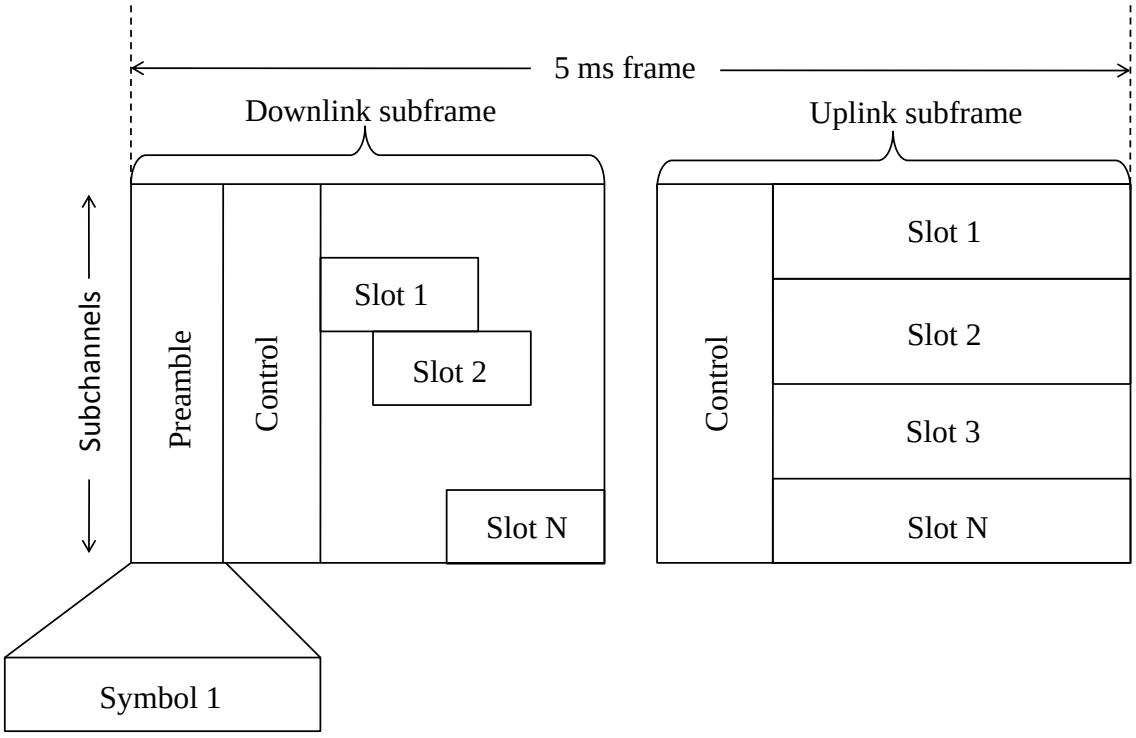

Figure 2.4. WiMAX frame structure

based on QoS attributes.

2.3 LTE Technology

LTE is an air interface technology based on 3GPP standard and was primarily designed for high-speed data services. LTE uses OFDMA as its multiple access mechanism in the downlink while it uses single carrier FDMA (SC-FDMA) in the uplink. Subcarriers are the smallest granular units in the frequency domain and OFDM symbol duration is the smallest granular unit in the time domain in OFDMA system. Mapping subcarriers to a particular subset can be referred to as resource mapping. A subcarrier bandwidth of 15 kHz is considered in LTE and 10% of the total system bandwidth is reserved for guard subcarriers and reference signals. In LTE the OFDM symbols can be organized into a number of physical resource blocks (PRB) consisting of 25 consecutive sub-carriers for a number of consecutive OFDM symbols that is equal to the number of OFDM symbols in a subframe. PRB is

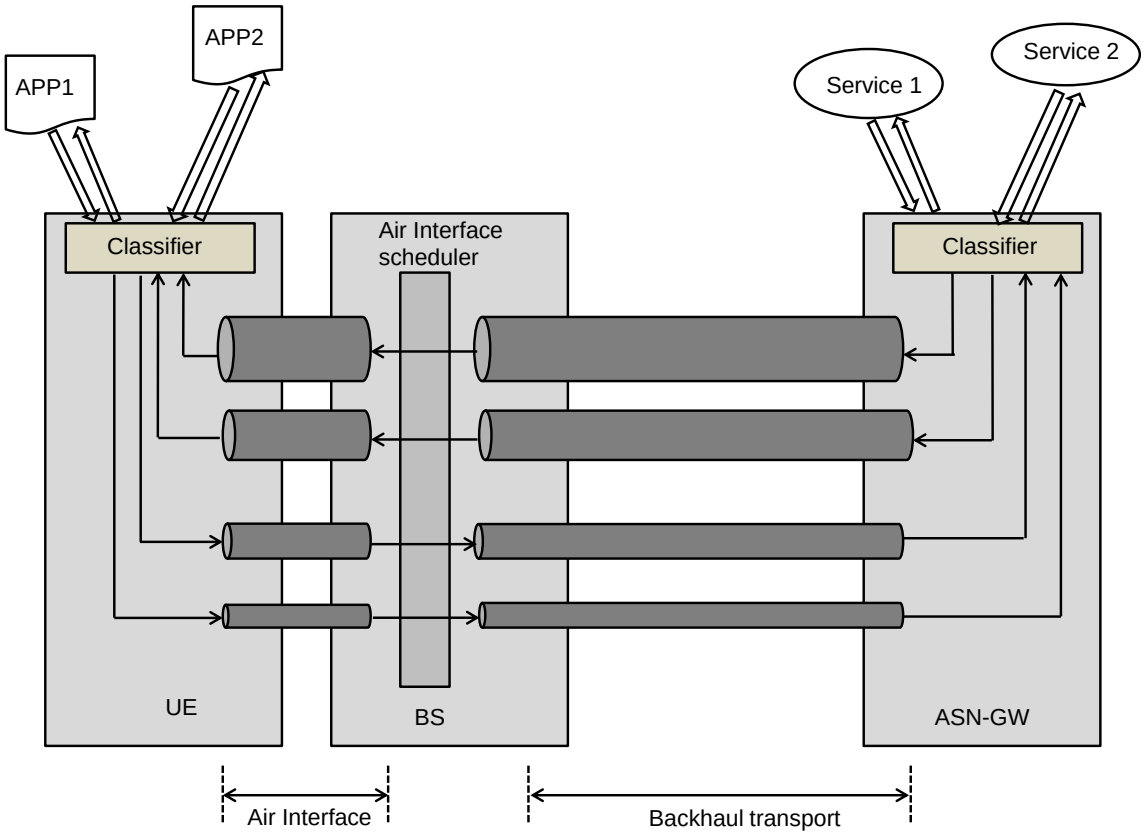

Figure 2.5. WiMAX QoS framework

a minimum resolution of scheduling in a frequency domain and its bandwidth is equal to 375kHz [16].

In LTE, the frame duration is fixed at 10ms and is divided into subframes of 1ms. Two slots of 0.5ms duration are formed out of a subframe. eNodeB schedules transmission every 1 ms known as transmission time interval (TTI). Fig. 2.6 depicts a typical LTE frame structure [17].

Unlike WiMAX, LTE uses primary synchronization sequence (PSS), which is sent twice in a LTE frame instead of a preamble symbol from frame synchronization. In addition, instead of using dense pilots in the time axis as in WiMAX, LTE embeds special reference signals in PRBs for channel estimation purpose. Hence, the overhead in LTE is less as compared to that in WiMAX. The assignment of reference signal in a subframe is depicted in Fig. 2.7. Reference signals (symbol "R" in figure) are transmitted during the first and fifth OFDM

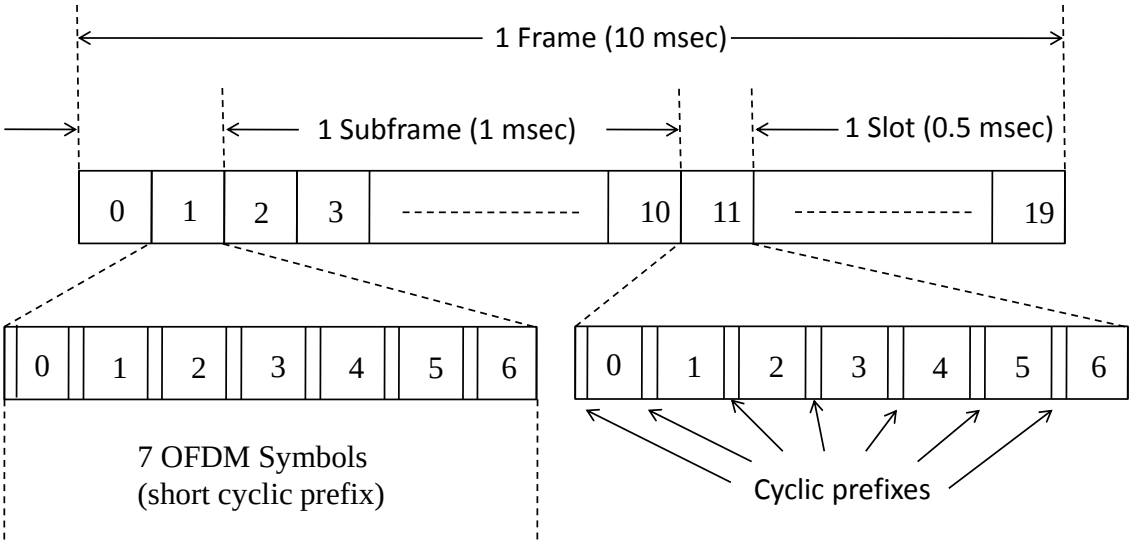

Figure 2.6. LTE frame structure

symbols of each slot when short CP is used and during the first and fourth OFDM symbols when the long CP is used [17]. A slot consists of 7 OFDM symbols when short CP is used and 6 OFDM symbols when long CP is used. Also it is important to note that the reference symbols are transmitted every sixth subcarrier. Frequency division duplexing (FDD) is the common duplexing deployment for LTE.

The LTE QoS framework is as shown in Fig. 2.8 [15]. QoS level granularity for LTE evolved packet system (EPS) is called bearer and is a packet flow between Packet Data Network Gateway (PDN-GW) and the UE. The traffic between client application and service can be separated into different service data flows (SDFs). SDFs mapped to the same bearer receive a common QoS treatment.

2.4 Adaptive Modulation and Coding Overview

AMC is a scheme where the advantage of the channel fluctuation over time and frequency is taken into account to adaptively select the set of modulation order and forward error correction (FEC) coding that best suits the channel condition while meeting the BER requirement. This is based on the fact that transmitting high data rate during good channel

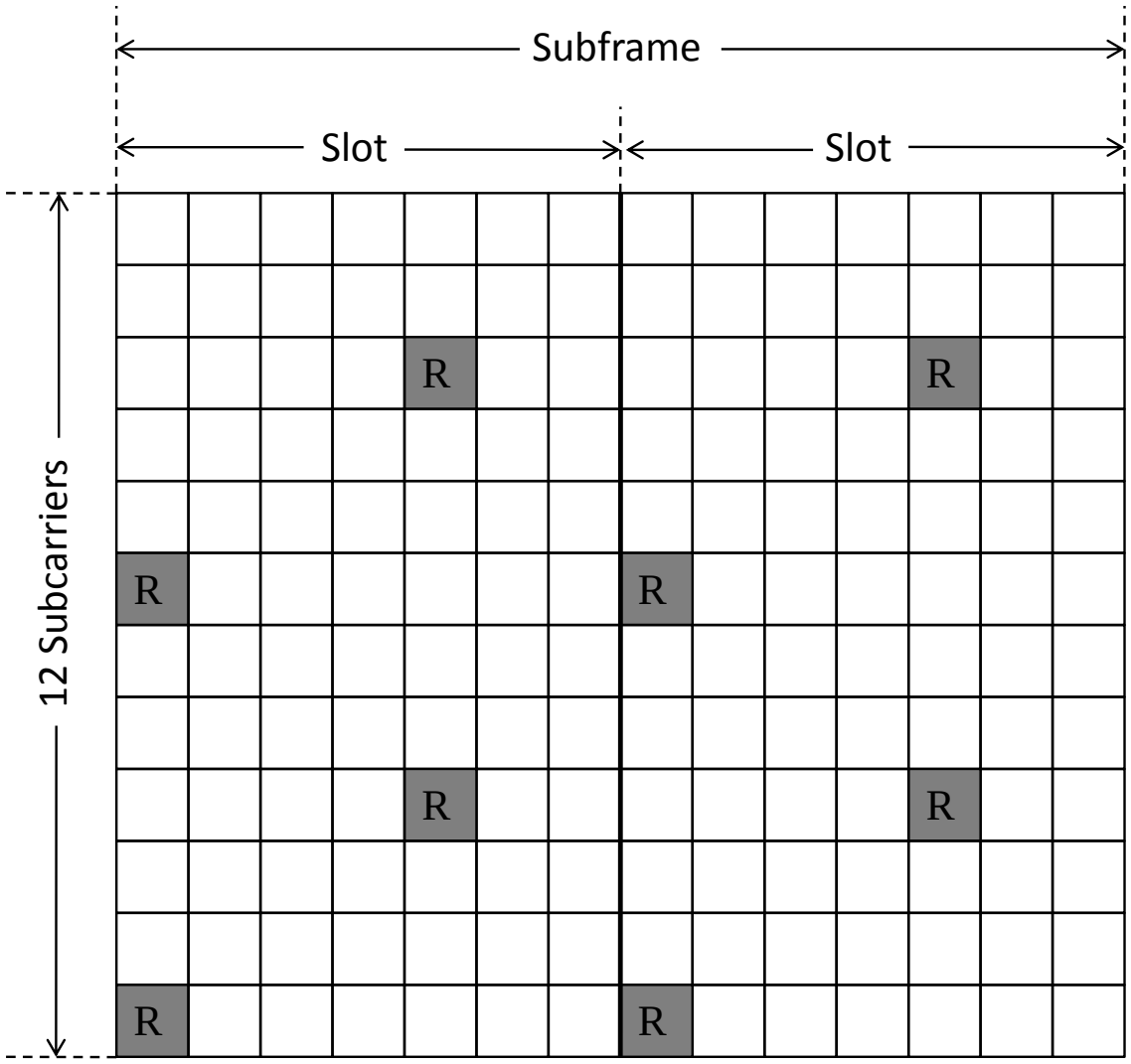

Figure 2.7. LTE reference signal in a subframe

conditions and low data rate during poor channel conditions achieves a significant performance gain. Hence, the dynamic adaptation of the modulation and coding set helps ensure the maximum system capacity. The concept of coded modulation that jointly optimize the channel coding and modulation was introduced by Ungerboeck [18]. The implementation of the coded-modulation achieved a significant coding gain without bandwidth expansion. It was later shown in [19] that an additional coding gain can be achieved by superimposing coset codes on top of the adaptive modulation. The general structure of adaptive coded modulation is as shown in Fig. 2.9. The channel coding segment of the structure a binary

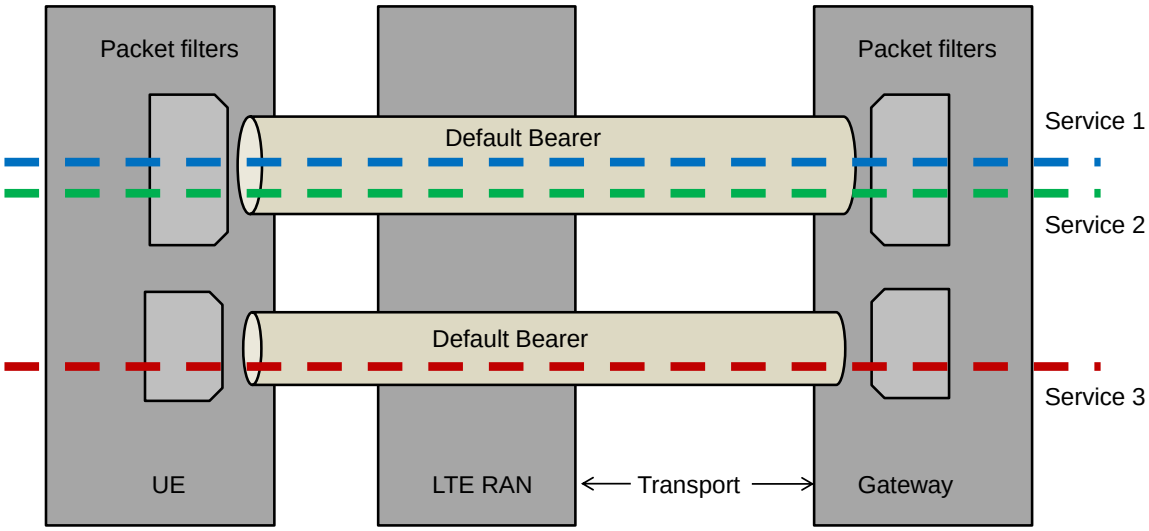

Figure 2.8. LTE QoS framework

encoder operates on k uncoded data bits to produce $k + r$ coded bits and then the coset (subset) selector uses these coded bits to choose one of the 2^{k+r} cosets from a partition of the signal constellation. While in the modulation segment, a signal point in the selected coset is chosen using $n - k$ uncoded data bits, where n is considered to be the function of the channel SNR. The selected point in the selected coset is one of the 2^{n+r} M-ary points in the transmit signal constellation.

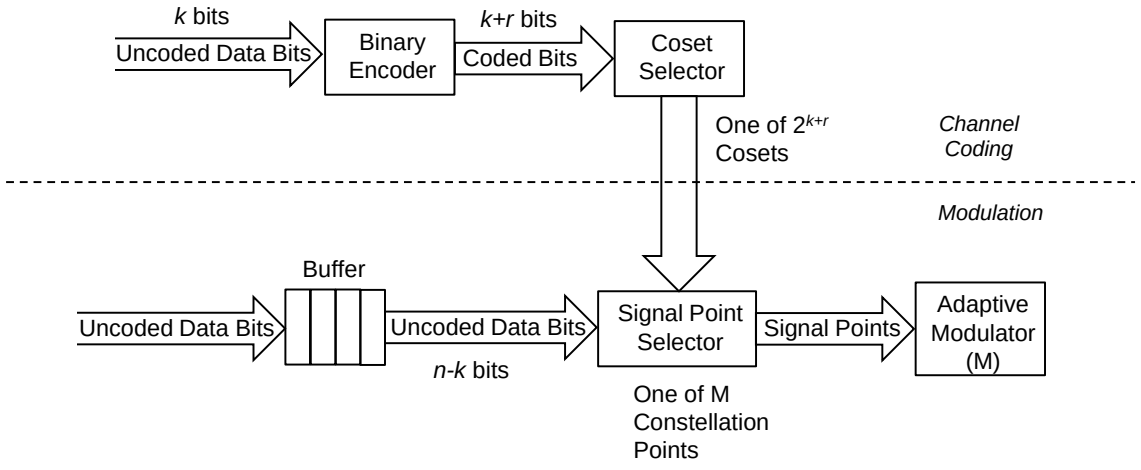

Figure 2.9. General structure of adaptive coded modulation

2.5 State-of-the-art

2.5.1 Maximum Fairness (MF) approach

While the maximum throughput approach maximizes the total rate from all users ignoring fairness among them, the maximum fairness algorithm [3] focuses on achieving fairness among users while maximizing the minimum data rate among them. Therefore, the maximum fairness is called a max-min problem - maximizing the minimum data rate. The optimization problem can be formulated as

$$\begin{aligned} \max_{P_{k,l},S_k} \min_{k} \quad & \sum_{n \subset S_k} \frac{B}{L} log_2 \left(1 + \frac{P_{k,l} h_{k,l}^2}{N_0 \frac{B}{L}}\right) \\ \text{subject to} \quad & \sum_{l=1}^{L} \sum_{k=1}^{K} P_{k,l} \leq P_{max} \\ & P_{k,l} \geq 0 \text{ for all } k, l \\ & S_1, S_2, ...S_k \text{ are disjoint} \\ & S_1 \cup S_2 \cup ... \cup S_k \subset \{1, 2, ..., K\} \end{aligned} \tag{2.5.1}$$

where $P_{k,l}$ is the power assigned to user $k's$ subchannel l, $h_{k,l}$ is the channel gain of user $k's$ subchannel l, S_k is the set of indices of subchannels assigned to user k, and N_0 is power of additive white Gaussian noise (AWGN). $S_1, S_2, ...S_k$ need to be disjoint since a subchannel is assigned to only one user. As set selection is involved with (2.5.1), it is not convex problem. It can be converted to convex problem by the introduction of parameter $w_{k,l}$, which represents portion of subchannel l assigned to a user k. The optimization problem

can then be represented as

$$\begin{aligned}
\max_{P_{k,l},w_{k,l}} \min_{k} \quad & \sum_{l=1}^{L} \frac{w_{k,l}B}{L} log_2\left(1+\frac{P_{k,l}h_{k,l}^2}{N_0\frac{w_{k,l}B}{L}}\right) \\
\text{subject to} \quad & \sum_{l=1}^{L}\sum_{k=1}^{K} P_{k,l} \leq P_{max} \\
& P_{k,l} \geq 0 \text{ for all } k,l \\
& \sum_{k=1}^{K} w_{k,l} \leq 1 \text{ for all } l \\
& w_{k,l} \geq 0 \text{ for all } k,l
\end{aligned} \tag{2.5.2}$$

This optimization problem can further be formulated into a standard convex optimization problem as

$$\begin{aligned}
& \max_{P_{k,l},w_{k,l}} t, \text{subject to} \\
& t \leq \sum_{l=1}^{L} \frac{w_{k,l}B}{L} log_2\left(1+\frac{P_{k,l}h_{k,l}^2}{N_0\frac{w_{k,l}B}{L}}\right) \text{ for all } k \\
& \sum_{l=1}^{L}\sum_{k=1}^{K} P_{k,l} \leq P_{max} \\
& P_{k,l} \geq 0 \text{ for all } k,l \\
& \sum_{k=1}^{K} w_{k,l} \leq 1 \text{ for all } l \\
& w_{k,l} \geq 0 \text{ for all } k,l
\end{aligned} \tag{2.5.3}$$

The optimal solution is difficult to determine, since the joint subchannel power allocation function is not concave [14]. Accordingly, a two-stage low-complexity suboptimal solution is preferred for simplicity. The subchannel and power allocation are done separately. The common approach is to assume equal power allocation at first to all subchannels. Then the first stage would be to iteratively allocate each subchannel to a low-rate user with the highest channel gain on the subchannel of interest [20], [3]. The next stage would be to apply waterfilling power-control method to the suboptimal solution in order to achieve the optimal one. In other words, more power to strong subchannels and less power to weak subchannelss. Following this suboptimal allocation, the final result is relatively close to the optimal result

in terms of fairness and total throughput [14].

2.5.2 Proportional Rate Constraint (PRC) approach

Although the UE with weak channels are deprived from resources in the maximum system throughput approach, they consume most of the resources in the maximum fairness approach. Moreover, the maximum fairness approach is not flexible enough to support services with varying QoS requirements like WiMAX. Therefore, a modification to it is the proportional rate constraints approach, in which the total throughput is maximized while maintaining some sort of proportionality among users' data rates and is presented in [5]. This proportionality is defined by a set of system parameters, $\{\beta_k\}_{k=1}^{K}$ and the constraint on the users' data rates is [1]

$$R_1 : R_2 : \ldots : R_K = \beta_1 : \beta_2 : \ldots : \beta_K \tag{2.5.4}$$

Let $P_{k,l}$ be the power allocated to a k^{th} user over subchannel l, N_0 be the additive white Gaussian noise (AWGN) power spectral density with zero mean, $h_{k,l}$ be the channel gain for a user k over subchannel l, and $\rho_{k,l} \in \{0, 1\}$ indicates whether or not a subchannel l is used by th user k. Then, the data rate of a k^{th} user is

$$R_k = \sum_{l=1}^{L} \frac{\rho_{k,l} B}{L} log_2(1 + \frac{P_{k,l} h_{k,l}^2}{N_0 \frac{B}{L}}) \tag{2.5.5}$$

Carefully looking at this approach it turns out that it is a general form for the maximum fairness approach. Alternatively, the maximum fairness is a special case of this approach when $\beta_k = 1 \ \forall K$. The derivation of the optimal solution is complicated since the objective function contains both continuous variables, $P_{k,l}$ and binary variables, $\rho_{k,l}$ and hence, a suboptimal solution is derived in [5] and a low-complexity implementation is developed in [20].

[1] This identity $x_1 : x_2 : \ldots : x_k = y_1 : y_2 : \ldots : y_k$ means $\frac{x_i}{y_i} = \frac{x_j}{y_j} \quad \forall i, j = 1, 2, \ldots, K$

2.5.3 Cross-layer weighted rate constraint (CLWRC) Approach: WiMAX

Cross-layer Weighted Rate Constraint (CLWRC) scheme, proposed in [1], is a resource allocation optimization scheme that takes into account, both the channel conditions and the queue status of each user as well as different QoS requirements to maximize system capacity. Two cross-layer parameters are introduced in this scheme and the weighted rate is defined based on these cross-layer parameters. These cross-layer parameters are also dependent on the WiMAX QoS classes. Each of the QoS classes defined in standards and the two cross-layer QoS parameters are discussed in detail one by one as follows

WiMAX QoS classes

WiMAX (IEEE 802.16e standard) defines five different service classes and associated QoS parameters. Different service classes support different applications that have some defined QoS parameters. The study of the traffic generated by the applications gives us a clear idea about the scheduling services. This section is focused on the details of different QoS classes defined in the standard.

- Unsolicited Grant Service (UGS): The UGS is designed to support real-time service flows that transport fixed-size data packets on a periodic basis, such as T1/E1 and Voice over IP (VoIP)traffic without silence suppression. As this service offers fixed size data grants on a real-time periodic basis, the overhead and latency associated with the UE's bandwidth request are eliminated and the availability of the data packets are assured to meet the flow's real-time requirements. The BS at optimal condition should provide the UGS data grants at periodic intervals based upon the Maximum Sustained Traffic Rate of the service flow. Generally, the VoIP packets are sent every 20 ms or 30 ms [21]. The interval between data grants for the UGS service is 20 ms since the BS, under optimal conditions, must allocate to the UGS connections data grants at

intervals equal to the UGS application packet generation rate [22]. The QoS parameters associated with this service are; Maximum Sustained Traffic Rate, Maximum Latency, Tolerance Jitter and Request/Transmission Policy [23], [14].

- Real Time Polling Service (rtPS): The rtPS traffic is designed to support real-time service flows that randomly transport variable size data packets on a periodic basis, such as moving pictures experts group (MPEG) video [23], [14]. The service offers real-time, periodic, unicast request opportunities, which meet the flow's real-time needs and allow the UE to specify the size of the desired data grant. The unicast polling opportunities are frequent enough to ensure that latency requirements of real-time services are met and which in turn requires higher overhead than UGS. However, rtPS is more efficient for the services that supports variable size data grant. The QoS parameters associated with this service are Minimum Reserved Traffic Rate, Maximum Sustained Traffic Rate, Maximum Latency and Request/Transmission Policy.

- Extended Real Time Polling Service (ErtPS): Extended rtPS is a scheduling mechanism which builds on the efficiency of both UGS and rtPS [23], [14]. ErtPS thus should involve the generation of data grants in an unsolicited manner as in UGS, saving the latency of a bandwidth request. However, the ErtPS allocations should be dynamic unlike UGS allocations. The QoS parameters associated with this service are the Maximum Sustained Traffic Rate, Minimum Reserved Traffic Rate, Maximum Latency, and Request/Transmission Policy. The Extended rtPS is designed to support real-time service flows that generate packets at variable bit rate (VBR) with changing bandwidth requirement, such as Voice over IP services with silence suppression. The voice model is an "ON" / "OFF" one. The duration of "ON" / "OFF" periods is exponentially distributed with the mean of distribution being the time for which the system is "ON" or "OFF" on average. In [22], [24] this average "ON" and "OFF" time is given as 1.2 seconds and 1.8 seconds, respectively. Similarly the time for which the system

resources will be dedicated only for voice processing is 20 milliseconds and the packet size is 66 Bytes.

- Non Real Time Polling Service (nrtPS): The nrtPS offers unicast polling opportunities on a regular basis, thereby enabling UE to use the contention-based polling in the uplink to request bandwidth [23], [14]. The QoS parameters associated with this service are; Minimum Reserved Traffic Rate, Maximum Sustained Traffic Rate, Traffic Priority and Request/Transmission Policy. The nrtPS traffic is modeled using File Transfer Protocol (FTP). FTP traffic is generated using an exponential distribution with a mean being the size of packet generated on average. In [22], the size of packet generated on average is 512Kbps.

- Best Effort (BE): The BE service supports the application that generates stream of data, such as web browsing with no strict QoS parameter [14]. The UE uses only the contention-based polling opportunity to request bandwidth and data are sent whenever resources are available. The QoS parameters associated with this service are; Maximum Sustained Traffic Rate, Traffic Priority and Request/Transmission Policy. Characteristic properties of Web client request based on statistical observation shows that the file size distributions are well modeled as hybrids having lognormal bodies, and power-law (i.e., Pareto) tails. In [22], [25], 88% of the total area of the probability density function (PDF) about the mean corresponds to the lognormal distribution and the remaining 12% of the total area near the tail region of pdf corresponds to pareto distribution. The mean rate values of lognormal distribution and pareto distribution are 7247 and 10558 bps, respectively i.e, the average size of the packets per second corresponding to the lognormal and pareto distribution is 7247 and 10558 bits, respectively.

Service Urgency

Service urgency as proposed in [11] is a cross-layer QoS parameter that is dependent on the information about the queues of the services in the data link layer. Every user can be associated with one of the five different service flows. Let x denotes any one of the five service flows such that x is any element in the set {UGS, rtPS, ErtPS, nrtPS and BE} and $x(k)$ denotes x associated with a user k. Now, let N be the total number of frames considered, then n is defined as the frame number being served such that $n \in \{1, 2, \ldots, N\}$. Also, let $A_k^x(n)$ be the number of bits arriving at the queue of a k^{th} user associated with an x service flow during a frame n, $Q_k^x(n)$ be the length of queue of the k^{th} user associated with an x service flow during a frame n and $B_k^x(n)$ be the number of bits the BS serves from the queue of the k^{th} user associated with an x service flow during a frame n. Then the queue length corresponding to the k^{th} user associated with an x service during a frame $n+1$ is given by

$$Q_k^x(n+1) = Q_k^x(n) + A_k^x(n) - B_k^x(n). \tag{2.5.6}$$

Also, let Ω^x be the set of all users associated with the same x. Then the set Ω^x is expressed as

$$\begin{aligned} \Omega^x = &\{1 \leq k \leq K : x(k) = x\} \\ &\forall x \in \{UGS, ErtPS, rtPS, nrtPS, BE\}, \end{aligned} \tag{2.5.7}$$

and let $Q^x(n)$ be the aggregate queue length corresponding to users associated with the same service class during frame n, then $Q^x(n)$ can be expressed as

$$Q^x(n) = \sum_{k \epsilon \Omega^x} Q_k^x(n). \tag{2.5.8}$$

Finally, the normalized queue length of the k^{th} user during frame n, $U_k^x(n)$, which will be called henceforth the Urgency Factor, can be defined as

$$U_k^x(n) = \begin{cases} \frac{Q_k^x(n)}{Q^x(n)}, & x \in \{\text{rtPS, nrtPS, BE}\} \\ 1, & x \in \{\text{UGS, ErtPS}\}. \end{cases} \tag{2.5.9}$$

It should be noted here that the Urgency Factor $U_k^x(n)$, is set to 1 for users with a UGS or ErtPS service flow type. It is known form the QoS requirements that the users associated

with UGS and ErtPS service classes should be allocated resources on a periodic basis and therefore the concept of urgency does not apply. It should also be noted here that $U_k^x(n) \in (0, 1]$, UGS and ErtPS service flows are thus assigned the highest urgency factor. However, the urgency factor for rtPS, nrtPS and BE are calculated using (2.5.9). Here the queue length associated with a given k^{th} user at a given frame number n corresponding to a given service flow type x is normalized by the total queue length of that particular x service flow in the system. This normalised factor, urgency factor, will be high for the service type with the longest queue associated with a given k^{th} user during frame n and hence this user will be served first. It is important to note that the concept of urgency factor does not apply if users do not have any queue. The significance of the urgency factor is two-fold. It gives indication about which user is being under-served relative to other users of the same service flow, and it also conveys information about the queue length of the user to the resource allocation algorithm. The higher the value of $U_k^x(n)$, the more it is urgent to allocate resources to the user.

Service Satisfaction

Service satisfaction based on different kinds of service flows depends on the information like data rate, delay satisfaction indicator or flow's coefficient as defined in [26]. Hence, service satisfaction can be deemed as the cross-layer QoS and is considered in this study. Let $\{\Gamma_{UGS}, \Gamma_{ErtPS}, \Gamma_{rtPS}, \Gamma_{nrtPS}, \Gamma_{BE}\}$ be defined as a set of configurable system parameters. Each Γ_x denotes a weighting factor that can be used to favor one service class over the other and be configurable depending on the system deployment. For example, if the priority order for different QoS classes is UGS>ErtPS>rtPS>nrtPS>BE, then the weighting factors can be set under the constraint $\Gamma_{UGS} > \Gamma_{ErtPS} > \Gamma_{rtPS} > \Gamma_{nrtPS} > \Gamma_{BE}$; e.g., $\Gamma_{UGS} = 0.8 > \Gamma_{ErtPS} = 0.6 > \Gamma_{rtPS} = 0.4 > \Gamma_{nrtPS} = 0.3 > \Gamma_{BE} = 0.2$. The same values for each of Γ_x are considered in the simulation as well and are listed in Table 2.1. Satisfaction factors for each service flow will be defined such that they are inversely proportional to the corresponding

weighting factors associated with each service flows. And based on the idea that the service flows of higher priority need to have lower service satisfaction, it is plausible to consider a fractional values for each Γ_x.

Now let $S_k^x(n)$ be the satisfaction factor corresponding to a k^{th} user associated with an x service flow during a frame n. For UGS service flows the satisfaction factor of a k^{th} at a given frame n is defined as

$$S_k^{UGS}(n) = \frac{1}{\Gamma_{UGS}}, \tag{2.5.10}$$

where Γ_{UGS} is the UGS class weighting factor. Therefore, the satisfaction factor is constant for all the users with UGS service flows and over all frames. As for ErtPS service flows, the satisfaction factor of a k^{th} at a given frame n is defined as

$$S_k^{ErtPS}(n) = \frac{1}{\Gamma_{ErtPS}}, \tag{2.5.11}$$

where Γ_{ErtPS} is the ErtPS class weighting factor. Also, the satisfaction factor is constant for all the users with ErtPS service flow and over all frames. ErtPS service flow generates constant size packets like UGS, but unlike UGS, packets are not generated periodically. Therefore, whenever data is available it is treated the same as UGS data. For rtPS service flows, if the waiting time of the packet in a queue exceeds a maximum allowed latency or the deadline T_k^{rtPS}, of a k^{th} user associated with rtPS service flow, then a timeout is set by the scheduler as T_k^{rtPS}. Hence the satisfaction factor of a k^{th} user associated with rtPS service flow is defined

$$S_k^{rtPS}(n) = \frac{DS_k^{rtPS}(n)}{\Gamma_{rtPS}}, \tag{2.5.12}$$

where Γ_{rtPS} is the rtPS class weighting factor, $DS_k^{rtPS}(n)$ is the delay satisfaction indicator of a k^{th} user associated with an rtPS service flow at a given frame n, which is defined as [26]

$$DS_k^{rtPS}(n) = max\{1, T_k^{rtPS} - \Delta T_k^{rtPS} - W_k^{rtPS}(n) + 1\} \tag{2.5.13}$$

where $W_k^{rtPS}(n) \in [0,\ T_k^{rtPS}]$ is the head of line (HOL) delay which is defined as the longest waiting time that a packet experiences for a k^{th} user associated with an rtPS service flow at a given frame n and $\Delta T_k^{rtPS} \in [0,\ T_k^{rtPS}]$, of a k^{th} user associated with an

rtPS service flow, is the guard time region ahead of the deadline T_k^{rtPS}, which indicates the time remaining before which the packet should be scheduled to avoid timeout. Now, if $W_k^{rtPS}(n) \in [0,\ T_k^{rtPS} - \Delta T_k^{rtPS}]$, then $DS_k^{rtPS}(n)$ will be greater than or equal to 1 indicating that the delay requirement of the service is satisfied, as the packet will be served before timeout [26]. And if $W_k(n) \in [T_k - \Delta T_k^{rtPS},\ T_k^{rtPS}]$, then $DS_k^{rtPS}(n)$ will be less than 1 indicating that the delay requirement of the service is not satisfied, as the packet can't be served before timeout. So a lower value of satisfaction factor will require a scheduling algorithm to allocate more resources to the service to meet the delay requirements. The satisfaction factor for users with an rtPS service flow has a minimum value of $\frac{1}{\Gamma_{rtPS}}$. For nrtPS service flows, the satisfaction factor of a k^{th} user at a given frame n is defined as

$$S_k^{nrtPS}(n) = \frac{RS_k^{nrtPS}(n)}{\Gamma_{nrtPS}}, \tag{2.5.14}$$

where Γ_{nrtPS} is the nrtPS class weighting factor, $RS_k^{nrtPS}(n)$, of a k^{th} user associated with an nrtPS service flow at a given frame n, is the rate satisfaction indicator which is defined as

$$RS_k^{nrtPS}(n) = max\{1, \hat{\eta}_k^{nrtPS}(n)/\eta_k^{nrtPS}\} \tag{2.5.15}$$

where η_k^{nrtPS} is the minimum reserved data rate of the k^{th} user associated with an nrtPS service flow, and $\hat{\eta}_k^{nrtPS}(n)$ is the exponentially weighted average data rate of the k^{th} user associated with an nrtPS service flow up to the frame n obtained by using the exponentially weighted low-pass filter [7]. Exponential weighted average or the exponential smoothing is performed on the time series data such that the data can be obtained in smooth and presentable form. In traditional weighted averaging, equal weights are assigned to the past observations however, in the exponential smoothing decreasing exponential weights are assigned to time series data. This exponential weighted averaging based on the exponential low-pass filter as applied to the data rate calculation over number of frames can be defined as

$$\hat{\eta}_k^{nrtPS}(n+1) = \begin{cases} C_k^{nrtPS}(n), & n = 0 \\ \hat{\eta}_k^{nrtPS}(n)(1 - \frac{1}{t_c}) + C_k^{nrtPS}(n)\frac{1}{t_c}, & n > 0 \end{cases} \tag{2.5.16}$$

where $C_k^{nrtPS}(n)$ is the user data rate allocated to a k^{th} user associated with nrtPS service flow during a frame n . The parameter t_c, window size, controls the latency of the system. If t_c is large, then the latency increases, with the benefit of higher data rate. If t_c is small, the latency decreases, since the average data rate values change more quickly, at the expense of some data rate [14, pp. 213]. This is because as latency increases, the data rate is averaged over a larger t_c and hence, the scheduler can afford to wait longer before scheduling a user when its channel hits a really high peak value and vice versa [7]. The satisfaction factor, $S_k^{nrtPS}(n)$, ensures that the user is receiving an average data rate above the minimum reserved rate, $\hat{\eta}_k^{nrtPS}(n) \geq \eta_k^{nrtPS}$. If $RS_k^{nrtPS}(n) \geq 1$, then the rate requirement is satisfied, which increases the satisfaction factor. Large values of $RS_k^{nrtPS}(n)$, therefore, indicate high degree of satisfaction. The minimum value for $RS_k^{nrtPS}(n)$ is 1, which is when the user is underserved and should be allocated more resources to meet the minimum rate requirements. The satisfaction factor for users with nrtPS service flow has a minimum value of $\frac{1}{\Gamma_{nrtPS}}$. For BE service flows, the satisfaction factor of a k^{th} user at a given frame n is defined as

$$S_k^{BE}(n) = \frac{1}{\Gamma_{BE}}, \tag{2.5.17}$$

where Γ_{BE} is the BE class weighting factor. Therefore, the satisfaction factor is constant for all the users with BE service flow and over all frames. The reason is that by definition of the QoS requirements, the users with BE service flow should be allocated resources after all other service flows are satisfied, and therefore, the concept of service satisfaction does not apply. The significance of the satisfaction factor is also two-fold. It allows for scalability, as when the system is overloaded, the performance of users with low-priority service classes will be degraded prior to those with high priority service classes, and it also allows users with low-priority service classes to lead when users with higher-priority service classes are well satisfied.

CLWRC Problem Formulation

Let $P_{k,l}(n)$ be the power allocated to a k^{th} user over a subchannel l during a frame n, N_0 denote the additive white Gaussian noise (AWGN) power spectral density with zero mean, $h_{k,l}$ be the channel gain for a user k over a subchannel l, and $\rho_{k,l} \in \{0,1\}$ indicates whether or not a subchannel l is used by the user k. Then the spectral efficiency or the channel capacity, in bits/symbol/Hz, for a k^{th} user associated with an x service flow during a frame n is expressed as

$$C_k^x(n) = \sum_{l=1}^{L} \frac{\rho_{k,l}}{L} log_2[1 + P_{k,l}(n)H_{k,l}(n)] \qquad bits/symbol/Hz \tag{2.5.18}$$

where

$$H_{k,l}(n) = \frac{h_{k,l}^2}{N_0 \frac{B}{L}}, \tag{2.5.19}$$

and the weighted capacity, $R_k^x(n)$, of a k^{th} user associated with an x service flow during a frame n, is then expressed as

$$R_k^x(n) = \frac{U_k^x(n)}{S_k^x(n)} C_k(n) \tag{2.5.20}$$

Now, the fairness constraint is defined as

$$R_i(n) = R_j(n) = R(n) \qquad \forall i, j \in [1, K]. \tag{2.5.21}$$

Based on the above discussion, the optimization problem can be expressed mathematically as

$$\max_{P_{k,l}, \rho_{k,l}} \quad C = \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{\rho_{k,l}}{L} \log_2 (1 + P_{k,l} H_{k,l}) \qquad bits/symbol/Hz \tag{2.5.22}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \sum_{l=1}^{L} P_{k,l} \leq P_{tot} \tag{2.5.23}$$

$$P_{k,l} \geq 0 \quad \forall \quad k, l \tag{2.5.24}$$

$$\rho_{k,l} = \{0,1\} \quad \forall \quad k, l \tag{2.5.25}$$

$$\sum_{k=1}^{K} \rho_{k,l} = 1 \quad \forall \quad l \tag{2.5.26}$$

$$R_i(n) = R_j(n) = R(n) \quad \forall \quad i, j \in [1, K], \tag{2.5.27}$$

where P_{tot} is the total available power. The first constraint (2.5.23) implies that the total power used by subchannels is not to exceed the total available system power. The second constraint (2.5.24) states that the power used by all subchannels should be non-negative. In the third constrain (2.5.25), $\rho_{k,l}$ is only allowed to be 0 or 1 which assures that a l^{th} subchannel is either used or not used by the k^{th} user. Furthermore, no sharing of subchannels among users is allowed, which is stated by the fourth constraint (2.5.26). The last constraint (2.5.27) is the fairness constraint presented in (2.5.20) and (2.5.21). A two phase greedy approach is then adopted to solve the optimization problem where subchannel and power allocation are made separately.

2.5.4 Cross-layer resource allocation schemes: LTE

The LTE standard don't define the ways in which the system resources, PRBs and power could be efficiently utilized. Resource allocation design approaches that determine the efficient way of scheduling the users, allocating the PRBs to the users and determining the appropriate power levels for each user on each PRB are left free to the LTE developers and vendors. Hence there are a large number of studies that propose the resource allocation algorithms and design approaches that aim to acquire an optimum balance between the system capacity with fairness among the users in the system. In most recent research studies reported in literature, authors are more focused on cross-layer design approaches. The authors in [27] present a novel LTE downlink MAC scheduling algorithm that differentiates between the different QoS classes and their requirement. The scheduler also considers the channel conditions and tries to maintain a proportional fairness among the QoS guarantees and the multi-user diversity. The authors, however, don't consider capacity optimization while guaranteeing the different QoS requirements. In [28], authors develop a Universal Terrestrial Radio Access Network (UTRAN) LTE downlink channel dependent scheduler framework that encompasses frequency domain packet scheduling, hybrid automatic repeat request (HARQ) management and inter-user fairness control. An effective control on user

fairness was achieved by dividing the packet scheduler into a time-domain and frequency-domain part. A performance evaluation in terms of cell throughput, coverage and capacity is performed. The proposed scheduler don't consider user queue status while making the scheduling decision. A QoS-guaranteed cross-layer resource allocation algorithm for multi-class services in downlink LTE system is proposed in [29]. The authors take into account the exponential (EXP) rule, channel quality variance, real-time services and non-real-time services and minimum transmission rate. The proposed algorithm only considers resource block allocation. In particular, it considers user queue status as a parameter for evaluating priority based on packet delay, not as a fairness parameter. The algorithm does not consider optimized power allocation, in the contrast an equal power distribution is considered instead. The authors in [30], propose a distributed protocol for radio resource allocation and network optimization that aims to achieve weighted proportional fairness among clients by jointly considering resource block scheduling, power control and client association. A user-dependent priority indicator is defined and an optimization problem is formulated so as to obtain the weighted proportional fairness, but the fairness so obtained don't consider user QoS and queue status. A cross-layer solution for real time service that allocates resource block for different services in order to meet their QoS requirements is proposed in [31]. Instantaneous downlink channel signal to interference plus noise ratio and service QoS information are utilized for resource block allocation. Fairness in resource block allocation based on the different QoS services and user queue status is not considered in that study. The authors in [32] propose a cross layer scheduling algorithm that aims to minimize the resource utilization for LTE downlink system. The algorithm takes into account the channel conditions, the size of transmission buffers and different QoS demands to make the scheduling decision. The algorithm, however, fails to consider fairness in scheduling based on user QoS requirement and queue status. An optimized power allocation among resource blocks is not considered in this algorithm; rather an equal power assumption is considered. In [33], the authors present an analytical framework for QoS-guaranteed cross-layer scheduling

and resource allocation that takes into account the accuracy of the available channel state information. An impact based on channel information accuracy on different scheduling algorithms is also studied in that paper. The proposed framework, however, don't address the scheduler fairness based on user QoS requirement and queue status.

Given the literature review herein and to the best of the author's knowledge none of the work reported in literature addresses the problem of cross-layer optimization in LTE downlink system by taking into account the channel conditions, queue status and QoS requirements simultaneously. This chapter considers the Cross-layer Weighted Rate Constraint (CLWRC) scheme presented in [11] and extend it accordingly such that it applies to LTE system. This thesis hence addresses the above mentioned issues and presents a resource allocation optimization scheme that takes into account, both the channel conditions and the queue status of each user as well as different QoS requirements to maximize LTE system capacity, which makes the proposed scheme unique to the the state-of-the-art research on LTE cross-layer optimization. This proposed scheme is termed as LTE-CLWRC scheme.

2.5.5 AMC-based cross-layer schemes

As mentioned in Section 2.5.4, the standards in general don't define the resource allocation strategy and the efficient way of scheduling the users, allocating the subchannels to the users, determining the appropriate power levels for each user on each subchannel and defining the constraints on making the scheduling are left free to the developers and vendors. Hence, the implementation of AMC in the scheduling algorithm is also dependent on the developers and the vendors. There are various other works found in literature based on adaptive modulation and coding. The authors in [34] propose various adaptive transmission algorithms that minimize both the downlink and uplink total power transmitted for the IEEE 802.16 OFDMA system. The optimization is subject to constraints including maximum transmission power and available time-frequency resource, while satisfying the QoS requirements for all downlink and uplink service flows scheduled for transmission. The authors also take into account the

MAC and PHY layers processing in case of automatic repeat request (ARQ). In this study, the authors don't present the queue lengths corresponding to the service flows as a fairness constraint and they also don't consider the non-error free Shannon capacity formulation. In [35], the authors propose two adaptive modulation and coding techniques for WiMAX systems with an aim to improve performances in non line-of-sight communications. The first technique, target Block Error Rate (BLER), aims to respect a maximum BLER imposed on the system based on the target QoS level being served. While the second technique, Maximum Throughput, aims to maximize the system throughput without any constraint on target BLER. The queue lengths and QoS requirements are not considered in this study as a fairness constraint in the optimization problem formulation. The authors in [36] introduce a new form of adaptive modulation with the ability of using high order modulation scheme such as 256-quadrature amplitude modulation (256-QAM) and 64-QAM to map the data onto the carriers at low signal to noise ratio (SNR) values. The authors then propose a new algorithm by combining together the new adaptive modulation along with clipping technique that is capable of reducing the peak-to-average power ratio (PAPR), enhance the data rate and improve the performance of the symbol error rate (SER) at low SNR values. In a separate study, a multiuser resource allocation optimization technique assuming adaptive modulation and low density parity check (LDPC) coding in OFDMA systems is presented in [37]. A low-complexity weight-adaptive mechanism is introduced that is capable of maximizing the achieved throughput while guaranteeing a given set of BER quality requirements. Application specific quality of experience (QoE) metrics are taken into account: however, fairness in terms of queue lengths and QoS parameters are not considered in this study.

Given the literature review herein and to the best of the author's knowledge none of the work reported in literature addresses the problem of AMC-based cross-layer optimization by taking into account the channel conditions, queue status and QoS requirements simultaneously. This thesis addresses this issue and presents an AMC-based resource allocation optimization scheme that takes into account, both the channel conditions and the queue status

of each user as well as different QoS requirements to maximize system capacity, which makes the proposed scheme unique to the the state-of-the-art research on AMC-based cross-layer optimization. This proposed scheme is an extension to the CLWRC scheduling algorithm introduced in [11] based on adaptive modulation and coding. The proposed scheme is hence termed as AMC-based CLWRC (AMC-CLWRC) algorithm.

2.6 Results based on CLWRC scheme presented in [1]

The numerical results in [1] based on the solution of the optimization problem in Section 2.5.3 adopting the CLWRC resource allocator algorithm is depicted in Fig. 2.10. The same wireless channel model as in [5] is considered in [1] which is a frequency-selective AWGN channel with zero mean consisting of six independent Rayleigh multipaths. Each multipath component is modeled by Clarke's flat fading model. It is assumed that the power delay profile is exponentially decaying with e^{-2l} rate, where l is the multipath index and $l \in \{0, 1, \ldots, 5\}$. Hence, the relative power of the six multipath components are $[0, -8.69, -17.37, -26.06, -34.74, -43.43]$ dB. A Doppler shift of 30 Hz is assumed and the total available bandwidth and transmit power are 5 MHz and 1 W, respectively. It is important to note that a subcarrier spacing of 10.94 kHz is chosen as a good balance between satisfying the delay spread and doppler spread requirements such that each subcarriers will experience relatively flat fading for operating between mixed, fixed and mobile environment [14, pp. 43]. Likewise, [1] considers five different traffic models (described in Section 2.5.3) and are used to simulate the arrival patterns of the five different service flows. Table 2.2 summarizes the characteristics of the 10 different active users' service flow in the system based on five different service flows, while the Table 2.1 shows the system parameters used for the simulation.

The results in Fig. 2.10 depict the total average system capacity versus frame number (1 to 1000). The system parameters listed in Table 2.1 are used in this case as well. The total

Table 2.1. Simulated System Parameters

Symbol	Parameter	Value
P_{tot}	Total system power	1 Watt
L	Number of subchannels	64
K	Number of users	10
N	Number of frames	1000
B	Total system bandwidth	5 MHz
E_s	Symbol Energy	1 Joule
T_f	Frame duration	5 ms
ΔT	Guard time ahead of deadline	20 ms
T_c	Moving average window size	1000 ms
Γ_{UGS}	UGS weighting factor	0.8
Γ_{rtPS}	rtPS weighting factor	0.6
Γ_{ErtPS}	ErtPS weighting factor	0.4
Γ_{nrtPS}	nrtPS weighting factor	0.3
Γ_{BE}	BE weighting factor	0.2

average system capacity achieved during each arrival time duration of a frame is depicted in the figure, for an average SNR of 12 dB, using the proposed CLWRC, PRC and MF algorithms. It is obvious from the figure that the proposed CLWRC algorithm remains superior as compared to the systems implementing PRC and MF algorithms.

2.7 Extension on CLWRC scheme

In [1] the system capacity optimization based on the CLWRC algorithm is studied over different number of frames and is compared with the PRC and MF algorithms. This section presents the extension of the work in [1] that includes a variety of case studies with new scenario resulting in new observations. The capacity performance of the system adopting the CLWRC algorithm is studied for a range of SNR values varying from -5 to 30 dB and is compared with the Shannon theoritical limit. This result along with the results in [1] is also presented in [11]. In addition, a case study based on the varying number of users is presented in this section and the multiuser diversity advantage of CLWRC algorithm is confirmed. A complexity comparison of the CLWRC algorithm as compared to other algorithms based on

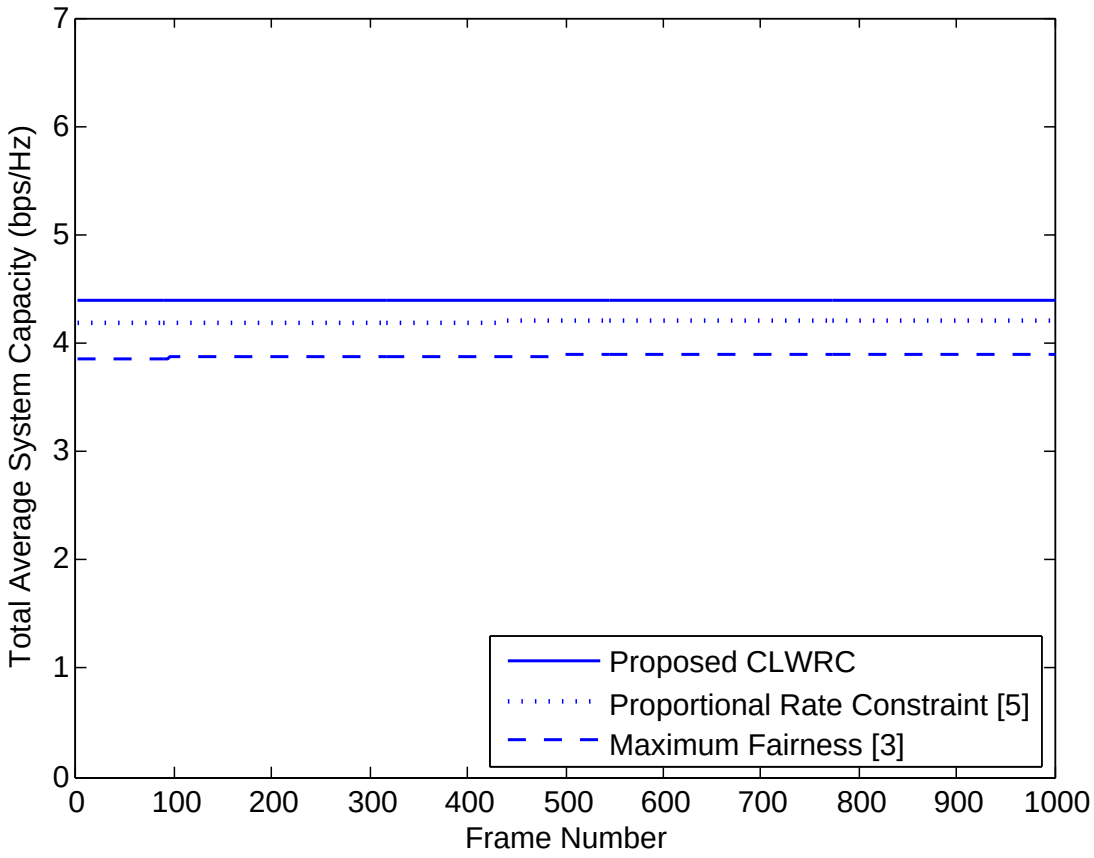

Figure 2.10. Total average system capacity (bps/Hz) vs. frame number (based on simulation parameters in Table 2.1)

the algorithm run time is also presented in this section as a separate case study to verify the system capacity performance.

The results in Fig. 2.11 show the total average system capacity versus average SNR based on the proposed CLWRC algorithm for the optimization problem formulated in section 2.5.3. In this figure the algorithm is executed for different values of average SNR, where for simplicity symbol energy is assumed to be 1 Joule and the system is assumed to be serving 10 users. So, for each value of the average SNR in Fig. 2.11, the corresponding power spectral density of the AWGN channel is evaluated and used in the optimization problem. The remaining of the system parameters needed in the computation are listed in Table 2.1. For performance comparison purposes Fig. 2.11 also shows results for PRC algorithm [5] and MF algorithm [3] along with the proposed CLWRC algorithm. A different approach

of power distribution among users is used in [5] and is discussed in detail in Appendix 5. The figure also shows the Shannon theoretical limit given by $\frac{C}{B} = log_2(1 + \overline{h^2}\overline{\gamma})$, where $\overline{h^2}$ is the average rayleigh fading channel gain and $\overline{\gamma}$ is the average signal to noise ratio, as the upper bound of the capacity curve. The optimized capacity curves for PRC and MF are based on the solution approach proposed in [5], where MF is explained as the special case of PRC. For PRC algorithm, as explained in [5] a set of predetermined capacity weighting factor values among all users are taken to ensure proportional fairness among users. Any of these predetermined values are less than 1 and assigning equal values to all of them results in MF. It can be seen from the figure that the proposed CLWRC algorithm achieves a higher total average system capacity throughout the observed average SNR range (-5 to 30 dB) as compared to PRC and MF algorithms and is closer to the Shannon limit.

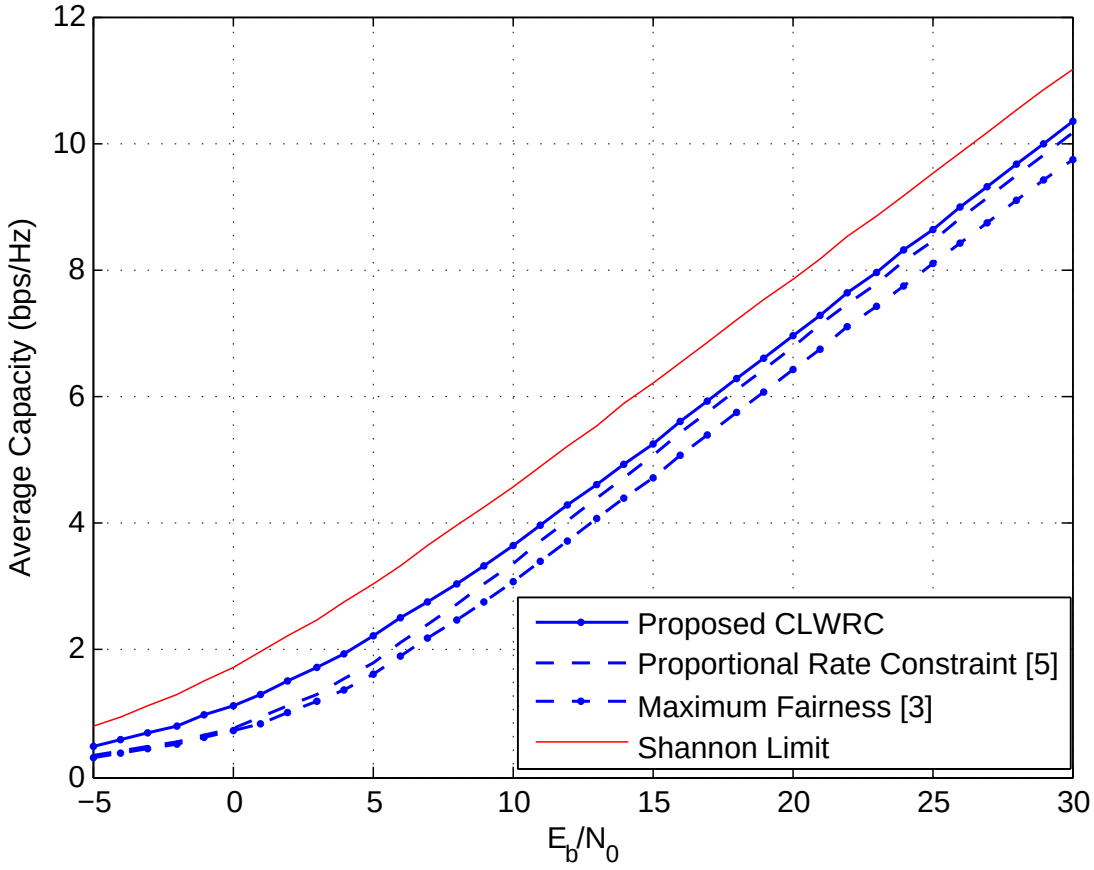

Figure 2.11. Average Capacity (bps/Hz) vs. average SNR per symbol (based on simulation parameters in Table 2.1)

The results in Fig. 2.12 depict the total average system capacity versus number of users. For each case of number of users in the system users are equally assigned over the different service classes with defined weighting factor $\Gamma_x(k)$ (e.g., if K = 30 users, 6 out of these 30 users are assigned to each of the 5 different service classes: i.e, $x(k) = UGS$ for $k = 1, 2, \ldots, 6$, $x(k) = ertPS$ for $k = 7, 8, \ldots, 12$, $\ldots$, $x(k) = BE$ for $k = 25, 26, \ldots, 30$). The values of γ_{SF^x} are taken from Table 2.1, and a system with an average SNR value of 10 dB is assumed. It can be seen from the graph that the proposed algorithm maintains its performance superiority to maximum optimum level as compared to other algorithms for all the range of the number of users. Furthermore, it is also observed in case of the proposed algorithm, the system capacity increases slightly with the increase in number of users but it stays within the Shannon capacity limit, which is 4.8 bits/symbol/Hz for the scenario in Fig 2.12 (as it is clear from Fig. 2.11 for SNR of 10 dB). The reason behind this increment is described as follows. The proposed algorithm maximizes the total system capacity while having constraint on the weighted capacity as governed by (2.5.20). For the user with high QoS requirement, the urgency factor is higher while the satisfaction factor is lower which increases the weighted capacity of the system that the algorithm tries to maintain. Hence, in our optimization process, since the different QoS classes are equally assigned to the users, as the number of users in the system increases the users that belong to QoS class with higher QoS service requirement will contribute in increased total average system capacity as shown in Fig. 2.12. On the contrary, TDMA algorithm compared with here, does not consider urgency and satisfaction factors and hence the capacity is independent of the number of users as is clear in the figure. In case of MF, the algorithm maximizes the system capacity while having constraint on the transmission rate itself and only a slight variation is observed for the scenario considered in Fig. 2.12. However, PRC algorithm considers the fairness parameter and hence the capacity of the system depends on the proportional factor assigned to each user in the system. The PRC algorithm tries to maximize the total system capacity while having constraint on the proportional rate. Therefor, in PRC, the proportional rate

of the user with lower rate is boosted while the one with higher rate is decreased, such that proportionality is maintained among users. So, for higher numbers of users in the system, the users with higher proportional factor will cause the system capacity to decrease and the same is reflected in the figure. If the proportional factor used in PRC for all users is assigned equal, then it is the case of MF and the capacity curves will coincide. It is clear form the results in Fig. 2.12, that the proposed CLWRC algorithm outperforms PRC and MF algorithms in terms of total average capacity maximization. As a conclusion, the superiority of the proposed cross-layer algorithm over PRC and MF in terms of total system capacity maximization is evident from all the results in Figs. 2.11 - 2.15.

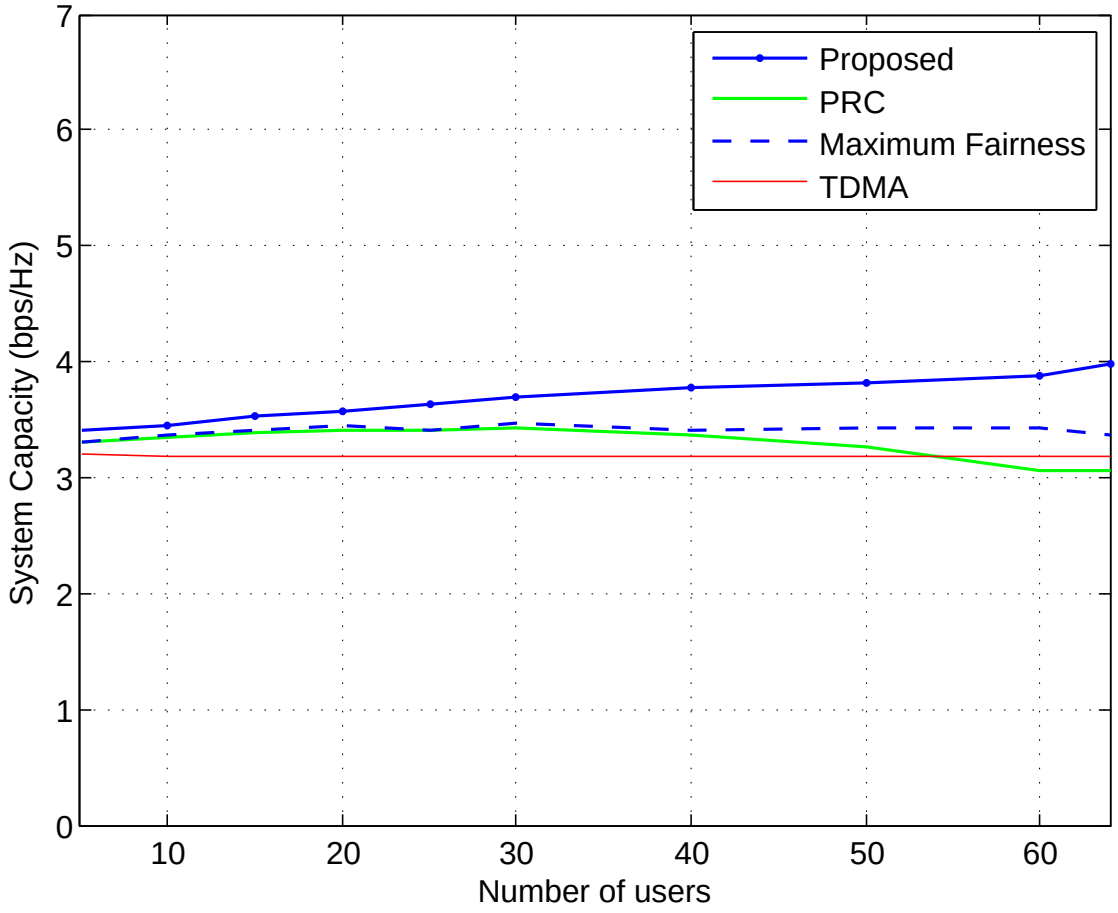

Figure 2.12. Total system capacity (bps/Hz) vs. no. of users (based on simulation parameters in Table 2.2)

The results in Fig. 2.13 depict a comparison between average system capacity for different number of users pertaining to the proposed CLWRC scheme. A scenario of varying number of

users in the system as discussed for the results in Fig. 2.12 is considered and the optimized average system capacity curves using proposed CLWRC algorithm for the system serving 10, 30 and 60 users are plotted in Fig. 2.13. It is evident from the figure that as the number of users in the system increases, there is an improvement in the system capacity and the system capacity gets closer to the Shannon Limit. These results confirm the multiuser diversity advantage in the proposed CLWRC scheme.

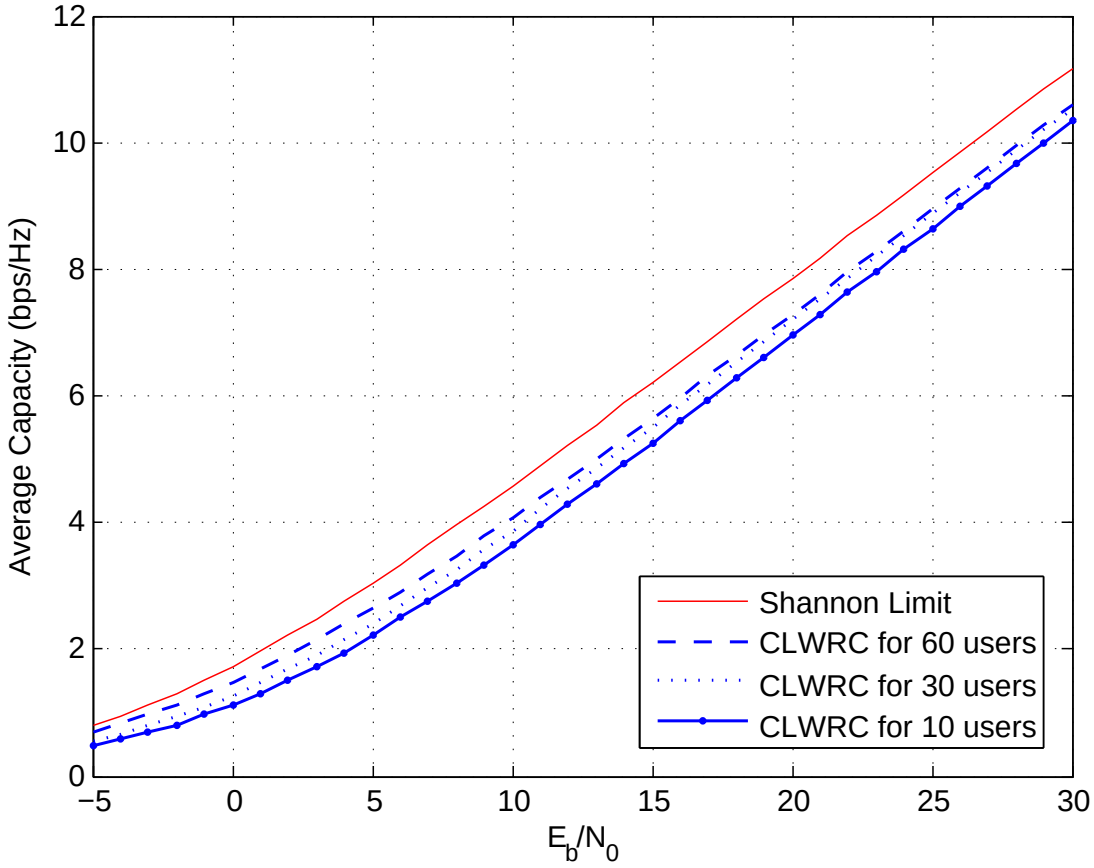

Figure 2.13. Average Capacity (bps/Hz) vs. average SNR per symbol (system serving 10, 30 and 60 users and implementing proposed CLWRC algorithm)

Figs. 2.14 and 2.15 also depict similar results as in Fig. 2.11, for a system serving 25 and 30 users, respectively. It is assumed here that the users are equally assigned over the different service classes for each case, similar as the scenario discussed for results in Fig. 2.12. It is observed from Figs. 2.14 and 2.15 that the proposed CLWRC algorithm achieves a higher total average system capacity throughout the observed SNR range as compared to

PRC and MF algorithms and is closer to the Shannon limit in both figures. There is slight improvement in the total average system capacity over the entire observed SNR range as the number of users increases for CLWRC algorithm. The total average system capacity for MF algorithm also shows an improvement with the increase in the number of users in the system. It is however interesting to note that for PRC algorithm, with the increase in number of users and average SNR, the total system capacity decreases. These results in Figs. 2.14 and 2.15 hence confirm the multiuser diversity in the case of the proposed CLWRC algorithm and is another powerful aspect of the proposed algorithm as compared to PRC algorithm (see the earlier discussions about Fig. 2.12).

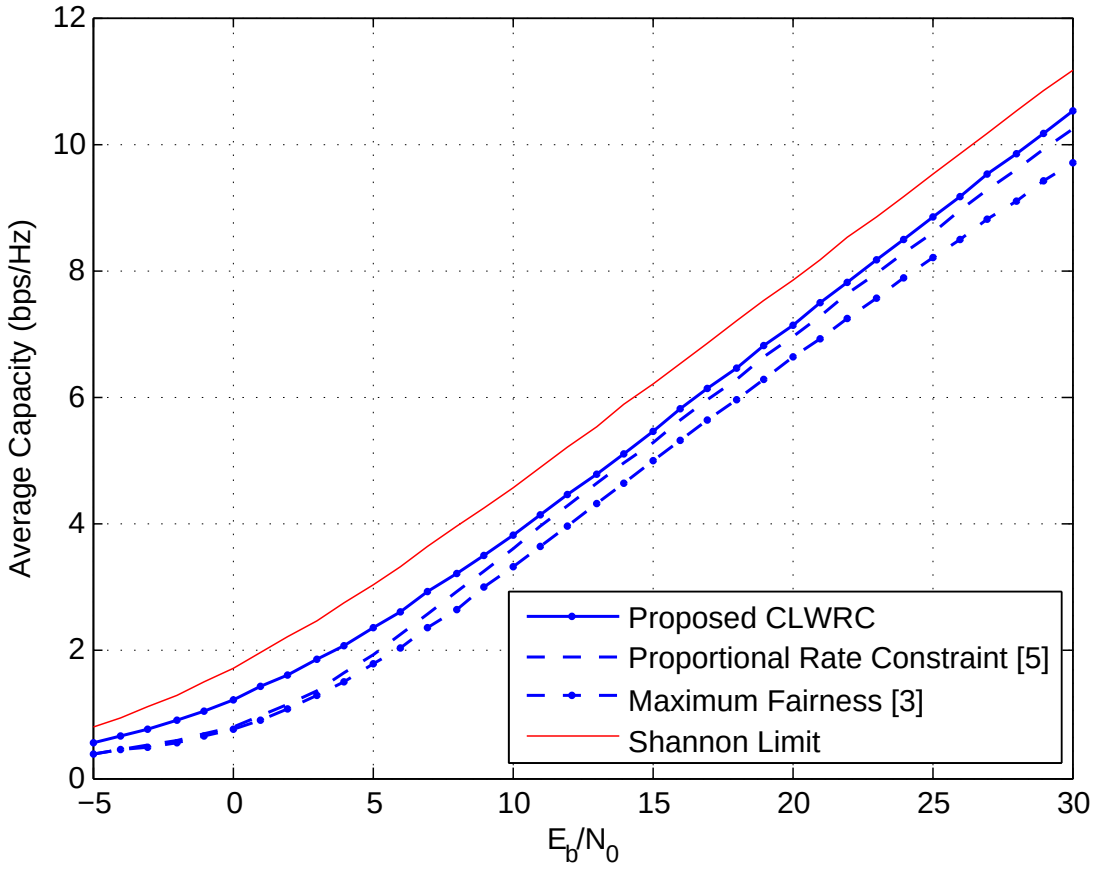

Figure 2.14. Average Capacity (bps/Hz) vs. average SNR per symbol (25 users are assumed to be served by the system)

Table 2.3 shows a comparison between execution times (in seconds) of the proposed CLWRC with the other known PRC and MF algorithms. The algorithm was executed

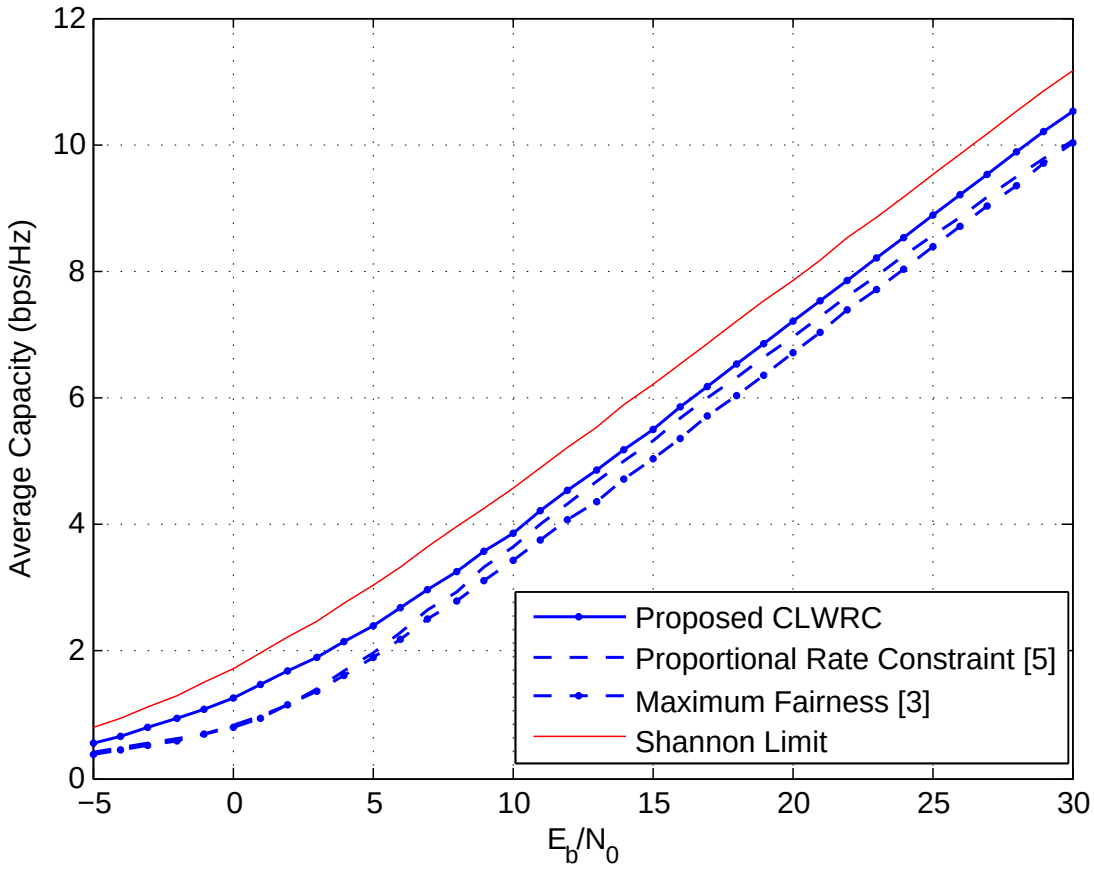

Figure 2.15. Average Capacity (bps/Hz) vs. average SNR per symbol (30 users are assumed to be served by the system)

on Intel(R) Core(TM) i5-2430M CPU @ 2.40 GHz 2.40 GHz processor for 10 MonteCarlo runs. It can be observed form the table that the proposed CLWRC algorithm has a faster execution time as compared to PRC and MF algorithms. This is due to the fact that in both PRC and MF algorithms, after subchannels are allocated to all the users, an initial power allocation algorithm is implemented before finalizing the power allocation based on water-filling approach. However, in the case of proposed CLWRC algorithm right after subchannel allocation, power allocation based on water-filling approach is performed without considering initial power allocation as in [5]. In the CLWRC algorithm, as an initial power allocation, an equal power distribution among each subchannel is considered and hence the execution time for the proposed algorithm is less. It is also interesting to note that, even without following the complex initial power allocation scheme as in [5], the proposed CLWRC algorithm has

better performance in terms of total average system capacity optimization as compared to PRC and MF algorithms.

2.8 Conclusion

A review on technologies like OFDMA and AMC and the air interface technologies like WiMAX and LTE is presented in this chapter. Similarly, various state-of-the-art works are reviewed in this chapter. In this chapter a CLWRC resource allocation scheme proposed in [1] is reviewed and an extension to it by considering different scenarios is also included. Besides evaluating the performance of the CLWRC algorithm for different frame numbers, this chapter presents an extension to the work by evaluating the a performance of the CLWRC algorithm for a range of average SNR (extension presented in [11]) and for different number of users. A capacity comparison between the CLWRC and the other know algorithms are then presented. Moreover, a complexity comparison between the CLWRC and the other algorithms is also performed in this chapter based on the algorithm execution time. From the numerical results, it is observed that the CLWRC scheme results in total average system capacity that is closer to the Shannon limit than other known resource allocation schemes. On the other hand, unlike other known techniques, the CLWRC algorithm not only maintains the optimum system capacity for different number of users in the system but also increases as the number of users increases, confirming the multiuser diversity advantage of the CLWRC algorithm. The CLWRC algorithm also maintains its superiority in terms of execution time as compared to the other algorithms. In particular, the CLWRC scheme outperforms other known approaches in four aspects; closeness to Shannon capacity limit, consistency in terms of maximum optimum capacity throughout the frames considered, consistency in maintaining maximum optimum system capacity for different number of users and fast execution time.

Table 2.2. Traffic Simulation Parameters

Users (k)	SF	Parameter	Value
1	UGS	CODEC	G.729
		Voice Processing Interval	20 ms
		Packet size	66 Bytes
2	UGS	CODEC	G.728
		Voice Processing Interval	30 ms
		Packet size	106 Bytes
3	rtPS	Bernoulli trial ($\bar{p}$)	0.4
		Mean rate	64 Kbps
		Maximum delay	30 ms
4	rtPS	Bernoulli trial ($\bar{p}$)	0.5
		Mean rate	16 Kbps
		Maximum delay	50 ms
5	ErtPS	CODEC	G.723.1
		Voice Processing Interval	30 ms
		Packet size	66 Bytes
		Mean ON period	1.2 sec
		Mean OFF period	1.8 sec
6	ErtPS	CODEC	G.711
		Voice Processing Interval	20 ms
		Packet size	206 Bytes
		Mean ON period	1.2 sec
		Mean OFF period	1.8 sec
7	nrtPS	Mean rate	512 Kbps
		Minimum rate	128 Kbps
8	nrtPS	Mean rate	1 Mbps
		Minimum rate	1 Mbps
9	BE	Pareto mean rate	10558 bps
		Lognormal mean rate	7247 bps
10	BE	Pareto mean rate	10558 bps
		Lognormal mean rate	7247 bps

Table 2.3. Execution time (in seconds) of different algorithms for different number of frames

	Number of Frames		
Algorithms	1000	100	10
CLWRC	107.2528	10.7973	1.1079
PRC	132.6554	13.0038	1.4211
MF	136.1923	13.5061	1.4498

CHAPTER 3

ERROR-FREE SHANNON CHANNEL CAPACITY OPTIMIZATION IN DOWNLINK LTE OFDMA SYSTEMS

3.1 Introduction

A cross-layer optimization design for PRB and power allocations in the downlink LTE system with the objective of maximizing the capacity (in bits/symbol/Hz) subject to fairness parameters and QoS requirements as constraints is presented in this chapter. The proposed scheme is termed as LTE-CLWRC scheme. Based on the literature review on resource allocation schemes for LTE presented in Section 2.5.4, it is known that none of the work reported in literature addresses the problem of cross-layer optimization in LTE downlink system by taking into account the channel conditions, queue status and QoS requirements simultaneously. To address these issues a resource allocation optimization scheme that takes into account, both the channel conditions and the queue status of each user as well as different QoS requirements to maximize the LTE system capacity is proposed. This proposed scheme considers the Cross-layer Weighted Rate Constraint (CLWRC) scheme presented in [11] and extend it accordingly such that it applies to LTE system. Based on the queue status and QoS requirements of users, cross-layer fairness parameters introduced in [11] that add a new dimension to the fairness concept: urgency and satisfaction factors are defined, respectively. Depending on the diverse QoS requirements of different users, resources can be allocated wisely; users that are well served and have no critical QoS requirements to schedule for

service immediately can lag for some time allowing underserved users to access the channel. Optimization of the system performance subject to the constraints on power and cross-layer fairness parameters are studied in this chapter as well. The significant improvement in the performance of the system in terms of maximization of LTE system capacity achieved with the implementation of the LTE-CLWRC approach is demonstrated by the extensive simulation results. In the numerical results section a comparison between the performance of the LTE system with the WiMAX system based on the results reported in [11] is also provided.

This chapter is organized as follows. A system model is presented in Section 3.2. The proposed cross-layer approach, including includes problem formulation and solution approach, is discussed in Section 3.4. Numerical results are presented in Section 3.5 and finally the chapter is concluded with some conclusions in Section 3.6.

3.2 Cross-Layer downlink LTE OFDMA System model

A multiuser downlink LTE OFDMA system is shown in Fig. 3.1. A total of K users sharing L PRBs are considered in the system and the total available transmit power is P_{tot}. The total available system bandwidth, B, is divided into L_{sc} subcarriers such that the bandwidth of each subcarrier is B/L_{sc} and the time slot duration corresponding to each subcarrier is $T_s = \frac{L_{sc}}{B}$. Subsets of these subcarriers form PRBs and they are the smallest allocation unit in LTE. Users can be assigned multiple PRBs at a certain time; however, a PRB can not be shared among multiple users. Data from users arrive from the MAC layer and is placed into an infinite buffer. These buffers follow a first in first out (FIFO) strategy. A channel fading that follows rayleigh distribution with envelope $h_{k,l}$ is assumed to be experienced by a user k over a PRB l. Based on the channel-state-information (CSI) and the information on QoS, the PRB and power allocation algorithm running in enhanced NodeB (eNodeB) optimizes the PRB and power allocation to maximize the error-free Shannon capacity while having a constraint P_{tot}. Moreover, the following assumptions are made: i) outgoing queues for every

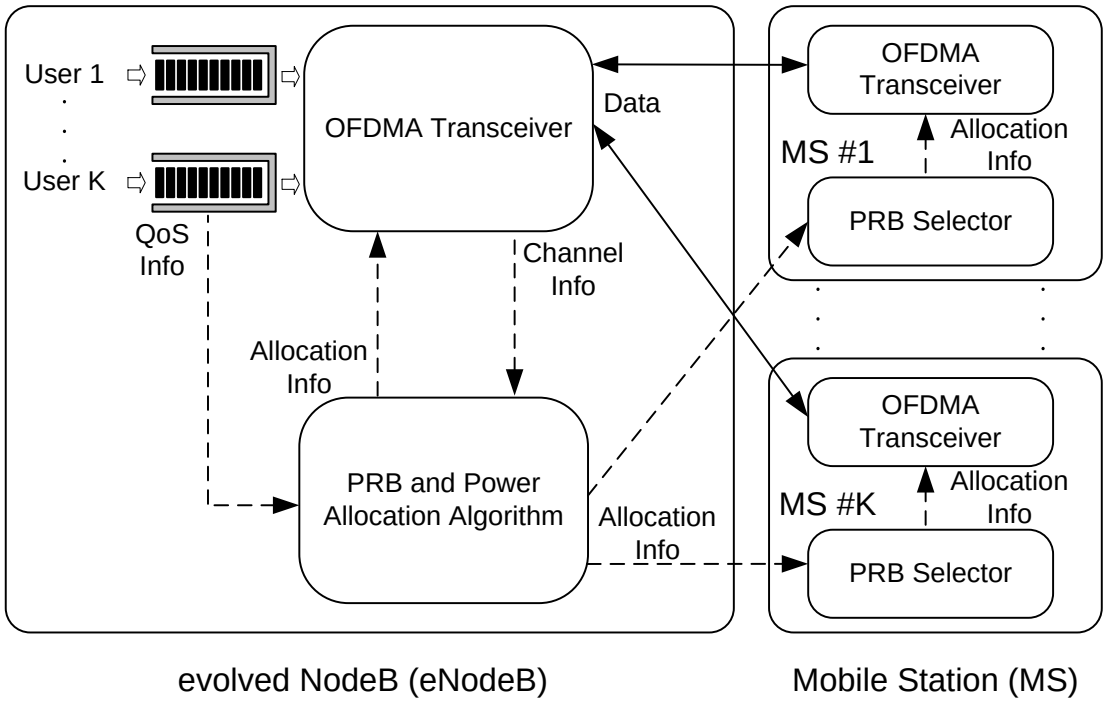

Figure 3.1. Cross-Layer LTE downlink OFDMA Resource-Allocation System

users are infinite; ii) the eNodeB has perfectly received the CSI from all UE; iii) the PRB and power allocation information is sent to each user on a separate channel; iv) coherent bandwidth of the channel is larger than $\frac{B}{L_{sc}}$, which means the channel response on each subcarrier is flat; v) the channel gain remains fixed during one time slot T_s; vi) the channel is varying in time slow enough that users can estimate the channel perfectly; vii) all system parameters and QoS parameters associated with all users are assumed to be made available to the eNodeB during the initial setup (signalling) session before the call takes place.

3.3 LTE QoS, Service Urgency and Service Satisfaction parameters

In this section various QoS class identifiers (QCIs) and the services associated with them, as defined in LTE standard [38, pp. 42], are discussed. This section also reviews the two cross-layer QoS parameters as discussed in [11] and presents the modifications to these parameters as they pertain to LTE system. Each of the QoS classes defined in standards and the two cross-layer QoS parameters, defined herein, are discussed in detail one by one as follows:

3.3.1 LTE QoS classes

LTE EPS bearer is a packet flow established and defined between the PDN-GW and the UE such that the end-to-end connectivity through the LTE network is achieved. EPS uses the concept of a bearer as the central element of QoS control. The user traffic that corresponds to a certain type of service can be differentiated into separate SDFs. SDFs mapped to a bearer receives a common QoS treatment as each bearer maps to a specific set of QoS parameters such as data rate, latency, and packet error rate. Signalling radio bearers carry the radio resource control signaling messages, while the data radio bearers carry the user plane data [12, pp. 355]. The bearers can broadly be divided into two classes:

- Guaranteed Bit Rate (GBR) bearers: A guarantee on a minimum available bit rate for a UE is defined and assured by these bearers. If resources are available in the system then the bit rates can go higher than the minimum allowed bit rate. GBR bearers are typically used for applications such as voice, streaming video, and real-time gaming [12, pp. 356].

- Non-GBR bearers: Unlike GBR bearers these bearers do not define or guarantee a minimum bit rate to the UE. The achieved bit rate depends on the system load, the number of UEs served by the eNode-B and the scheduling algorithm. Non-GBR bearers are used for applications such as web browsing, e-mail, FTP, and peer to peer (P2P) file sharing [12, pp. 356].

Each bearer is associated with a QCI that refers to a packet forwarding treatments including the priority, packet delay budget, acceptable packet error loss rate and the GBR/non-GBR classification. The nine standardized QCIs defined in LTE standard [38, pp. 42] are shown in Table 3.1. The LTE standard [39, pp. 74] suggests FTP and VoIP models for system performance evaluation. It also specifies that the system throughput studies should be assessed using full-buffer traffic model. Therefore, in this chapter, FTP and VoIP traffic models are considered for system performance evaluation. In addition to FTP and VoIP

traffic models, hyper-text transfer protocol (HTTP) and real time video traffic models are also considered in this chapter, which are discussed as follows

- VoIP without silence suppression: This service is associated with $QCI1$ and offers fixed size data packets on a real-time periodic basis such that the overhead and latency associated with the UE's bandwidth request are eliminated. Generally, the VoIP packets are sent every 20 ms or 30 ms [21].

- Moving Pictures Experts Group (MPEG) video: This service is associated with $QCI3$ and offers a real time, variable size data packets on a periodic basis. This allows the UE to specify the size of the desired data packet.

- VoIP with silence suppression: This service is also associated with $QCI1$ and offers real time variable bit rate (VBR) packets with changing bandwidth requirement. An ON/OFF voice model is used to generate this service. The duration of "ON" / "OFF" periods is exponentially distributed with the mean of distribution being the time for which the system is "ON" or "OFF" on average. In [22], [24] this average "ON" and "OFF" time is given as 1.2 seconds and 1.8 seconds, respectively. Similarly the time for which the system resources will be dedicated only for voice processing is 20 milliseconds and the packet size is 66 Bytes.

- File Transfer Protocol (FTP): This service is associated with $QCI8$ and offers data packets on a regular basis. This enables UE to use the contention-based polling in the uplink to request bandwidth [14]. This service is modeled using an exponential distribution with a mean being the size of packet generated on average. In [22], the size of packet generated on average is 512Kbps.

- Hyper-Text Transfer Protocol (HTTP): This service is associated with $QCI9$ and offers a stream of data with no strict QoS parameter [14]. The UE uses only the contention-based polling opportunity to request bandwidth and data are sent whenever resources

are available. This service is well modeled as hybrids having lognormal bodies, and power-law (i.e., Pareto) tails. In [22], [25], 88% of the total area of the PDF about the mean corresponds to the lognormal distribution and the remaining 12% of the total area near the tail region of pdf corresponds to pareto distribution. The mean rate values of lognormal distribution and pareto distribution are 7247 and 10558 bps, respectively i.e, the average size of the packets per second corresponding to the lognormal and pareto distribution is 7247 and 10558 bits, respectively.

3.3.2 Service Urgency

Service urgency discussed in [11] is a cross-layer QoS parameter that is dependent on the information about the queues of the services in the data link layer. Every user in the can be associated with one of the nine different QCIs listed in Table 3.1. Let x denotes any one of the nine QCIs such that x is any element in the set $\{QCI1, QCI2, \ldots, QCI9\}$ and $x(k)$ denotes x associated with a user k. Now, let N be the total number of frames considered, then n is defined as the frame number being served such that $n \in \{1, 2, \ldots, N\}$. Also, let $A_k^x(n)$ be the number of bits arriving at the queue of a k^{th} user associated with an x QCI during a frame n, $Q_k(n)$ be the length of queue of the k^{th} user associated with an x QCI during a frame n and $B_k^x(n)$ be the number of bits the eNodeB serves from the queue of the k^{th} user associated with an x QCI during a frame n. Then the queue length corresponding to the k^{th} user associated with an x QCI during a frame $n+1$ is given by

$$Q_k^x(n+1) = Q_k^x(n) + A_k^x(n) - B_k^x(n) \tag{3.3.1}$$

Also, let Ω^x be the set of all users associated with the same x. Then the set Ω^x is expressed as

$$\Omega^x = \{1 \leq k \leq K : x(k) = x\} \qquad \forall\ x \in \{QCI1, QCI2, \ldots, QCI9\} \tag{3.3.2}$$

and let $Q^x(n)$ be the aggregate queue length corresponding to users associated with the same QCI during frame n, then $Q^x(n)$ can be expressed as

$$Q^x(n) = \sum_{k \in \Omega^x} Q_k(n) \tag{3.3.3}$$

Finally, the normalized queue length of the k^{th} user associated with an x QCI during a frame n, $U_k^x(n)$, which will be called henceforth the Urgency Factor, can be defined as

$$U_k^x(n) = \begin{cases} \frac{Q_k^x(n)}{Q^x(n)} & \forall\ x \in \{QCI2, QCI3, \ldots, QCI9\} \\ 1 & \forall\ x \in \{QCI1\} \end{cases} \tag{3.3.4}$$

It should be noted here that the Urgency Factor $U_k^x(n)$, is set to 1 for bearer associated with QCI that serves VoIP traffic. It is known form the QoS requirements that the users associated with VoIP traffic should be allocated resources on a periodic basis and therefore the concept of urgency does not apply. It should also be noted here that $U_k^x(n) \in (0, 1]$, and QCI associated with VoIP traffic is thus assigned the highest urgency factor. However, the urgency factor for other QCI corresponding to various other traffic is calculated using (3.3.4). Here the queue length associated with a given k^{th} user at a given frame number n corresponding to a given QCI is normalized by the total queue length of that particular QCI in the system. This normalised factor, urgency factor, will be high for the QCI with the longest queue associated with a given k^{th} user during a frame n and hence this user will be served first. The significance of the urgency factor is two-fold: firstly, it gives indication about which user is being under-served relative to other users of the same QCI; secondly, it also conveys information about the queue length of the user to the resource allocation algorithm. The higher the value of $U_k^x(n)$, the more it is urgent to allocate resources to the user.

3.3.3 Service Satisfaction

One of the important QoS parameters defined in LTE standard is the packet delay budget for a given QCI which corresponds to the maximum allowable latency for the service corresponding to the given QCI. If the waiting time of the packet in a queue exceeds a maximum

Table 3.1. LTE standardize QoS Class Identifier (QCI)

QCI	Resource Type	Priority	Packet Delay Budget (ms)	Packet Error Loss Rate	Example services
1	GBR	2	100ms	10^{-2}	Conversational voice
2		4	150ms	10^{-3}	Conversational video (live streaming)
3		3	50ms	10^{-3}	Real time gaming
4		5	300ms	10^{-4}	Non-conversational video (buffered streaming)
5	Non-GBR	1	100ms	10^{-6}	IMS signalling
6		5	300ms	10^{-6}	Video (buffered streaming)
7		7	100ms	10^{-3}	Voice Video Interactive gaming
8		8	100ms	10^{-6}	Voice Streaming, TCP based (FTP)
9		9			

allowed latency or the deadline T_k^x, of a k^{th} user associated with an x QCI, then a timeout is set by the scheduler as T_k^x. Hence, a delay satisfaction indicator corresponding to an x QCI for a given k^{th} user at a given frame number n, denoted by $DS_k^x(n)$, based on the delay budget parameter defining the maximum allowed latency, is defined as [26]

$$DS_k^x(n) = max\{1, T_k^x - \Delta T_k^x - W_k^x(n) + 1\} \quad \forall\ x \in \{QCI1, QCI2, \ldots, QCI9\} \tag{3.3.5}$$

where $W_k^x(n) \in [0,\ T_k^x]$ is the head of line (HOL) delay which is defined as the longest waiting time that a packet experiences for a k^{th} user associated with an x QCI at a given frame number n and $\Delta T_k^x \in [0,\ T_k^x]$, for a k^{th} user associated with an x QCI, is the guard time region ahead of the deadline T_k^x, which indicates the time remaining before which the packet should be scheduled to avoid timeout. Now, if $W_k^x(n) \in [0,\ T_k^x - \Delta T_k^x]$, then $DS_k^x(n)$ will be greater than or equal to 1 indicating that the delay requirement of the service is satisfied, as the packet will be served before timeout [26]. And if $W_k^x(n) \in [T_k^x - \Delta T_k^x,\ T_k^x]$, then $DS_k^x(n)$ will be less than 1 indicating that the delay requirement of the service is not satisfied, as the packet can't be served before timeout. So a lower value of delay satisfaction factor will require a scheduling algorithm to allocate more resources to the service to meet the delay requirements.

Furthermore, other QoS parameter defined in LTE standard is the priority of the service. Different priorities have been assigned to different QCI such that the services that has highest

priority shall be served first. Let PR_x be the priority defined in the standard for x^{th} QCI, where PR_x takes the values in the set of $\{1, 2, \ldots, 9\}$. A priority of value 1 corresponding to a QCI signifies that the service associated with that QCI should have the first priority, likewise a priority of value 2 signifies that the service should have second priority. A priority factor corresponding to a QCI, Γ_x, dependent on the priority parameter is defined as

$$\Gamma_x = \frac{1}{PR_x} \quad \forall\ x \in \{QCI1, QCI2, \ldots, QCI9\} \tag{3.3.6}$$

It is to be noted that $\Gamma_x \in [0, 1]$, such that as the value of PR_x is high, that is, the service has the lowest priority, the corresponding priority factor Γ_x is also lower.

Now let $S_k^x(n)$ be the satisfaction factor corresponding to an x^{th} QCI and associated with a k^{th} user during a frame n such that $x \in \{QCI1, QCI2, \ldots,\ QCI9\}$. Based on the delay satisfaction indicator and the priority factor defined above, the satisfaction factor is defined as,

$$S_k^x(n) = \frac{DS_k^x(n)}{\Gamma_x} \quad \forall\ x \in \{QCI1, QCI2, \ldots, QCI9\} \tag{3.3.7}$$

It is important to note that the service satisfaction is inversely proportional to the priority factor, such that the services with higher priority will have the lowest service satisfaction and hence these services are more resource hungry. Therefore, a scheduling algorithm is accordingly developed such that it prioritises the services with lowest service satisfaction. The significance of the satisfaction factor is also two-fold: firstly, it allows for scalability as when the system is overloaded the performance of the users with low-priority service classes will be degraded prior to those with high priority service classes; secondly, it also allows users with low-priority service classes to lead when users with higher-priority service classes are well satisfied.

3.4 Proposed Cross-layer Algorithm

3.4.1 Proposed Algorithm: Optimization Problem Formulation

Let $P_{k,l}(n)$ be the power allocated to a k^{th} user over a PRB l during a frame n, N_0 denote the additive white Gaussian noise (AWGN) power spectral density with zero mean, $h_{k,l}$ be the channel gain for a user k over a PRB l, and $\rho_{k,l} \in \{0, 1\}$ indicates whether or not a PRB l is used by the user k. Then the spectral efficiency or the channel capacity, in bits/symbol/Hz, for a k^{th} user associated with an x QCI during a frame n is expressed as

$$C_k^x(n) = \sum_{l-1}^{L} \frac{\rho_{k,l}}{L} log_2(1 + P_{k,l}(n)H_{k,l}(n)) \qquad bits/symbol/Hz \tag{3.4.1}$$

where

$$H_{k,l}(n) = \frac{h_{k,l}^2}{N_0 \frac{B}{L}} \tag{3.4.2}$$

and the weighted capacity, $R_k^x(n)$, for a k^{th} user associated with an x QCI during a frame n is then expressed as

$$R_k^x(n) = \frac{U_k^x(n)}{S_k^x(n)} C_k^x(n) \tag{3.4.3}$$

The weighted capacity in (3.4.3) incorporates both the urgency factor and satisfaction factor. The priority factor is considered within the satisfaction factor. Urgency corresponds to the scheduling priority of a QCI based on the queue status of that QCI and it is known from the explanation of urgency factor that the QCIs with higher queue lengths have higher urgency factor except for the $QCI1$ where it has highest urgency irrespective of the queue lengths. So, the QCI with highest service urgency requirement needs to be scheduled first and hence the weighted capacity is so defined that it is directly proportional to the urgency factor. However, the satisfaction factor indicates the satisfaction level of the QCIs. If a specified data rate requirement, a delay requirement or any other requirements specific to the QCI is met, then the satisfaction is high, also the QCI with lower priority increases the satisfaction. So, the QCI with highest satisfaction can be scheduled later and hence the weighted capacity

is defined to be inversely proportional to the service satisfaction. Since a k^{th} user is always associated with any one of the x QCIs, and hence for simplicity, from this point onward the superscript x is dropped in the mathematical formulations. Having said that, now the fairness constraint is defined as

$$R_i(n) = R_j(n) = R(n) \qquad \forall i, j \in \{1, 2, \ldots, K\} \tag{3.4.4}$$

Based on the above discussion, the optimization problem can be expressed mathematically as

$$\max_{P_{k,l}, \rho_{k,l}} \quad C = \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{\rho_{k,l}}{L} \log_2 \left(1 + P_{k,l} H_{k,l}\right) \; bits/symbol/Hz \tag{3.4.5}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \sum_{l=1}^{L} P_{k,l} \leq P_{tot} \tag{3.4.6}$$

$$P_{k,l} \geq 0 \quad \forall \quad k, l \tag{3.4.7}$$

$$\rho_{k,l} = \{0, 1\} \quad \forall \quad k, l \tag{3.4.8}$$

$$\sum_{k=1}^{K} \rho_{k,l} = 1 \quad \forall \quad l \tag{3.4.9}$$

$$R_i(n) = R_j(n) = R(n) \, \forall \, i, j \in [1, K], \tag{3.4.10}$$

where P_{tot} is the total available power. The first constraint (3.4.6) implies that the total power used by PRBs is not to exceed the total available system power. The second constraint (3.4.7) states that the power used by all PRBs should be non-negative. In the third constrain (3.4.8), $\rho_{k,l}$ is only allowed to be 0 or 1 which assures that a l^{th} PRB is either used or not used by the k^{th} user. Furthermore, no sharing of PRBs among users is allowed, which is stated by the fourth constraint (3.4.9). The last constraint (3.4.10) is the fairness constraint presented in (3.4.3) and (3.4.4).

It should be noted that a similar formulation was presented in [5] to address different constraints. In [5], the weighted user data rate, R_k, is defined based on the predetermined values, β_k, as $R_k = \frac{C_k}{\beta_k}$, where C_k is the capacity for a k^{th} user defined as in (3.4.1). These weighted user data rates are then used to ensure proportional fairness among users such that

$R_1 = R_2 = \ldots = R_k$ which is also equivalent to $C_1 : C_2 : \ldots : C_K = \beta_1 : \beta_2 : \ldots : \beta_K$ [1]. While for the LTE-CLWRC algorithm, herein, the weighted user data rate, R_k, is defined in (3.4.3) that is based on the cross-layer fairness parameters: service urgency and service satisfaction that are discussed in Sections 3.3.2 and 3.3.3, respectively.

3.4.2 Proposed Algorithm: Problem Solution and Implementation

The optimization problem given in (3.4.5) is very hard to solve. It is a mixed binary integer programming problem. The problem has nonlinear constraints as well as continuous variables, $P_{k,l}$, and binary variables, $\rho_{k,l}$. An optimum solution for this optimization problem exists which is highly computationally complex and is not favored considering the high frequency of executing the schedulers in practical systems. The PRB and power allocation decisions are to be taken for every frame, therefore, it is usually the case that a suboptimal solution is adopted which approaches the optimal results.

For a system with K users and L PRBs, there are K^L possible PRB allocations and for every allocation the optimal power allocation can be computed. Even though it is possible to compute the global maximum solution, a suboptimal greedy approach is presented in this work and optimality is compromised for complexity reduction. An analytical solution to the optimization problem in (3.4.5) can be obtained using the method of Lagrange multipliers as follows. For a given PRB allocation, Π_k, such that Π_k is the set of PRBs allocated to the user k, the capacity of the user k during a frame n, in bits/symbol/Hz, is expressed as

$$C_k(n) = \sum_{l \in \Pi_k} \frac{1}{L} log_2(1 + P_{k,l} H_{k,l}), \tag{3.4.11}$$

[1] The identity $x_1 : x_2 : \ldots : x_k = y_1 : y_2 : \ldots : y_k$ means $\frac{x_i}{y_i} = \frac{x_j}{y_j} \quad \forall i, j = 1, 2, \ldots, K$

Then the optimization problem in (3.4.5) is reformulated

$$\max_{P_{k,l}} \quad C = \sum_{k=1}^{K} \sum_{l \in \Pi_k} \frac{1}{L} log_2 \left(1 + \frac{P_{k,l} h_{k,l}^2}{N_0 \frac{B}{L}} \right) \; bits/symbol/Hz \tag{3.4.12}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \sum_{l \in \Pi_k} P_{k,l} \leq P_{tot} \tag{3.4.13}$$

$$P_{k,l} \geq 0 \;\; \forall \;\; k, l \tag{3.4.14}$$

$$\Pi_i \cap \Pi_j = \Phi \;\; \forall \;\; i \neq j \tag{3.4.15}$$

$$\Pi_1 \cup \Pi_2 \cup \ldots \cup \Pi_K \subseteq \{1, 2, \ldots, L\} \tag{3.4.16}$$

$$R_i(n) = R_j(n) = R(n) \; \forall \; i, j \in [1, 2, \ldots, K], \tag{3.4.17}$$

The solution to the optimization problem in (3.4.12) results in

$$P_{k,l} = P_{k,1} + \frac{H_{k,i} - H_{k,1}}{H_{k,i} H_{k,1}} \tag{3.4.18}$$

for $k \in \{1, 2, \ldots, K\}$ and $l \in \{1, 2, \ldots, L\}$. This result is obtained by solving optimization problem in (3.4.12) using the method of Lagrange multipliers. The derivation is provided in Appendix 5. The expression in (3.4.18) is the water-filling equation, which means that PRBs with higher SNR receive more power in order to maximize the capacity. A similar equation was obtained in [5] for different constraints (as indicated in Section 3.4.1).

Although the optimal solution is to jointly allocate PRBs and power, a less complex two-phase greedy approach is adopted here that results in a suboptimal solution of the optimization problem formulated in this study. The two-phase greedy approach starts off by allocating PRBs to users assuming equal power allocation to all the PRBs. After PRB allocation is complete, power allocation is performed in order to maximize the total system capacity while maintaining fairness and QoS support. In the following section, resource allocator algorithm is discussed in detail. The solution altogether offers a low delay cost effective approach.

Resource Allocator

The proposed resource allocator algorithm based on the two-phase greedy approach is shown in Algorithm 1. The terms used in this algorithm are defined as follows: T_f is the frame duration and T is the total traffic duration, such that $T = N \times T_f$, T_c & T_s are the in-phase & quadrature phase E-field components of Rayleigh fading channel, respectively. $P_{k(tot)}$ is the initial total power allocation to a k^{th} user, $\bar{C}$ is the exponentially weighted average capacity.

Algorithm 1 LTE Resource Allocator

Input: $K,\ L,\ P_{tot},\ B,\ N,\ T_f,\ N_0,\ \gamma_x,\ \Delta T,\ t_c,\ T_k$
Initialize Array: $\bar{C}_k \Leftarrow 0,\ Q_k \Leftarrow 0$
generate $Q_k^x(n) \quad \forall\ n \in \{1,\ 2,\ \ldots,\ N\}$ and $\forall\ k$ with associated x QCI
for $n = 1 \rightarrow N$ **do**
    $find\ Q^x(n)$ // (3.3.3)
    $find\ U_k(n)$ // (3.3.4)
    $find\ S_k(n)$ // (3.3.7)
    generate T_c & T_s
    $h_{k,l} \Leftarrow T_c + jT_s$
    invoke PRB Allocator // *assigns* $\rho_{k,l}$
    for $k = 1 \rightarrow K$ **do**
        $P_{k(tot)} \Leftarrow \frac{P_{tot}}{L} \sum_i \rho_{k,i}$
        invoke Power Allocator // *assigns* C_k
    end for
    $C(n) \Leftarrow \sum_{k=1}^{K} C_k$
    if $n = 1$ **then**
        $\bar{C}(n) \Leftarrow C(n)$
    else
        $\bar{C}(n) \Leftarrow \bar{C}(n-1) * (1 - \frac{1}{t_c}) + C(n) * \frac{1}{t_c}$
    end if
end for

The working of the resource allocator algorithm depicted in Algorithm 1 is described in detail in the following. It first reads the queues lengths, $Q_k(n)$, QCIs associated with each user, x and the maximum delay accepted for QCIs, T_k^x from the MAC layer. Likewise, K, L, P_{tot}, B, N_0, γ_x, t_c, and ΔT are configured by the allocator. With all the information in hand

traffic corresponding to different types of QCIs are simulated as explained in Appendix 5 and queue lengths corresponding to each QCIs are calculated as explained by (3.3.3). Service urgency and satisfaction parameters are then evaluated using (3.3.4)-(3.3.7) and a Rayleigh fading channel based on Clarks's model is simulated as explained in Appendix 5.

The algorithm then proceeds forward by invoking the PRB allocation, which starts with assigning the PRB with maximum channel gain for each user in Rayleigh fading environment. The total available system power is equally divided among PRBs and the weighted data rates for each user dependent on the urgency and satisfaction factor are then calculated. The weighted data rates so generated are then evaluated to allocate the remaining PRBs to the users such that the fairness among users in terms of weighted data rate is maintained. The details of PRB allocator is discussed in Appendix 5. Once PRB allocation is completed, power allocator algorithm based on the derivation in Section 3.4.2 is invoked, the detailed description of this algorithm is provided in Appendix 5. Finally the overall exponentially weighted total average system capacity is evaluated.

3.5 Simulations and Numerical Results

In this section we numerically implement and simulate the solution described in Section 3.4.2 for the optimization problem presented in Section 3.4.1 based on Algorithm 1. Table 3.2 shows the system parameters used for the simulation. In this chapter, the same wireless channel model as in [5] is considered which is a frequency-selective AWGN channel with zero mean consisting of six independent Rayleigh multipaths. Each multipath component is modeled by Clarke's flat fading model (see Appendix 5). It is assumed that the power delay profile is exponentially decaying with e^{-2l} rate, where l is the multipath index and $l \in \{0, 1, \ldots, 5\}$. Hence, the relative power of the six multipath components are $[0, -8.69, -17.37, -26.06, -34.74, -43.43]$ dB. A Doppler shift of 30 Hz is assumed and the total available bandwidth and transmit power are 20 MHz and 1 W, respectively. It is

important to note that the subcarrier spacing of 15 kHz provides a larger OFDM symbol duration and is able to combat the large delay spread such that each subcarriers will experience relatively flat fading [12, pp. 242]. Likewise, five different traffic models (described in Section 3.3.1) were used to simulate the arrival patterns of the five different service flows. Table 3.3 summarizes the characteristics of the 10 different active users' service flow in the system based on five different service flows.

Table 3.2. Simulated System Parameters

Symbol	Parameter	Value
P_{tot}	Total system power	1 Watt
L	Number of PRBs	48
K	Number of users	10
N	Number of frames	1000
B	Total system bandwidth	20 MHz
E_s	Symbol Energy	1 Joule
T_f	Frame duration	1 ms
ΔT	Guard time ahead of deadline	20 ms
T_c	Moving average window size	1000 ms

3.5.1 Performance Comparison (LTE vs. WiMAX)

The results in Fig. 3.2 show a capacity comparison between the LTE and WiMAX systems both implementing the CLWRC algorithm. The available system resources are as follows: bandwidth of 5 MHz and power of 1 Watt are assumed for both systems; and a total of 10 users are also assumed to be served by each system. The same channel model discussed in Section 2.6 is considered for both systems. The result corresponding to WiMAX is taken from [11]. It is observed from the results that the total achievable system capacity throughout the observed SNR range is higher for the LTE system as compared to WiMAX. This superiority of LTE over WiMAX is due to the fact that the transmission time interval for scheduling in LTE is small (TTI = 1 ms) which makes the resource allocation more dynamic and frequent as compared to WiMAX where TTI is 5 ms, which makes the system resources be handled more efficiently in the case of LTE. Moreover, there is less overhead in case of LTE as

compared to WiMAX. The overhead evaluation for both LTE and WiMAX, based on the standards, is discussed in the sequel.

A subcarrier bandwidth of 10.94 kHz is considered by WiMAX standard [23]. Hence, a total system spectrum bandwidth of 5 MHz can handle 457 subcarriers. In PUSCs sub-channelization implementation, an OFDMA symbol consists of 360 data-bearing subcarriers and 60 pilot subcarriers, which adds up to 420. Also, the PUSCs implementation consists of additional 92 guard band subcarriers, which are the remaining 37 subcarries of the real 457 subcarriers plus 55 dummy subcarriers from the zero padding used in IFFT/FFT implementation of the OFDMA system. Since the 5 MHz bandwidth consists of 457 real subcarriers the input and the output of the IFFT/FFT process will take $2^n = 512$, n is an integer, as input subcarriers (457 real and 55 dummy). Hence, for the WiMAX system of 5 MHz bandwidth 3.94 MHz (360×10.94 kHz) is available for data and remaining 1.06 MHz ($(60 + 37) \times 10.94$ kHz) is an overhead for the system. Hence, the total overhead in the case of WiMAX occupies 21.2% of the available bandwidth. On the other hand, a subcarrier bandwidth of 15 kHz is considered by LTE standard [16], and for a system bandwidth of 5 MHz, an LTE OFDMA symbol consists of 300 data-bearing subcarriers. LTE allocates only 10% of the total system bandwidth for guard band subcarriers and reference signals (as explained in Section 2.3 LTE uses reference signal for channel estimation unlike pilots in WiMAX). Hence, for a LTE system of 5 MHz, only 0.5 MHz, as an overhead to the system, are reserved for guard band subcarrier and reference signals while the remaining 4.5 MHz is allocated for data-bearing subcarriers. This shows that the LTE has less overhead as compared to WiMAX; one of the possible reasons that LTE outperforms WiMAX.

Remark: *It is however, important to note here that the results are based on the assumption that both the LTE and WiMAX systems are serving* 10 *users, that is, none of the systems are overloaded and both the systems are capable of handling those users simultaneously. The maximum number of users that the LTE and WiMAX systems can support*

are different. In PUSCs subchannelization implementation of WiMAX, 6 *subcarriers distributed pseudo-randomly across the frequency spectrum constitute a subchannel. Therefore, a total of* 60 *subchannels are available in the system with* 5 *MHz. While, in LTE,* 12 *PRBs each consisting of* 25 *consecutive data-bearing subcarriers are available in* 5 *MHz bandwidth [16, pp. 26]. Hence, it is known from the above discussion that for a total bandwidth of* 5 *MHz, in WiMAX, up to* 60 *users can be served at the same time while in LTE at most 12 users can be served simultaneously. In LTE, if there are more than* 12 *users in the system then the remaining users are not served at all. Hence, for the system with more than* 12 *users, it is likely that the WiMAX system will outperform the LTE system. The proposed algorithm in this thesis does not consider the congested system. For the congested system, the algorithm needs to be changed accordingly by introducing new parameters. LTE-Advance (LTE-A) introduces a new parameter called allocation and retention priority (ARP) specifically for congested system; however, this parameter is not considered in this study and is left for a future work.*

3.5.2 Capacity Comparison

The results in Fig. 3.3 show the total average system capacity versus average SNR based on the proposed LTE-CLWRC algorithm that implements the solution approach presented in Section 3.4.2 for the optimization problem formulated in Section 3.4.1. In this figure the algorithm is executed for different values of average SNR, where for simplicity symbol energy is assumed to be 1 Joule and the system is assumed to be serving 10 users. So, for each value of the average SNR in Fig. 3.3, the corresponding power spectral density of the AWGN channel is evaluated and used in the optimization problem. Furthermore, the other system parameters needed in the computation are listed in Table 3.2. For performance comparison purposes Fig. 3.3 also shows results for the PRC [5] and the MF [3] algorithms along with the LTE-CLWRC. A different approach of power distribution among users is used and is discussed in The power distribution approach among users that is used in [5] and MF

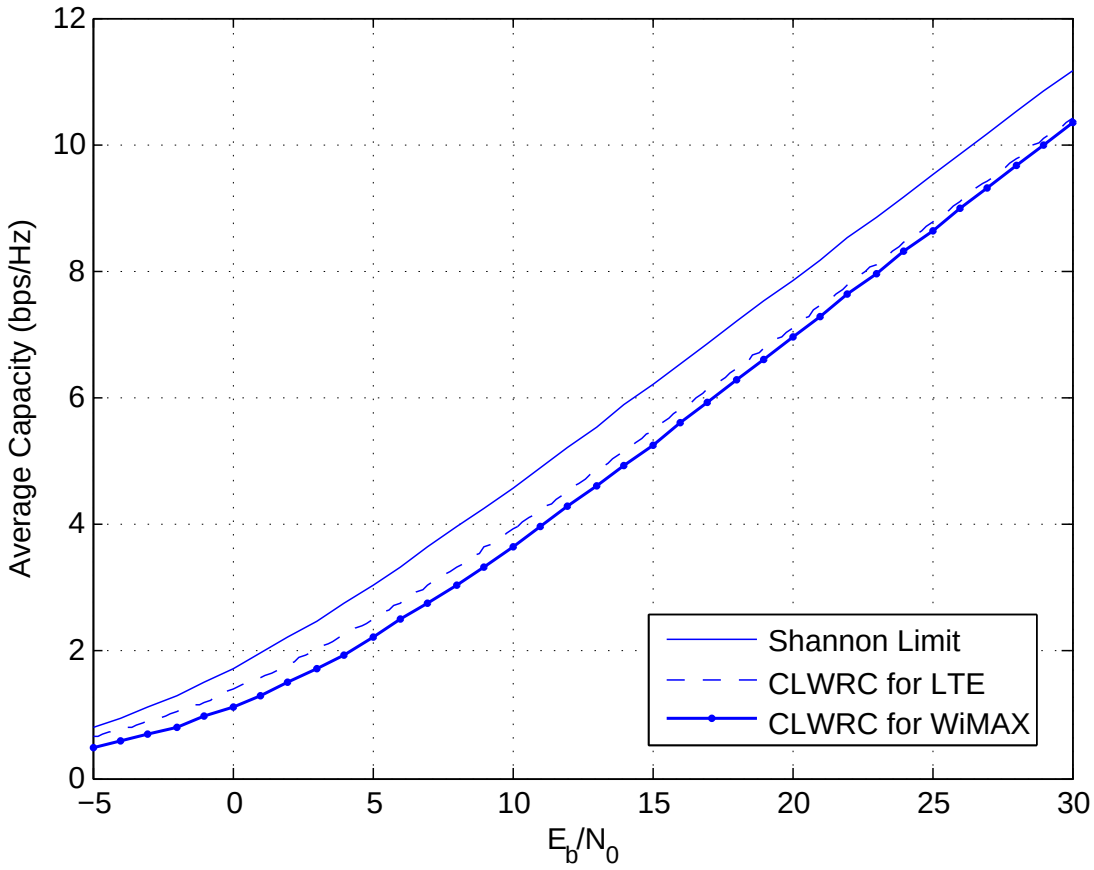

Figure 3.2. Capacity comparison between LTE and WiMAX

[3] is referenced here in Appendix 5). The figure also shows the Shannon theoretical limit given by $\frac{C}{B} = log_2(1 + \overline{h^2}\overline{\gamma})$, where $\overline{h^2}$ is the average rayleigh fading channel gain and $\overline{\gamma}$ is the average signal to noise ratio, as the upper bound of the capacity curve. The optimized capacity curves for the PRC and MF are based on the solution approach proposed in [5], where MF is explained as the special case of the PRC. For the PRC algorithm, as explained in [5] a set of predetermined capacity weighting factor values among all users are taken to ensure proportional fairness among users. Any of these predetermined values are less than 1 and assigning equal values to all of them results in MF. It can be seen from the figure that the LTE-CLWRC algorithm achieves a higher total average system capacity throughout the observed average SNR range (-5 to 30 dB) as compared to the PRC and MF algorithms and is closer to the Shannon limit.

Likewise, the result in Fig. 3.4 depicts the total average system capacity versus frame

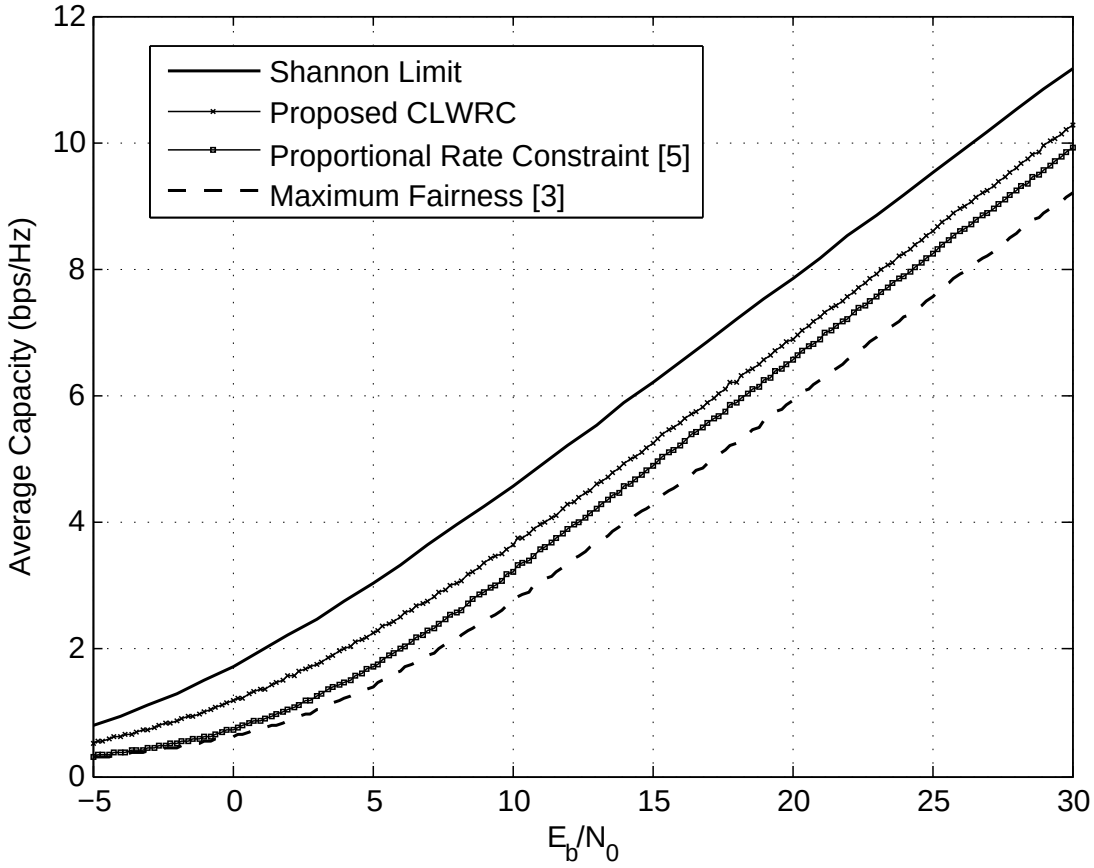

Figure 3.3. Average Capacity (bps/Hz) vs. average SNR per symbol (based on simulation parameters in Table 3.2)

number (1 to 1000). The system parameters listed in Table 3.2 are used in this case as well. The total average system capacity achieved during each arrival time duration of a frame is depicted in the figure, for an average SNR of 10 dB, using the LTE-CLWRC, PRC and MF algorithms. It is obvious from the figure that the LTE-CLWRC algorithm remains superior as compared to the systems implementing the PRC and MF algorithms.

The results in Fig. 3.5 depict the total average system capacity versus number of users. For each case of number of users in the system users are equally assigned over the different QoS classes with associated QCI x (e.g., if K = 30 users, 6 out of these 30 users are assigned to each of the 5 different QoS classes with associated QCI: i.e, $x(k) = QCI1$ for $k = 1, 2, \ldots, 6$, $x(k) = QCI3$ for $k = 7, 8, \ldots, 12$, $x(k) = QCI1$ for $k = 13, 14, \ldots, 18$, $x(k) = QCI8$ for $k = 19, 20, \ldots, 24$, $x(k) = QCI9$ for $k = 25, 26, \ldots, 30$). The traffic parameters corresponding to

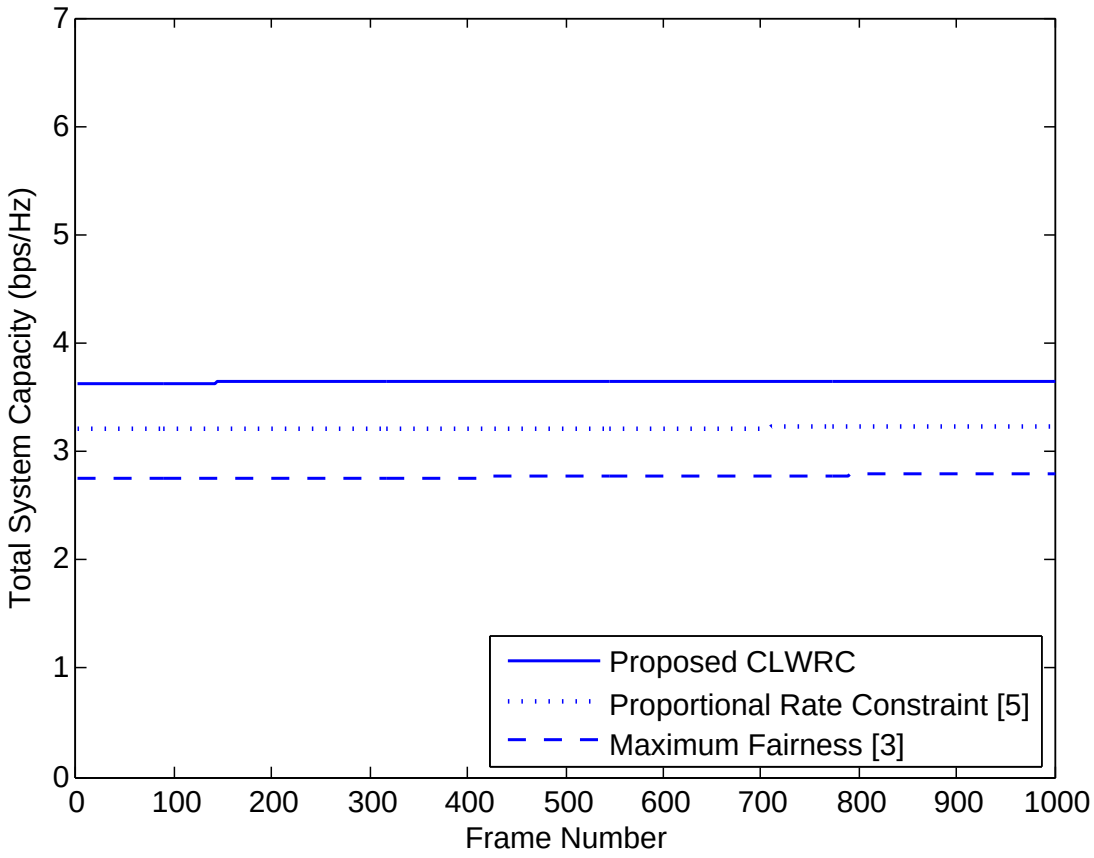

Figure 3.4. Total average system capacity (bps/Hz) vs. frame number (based on simulation parameters in Table 3.2)

users associated with a QCI is taken from Table 3.3, and a system with an average SNR value of 10 dB is assumed. It can be seen from the graph that the proposed algorithm maintains its performance superiority to maximum optimum level as compared to other algorithms for all the range of the number of users. Furthermore, it is also observed in case of the proposed algorithm, the system capacity increases slightly with the increase in number of users but it stays within the Shannon capacity limit, which is around 4.8 bits/symbol/Hz for the scenario in Fig. 3.5 (as it is clear from Fig. 3.3 for SNR of 10 dB). This behavior confirms the multiuser diversity advantage in the case of the LTE-CLWRC algorithm and is another powerful aspect as compared to the PRC and MF algorithms. The reason behind this increase in system capacity is described as follows. The proposed algorithm maximizes the total system capacity while having constraint on the weighted capacity as governed by

(3.4.3). For the user with high QoS requirement, the urgency factor is higher while the satisfaction factor is lower which increases the weighted capacity of the system that the algorithm tries to maintain. Hence, in the proposed optimization process, since the different QoS classes are equally assigned to the users, as the number of users in the system increases the users that belong to QoS class with higher QoS service requirement will contribute in increased total average system capacity as shown in Fig. 3.5. On the contrary, TDMA algorithm compared with here, does not consider urgency and satisfaction factors and hence the capacity is independent of the number of users as is clear in Fig. 3.5. In case of MF, the algorithm maximizes the system capacity while having constraint on the transmission rate itself and only a slight variation is observed; however, the PRC algorithm considers the fairness parameter and hence the capacity of the system depends on the proportional factor assigned to each user in the system. The PRC algorithm tries to maximize the total system capacity while having constraint on the proportional rate. Therefor, in the PRC algorithm, the proportional rate of the user with lower rate is boosted while the one with higher rate is decreased, such that proportionality is maintained among users. So, for higher numbers of users in the system, the users with higher proportional factor will cause the system capacity to decrease and the same is reflected in the figure. If the proportional factor used in the PRC algorithm for all users is assigned equal, then it is the case of the MF algorithm and the capacity curves will coincide. It is clear form the results in Fig. 3.5, that the LTE-CLWRC algorithm outperforms the PRC and MF algorithms in terms of total average capacity maximization.

The results in Fig. 3.6 depict a comparison between average system capacity for different number of users pertaining to the LTE-CLWRC scheme. A scenario of varying number of users in the system as discussed for the results in Fig. 3.5 is considered and the optimized average system capacity curves using the LTE-CLWRC algorithm for the system serving 10, 30 and 48 users are plotted in Fig. 3.6. It is evident from the figure that as the number of users in the system increases, there is an improvement in the system capacity and the

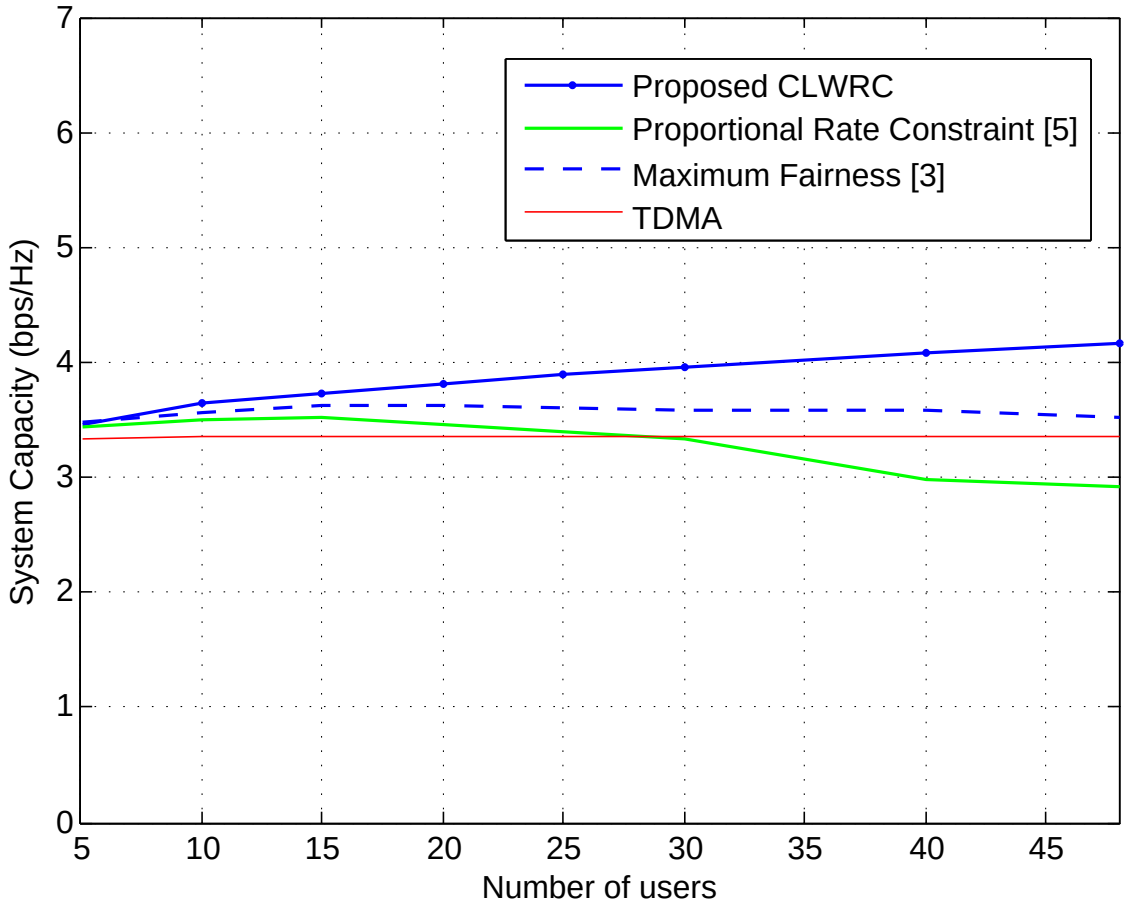

Figure 3.5. Total system capacity (bps/Hz) vs. no. of users (based on simulation parameters in Table 3.3)

system capacity gets closer to the Shannon Limit. These results confirm the multiuser diversity advantage in the LTE-CLWRC scheme.

Figs. 3.7 and 3.8 also depict similar results as in Fig. 3.3, for a system serving 15 and 20 users, respectively. It is assumed here that the users are equally assigned over the different service classes for each case, similar as the scenario discussed for results in Fig. 3.5. It is observed from Figs. 3.7 and 3.8 that the LTE-CLWRC algorithm achieves a higher total average system capacity throughout the observed SNR range as compared to the PRC and the MF algorithms and is closer to the Shannon limit in both figures. There is an improvement in the total average system capacity over the entire observed SNR range as the number of users increases for LTE-CLWRC algorithm. The total average system capacity for MF algorithm also shows an improvement with the increase in the number of users in

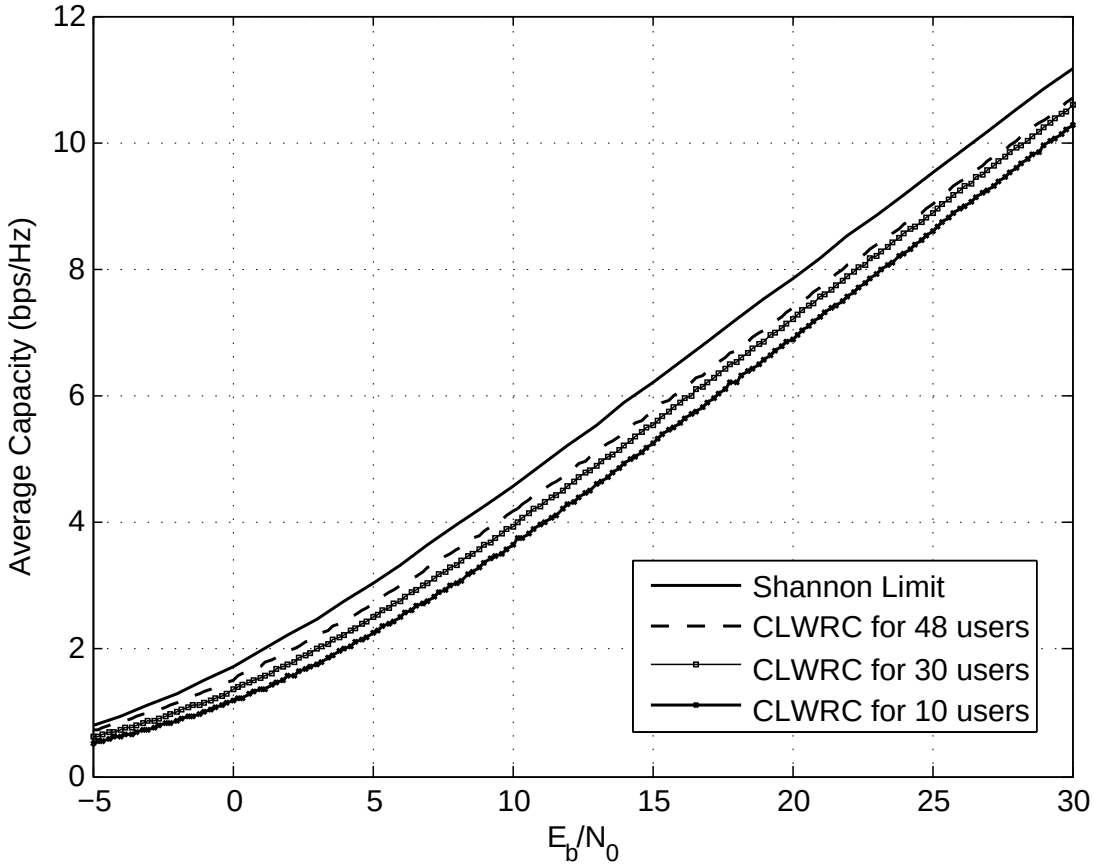

Figure 3.6. Average Capacity (bps/Hz) vs. average SNR per symbol (system serving 10, 30 and 48 users and implementing the proposed LTE-CLWRC algorithm)

the system. It is however interesting to note that for PRC algorithm, with the increase in number of users and average SNR, the total system capacity decreases. These results in Figs. 3.7 and 3.8 hence confirm the multiuser diversity in the case of the LTE-CLWRC algorithm and is another powerful aspect of the proposed algorithm as compared to PRC algorithm (see the earlier discussions about Fig. 3.5). As a conclusion, the superiority of the proposed cross-layer algorithm over the PRC and the MF in terms of total system capacity maximization is evident from all the results in Figs. 3.3 - 3.8.

3.5.3 Complexity Comparison

As a case study, time complexity is considered here for comparison between the proposed LTE-CLWRC algorithm with the PRC and MF algorithms. Table 3.4 shows a comparison

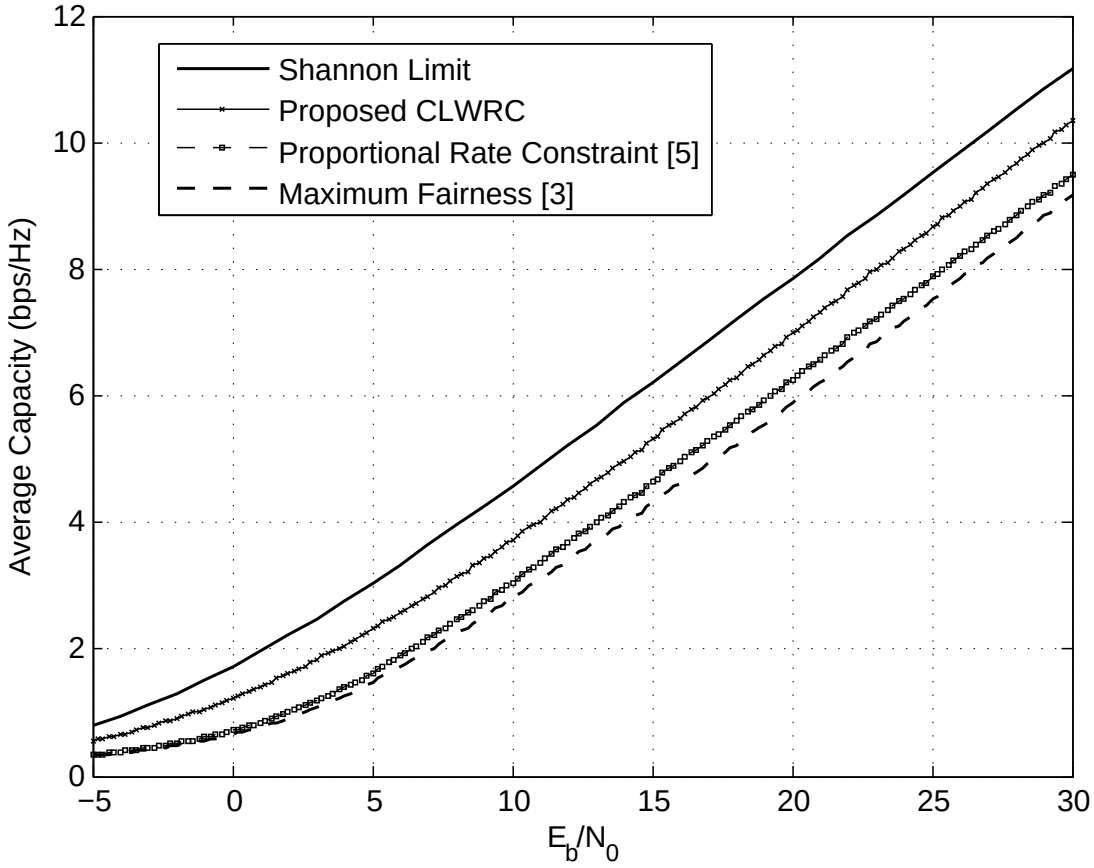

Figure 3.7. Average Capacity (bps/Hz) vs. average SNR per symbol (15 users are assumed to be served by the system)

between execution times (in seconds) of the LTE-CLWRC with other known algorithms for different number of frames. The algorithm was executed on Intel(R) Core(TM) i5-2430M CPU @ 2.40 GHz 2.40 GHz processor for 10 MonteCarlo runs. It can be observed form the table that the LTE-CLWRC algorithm has a faster execution time as compared to the PRC and MF algorithms. This is due to the fact that in both the PRC and MF algorithms, after PRBs are allocated to all the users, an initial power allocation algorithm is implemented before finalizing the power allocation based on water-filling approach. However, in the case of the LTE-CLWRC algorithm right after PRB allocation, power allocation based on water-filling approach is performed without considering initial power allocation as in [5]. In the LTE-CLWRC algorithm, as an initial power allocation, an equal power distribution among each PRB is considered and hence the execution time for the proposed algorithm is less. It

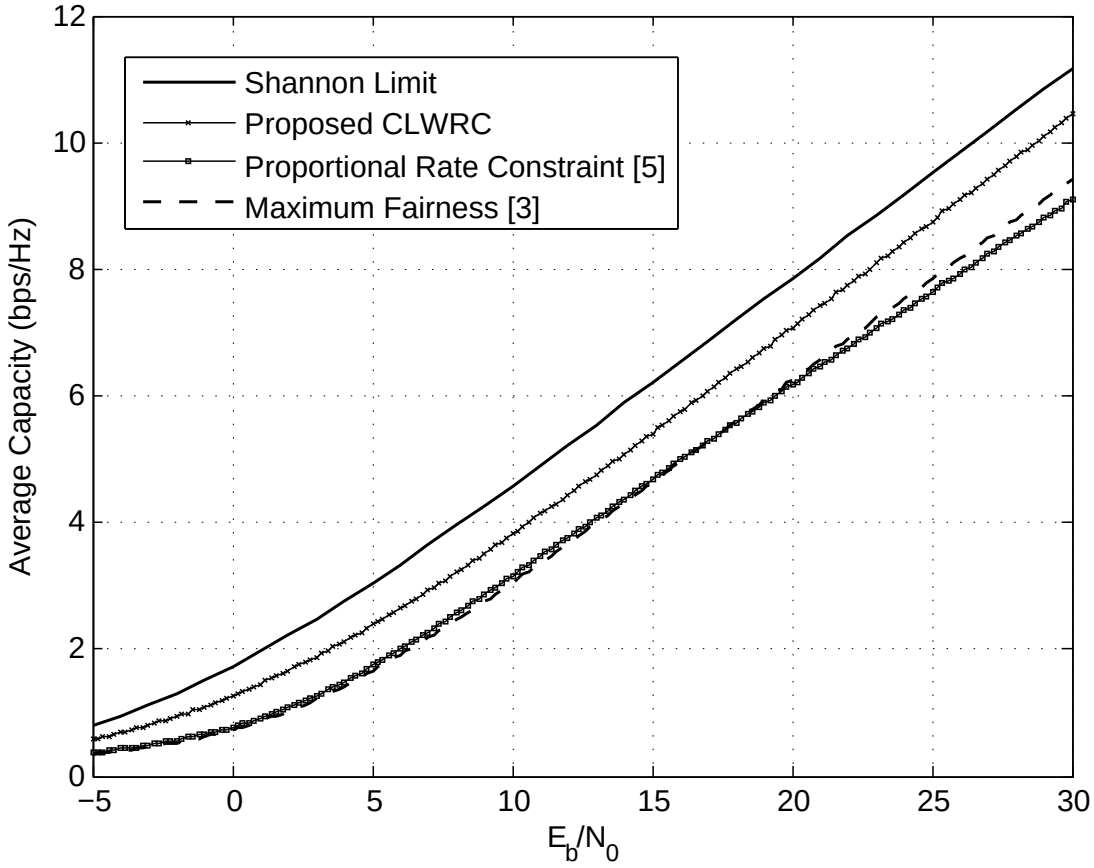

Figure 3.8. Average Capacity (bps/Hz) vs. average SNR per symbol (20 users are assumed to be served by the system)

is also interesting to note that, even without adopting the complex initial power allocation scheme as in [5], the LTE-CLWRC algorithm has better performance in terms of total average system capacity optimization as compared to PRC and MF algorithms.

3.6 Conclusion

In this chapter a detailed analysis on the CLWRC algorithm as it pertain to LTE system, LTE-CLWRC algorithm, is presented. A subsequent update on the cross-layer fairness parameters; service urgency and service satisfaction is presented in this chapter so that they address the LTE QoS definition. A weighted capacity is then evaluated, where the weights are dependent on the cross-layer fairness parameters. Then subject to the constraint on this

weighted capacity, an error-free Shannon channel capacity optimization problem is solved using a suboptimal solution. In this chapter the capacity performance of the LTE-CLWRC algorithm is evaluated subject to different observation scenarios like varying average SNR, different number of users and different frame numbers. This chapter then presents a capacity comparison between the LTE-CLWRC and the other know algorithms. Moreover, a complexity comparison between the LTE-CLWRC and the other algorithms is also performed in this chapter based on the algorithm execution time. From the numerical results, it has been observed that the LTE-CLWRC scheme results in total average system capacity that is closer to the Shannon limit as compared to the other known resource allocation schemes. On the other hand, unlike other known techniques, the LTE-CLWRC algorithm not only maintains the optimum system capacity for different number of users in the system but also increases as the number of users increases, confirming the multiuser diversity advantage of the LTE-CLWRC algorithm. The LTE-CLWRC algorithm also maintains its superiority in terms of execution time as compared to the other algorithms. In particular, the CLWRC scheme outperforms other known approaches in four aspects; closeness to Shannon capacity limit, consistency in terms of maximum optimum capacity throughout the frames considered, consistency in maintaining maximum optimum system capacity for different number of users and fast execution time. This chapter also presents a fair, in terms of available system resource, capacity performance comparison between LTE and WiMAX implementing the CLWRC algorithm. It is observed from the results that the total system capacity throughout the observed SNR range is higher for the LTE system.

Table 3.3. Traffic Simulation Parameters

Users (k)	Traffic Type	Parameter	Value
1	VoIP w/o SS	CODEC	G.729
		Voice Processing Interval	20 ms
		Packet size	66 Bytes
2	VoIP w/o SS	CODEC	G.728
		Voice Processing Interval	30 ms
		Packet size	106 Bytes
3	MPEG vedio	Bernoulli trial ($\bar{p}$)	0.4
		Mean rate	64 Kbps
		Maximum delay	30 ms
4	MPEG vedio	Bernoulli trial ($\bar{p}$)	0.5
		Mean rate	16 Kbps
		Maximum delay	50 ms
5	VoIP /w SS	CODEC	G.723.1
		Voice Processing Interval	30 ms
		Packet size	66 Bytes
		Mean ON period	1.2 sec
		Mean OFF period	1.8 sec
6	VoIP /w SS	CODEC	G.711
		Voice Processing Interval	20 ms
		Packet size	206 Bytes
		Mean ON period	1.2 sec
		Mean OFF period	1.8 sec
7	FTP	Mean rate	512 Kbps
		Minimum rate	128 Kbps
8	FTP	Mean rate	1 Mbps
		Minimum rate	1 Mbps
9	HTTP	Pareto mean rate	10558 bps
		Lognormal mean rate	7247 bps
10	HTTP	Pareto mean rate	10558 bps
		Lognormal mean rate	7247 bps

Table 3.4. Execution time (in seconds) of different algorithms for different number of frames

	Number of Frames		
Algorithms	1000	100	10
LTE-CLWRC	5957.69	598.87	59.34
PRC	13829.26	1374.97	137.07
MF	11975.37	1189.31	167.43

CHAPTER 4

NON ERROR-FREE SHANNON CHANNEL CAPACITY OPTIMIZATION IN WIMAX OFDMA SYSTEMS: ADAPTIVE MODULATION AND CODING

4.1 Introduction

An AMC-based cross-layer design optimization for subchannel and power allocations with the objective of maximizing the capacity (in bits/symbol/Hz) subject to fairness parameters and QoS requirements as constraints is presented in this chapter. The proposed scheme is termed as AMC-based CLWRC (AMC-CLWRC) scheme. AMC scheme has been adopted in this chapter to realize a practical and viable resource allocation. AMC is a scheme where the advantage of the channel fluctuation over time and frequency is taken into account to adaptively select the set of modulation and coding that best suits the channel condition while meeting the BER requirement. Based on the literature review on resource allocation schemes for AMC presented in Section 2.5.5, it is known that none of the work reported in literature addresses the problem of AMC-based cross-layer optimization by taking into account the channel conditions, queue status and QoS requirements simultaneously. To address this issue an AMC-based resource allocation optimization scheme scheme that takes into account, both the channel conditions and the queue status of each user as well as different QoS requirements to maximize system capacity is proposed. The study presented

in this chapter is an extension to the CLWRC scheduling algorithm introduced in [11] based on adaptive modulation and coding. Based on the queue status and QoS requirements of users, cross-layer fairness parameters [11], urgency and satisfaction factors are used in this study. This adds a new dimension to the fairness concept. Depending on the diverse QoS requirements of different users, resources can be allocated wisely; users that are well served and have no critical QoS requirements to schedule for service immediately can lag for some time allowing underserved users to access the channel. An AMC-based optimization of the system performance subject to the constraints on power and cross-layer fairness parameters are studied in this chapter as well. The significant improvement in the performance of the system in terms of maximization of system capacity achieved with the implementation of the proposed AMC-CLWRC approach is demonstrated by the extensive simulation results.

This chapter is organized as follows. A system model is presented in Section 4.2. The proposed AMC-based cross-layer approach is discussed in Section 4.3 and includes problem formulation and solution approach. Numerical results are presented in Section 4.4 and finally the chapter is concluded with some conclusions in Section 4.5.

4.2 AMC-based Cross-Layer OFDMA System model

A multiuser downlink OFDMA system is shown in Fig. 4.1 [1]. A total of K users sharing L subchannels are considered in the system and the total available transmit power is P_{tot}. The total available system bandwidth, B, is divided into L_{sc} subcarriers such that the bandwidth of each subcarrier is B/L_{sc} and the time slot duration corresponding to each subcarrier is $T_s = \frac{L_{sc}}{B}$. Subsets of these subcarriers form subchannels and they are the smallest allocation unit in WiMAX. Users can be assigned multiple subchannels at a certain time; however, a subchannel can not be shared among multiple users. Data from users arrive from the MAC layer and is placed into an infinite buffer. These buffers follow a FIFO strategy. A channel

[1]The system model is similar to the one presented for LTE system as in Fig. 3.1 except for the resource allocator block.

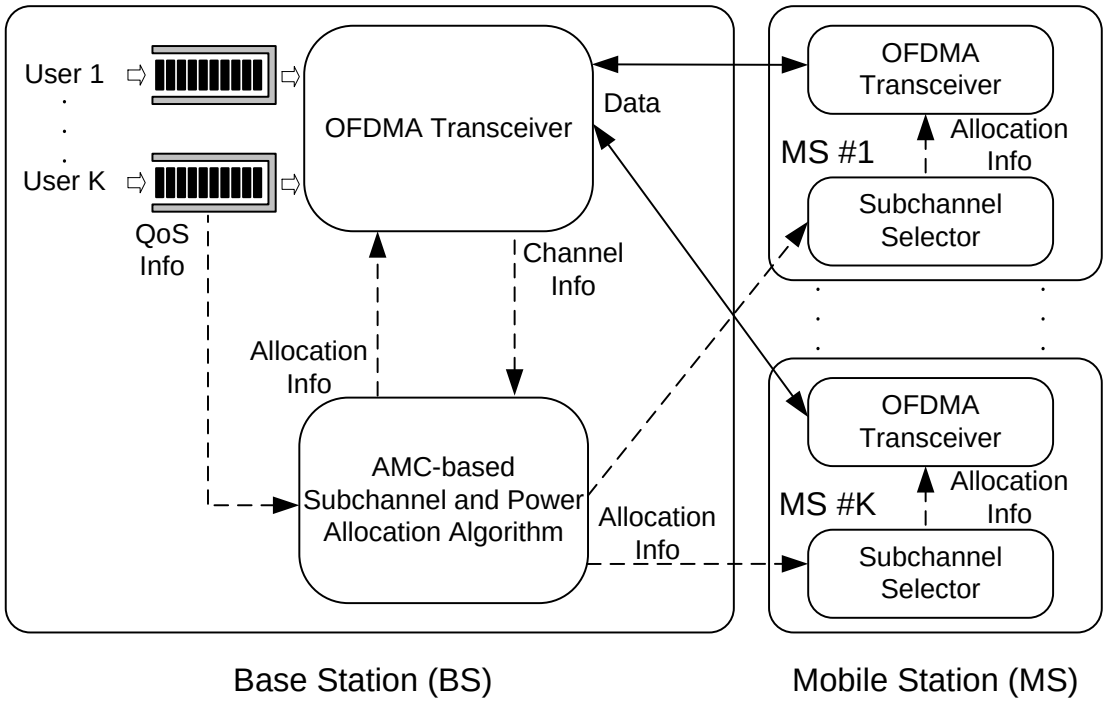

Figure 4.1. Cross-Layer downlink OFDMA Resource-Allocation System

fading that follows rayleigh distribution with envelope $h_{k,l}$ is assumed to be experienced by a user k over a subchannel l. Based on the CSI and the information on QoS, the subchannel and power allocation algorithm running in BS optimizes the subchannel and power allocation to maximize the error-free Shannon capacity while having a constraint P_{tot}. Moreover, the following assumptions are made: i) outgoing queues for every users are infinite; ii) the BS has perfectly received the CSI from all UE; iii) the subchannel and power allocation information is sent to each user on a separate channel; iv) coherent bandwidth of the channel is larger than $\frac{B}{L_{sc}}$, which means the channel response on each subcarrier is flat; v) the channel gain remains fixed during one time slot T_s; vi) the channel is varying in time slow enough that users can estimate the channel perfectly; vii) all system parameters and QoS parameters associated with all users are assumed to be made available to the BS during the initial setup (signalling) session before the call takes place.

4.3 Proposed AMC-based Cross-layer Algorithm

4.3.1 Proposed Algorithm: Optimization Problem Formulation

Let P_b be the bit error probability, M be the number of points in each signal constellation, R_c be the coding rate, G_c be the coding gain of the codes implemented (convolution codes) independent of the modulation and let γ be the instantaneous SNR, then the BER expression for M-QAM adjusted for coding gain is given by [40, pp. 281]

$$P_b \approx 0.2 \exp\left(-\frac{3\ G_c\ \gamma}{2(M-1)}\right). \tag{4.3.1}$$

Based on (4.3.1) SNR can be expressed in terms of P_b and M as

$$\gamma = -\frac{2}{3\ G_c}(M-1)\ln(5\ P_b). \tag{4.3.2}$$

Also, M as a function of γ can be expressed as

$$M(\gamma) = 1\ +\ \frac{1.5\ G_c\ \gamma}{-\ln(5\ P_b)}. \tag{4.3.3}$$

Air interface technologies that are based on AMC services use finite-size set of modulation and coding schemes (MCSs). MCSs are assigned to different users based on their service demand (voice or data) and channel characteristics. These MCSs consist of parameters that are associated with the PHY layer.

Each MCS set defines a particular digital modulation and coding rate. It is important to note that the coding gain is dependent on the coding rate that is defined in the MCS set and is also dependent on the coding schemes implemented. The appropriate pair of digital modulation and coding rate assigned to a user that meets the users' service demand (BER requirement either for voice or data service) is controlled by the selection of a prescribed MCS set based on the user's SNR; i.e. each MCS requires a minimum SNR. Let $i = 1, 2, \ldots, M$ where M is the number of MCS sets, then MCS_i is associated with an SNR_i (γ_i) threshold that a typical user must have in order to use the specific modulation and coding pair provided

by that MCS_i to meet the BER requirement of the user service. A range of instantaneous SNR, γ, is defined for each MCS set, such that $\gamma \in [\gamma_i,\ \gamma_{i+1})$. The γ_i threshold associated with a MCS_i set for a given BER requirement is determined by (4.3.2) where the coding gain specific to the MCS_i is represented by G_{c_i}. The spectral efficiency or channel capacity, in bits/symbol/Hz, of an AMC-based system for a given MCS_i can then be represented as [2]

$$\begin{aligned} C &= R_c\ log_2 M(\gamma_i) \\ &= R_c\ log_2 \left[1 + \frac{1.5\ G_{c_i}\ \gamma_i}{-ln(5\ P_b)}\right]. \end{aligned} \tag{4.3.4}$$

A convolution encoder with a constraint length of 5 is assumed and the value of coding gains corresponding to different code rates are taken from [41]. Figs. 4.2 and 4.3 depict the BER versus SNR plot for various MCSs defined in Table 4.1 for system serving voice and data services, respectively. In these figures the γ_i value corresponding to a given MCS_i is marked.

Table 4.1. MCS based on IEEE 802.16e Standard

MCS_i	Modulation Type	Modulation Index (M)	Code Rate (R_c)
MCS_1	QPSK	4	1/2
MCS_2	QPSK	4	3/4
MCS_3	16-QAM	16	1/2
MCS_4	16-QAM	16	3/4
MCS_5	64-QAM	64	2/3
MCS_6	64-QAM	64	3/4

As can be observed in these figures for six different MCSs, six different γ_i: $\gamma_1,\ \gamma_2,\ \ldots,\ \gamma_6$ are marked. Now for any instantaneous γ that is greater than or equal to γ_1 and less than γ_2, the AMC-based scheduling will select MCS_1. Similarly MCS_2 is selected if γ is greater than or equal to γ_2 and less than γ_3, and so on. Based on the curves depicted in Figs. 4.2 and 4.3, the range of γ corresponding to each MCS_i is tabulated in Tables 4.2 and 4.3 for a system serving voice and data, respectively.

For an OFDMA system with K users and L subchannels sharing a system bandwidth B, let $\gamma_{k,l}$ be the instantaneous SNR corresponding to a k^{th} user using l^{th} subchannel and

[2]The expression in (4.3.4) can be interpreted as the non error-free channel capacity version of the well known error-free Shannon Capacity defined in [40, pp. 98]

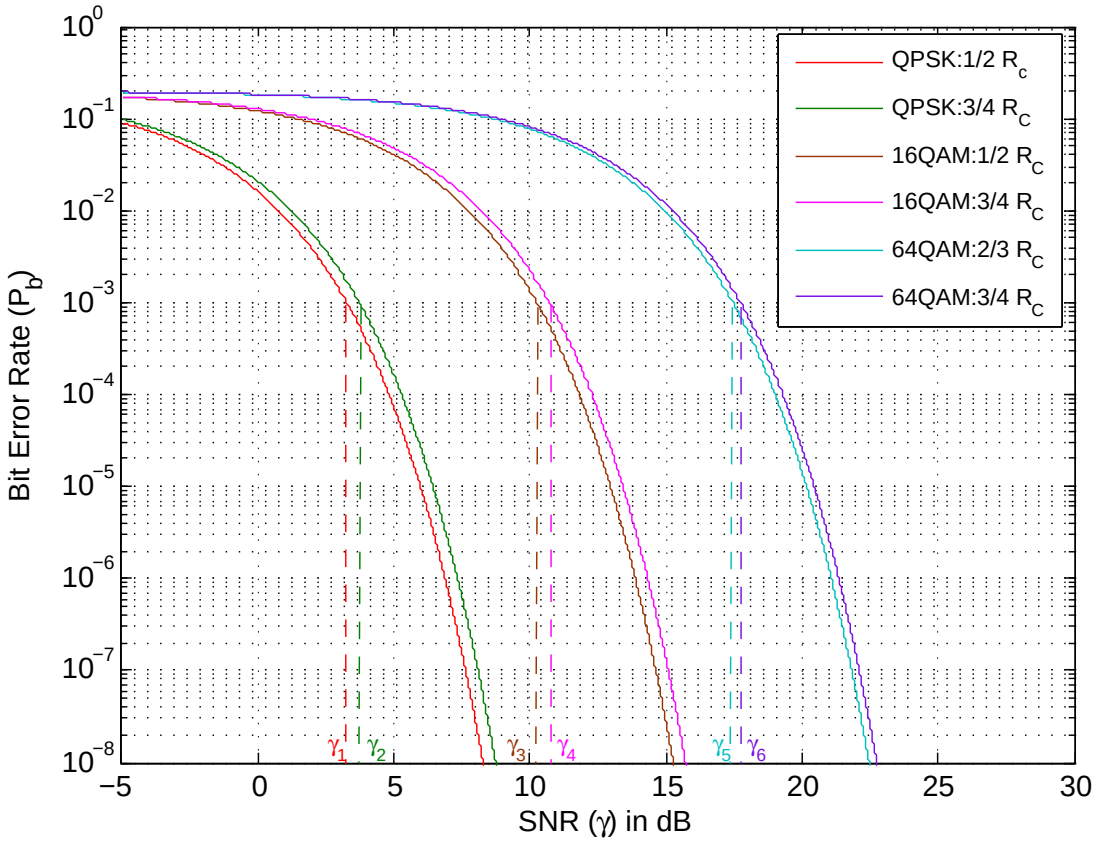

Figure 4.2. Bit error rate (BER) vs. signal to noise ratio (SNR) for voice service

define $\gamma^i_{k,l} = \gamma_i$ when $\gamma_{k,l} \in [\gamma_i,\ \gamma_{i+1})$. Let $P_{k,l}$ be the power allocated to a user k over subchannel l, then $\gamma_{k,l}$ is defined as $\frac{P_{k,l}}{N_0 \frac{B}{L}}$, where N_0 represents the power spectral density of an additive white Gaussian noise (AWGN). Also, let $h_{k,l}$ be the channel gain of a kth user over subchannel l, and $\rho_{k,l} \in \{0, 1\}$ indicates whether or not a subchannel l is used by the k^{th} user. Then, the spectral efficiency or channel capacity, in bits/symbol/Hz, of a k^{th} user associated with an x service flow for a given MCS_i during a given frame n is expressed as

$$C^x_k(n) = \sum_{l=1}^{L} \frac{\rho_{k,l}}{L}\ R_c\ log_2 \left[1 + \frac{1.5\ G_{c_i}\ \gamma^i_{k,l}\ h^2_{k,l}}{-ln(5\ P_b)}\right]\ bits/symbol/Hz. \tag{4.3.5}$$

The QoS classes adopted by WiMAX standard and as discussed in Section 2.5.3 is considered as a case study. The cross-layer QoS parameters, service urgency and service satisfaction, introduced in [11] and reviewed in Section 2.5.3 and 2.5.3, respectively are considered in this section as a part of problem formulation. The weighted capacity, $R^x_k(n)$ of a k^{th} user

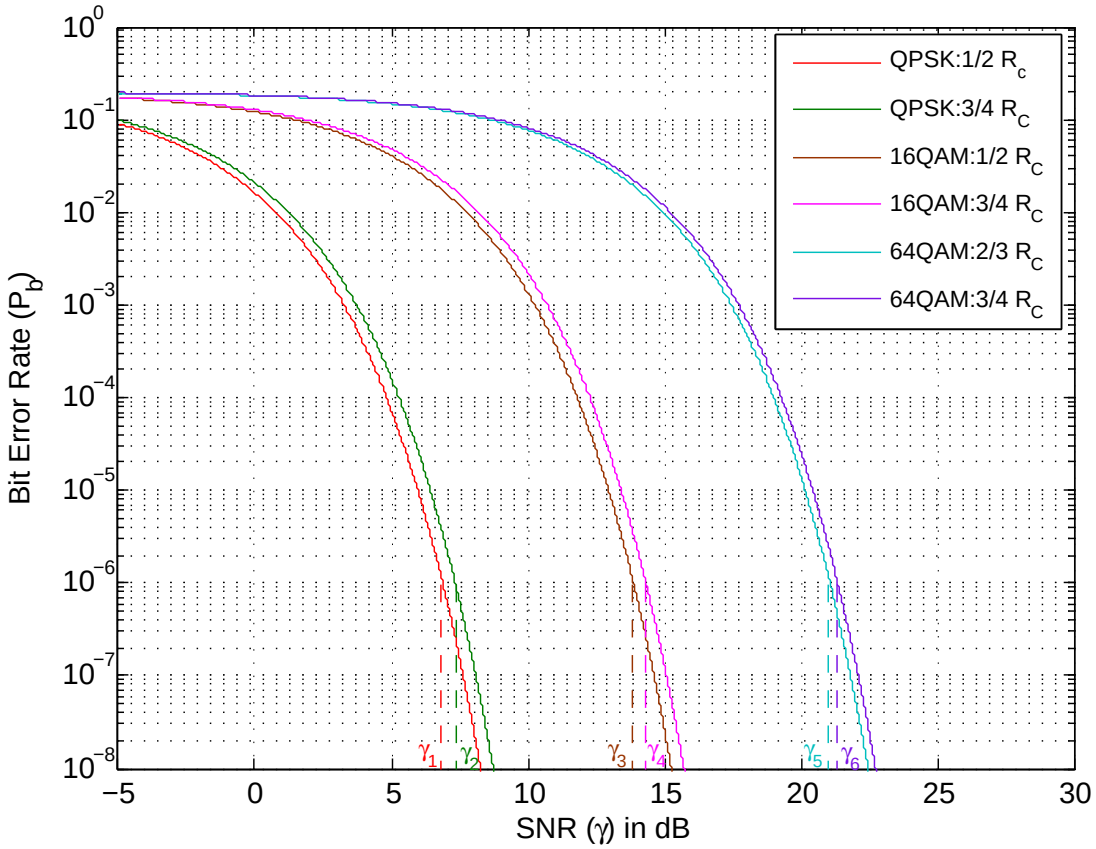

Figure 4.3. Bit error rate (BER) vs. signal to noise ratio (SNR) for data service

associated with an x service flow during a given frame n, can then be expressed as

$$R_k^x(n) = \frac{U_k^x(n)}{S_k^x(n)} C_k^x(n). \tag{4.3.6}$$

The weighted capacity in (4.3.6) incorporates both the urgency factor and the satisfaction factor. The weighted rate is directly proportional to the service urgency and inversely proportional to the service satisfaction (refer to the discussion in Section 3.4.1; service flows in WiMAX correspond to QCIs in LTE). Similarly as discussed in Section 3.4.1 the superscript x is dropped and the fairness constraint is defined as

$$R_u(n) = R_v(n) = R(n) \qquad \forall u, v \in [1, K]. \tag{4.3.7}$$

Based on the above discussion, the optimization problem can be expressed mathematically

Table 4.2. MCS with SNR threshold for Voice Service

MCS_i	γ_i (dB)	γ_{i+1} (dB)	Range
No service	-∞	3.3	$(-\infty, \gamma_1)$
MCS_1	3.3	3.8	$[\gamma_1, \gamma_2)$
MCS_2	3.8	10.3	$[\gamma_2, \gamma_3)$
MCS_3	10.3	10.8	$[\gamma_3, \gamma_4)$
MCS_4	10.8	17.5	$[\gamma_4, \gamma_5)$
MCS_5	17.5	17.8	$[\gamma_5, \gamma_6)$
MCS_6	17.8	$+\infty$	$[\gamma_6, +\infty)$

Table 4.3. MCS with SNR threshold for Data Service

MCS_i	γ_i (dB)	γ_{i+1} (dB)	Range
No service	-∞	6.9	$(-\infty, \gamma_1)$
MCS_1	6.9	7.4	$[\gamma_1, \gamma_2)$
MCS_2	7.4	13.9	$[\gamma_2, \gamma_3)$
MCS_3	13.9	14.4	$[\gamma_3, \gamma_4)$
MCS_4	14.4	21.1	$[\gamma_4, \gamma_5)$
MCS_5	21.1	21.4	$[\gamma_5, \gamma_6)$
MCS_6	21.4	$+\infty$	$[\gamma_6, +\infty)$

as

$$\max_{P_{k,l},\rho_{k,l},M} \quad C = \sum_{k=1}^{K}\sum_{l=1}^{L} \frac{\rho_{k,l}}{L}\ R_c\ log_2\left[1 + \frac{1.5\ G_{c_i}\ \gamma_{k,l}^{i}\ h_{k,l}^{2}}{-ln(5\ P_b)}\right]\ bits/symbol/Hz \tag{4.3.8}$$

$$\text{subject to} \quad \sum_{k=1}^{K}\sum_{l=1}^{L} P_{k,l} \leq P_{tot} \tag{4.3.9}$$

$$P_{k,l} \geq 0 \quad \forall \quad k,l \tag{4.3.10}$$

$$\rho_{k,l} = \{0,1\} \quad \forall \quad k,l \tag{4.3.11}$$

$$\sum_{k=1}^{K} \rho_{k,l} = 1 \quad \forall \quad l \tag{4.3.12}$$

$$R_u = R_v = R\ \forall\ u,v \in [1,K], \tag{4.3.13}$$

where P_{tot} is the total available power. The first constraint implies that the total power over all subchannels is not to exceed the total available power. The second constraint states that the power for all subchannels should be positive or zero. In the third constrain, $\rho_{k,l}$ is only allowed to be 0 or 1 which means a user is not allowed to use a portion of a subchannel. Furthermore, no sharing of subchannel is allowed, which is stated by the fourth constraint. The last constraint is the fairness constraint presented in (4.3.7) and (4.3.7).

4.3.2 Proposed Algorithm: Problem Solution and Implementation

The optimization problem given in (4.3.8) is very hard to solve. It is a mixed binary integer programming problem. The problem has nonlinear constraints as well as continuous variables, $P_{k,l}$, and binary variables, $\rho_{k,l}$. An optimum solution for this optimization problem exists which is highly computationally complex and is not favored considering the high frequency of executing the schedulers in practical systems. The subschannel and power allocation decisions are to be taken for every frame, therefore, it is usually the case that a suboptimal solution is adopted which approaches the optimal results.

For a system with K users and L subchannels, there are K^L possible subchannel allocations and for every allocation the optimal power allocation can be computed. Even though it is possible to compute the global maximum solution, a suboptimal greedy approach is presented in this work and optimality is compromised for complexity reduction. An analytical solution to the optimization problem in (4.3.8) is obtained by adopting the approach presented in [5]. For a given subchannel allocation, Π_k, such that Π_k is the set of subchannels allocated to the k^{th} user, the capacity of the kth user during a given frame n, in bits/symbol/Hz, before adaptive selection of modulation and coding is considered is expressed as

$$C_k(n) = \sum_{l \in \Pi_k} \frac{\rho_{k,l}}{L} \; R_c \; log_2 \left[1 + a \; P_{k,l} \; H_{k,l}\right], \tag{4.3.14}$$

where

$$H_{k,l} = \frac{h_{k,l}^2}{N_0 \frac{B}{L}} \qquad \& \qquad a = \frac{1.5 \; G_c}{-ln(5 \; P_b)}, \tag{4.3.15}$$

Then the optimization problem in (4.3.8) can be reformulated as

$$\max_{P_{k,l}} \quad C = \sum_{k=1}^{K} \sum_{l \in \Pi_k} \frac{1}{L} \; R_c \; log_2 \left[1 + a \; P_{k,l} \; H_{k,l}\right] \; bits/symbol/Hz \tag{4.3.16}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \sum_{l \in \Pi_k} P_{k,l} \leq P_{tot} \tag{4.3.17}$$

$$P_{k,l} \geq 0 \;\; \forall \;\; k, l \tag{4.3.18}$$

$$\Pi_u \cap \Pi_v = \Phi \;\; \forall \;\; u \neq v \tag{4.3.19}$$

$$\Pi_1 \cup \Pi_2 \cup \ldots \cup \Pi_K \subseteq \{1, 2, \ldots, L\} \tag{4.3.20}$$

$$R_u(n) = R_v(n) = R(n) \; \forall \; u, v \in [1, 2, \ldots, K], \tag{4.3.21}$$

The solution to the optimization problem in (4.3.16) results in

$$P_{k,x} = P_{k,1} + \frac{-ln(5 \; P_b)(H_{k,x} - H_{k,1})}{1.5 \; G_c \; H_{k,x} \; H_{k,1}} \tag{4.3.22}$$

for $k \in \{1, 2, \ldots, K\}$ and $x \in \{1, 2, \ldots, |\Pi_k|\}$. This result is obtained by solving optimization problem in (4.3.16) using the method of Lagrange multipliers. The derivation is provided in Appendix 5. The expression in (4.3.22) is the water-filling equation, which means that subchannels with higher SNR receive more power in order to maximize the capacity. A similar equation was obtained in [5] for different constraints (as indicated in Section 4.3.1).

Although the optimal solution is to jointly allocate subchannels and power, a less complex approach, two-phase greedy approach, that results in a suboptimal solution of the optimization problem formulated in this study is adopted here. The two-phase greedy approach starts off with allocating equal power to all the subchannels. Later power is allocated in order to maximize the total system capacity while maintaining fairness and QoS support. In the following section, resource allocator, subchannel allocator and power allocator are discussed in detail. The solution altogether offers a low delay cost effective approach.

Resource Allocator

The proposed resource allocator algorithm based on the two-phase greedy approach is shown in Algorithm 2. The terms used in this algorithm are defined as follows: T_f is the frame

duration and T is the total traffic duration, such that $T = N \times T_f$, T_c & T_s are the in-phase & quadrature phase E-field components of Rayleigh fading channel, respectively, K_{tot} is the total number of users considered while scheduling, K_{voice} is the total number of users with voice service, γ_{avg} is the average SNR of the system, γ_i^{voice} and γ_i^{data} are the minimum SNR values for which voice and data services are allowed for MCS_i, respectively, $P_{k(tot)}$ is the initial total power allocation to a k^{th} user, C_k is the capacity corresponding to a k^{th} user and $\bar{C}$ is the exponentially weighted average capacity.

The working of the resource allocator algorithm depicted in Algorithm 2 is described in detail in the following. It first reads the queues lengths, $Q_k(n)$, service flows associated with each user, $SF^x(k)$, the maximum delay accepted for every rtPS user, T_k, the minimum data rate accepted for every nrtPS, $\eta_k(n)$, from the MAC layer. Likewise, K, L, P_{tot}, B, N_0, Γ_x, t_c, and ΔT are configured by the allocator. With all the information in hand traffic corresponding to different types of QoS classes are simulated as explained in Appendix 5 and queue lengths corresponding to a particular service flow are calculated as explained by equation (2.5.8). Service urgency and satisfaction parameters are then evaluated using (2.5.8)-(2.5.17) and a Rayleigh fading channel based on Clarks's model is simulated as explained in Appendix 5.

The algorithm then proceeds forward by invoking the subchannel allocation, which starts with assigning the subchannel with maximum channel gain for each user in Rayleigh fading environment. The total available system power is equally divided among channels and the weighted data rates for each user dependent on the urgency and satisfaction factor are then calculated. The weighted data rates so generated are then evaluated to allocate the remaining subchannels to the users such that the fairness among users in terms of weighted data rate is maintained. Since adaptive modulation and coding is considered, based on the users being scheduled, the maximum acceptable BER is picked adaptively so as to meet the users QoS requirement. The details of subchannel allocator is discussed in Appendix 5. Once subchannel allocation is completed, power allocator algorithm based on the derivation

in Section 4.3.2 is invoked, the detailed description of this algorithm is provided in Appendix 5. Finally the overall exponentially weighted total average system capacity is evaluated.

4.4 Simulations and Numerical Results

In this section we numerically implement and simulate the solution described in Section 4.3.2 for the optimization problem presented in Section 4.3.1 based on Algorithm 2. This section considers the same system parameters, the traffic models and the wireless channel model as discussed in Section 2.6.

4.4.1 Capacity Comparison

The results in Fig. 4.4 show the total average system capacity versus average SNR based on the proposed AMC-CLWRC algorithm that implements the solution approach presented in Section 4.3.2 for the optimization problem formulated in Section 4.3.1. In this figure the algorithm is executed for different values of average SNR, where for simplicity symbol energy is assumed to be 1 Joule. So, for each value of the average SNR in Fig. 4.4, the corresponding power spectral density of the AWGN channel is evaluated and used in the optimization problem. The remaining of the system parameters needed in the computation are listed in Table 2.1. The proposed AMC-based optimization algorithm is used to optimize the system serving mixed traffic, i.e, voice and data. A total of 10 users are considered, where 4 of them are voice users and 6 of them are data users. For performance comparison purposes, Fig. 4.4 also shows results for the PRC algorithm [5] and the MF algorithm [3], modified accordingly to support adaptive modulation and coding, along with the proposed AMC-based CLWRC algorithm. The power distribution approach among users that is used in [3] and [5] is referenced here in Appendix 5. In Fig. 4.4 , the AMC-based optimized capacity curves for the PRC and MF are based on the solution approach proposed in [5], which is modified accordingly to address adaptive modulation and coding, where MF is a special case of PRC.

Algorithm 2 AMC Resource Allocator

Input: $K,\ L,\ P_{tot},\ B,\ N,\ T_f,\ N_0,\ \gamma_{avg},\ \gamma_i^{voice},\ \gamma_i^{data},\ \Gamma_x,\ \Delta T,\ t_c,\ T_k,\ \eta_k$
Initialize Array: $\bar{C}_k \Leftarrow 0,\ Q_k \Leftarrow 0$
generate $Q_k^x(n) \quad \forall\ n \in \{1,\ 2,\ \ldots,\ N\}$ and $\forall\ k$ with respective SF^x
for $n = 1 \rightarrow N$ **do**
    find $Q^{SF^x}(n)$ // (2.5.8)
    find $U_k(n)$ // (2.5.9)
    find $S_k(n)$ // (2.5.10)-(2.5.17)
    generate T_c & T_s
    $h_{k,l} \Leftarrow T_c + jT_s$
    if $\gamma_{avg} \leq \gamma_1^{voice}$ **then**
        if $\gamma_{avg} \leq \gamma_1^{data}$ **then**
            $K_{tot} \Leftarrow K_{voice}$
        else
            $K_{tot} \Leftarrow K$
        end if
        invoke Subchannel Allocator // *assigns* $\rho_{k,l}$
        for $k = 1 \rightarrow K$ **do**
            // *SELECT BER*
            if $k \in \{1, 2, \ldots, K_{voice}\}$ **then**
                $P_b = 10^{-3}$
            else
                $P_b = 10^{-6}$
            end if
            invoke Power Allocator // *assigns* C_k
        end for
        $C(n) \Leftarrow \sum_{k=1}^{K} C_k$
        if $n = 1$ **then**
            $\bar{C}(n) \Leftarrow C(n)$
        else
            $\bar{C}(n) \Leftarrow \bar{C}(n-1) * (1 - \frac{1}{t_c}) + C(n) * \frac{1}{t_c}$
        end if
    else
        $\bar{C} \Leftarrow 0$
    end if
end for

For the PRC algorithm, as explained in [5], a set of predetermined capacity weighting factor values among all users are taken to ensure proportional fairness among users. Any of these predetermined values are less than 1 and assigning equal values to all of them results in the MF. It can be seen from the figure that the proposed AMC-CLWRC algorithm achieves higher total average system capacity throughout the observed average SNR range (-5 to 30 dB) with slight improvement when the SNR is higher than 22 dB as compared to the other algorithms. This slight improvement is due to the fact that only one MCS set is available for the average SNR greater than 22 dB and hence the concept of adaptive modulation and coding does not apply in this SNR range and the total system capacity for all the algorithms almost remains the same. A glitch is observed in the simulation curve at the SNR value of 6.9 dB, this is because of the fact that there is no service to the data users unless the SNR of 6.9 dB is reached. Before this SNR value only voice users were served such that all the available resources were dedicated to the users with voice traffic. So, with addition of users with higher BER requirement, it is obvious that the overall system capacity will be lowered.

Figs. 4.5 and 4.6 also depict similar results as in Fig. 4.4, for a system serving 25 (10 voice users and 15 data users) and 30 (system serving 12 voice users and 18 data users) users, respectively. It is assumed here that for each case of number of users in the system users are equally assigned over the different service classes with defined weighting factor $\Gamma_x(k)$ (e.g., if K = 30 users, 6 out of these 30 users are assigned to each of the 5 different service classes: i.e, $x(k) = UGS$ for $k = 1, 2, \ldots, 6$; $x(k) = ertPS$ for $k = 7, 8, \ldots, 12$; $\ldots$; $x(k) = BE$ for $k = 25, 26, \ldots, 30$). The values of Γ_{SF^x} are taken from Table 2.1. It can be observed in Figs. 4.5 and 4.6 that the proposed AMC-CLWRC algorithm achieves higher total average system capacity throughout the observed average SNR range (-5 to 30 dB) with slight improvement when the SNR is higher than 22 dB as compared to the other algorithms. The behaviour is similar to the one observed and discussed in Fig. 4.4. There is an improvement in the system capacity over the entire observed SNR range as the number of users increases for

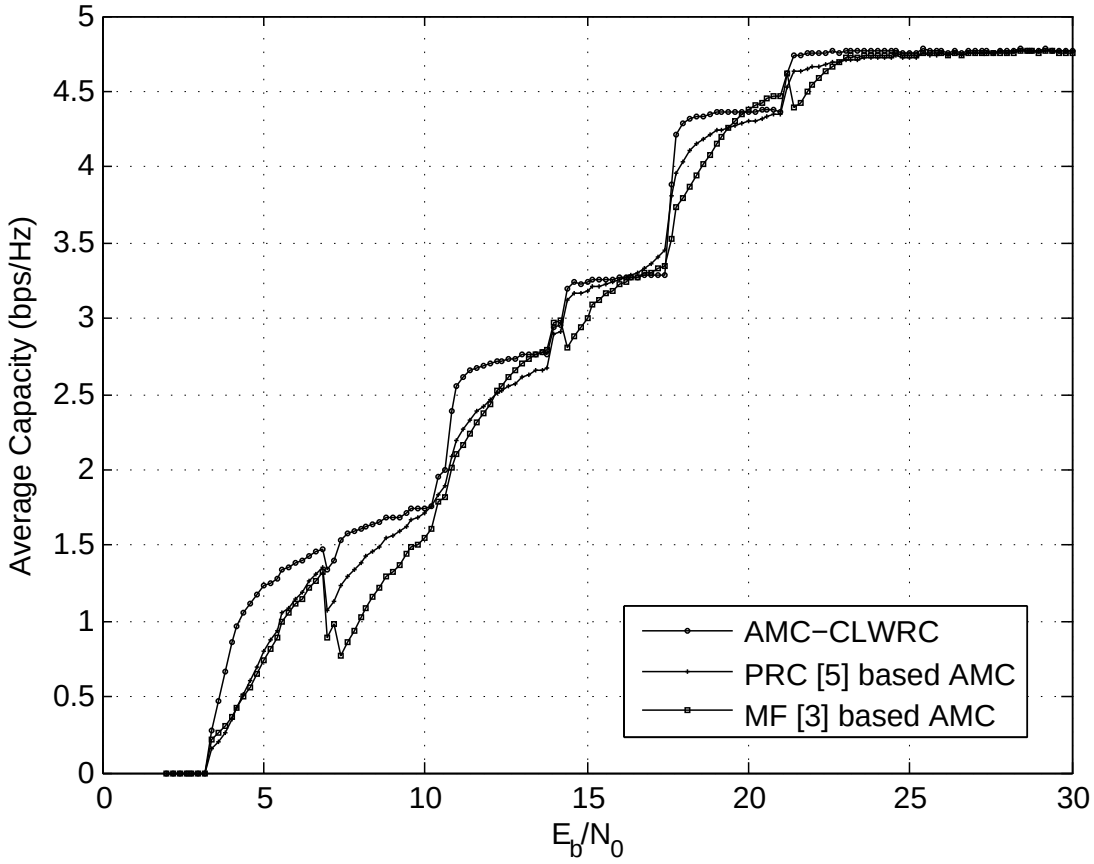

Figure 4.4. Average Capacity (bits/symbol/Hz) Vs. Average SNR (system serving 4 voice users and 6 data users)

AMC-CLWRC algorithm. The increase in the number of users in the system however, has no significant effect on the system capacity for the PRC algorithm. It is interesting to note that the system capacity for the MF algorithm decreases at some SNR with the increase in number of users in the system. These results in Figs. 4.5 and 4.6 hence confirm the multiuser diversity advantage in the case of the proposed AMC-CLWRC algorithm and is another powerful aspect of the proposed algorithm as compared to the other algorithm.

As clear from Figs. 4.4, 4.5 and 4.6 that there are some overlap between capacity curves, over some SNR ranges, when a comparison is made between the proposed AMC-CLWRC algorithm and other algorithms. There are few ranges of SNR where other algorithms perform better than the proposed AMC-CLWRC algorithm, in terms of system capacity maximization. Hence, for a fair quantitative comparison between different algorithms over the whole

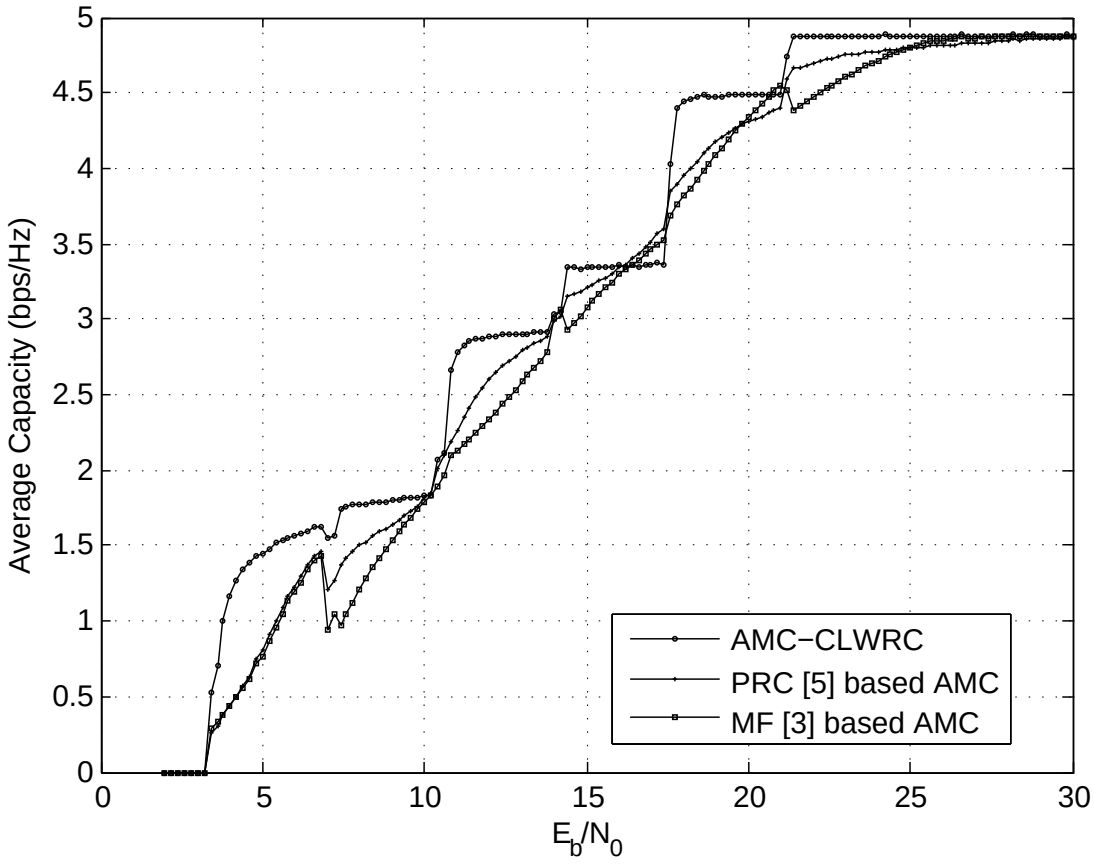

Figure 4.5. Average Capacity (bits/symbol/Hz) Vs. Average SNR (system serving 10 voice users and 15 data users)

SNR range, a new performance metric, named Sum SNR-Capacity Product (SSCP) is introduced. This metric reveals the amount of successfully transmitted data associated with the supporting set of the SNR range for a system serving multi-user demanding multi-class QoS services. This metric is important to compare between the proposed AMC-based scheduling algorithm with other scheduling algorithms. Mathematically, we define the SSCP as

$$\Sigma_{SCP} \;=\; \int_{\gamma \in \mathcal{S}(\gamma)} \mathcal{C}(\gamma)\; d\gamma \tag{4.4.1}$$

where $\mathcal{S}(\gamma)$ is the supporting set of the SNR values for a system supporting multi-class QoS services.

The SSCP factor provides two important indicators for such a system model, the first indicator is associated with having a non-zero value of the SSCP that shows the range of SNR capable of receiving the service for the system serving mixed QoS services. The second

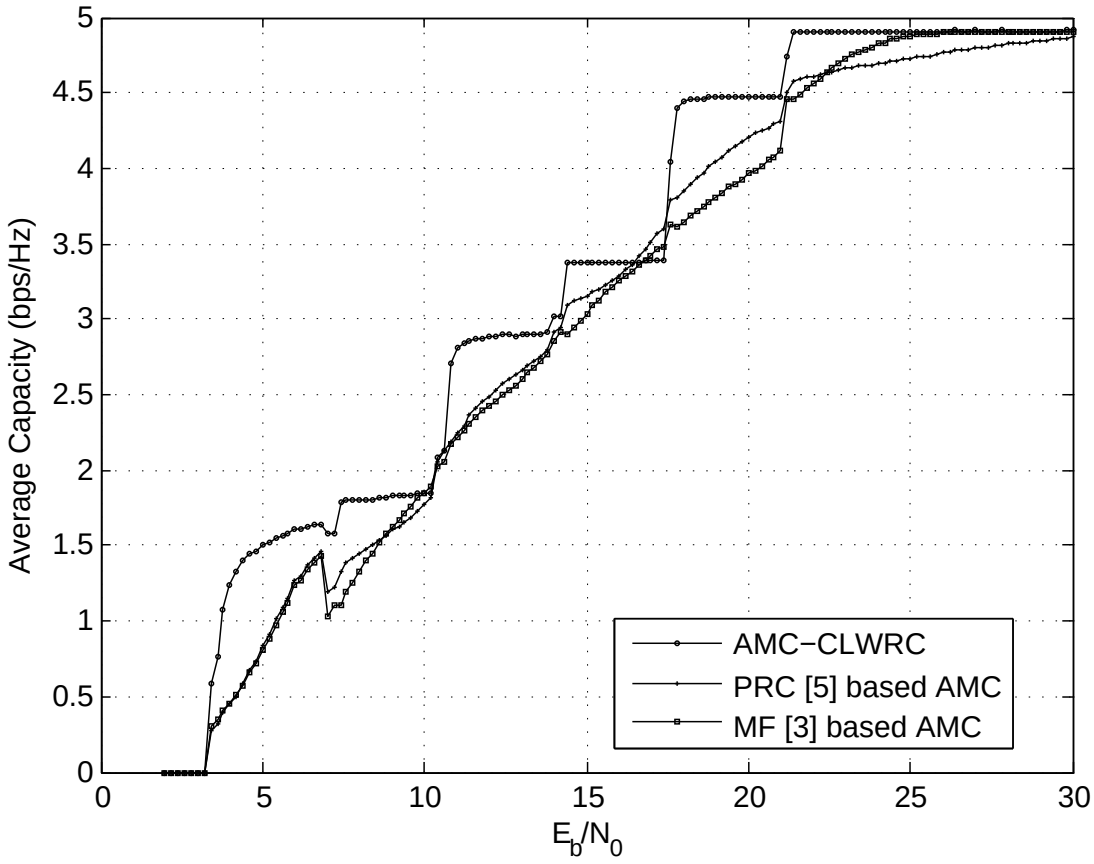

Figure 4.6. Average Capacity (bits/symbol/Hz) Vs. Average SNR (system serving 12 voice users and 18 data users)

indicator associated with the greater value of the SSCP among all algorithms indicates that, greater the value of SSCP is, higher is the overall achievable system capacity.

Table 4.4. Comparison of the SSCP for different scheduling algorithms with different number of users

Number of users	AMC-CLWRC	PRC-based AMC	MF-based AMC
10	8.8225×10^3	8.6493×10^3	8.5612×10^3
25	9.0982×10^3	8.8005×10^3	8.8665×10^3
30	9.1457×10^3	8.6719×10^3	8.6434×10^3

Table 4.4 shows a SSCP comparison between the AMC-CLWRC algorithm and the other algorithms for different number of users based on the results in Figs. 4.4, 4.5 and 4.6. It can be seen from the table that the SSCP values for the proposed AMC-CLWRC algorithm are higher than those for the other algorithms. So, based on the SSCP performance metric, it can be concluded that the proposed AMC-CLWRC algorithm has superior performance over

the other algorithms. However, SNR range supporting the service for mixed traffic scenario is the same for all algorithms as is evident from the referenced figures.

The results in Figs. 4.7 and 4.8 depict a comparison between average system capacity for different number of users pertaining to the proposed AMC-CLWRC scheme. A scenario of different number of users in the system as discussed for the results in Figs. 4.5 and 4.6 is considered and the optimized average system capacity curves using the proposed AMC-CLWRC algorithm for the system serving 10 [4 voice and 6 data users], 20 [8 voice and 12 data users] and 30 [12 voice and 18 data users] users are plotted in Fig. 4.7 and the system serving 40 [16 voice and 24 data users], 50 [20 voice and 30 data users] and 60 [24 voice and 36 data users] users are plotted in Fig. 4.8. It is evident from Fig. 4.7 that as the number of voice and data users in the system increases in equal proportion, there is an improvement in the total system capacity throughout the observed SNR range. These results confirm the multiuser diversity advantage in the proposed AMC-CLWRC scheme. However in case of Fig. 4.8, the increase in number of users being served by the system is considered in such a way that more data users with higher BER requirement are served as compared to voice users. In this scenario, it is likely that the system capacity decreases during some SNR ranges(11-14 and 17-22 dB) with an increase in the number of users which is evident in Fig. 4.8. Therefore, we incline to depend on the SSCP for the performance comparison among the different users as shown in Table 4.5 which is associated with Figs. 4.7 and 4.8. It can be seen from the table that the SSCP values for the proposed AMC-CLWRC algorithm pertaining to the scenario in Fig. 4.7 increases with the increase in number of users while it decreases with the increase in number of users for the scenario pertaining to that Fig. 4.8. This is due to the fact that in case of Fig. 4.7 the number of voice and data users are comparable while in case of 4.8 the number of data users are significantly higher as compared to voice users. Since the data users have a higher QoS requirement and demand more resources the increase in data users in the system results in overall decrease in the system capacity.

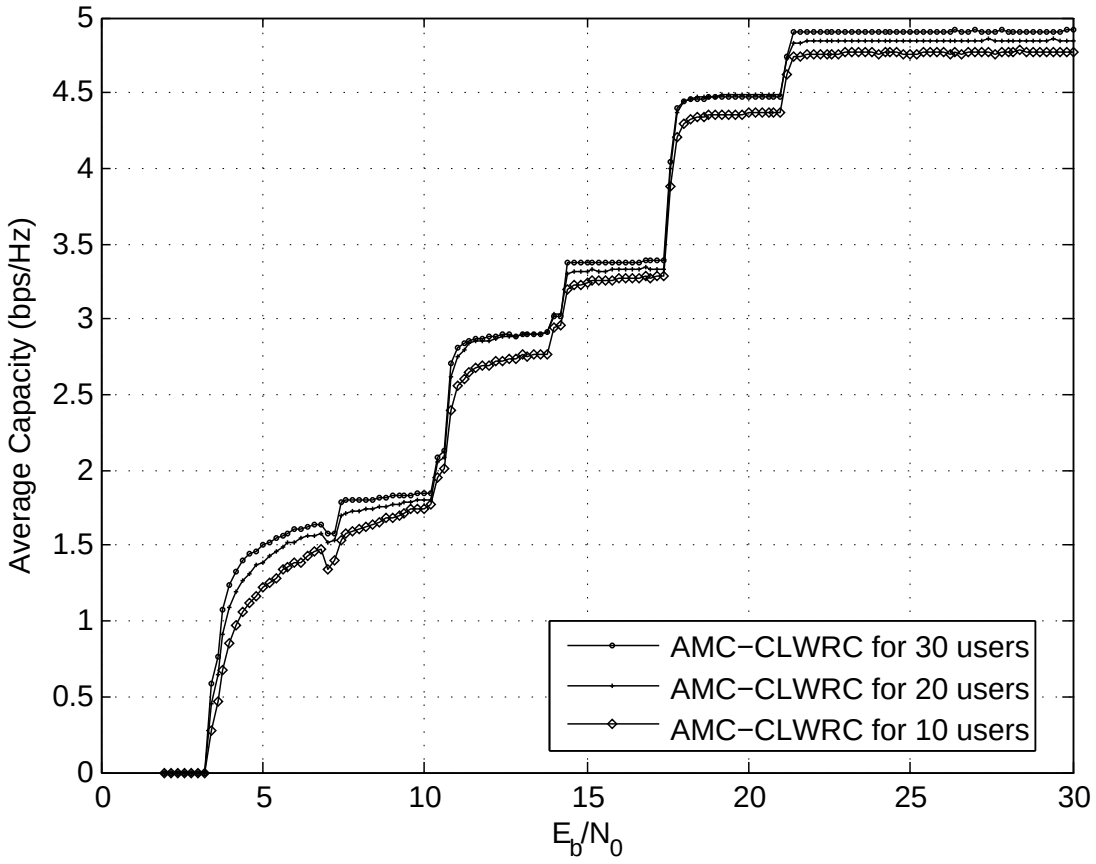

Figure 4.7. Average Capacity (bps/Hz) vs. average SNR per symbol (system serving 10 [4 voice and 6 data users], 20 [8 voice and 12 data users] and 30 [12 voice and 18 data users] users and implementing the AMC-CLWRC algorithm)

Table 4.5. Comparison of the SSCP for the proposed AMC-CLWRC scheme with different number of users

Number of users	10	20	30	40	50	60
SSCP ($\times 10^3$)	8.8225	9.0418	9.1457	9.1639	9.1426	9.0997

4.4.2 Complexity Comparison

As a case study, time complexity is considered here for comparison between the proposed AMC-CLWRC algorithm with the PRC and MF algorithms modified to support AMC. Table 4.6 shows a comparison between execution times (in seconds) of the proposed AMC-CLWRC with other known algorithms for different number of frames. The algorithms were executed on Intel(R) Core(TM) i5-2430M CPU @ 2.40 GHz 2.40 GHz processor for 10 MonteCarlo runs. It can be observed from the table that AMC-CLWRC algorithm has

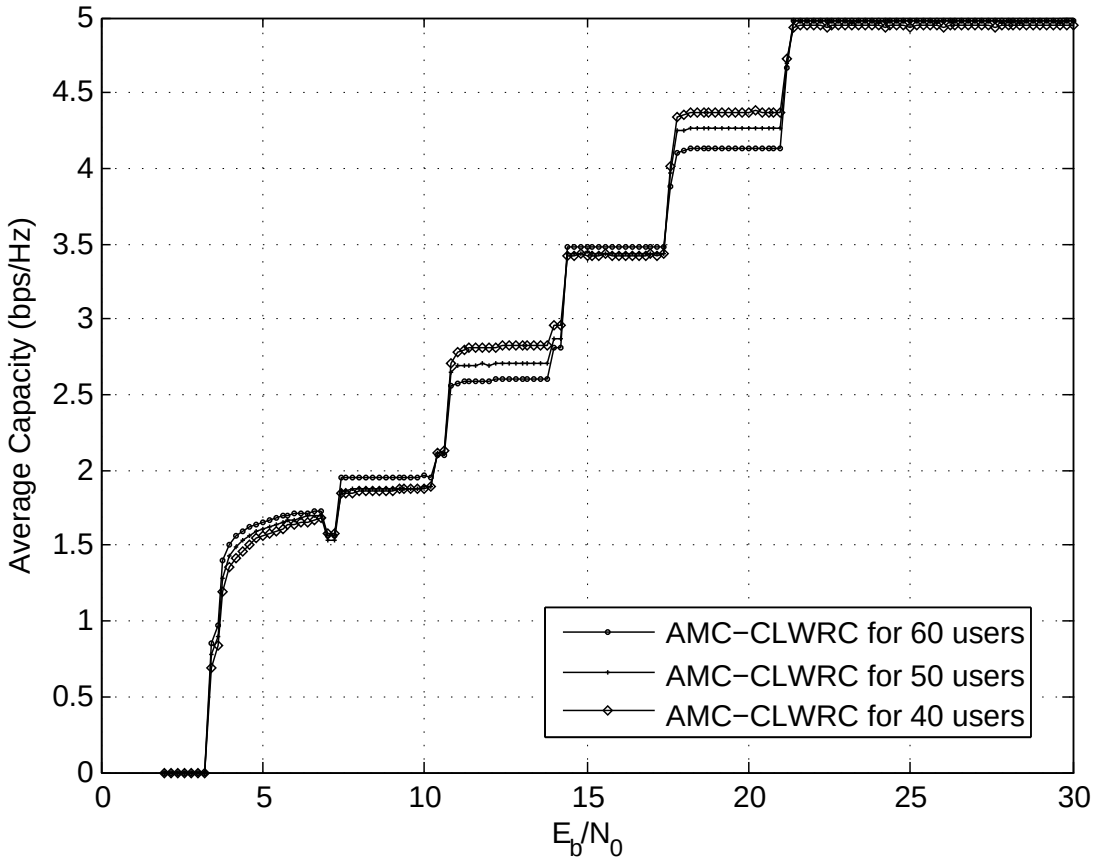

Figure 4.8. Average Capacity (bps/Hz) vs. average SNR per symbol (system serving 40 [16 voice and 24 data users], 50 [20 voice and 30 data users] and 60 [24 voice and 36 data users] users and implementing AMC-CLWRC algorithm)

faster execution time as compared to the other algorithms. This is due to the fact that in both the PRC and MF algorithms, after subchannels are allocated to all users, an initial power allocation algorithm is implemented before finalizing the power allocation based on water-filling approach. However, unlike as in [5], in the case of AMC-CLWRC algorithm right after subchannel allocation, power allocation based on water-filling approach is performed without considering initial power allocation. In the AMC-CLWRC algorithm, as an initial power allocation, an equal power distribution among each subchannel is considered and hence the execution time for the proposed algorithm is less. It is also interesting to note that, even without adopting the complex initial power allocation scheme as in [5], AMC-CLWRC algorithm has better performance in terms of total average system capacity optimization as compared to the other algorithms.

Table 4.6. Execution time (in seconds) for different algorithms and different number of frames

	Number of Frames		
Algorithms	1000	100	10
AMC-CLWRC	6079.74	605.46	61.42
PRC	9873.56	987.08	97.73
MF	10097.37	1005.99	101.02

4.5 Conclusion

In this chapter a detailed analysis on the CLWRC algorithm taking into account the adaptive modulation and coding, AMC-CLWRC algorithm, is presented. All the formulations in this chapter are based on the WiMAX QoS classes. A capacity comparison between the AMC-CLWRC and the other know algorithms over a range of average SNR is presented in this chapter. Moreover, a complexity comparison between the CLWRC and the other algorithms, modified accordingly to support AMC, is also performed here based on the algorithm execution time. The chapter also introduces a new performance metric SSCP for a fair quantitative comparison between different algorithms over the entire observed SNR range. A comparison between different algorithms based on SSCP metric is also presented here. From the numerical results and the SSCP evaluation, it has been observed that the AMC-CLWRC scheme supports higher system capacity as compared to the other algorithms. The AMC-CLWRC algorithm also maintains its superiority in terms of execution time as compared to the other algorithms.

CHAPTER 5

CONCLUSION

In this thesis a cross-layer resource allocation scheme for an OFDMA systems is presented. Air interface technologies like WiMAX and downlink LTE systems both use OFDMA as their multiple access mechanism. Both of these technologies are considered here and a detailed analysis on the resource allocation scheme as applied to these technologies are studied. The resource allocation scheme, CLWRC introduced in [11] for WiMAX air interface technology is reviewed and extended here. Similarly an extension to the CLWRC algorithm as applied to LTE air interface technology, LTE-CLWRC algorithm, is also presented in this thesis. Besides, the thesis also includes a detailed analysis on the CLWRC algorithm taking into account the adaptive modulation and coding, AMC-CLWRC. All the formulations required for AMC-CLWRC algorithm performance evaluation are based on the WiMAX QoS. A suboptimal solution to an optimization problem subject to the weighted capacity constraint is presented. The weighted capacity compensates for various cross-layer parameters and multi-class QoS requirements. The capacity optimization based on the CLWRC algorithm for an error-free Shannon channel condition for both the WiMAX and LTE systems is studied in this thesis. On the other hand, a non error-free channel condition and the subsequent implementation of AMC as applied to WiMAX is also studied in this thesis. When LTE-CLWRC algorithm is considered, an update on the cross-layer fairness parameters; service urgency and service satisfaction is presented accordingly so that they address the LTE QoS definition. A new performance metric SSCP for a fair quantitative comparison between

different algorithms over the entire observed SNR range is also introduced in this thesis.

The capacity performance of the CLWRC algorithm is evaluated subject to different observation scenarios like varying average SNR, different number of users and different frame numbers. A capacity comparison between the CLWRC algorithm and the other know algorithms is then presented for both the WiMAX and LTE systems. Moreover, a complexity comparison, based on the algorithm execution time, between the CLWRC and the other algorithms is also performed. It is observed from the numerical results that the CLWRC scheme results in total average system capacity that is closer to the Shannon limit as compared to the other known resource allocation schemes. On the other hand, unlike other known techniques, the CLWRC algorithm not only maintains the optimum system capacity for different number of users in the system but also increases as the number of users increases, confirming the multiuser diversity advantage of the CLWRC algorithm. The CLWRC algorithm also maintains its superiority in terms of execution time as compared to the other algorithms. In particular, the CLWRC scheme outperforms other known approaches in four aspects; closeness to Shannon capacity limit, consistency in terms of maximum optimum capacity throughout the frames considered, consistency in maintaining maximum optimum system capacity for different number of users and fast execution time. Similar observation scenarios are also considered for the capacity performance evaluation of the LTE-CLWRC algorithm. Based on the performance evaluation results obtained for the LTE-CLWRC algorithm, it can be concluded that the CLWRC algorithm has a similar impact on the LTE system performance as it had on the WiMAX system. An optimum system capacity that remain closer to the Shannon limit, multiuser diversity advantage and a superiority in terms of execution time, all are also observed in case of the LTE-CLWRC algorithm. The results obtained from a fair, in terms of available system resource, comparison shows that the LTE system supports higher system capacity as compared to the WiMAX system. A capacity and complexity comparison between the AMC-CLWRC and the other know algorithms, modified accordingly to support AMC, over a range of average SNR is also presented. The SSCP evaluation on the

capacity curves indicate that the AMC-CLWRC scheme supports higher system capacity as compared to the other algorithms. Similarly the AMC-CLWRC algorithm also maintains its superiority in terms of execution time as compared to the other algorithms. It can hence be concluded that the CLWRC presented in this thesis has a significant improvement in the system capacity performance for all the diverse scenarios considered.

An extension to this thesis work can be made on various areas. While IEEE 802.16e (WiMAX) QoS classes have been utilized in developing the work for AMC-CLWRC algorithm, it can be extended to LTE standard QoS classes where a common QoS treatment is offered to service data flows mapped to the same bearer. A comparison on the system performance while implementing the AMC-CLWRC scheme between WiMAX and LTE will also be an interesting extension to this work. It would also be interesting to consider the scenario with a majority of users traffic demanding the same QoS class and observe the performance of the algorithm. The analysis of the LTE-CLWRC algorithm could also be extended to implement allocation and retention priority (ARP) scheme for a congested system scenario. Moreover, the algorithm can be extended to support the control plane besides the data plane and also can be enhanced by supporting multiple users sharing subchannels/PRB in time, adding another dimension to multiuser diversity.

BIBLIOGRAPHY

BIBLIOGRAPGY

[1] O. M. Hammouri, "A cross-layer approach to resource-allocation with qos support in wimax networks," Master's thesis, University of Mississippi, University, MS, USA. www.standford.edu/group/coiffi/book/chap10.pdf, 2009.

[2] Y. J. Zhang and K. B. Letaief, "Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for ofdm systems," *Wireless Communications, IEEE Transactions on*, vol. 3, pp. 1566–1575, Sept 2004.

[3] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser ofdm system using dynamic subchannel allocation," in *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*, vol. 2, pp. 1085–1089 vol.2, 2000.

[4] Z. Shen, J. G. Andrews, and B. L. Evans, "Optimal power allocation in multiuser ofdm systems," in *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, vol. 1, pp. 337–341 Vol.1, Dec 2003.

[5] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser ofdm systems with proportional rate constraints," *Wireless Communications, IEEE Transactions on*, vol. 4, pp. 2726–2737, Nov 2005.

[6] D. N. C. Tse, "Forward link multiuser diversity through proportional fair scheduling," *presentation at Bell Labs*, Aug 1999.

[7] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *Information Theory, IEEE Transactions on*, vol. 48, pp. 1277–1294, Jun 2002.

[8] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Communications, 1995. ICC '95 Seattle, 'Gateway to Globalization', 1995 IEEE International Conference on*, vol. 1, pp. 331–335 vol.1, Jun 1995.

[9] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer design for wireless networks," *Communications Magazine, IEEE*, vol. 41, pp. 74–80, Oct 2003.

[10] K. Wongthavarawat and A. Ganz, "Packet scheduling for qos support in ieee 802.16 broadband wireless access systems," *International Journal of Communication Systems*, vol. 16, pp. 81–96, Feb 2003.

[11] M. M. Matalgah, O. M. Hammouri, and B. Paudel, "Cross-layer resource allocation approach in ofdma systems with multi-class qos services and users queue status," in *Global Telecommunications Conference, 2013 IEEE*, Dec. 2013.

[12] A. Ghosh, J. Ahang, J. Andrews, , and R. Muhamed, *Fundamentals of LTE*. Prentice-Hall, Aug. 2010.

[13] S. Srikanth and P. A. M. Pandian, "Orthogonal frequency division multiple access in wimax and lte; a comparison," in *Communications (NCC), 2010 National Conference on*, pp. 1–5, Jan 2010.

[14] J. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of Wimax: Understanding Broadband Wireless Networking*. Prentice-Hall, Feb. 2007.

[15] M. Alasti, B. Neekzad, J. Hui, and R. Vannithamby, "Quality of service in wimax and lte networks [topics in wireless communications]," *Communications Magazine, IEEE*, vol. 48, pp. 104–111, May 2010.

[16] "Physical layer aspects for evolved universal terrestrial radio access (utra) (release 7)," tech. rep. 3rd Generation Partnership Project: Technical Specification Group Radio Access Network: 3GPP TR 25.814 V7.1.0: 2006.

[17] W. McCoy, *Overview of 3GPP LTE Physical Layer*. White Paper, 2007.

[18] G. Ungerboeck, "Channel coding with multilevel/phase signals," *Information Theory, IEEE Transactions on*, vol. 28, pp. 55–67, Jan 1982.

[19] A. J. Goldsmith and S.-G. Chua, "Adaptive coded modulation for fading channels," *Communications, IEEE Transactions on*, vol. 46, pp. 595–602, May 1998.

[20] I. C. Wong, Z. Shen, B. L. Evans, and J. G. Andrews, "A low complexity algorithm for proportional resource allocation in ofdma systems," in *Signal Processing Systems, 2004. SIPS 2004. IEEE Workshop on*, pp. 1–6, Oct 2004.

[21] B. Xi, H. Chen, W. S. Cleveland, and T. Telkamp, "Statistical analysis and modeling of internet voip traffic for network engineering," *Electronic Journal of Statistics*, vol. 4, pp. 58–116, 2010.

[22] J. Freitag and N. L. S. da Fonseca, "Uplink scheduling with quality of service in ieee 802.16 networks," in *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, pp. 2503–2508, Nov 2007.

[23] "Physical and medium access control layers for combined fixed and mobile operation in licensed bands and corrigendum 1, amendment 2," tech. rep. IEEE Standard for Local and Metropolitan Area Networks: Air Interface for Fixed Broadband Wireless Access Systems : Part 16 : 2005.

[24] P. T. Brady, "A statistical analysis of on-off patterns in 16 conversations," *The Bell System Technical Journal*, vol. 47, pp. 73–91, Jan. 1968.

[25] P. e. a. Barford, "Changes in web client access patterns: Characteristics and caching implications," *World Wide Web 2*, pp. 15–28, 1999.

[26] Q. Liu, X. Wang, and G. Giannakis, "A cross-layer scheduling algorithm with qos support in wireless networks," *Vehicular Technology, IEEE Transactions on*, vol. 55, pp. 839–847, May 2006.

[27] Y. Zaki, T. Weerawardane, C. Gorg, and A. Timm-Giel, "Multi-qos-aware fair scheduling for lte," in *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, pp. 1–5, May 2011.

[28] A. Pokhariyal, K. I. Pedersen, G. Monghal, I. Z. Kovacs, C. Rosa, T. E. Kolding, and P. E. Mogensen, "Harq aware frequency domain packet scheduler with different degrees of fairness for the utran long term evolution," in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pp. 2761–2765, April 2007.

[29] Q. Ai, P. Wang, F. Liu, Y. Wang, F. Yang, and J. Xu, "Qos-guaranteed cross-layer resource allocation algorithm for multiclass services in downlink lte system," in *Wireless Communications and Signal Processing (WCSP), 2010 International Conference on*, pp. 1–4, Oct 2010.

[30] I.-H. Hou and C. S. Chen, "Self-organized resource allocation in lte systems with weighted proportional fairness.," in *ICC*, pp. 5348–5353, IEEE, 2012.

[31] L. Zhao, Y. Qin, M. Ma, X. Zhong, and L. Li, "Qos guaranteed resource block allocation algorithm in lte downlink," in *Communications and Networking in China (CHINACOM), 2012 7th International ICST Conference on*, pp. 425–429, Aug 2012.

[32] A. P. Avramova, Y. Yan, and L. Dittmann, "Cross layer scheduling algorithm for lte downlink," in *Telecommunications Forum (TELFOR), 2012 20th*, pp. 307–310, Nov 2012.

[33] G. Femenias and F. Riera-Palou, "Cross-layer resource allocation and scheduling in lte systems under imperfect csit," in *Wireless Days (WD), 2013 IFIP*, pp. 1–8, Nov 2013.

[34] S. A. Filin, S. N. Moiseev, M. S. Kondakov, A. V. Garmonov, A. Y. Savinkov, Y. S. Park, D. H. Yim, J. H. Lee, S. H. Cheon, and K. L. Han, "Qos-guaranteed cross-layer adaptive transmission algorithms with selective arq for the ieee 802.16 ofdma system," in *Vehicular Technology Conference, 2006. VTC-2006 Fall. 2006 IEEE 64th*, pp. 1–5, Sept 2006.

[35] D. Marabissi, D. Tarchi, R. Fantacci, and F. Balleri, "Efficient adaptive modulation and coding techniques for wimax systems," in *Communications, 2008. ICC '08. IEEE International Conference on*, pp. 3383–3387, May 2008.

[36] S. M. B. Taslim, S. M. R. Tousif, and M. Tareq, "A novel algorithm with a new form of adaptive modulation for mobile wimax performance improvement," in *Advanced Communication Technology (ICACT), 2011 13th International Conference on*, pp. 300–305, Feb 2011.

[37] M. Mazzotti, S. Moretti, and M. Chiani, "Multiuser resource allocation with adaptive modulation and ldpc coding for heterogeneous traffic in ofdma downlink," *Communications, IEEE Transactions on*, vol. 60, pp. 2915–2925, October 2012.

[38] "Policy and charging control architecture (release 12)," tech. rep. 3rd Generation Partnership Project: Technical Specification Group Services and System Aspects: 3GPP TS 23.203 V12.3.0: Dec. 2013.

[39] "Further advancements for e-utra physical layer aspects(release 7)," tech. rep. 3rd Generation Partnership Project: Technical Specification Group Radio Access Network: Evolved Universal Terrestrial Radio Access (E-UTRA):3GPP TR 36.814 V9.0.0: Mar. 2010.

[40] A. J. Goldsmith, *Wireless Communication: Meet the Challenges and Opportunities in Wireless Commnication.* Cambridge University Press, 2005.

[41] J. M. Cioffi, *Chapter 10: System Desigh with Codes.* Department of Electrical Engineering, Standford University, 2009.

LIST OF APPENDICES

APPENDIX A: TRAFFIC GENERATION

TRAFFIC GENERATION

Details on the simulation of various kinds of traffic corresponding to different WiMAX service classes and LTE QCIs as discussed in Sections 2.5.3 and 3.3.1 and based on Tables 2.2 and 3.3, respectively is presented this section. Various traffic models are used to generate the WiMAX traffic. VoIP without silence suppression traffic model is considered to generate the UGS traffic, MPEG video traffic model to generate the rtPS, VoIP with silence suppression traffic model to generate the ErtPS, FTP traffic model to generate the nrtPS and HTTP traffic model to generate the BE traffic. The same traffic models are also used to generate the LTE traffic corresponding to $QCI1$, $QCI3$, $QCI1$, $QCI8$ and $QCI9$ traffic, respectively. The only difference in WiMAX and LTE traffic generation is the frame duration consideration. The frame duration of WiMAX is 5ms while that for LTE is 1ms.

VoIP without silence suppression (VoIPw/oSS) Traffic generation

VoIPw/oSS traffic is associated with $QCI1$ and is continuously generated on a periodic basis based on the voice processing interval (VPI). Therefore, there is no random distribution function associated with this kind of traffic generation. The simulation of this kind of traffic is depicted in Algorithm 3 and the terms used in this algorithm are defined as follows: VPI represents voice processing interval in miliseconds, N is total number of frames, n is specific n^{th} frame number, T_f is the frame duration, T is the total time for which traffic is generated in miliseconds, PS is the packet size in Bytes, VoIPw/oSS_traffic represents a 1 by N matrix that stores VoIPw/oSS traffic packet size, $\bigotimes X$ represents multiple of X and $\lceil x \rceil$ is a ceiling function that returns smallest integer larger than x.

The working of VoIPw/oSS traffic simulation depicted in Algorithm 3 is described in detail in the following. First, all the required parameters are initialized. The condition for

Algorithm 3 VoIPw/oSS traffic generation algorithm

1: **initialize parameters:** $t \Leftarrow 1,\ VPI \Leftarrow 20,\ N \Leftarrow 1000,\ T_f \Leftarrow 5/1,\ PS \Leftarrow 66$
2: **initialize array:** VoIP_traffic $\Leftarrow 0$
3: $T \Leftarrow N \times T_f$
4: **for** $t = 1 \rightarrow T$ **do**
5: **if** $(t-1) = 0$ or $\bigotimes uVPI$ **then**
6: $n \Leftarrow \lceil t/T_f \rceil$
7: VoIP_traffic$(1, n) \Leftarrow uPS$
8: **end if**
9: **end for**

which traffic could be generated is checked every instant of time, the smallest unit of time for our case is one millisecond. VoIPw/oSS traffic is generated at a regular interval defined by VPI, so this condition is checked and the traffic or the packet to be transmitted during a frame is updated. The traffic so generated should be stored on per frame basis not at every instant, this is why the frame duration defined by T_f comes into play in the simulation. All these processes of traffic generation are repeated every millisecond until they are repeated T number of times.

Moving Pictures Experts Group video (MPEGV) Traffic

The arrival time for MPEGV traffic is a random process such that the probability of receiving a packet at a given time follows a Bernoulli distribution and MPEGV traffic is associated with $QCI3$. For every i^{th} traffic connection, the packet arrival process to the queue is Bernoulli distributed with a given average rate of μ_{MPEGV} (bps) and probability of failure $\bar{p}$ [26]. From the definition of mean we have that the mean of x (set of elements) is equal to the summation of elements of x times the probability of occurrence of that element. Using the same analogy instantaneous arriving rate is evaluated. As a result, the instantaneous arriving rate at time t, A_i(t) can be expressed as:

$$A_i(t) = \begin{cases} 0, & \text{with probability } \bar{p} \\ \frac{\mu_{MPEGV}}{(1-\bar{p})}, & \text{with probability (1 - } \bar{p}) \end{cases} \quad (.0.1)$$

In [22], [26], μ_{MPEGV} and $\bar{p}$ are 64Kbps and 0.4, respectively. The simulation of MPEGV traffic is depicted in Algorithm 4 and the terms used in this algorithm are defined as follows: $\bar{p}$ represents the probability of failure for bernoulli distribution, μ_{MPEGV} is the average number

of bits per second for the required MPEGV traffic and MPEGV_traffic represents a 1 by N matrix that stores MPEGV traffic packet size in bits.

Algorithm 4 MPEGV traffic generation algorithm

1: **initialize parameters:** $t \Leftarrow 1,\ N \Leftarrow 1000, T\_f \Leftarrow 5,\ \bar{p} \Leftarrow 0.4,\ \mu_{MPEGV} \Leftarrow 64 \times 1024$
2: **initialize array:** MPEGV_traffic $\Leftarrow 0$
3: $T \Leftarrow N \times T_f$
4: **for** $t = 1 \rightarrow T$ **do**
5: **if** $(t-1) = 0$ or $\bigotimes T_f$ **then**
6: $AE \Leftarrow binomialRV(1, 1-\bar{p})$
7: **if** $AE > 0$ **then**
8: $n \Leftarrow \lceil t/T_f \rceil$
9: MPEGV_traffic$(1, n) \Leftarrow \frac{T_f}{1000} \times \frac{\mu_{MPEGV}}{1-\bar{p}}$
10: **end if**
11: **end if**
12: **end for**

The working of MPEGV traffic simulation depicted in Algorithm 4 is described in detail in the following. First all the required parameters are initialized. A binomial random variable corresponding to the packet arrival event with parameters: number of trails as 1 and probability of success as $1 - \mu_{MPEGV}$ is generated. The random variable is generated once during a frame duration time T_f and if the random variable so generated is greater than 0 then MPEGV traffic is generated else not. Once again it is to be noted that the traffic so generated should be stored on per frame basis not at every instant. All these processes of traffic generation are repeated every millisecond until they are repeated T number of times.

VoIP with silence suppression (VoIP/wSS) Traffic

The VoIP/wSS traffic generation period is a random process such that the time interval for which the traffic packets are generated or not follows the exponential random distribution and VoIP/wSS traffic is also associated with $QCI1$. The simulation of VoIP/wSS traffic is depicted in Algorithm 5 and the terms used in this algorithm are defined as follows: μ_{on} is the mean time in milliseconds during which traffic generation is valid (ON duration) while μ_{off} is the mean time in milliseconds during which traffic generation is not valid (OFF duration) corresponding to the silence suppression, *Period* is a variable counting the ON and OFF time duration, *State* is a variable corresponding to ON and OFF state of traffic generation,

VPI is the voice processing interval in milliseconds, PS represents a packet size in Bytes and VoIP/wSS_traffic represents a 1 by N matrix that stores VoIP/wSS traffic packet size.

Algorithm 5 VoIP/wSS traffic generation algorithm

1: **initialize parameters:** $t \Leftarrow 1,\ N \Leftarrow 1000,\ T_f = 5,\ Period = 0,\ State = 0,\ \mu_{on} \Leftarrow 1.2,\ \mu_{off} \Leftarrow 1.8,\ VPI \Leftarrow 30,\ PS = 66$
2: **initialize array:** VoIP/wSS_traffic $\Leftarrow 0$
3: $T \Leftarrow N \times T_f$
4: **for** $t = 1 \rightarrow T$ **do**
5: **if** Period $= 0$ **then**
6: **if** State $= 0$ **then**
7: $Period \Leftarrow \lceil exponentialRV \sim X(\mu_{on}) \rceil$
8: $State \Leftarrow 1$
9: **else**
10: $Period \Leftarrow \lceil exponentialRV \sim X(\mu_{off}) \rceil$
11: $State \Leftarrow 0$
12: **end if**
13: **else**
14: $Period \Leftarrow Period - 1$
15: **if** State $= 1$ **then**
16: **if** VPI $= 0$ **then**
17: $n \Leftarrow \lceil t/T_f \rceil$
18: VoIP/wSS_traffic$(1, n) \Leftarrow PS$
19: $VPI \Leftarrow 30$
20: **else**
21: $VPI \Leftarrow VPI - 1$
22: **end if**
23: **end if**
24: **end if**
25: **end for**

The working of VoIPwSS traffic simulation depicted in Algorithm 5 is described in detail in the following. First, all the required parameters are initialized. If the variables *Period* and *State* are 0 (system ready for ON duration) then a random variable that takes the parameter μ_{on} as the mean of exponential distribution is generated, the ceiling value of which corresponds to the ON duration. The *State* variable is then changed to 1 such that when ON duration is complete, i.e, the variable *Period* is decreased down to 0, system can jump to the OFF duration. So, when *Period* is 0 and *State* is 1 (end of ON duration), a random variable that takes the parameter μ_{off} as the mean of exponential distribution is generated, the ceiling value of which corresponds to the OFF duration. The *State* variable

is then changed back to 0 such that the system will be ready to switch back to the ON duration when the variable *Period* is decreased down to 0. Now, if the system is in ON duration and is not within the VPI limit, then the VoIPwSS traffic is generated. Then the traffic so generated is stored on per frame basis not at every instant. All these processes of traffic generation are repeated every millisecond until they are repeated T number of times.

File Transfer Protocol (FTP) Traffic

FTP packet size is a random process such that it follows the exponential distribution and the FTP traffic is associated with $QCI8$. The simulation of FTP traffic is depicted in Algorithm 6 and the terms used in this algorithm are defined as follows: μ_{FTP} represents the average packet size per second corresponding to the mean of exponential distribution as the traffic follows exponential distribution and FTP_traffic represents a 1 by N matrix that stores FTP traffic packet size.

Algorithm 6 FTP traffic generation algorithm

1: **initialize parameters:** $t \Leftarrow 1,\ N \Leftarrow 1000,\ T_f \Leftarrow 5,\ \mu_{FTP} \Leftarrow 512 \times 1024$
2: **initialize parameters:** FTP_traffic $\Leftarrow 0$
3: $T \Leftarrow N \times T_f$
4: **for** $t = 1 \rightarrow T$ **do**
5: **if** $(t-1) = 0$ or $\bigotimes T_f$ **then**
6: $n \Leftarrow \lceil t/T_f \rceil$
7: FTP_traffic $(1, n) \Leftarrow exponentialRV \sim X(\mu_{FTP} \times \frac{T_f}{1000})$
8: **end if**
9: **end for**

The working of FTP traffic simulation depicted in Algorithm 6 is described in detail in the following. First, all the required parameters are initialized. FTP traffic is then generated as a random variable that takes the parameter $\mu_{FTP} \times \frac{T_f}{1000}$ as the mean of the exponential distribution. The traffic so generated as always is stored on per frame basis. All these processes of traffic generation are repeated every millisecond until they are repeated T number of times.

Hyper-Text Transfer Protocol (HTTP) Traffic

The HTTP traffic is modeled using Lognormal/Pareto distribution and is associated with $QCI9$. The simulation of HTTP traffic is depicted in Algorithm 7 and the terms used in this algorithm are defined as follows: μ_{pareto} represents mean packet size per second corresponding to the mean of pareto distribution, where traffic follows pareto distribution, $\mu_{lognormal}$ represents mean packet size per second corresponding to the mean of lognormal distribution, where traffic follows lognormal distribution, PL_area is the part of the distribution of traffic where the lognormal distribution is considered, L is a logarithmic mean that defines the lognormal distribution based on $\mu_{lognormal}$, σ is the standard deviation of the lognormal distribution, α is the shape parameter of the pareto distribution, A is the random variable from the standard normal distribution on the open interval of (0,1) and HTTP_traffic represents a 1 by N matrix that stores HTTP traffic packet size.

Algorithm 7 HTTP traffic generation algorithm

1: **initialize parameters:** $t \Leftarrow 1,\ n \Leftarrow 1000,\ T_f \Leftarrow 5, \mu_{pareto} \Leftarrow 10558,\ \mu_{lognormal} \Leftarrow 7247,\ PL\_area \Leftarrow 0.88, \sigma \Leftarrow 2, \alpha \Leftarrow 2$
2: **initialize parameters:** HTTP_traffic $\Leftarrow 0$
3: $T \Leftarrow N \times T_f$
4: $L \Leftarrow ln(\frac{\mu_{lognormal} \times T_f}{1000} - 2)$
5: **for** $t = 1 \rightarrow T$ **do**
6: **if** $(t-1) = 0$ or $\bigotimes T_f$ **then**
7: $A \Leftarrow normalRV$
8: $n \Leftarrow \lceil t/T_f \rceil$
9: **if** A $>$ PL_area **then**
10: HTTP_traffic $(1, n) \Leftarrow \frac{\mu_{pareto} \times T_f}{1000 \times 2 \times (normalRV)^{(1/\alpha)}}$
11: **else**
12: HTTP_traffic $(1, n) \Leftarrow lognormalRV \sim X(L, \sigma)$
13: **end if**
14: **end if**
15: **end for**

The working of HTTP traffic simulation depicted in Algorithm 7 is described in detail in the following. First, all the required parameters are initialized. Then a random variable from the standard normal distribution on the open interval of (0,1) is generated such that it corresponds to the % of area of the total pdf. If the random variable thus generated is greater than 0.88 we generate HTTP traffic considering the Pareto distribution else we consider a

lognormal distribution to generate the HTTP traffic. Once again it is important to note that the traffic so generated is stored on per frame basis. All these processes of traffic generation are repeated every millisecond until they are repeated T number of times.

APPENDIX B: RAYLEIGH CHANNEL SIMULATION

RAYLEIGH FADING CHANNEL SIMULATION

The channel conditions as discussed in Section 2.6 and as explained in [5] is simulated using Clark's Model. The procedure that enumerates the steps to implement Clark's Model is depicted in Algorithm 8. Following are the terms that are used in the algorithm. N_w is the number of azimuthal plane waves each with arbitrary carrier phase and arbitrary azimuthal angle of arrival. These waves are random with same average amplitudes. t is the time at which multipaths occurs, f_n is the Doppler shift for the arriving n^{th} path, I is the total number of multipath waves considered, $\mathcal{NR}(i,j)$ generates a i by j matrix consisting of the values from the standard normal distribution each with zero mean and unit variance, c_n^k is the real random variable representing the amplitude of individual waves and A_n^k is normalized a_n^k for a k^{th} user, Φ_n^k is the phase angle of the n^{th} arriving component and ϕ_n^k is normalized Φ_n for a k^{th} user, where the phase angles are assumed to have a uniform pdf on the interval $(0, 2\pi]$, E_o represents real amplitude of local average E-field which is assumed to be exponentially decaying with e^{-2l_m}, where l_m is the multipath index and T_c^k is the in phase component of E-field while T_s^k is the quadrature phase component of E-field for a k^{th} user and are Gaussian random process with zero mean and unit variance.

The working of Algorithm 8 is described in detail in the following. The algorithm depicts the Rayleigh fading channel simulation based on Clark's model which takes the parameters like: number of users, time delay spread, and Doppler Shift as inputs when invoked in the resource allocator algorithm. The amplitude and phase of azimuthal plane waves are then randomly generated. The amplitudes of the E-field are so normalized that the ensemble average of a_n's is 1. The E-field can be approximated as Gaussian random variables if total number of plane waves is sufficiently large. In-phase and quadrature phase components of E-field corresponding to a k^{th} user is then evaluated as

$$T_c^k(i) = E_o(i) \sum_{n=1}^{N_w} A_n cos(2\pi f_n t + \phi_n) \tag{B-1}$$

$$T_s^k(i) = E_o(i) \sum_{n=1}^{N_w} A_n sin(2\pi f_n t + \phi_n) \tag{B-2}$$

The envelope of the received E-field corresponding to a k^{th} user can then be found as

$$E(k) = \sqrt{(T_c^k)^2 + (T_s^k)^2} \tag{B-3}$$

This envelope corresponds to the modeling of a frequency selective channel consisting of six independent Rayleigh multipaths.

Algorithm 8 Clark's Model

Input: K, t, f_n, I
Initialize: $N_w \Leftarrow 200$
for $k = 1 \rightarrow K$ **do**
  $c_n^k \Leftarrow \mathcal{NR}(1, N)$
  $C_n^k \Leftarrow \frac{c_n^k}{\sqrt{\sum_n (c_n^k)^2}}$
  $\Phi_n^k \Leftarrow \mathcal{NR}(1, N)$
  $\phi_n^k \Leftarrow \frac{2\pi\Phi_n^k}{max(\Phi_n^k)}$
  $E_o(i) \Leftarrow e^{-2l_m} \qquad \forall\ 2l_m \in [0,\ I-1]$
  for $i = 1 \rightarrow I$ **do**
    $T_c^k(i) \Leftarrow E_o(i) \sum_{n=1}^{N_w} C_n^k cos(2\pi f_n t + \phi_n^k)$
    $T_s^k(i) \Leftarrow E_o(i) \sum_{n=1}^{N_w} C_n^k sin(2\pi f_n t + \phi_n^k)$
  end for
end for

APPENDIX C: WATER-FILLING DERIVATION

DERIVATION OF WATER-FILLING EQUATION

Derivation of equation in (3.4.18)

The optimization problem in (3.4.12) is as follow

$$\max_{P_{k,l}} \quad C = \sum_{k=1}^{K} \sum_{l \in \Pi_k} \frac{1}{L} log_2 \left(1 + \frac{P_{k,l} h_{k,l}^2}{N_0 \frac{B}{L}} \right) \; bits/symbol/Hz \tag{C-1}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \sum_{l \in \Pi_k} P_{k,l} \leq P_{tot} \tag{C-2}$$

$$P_{k,l} \geq 0 \;\; \forall \;\; k, l \tag{C-3}$$

$$\Pi_i \cap \Pi_j = \Phi \;\; \forall \;\; i \neq j \tag{C-4}$$

$$\Pi_1 \cup \Pi_2 \cup \ldots \cup \Pi_K \subseteq \{1, 2, \ldots, L\} \tag{C-5}$$

$$R_i(n) = R_j(n) = R(n) \; \forall \; i, j \in [1, 2, \ldots, K], \tag{C-6}$$

This optimization problem can be represented with a cost function

$$\begin{aligned} f_L \;\; = \;\; & \sum_{k=1}^{K} \sum_{l \in \Pi_k} (\frac{1}{L} log_2(1 + P_{k,l} H_{k,l})) \\ & + \lambda_1 \sum_{k=1}^{K} \sum_{l \in \Pi_k} (P_{k,l} - P_{tot}) \\ & + \lambda_k \sum_{k=2}^{K} \sum_{l \in \Pi_1} (\frac{1}{L} log_2(1 + P_{1,l} H_{1,l})) \\ & - \lambda_k \sum_{k=2}^{K} \sum_{l \in \Pi_k} (\frac{S_1}{U_1} \frac{U_k}{S_k} \frac{1}{L} log_2(1 + P_{k,l} H_{k,l})) \end{aligned} \tag{C-7}$$

where $\{\lambda_k\}_{k=1}^{K}$ are the Lagrangian multipliers. The objective is to maximize the cost, f_L. Differentiating (C-7) with respect to $P_{k,l}$

$$\begin{aligned}\frac{\partial f_L}{\partial P_{1,l}} &= \frac{1}{L\ ln\ 2}\frac{H_{1,l}}{1+P_{1,l}H_{1,l}} \\ &\quad +\lambda_1+\sum_{k=2}^{K}\lambda_k\frac{1}{L\ ln\ 2}\frac{H_{1,l}}{1+P_{1,l}H_{1,l}}=0 \end{aligned} \tag{C-8}$$

$$\begin{aligned}\frac{\partial f_L}{\partial P_{k,l}} &= \frac{1}{L\ ln\ 2}\frac{H_{k,l}}{1+P_{k,l}H_{k,l}} \\ &\quad +\lambda_1-\lambda_k\frac{S_1}{U_1}\frac{U_k}{S_k}\frac{1}{L\ ln\ 2}\frac{H_{k,l}}{1+P_{k,l}H_{k,l}}=0 \end{aligned} \tag{C-9}$$

for $k \in \{2,3,\ldots,K\}$ and $l \in \Pi_k$.

From either (C-8) or (C-9)

$$\frac{H_{k,x}}{1+P_{k,x}H_{k,x}} = \frac{H_{k,y}}{1+P_{k,y}H_{k,y}} \tag{C-10}$$

for $x, y \in \Pi_k$ and $k \in \{1,2,\ldots,L\}$. Without loss of generality, if the PRBs were recorded in ascending order such that $H_{k,1} \leq H_{k,2} \leq \ldots \leq H_{k,|\Pi_k|}$, where $|\Pi_k|$ indicates the total number of elements of vector Π_k, then (C-10) can be rewritten as

$$P_{k,x} = P_{k,1} + \frac{H_{k,x}-H_{k,1}}{H_{k,x}H_{k,1}} \tag{C-11}$$

for $k \in \{1,2,\ldots,K\}$ and $x \in \{1,2,\ldots,|\Pi_k|\}$.

Derivation of equation in (4.3.22)

The optimization problem in (4.3.16) is as follow

$$\max_{P_{k,l}} \quad C=\sum_{k=1}^{K}\sum_{l\in\Pi_k}\frac{1}{L}\ R_c\ log_2\left[1+a\ P_{k,l}\ H_{k,l}\right]\ bits/symbol/Hz \tag{C-12}$$

$$\text{subject to} \quad \sum_{k=1}^{K}\sum_{l\in\Pi_k}P_{k,l} \leq P_{tot} \tag{C-13}$$

$$P_{k,l} \geq 0\ \ \forall\ \ k,l \tag{C-14}$$

$$\Pi_u \cap \Pi_v = \Phi\ \ \forall\ \ u \neq v \tag{C-15}$$

$$\Pi_1 \cup \Pi_2 \cup \ldots \cup \Pi_K \subseteq \{1,2,\ldots,L\} \tag{C-16}$$

$$R_u(n) = R_v(n) = R(n)\ \forall\ u,v \in [1,2,\ldots,K], \tag{C-17}$$

This optimization problem can be represented with a cost function

$$f_L = \sum_{k=1}^{K}\sum_{l\in\Pi_k}\left(\frac{1}{L}\ R_c\ log_2(1+aP_{k,l}H_{k,l})\right) \tag{C-18}$$

$$+\lambda_1\sum_{k=1}^{K}\sum_{l\in\Pi_k}(P_{k,l}-P_{tot}) \tag{C-19}$$

$$+\lambda_k\sum_{k=2}^{K}\sum_{l\in\Pi_1}\left(\frac{1}{L}\ R_c\ log_2(1+aP_{1,l}H_{1,l})\right) \tag{C-20}$$

$$-\lambda_k\sum_{k=2}^{K}\sum_{l\in\Pi_k}\left(\frac{S_1}{U_1}\frac{U_k}{S_k}\frac{1}{L}\ R_c\ log_2(1+aP_{k,l}H_{k,l})\right) \tag{C-21}$$

$$\tag{C-22}$$

where $\{\lambda_k\}_{k=1}^{K}$ are the Lagrangian multipliers. The objective is to maximize the cost, f_L. Differentiating (C-7) with respect to $P_{k,l}$

$$\frac{\partial f_L}{\partial P_{1,l}} = \frac{R_c}{L\ ln\ 2}\frac{aH_{1,l}}{1+aP_{1,l}H_{1,l}}$$

$$+\lambda_1+\sum_{k=2}^{K}\lambda_k\frac{R_c}{L\ ln\ 2}\frac{aH_{1,l}}{1+aP_{1,l}H_{1,l}}=0 \tag{C-23}$$

$$\frac{\partial f_L}{\partial P_{k,l}} = \frac{R_c}{L\ ln\ 2}\frac{aH_{k,l}}{1+aP_{k,l}H_{k,l}}$$

$$+\lambda_1-\lambda_k\frac{S_1}{U_1}\frac{U_k}{S_k}\frac{R_c}{L\ ln\ 2}\frac{aH_{k,l}}{1+aP_{k,l}H_{k,l}}=0 \tag{C-24}$$

for $k\in\{2,3,\ldots,K\}$ and $l\in\Pi_k$.

From either (C-23) or (C-24)

$$\frac{H_{k,x}}{1+aP_{k,x}H_{k,x}}=\frac{H_{k,y}}{1+aP_{k,y}H_{k,y}} \tag{C-25}$$

for $x,y\in\Pi_k$ and $k\in\{1,2,\ldots,K\}$. Without loss of generality, if the subchannels were recorded in ascending order such that $H_{k,1}\le H_{k,2}\le\ldots\le H_{k,|\Pi_k|}$, where $|\Pi_k|$ indicates the total number of elements of vector Π_k, then (C-25) can be rewritten as

$$P_{k,x}=P_{k,1}+\frac{H_{k,x}-H_{k,1}}{aH_{k,x}H_{k,1}} \tag{C-26}$$

or,

$$P_{k,x}=P_{k,1}+\frac{-ln(5\ P_b)(H_{k,x}-H_{k,1})}{1.5\ G_c\ H_{k,x}\ H_{k,1}} \tag{C-27}$$

for $k\in\{1,2,\ldots,K\}$ and $x\in\{1,2,\ldots,|\Pi_k|\}$.

APPENDIX D: SUBCHANNEL/PRB ALLOCATOR

SUBCHANNEL/PRB ALLOCATOR

Subchannel/PRB Allocator for WiMAX/LTE

The subchannel/PRB allocator algorithm assigns subchannels/PRB to a user based on the channel gain offered to the user by a particular subchannel/PRB under Rayleigh fading channel condition. The subchannel/PRB allocator algorithm is shown in Algorithm 9 and the terms used in this algorithm are defined as follows: X_{init} is the initial value of variable X for the purpose of comparison in simulation, λ_l indicates the occupancy of l^{th} subchannel and $max(X)$ represents the element with maximum value in X matrix.

The working of subchannel/PRB allocator algorithm is described in detail in the following. The subchannel/PRB allocation algorithm takes the parameters K, L, P_{tot}, B, $H_{k,l}$, and $\gamma_{SF^x(k)}$ as an input when invoked in the resource allocator algorithm. First, the total available system power is divided equally among subchannels/PRBs and is denoted by p. The subchannel/PRB with maximum channel gain in the Rayleigh fading environment for a particular user is then found. This subchannel/PRB with maximum gain is then assigned to the user so found and $\rho_{k,l}$ and λ_l are updated. User data rate supported by the assigned subchannel/PRB is then calculated. A fairness compensated data rate R_k is calculated using the weighting factor for all the users. The user with the minimum fairness compensated data rate is then found. Since the main purpose here is to maximize the capacity of the system, user supporting minimum R_k found in preceding step is given first priority and next subchannel/PRB is allocated to that user such that the total capacity for the user will increase. For assigning the next subchannel/PRB, from among the unassigned subchannels/PRB, the subchannel/PRB with maximum gain is found for the user supporting minimum R_k. This subchannel/PRB with maximum gain is then allocated to the user so found that supports minimum R_k. $\rho_{k,l}$ and λ_l are then updated and the sum capacity for that user is calculated.

This process is continued in a loop till all the subchannels/PRBs are allocated to users.

Subchannel Allocator for AMC implementation

The subchannel allocator algorithm for AMC implementation is shown in Algorithm 10 and the terms used in this algorithm are defined as follows: X_{init} is the initial value of variable X for the purpose of comparison in simulation, λ_l indicates the occupancy of l^{th} subchannel and $max(X)$ represents the element with maximum value in X matrix.

The working of subchannel allocator algorithm is described in detail in the following. The subchannel allocation algorithm takes the parameters K, L, P_{tot}, B, $H_{k,l}$, γ_i^{voice} and γ_i^{data} and $\Gamma_{SF^x(k)}$ as an input when invoked in the resource allocator algorithm. First, the total available system power is divided equally among subchannels and is denoted by p. The subchannel with maximum channel gain in the Rayleigh fading environment for a particular user is then found. This subchannel with maximum gain is then assigned to the user so found. Based on the QoS requirement of the user being served, the required BER is selected. BER so selected and average system SNR are then considered to pick the best γ_i and the corresponding MCS_i. User data rate supported by the assigned subchannel is then calculated and $\rho_{k,l}$ and λ_l are updated. A fairness compensated data rate R_k is calculated using the weighting factor for all the users. The user with the minimum fairness compensated data rate is then found. Since the main purpose here is to maximize the capacity of the system, user supporting minimum R_k found in preceding step is given first priority and next subchannel is allocated to that user such that the total capacity for the user will increase. For assigning the next subchannel, from among the unassigned subchannels, the subchannel with maximum gain is found for the user supporting minimum R_k. This subchannel with maximum gain is then allocated to the user so found that supports minimum R_k. $\rho_{k,l}$ and λ_l are then updated and the sum capacity for that user is calculated as explained earlier. This process is continued in a loop till all the subchannels are allocated to users.

Algorithm 9 Subchannel/PRB Allocator for WiMAX/LTE

Input: $K,\ L,\ B,\ N_0,\ P_{tot},\ H_{k,l},\ U_k,\ S_k$, where $k \in \{1, 2, .., K\}, l \in \{1, 2, .., L\}$
Initialize array: $\rho_{k,l} \Leftarrow 0, C_k \Leftarrow 0, \lambda_l \Leftarrow 0$
Initialize: $p \Leftarrow \frac{P_{tot}}{L},\ noise \Leftarrow N_0 \frac{B}{L}$
for $k = 1 \rightarrow K$ **do**
  $H_{init} \Leftarrow 0$
  for $l = 1 \rightarrow L$ **do**
    if $\lambda_l = 0$ & $H_{k,l} > H_{init}$ **then**
      $H_{init} \Leftarrow H_{k,l}$
      $l^* \Leftarrow l$
    end if
  end for
  $C_k \Leftarrow \frac{1}{L} log_2(1 + p \times H_{k,l^*})$
  $\lambda_{l^*} \Leftarrow 1,\ \rho_{k,l^*} \Leftarrow 1$
end for
while $\sum_i \lambda_i < L$ **do**
  $R_k \Leftarrow \frac{U_k}{S_k} C_k \qquad \forall k \in [1, 2, \ldots, K]$
  $R(max) \Leftarrow max(R_k)$
  for $k = 1 \rightarrow K$ **do**
    if $\frac{U_k}{S_k} C_k < R(max)$ **then**
      $R(max) \Leftarrow \frac{U_k}{S_k} C_k$
      $\hat{k} \Leftarrow k$
    end if
  end for
  $\bar{H}(init) \Leftarrow 0$
  for $l = 1 \rightarrow L$ **do**
    if $\lambda_l = 0$ & $H_{\hat{k},l} > \bar{H}(init)$ **then**
      $\bar{H}(init) \Leftarrow H_{\hat{k},l}$
      $\hat{l} \Leftarrow l$
    end if
  end for
  $C_{\hat{k}} \Leftarrow C_{\hat{k}} + \frac{1}{N} log_2(1 + p \times H_{\hat{k},\hat{l}})$
  $\lambda_{\hat{l}} \Leftarrow 1,\ \rho_{\hat{k},\hat{l}} \Leftarrow 1$
end while
return: // $\rho_{k,l}$

Algorithm 10 Subchannel Allocator for AMC implementation

Input: $K,\ L,\ B,\ N_0,\ P_{tot},\ H_{k,l},\ \Gamma_{SF^x(k)},\ \gamma_i^{voice},\ \gamma_i^{data}$ where $k \in \{1,2,..,K\}, l \in \{1,2,..,L\}$
Initialize array: $\rho_{k,l} \Leftarrow 0, C_k \Leftarrow 0, \lambda_l \Leftarrow 0$
Initialize: $p \Leftarrow \frac{P_{tot}}{L},\ noise \Leftarrow N_0 \frac{B}{L}$
for $k = 1 \rightarrow K$ **do**
  $H_{init} \Leftarrow 0$
  for $l = 1 \rightarrow L$ **do**
    if $\lambda_l = 0$ & $H_{k,l} > H_{init}$ **then**
      $H_{init} \Leftarrow H_{k,l}$, $l^* \Leftarrow l$
    end if
  end for
    // *SELECT BER*
  if $(\gamma_{avg} \leq \gamma_1^{data} and P_b = 10^{-3})$ *or* $(\gamma_{avg} > \gamma_1^{data})$ **then**
    if $p/noise \in [\gamma_i^{voice},\ \gamma_{i+1}^{voice})$ *or* $\in [\gamma_i^{data},\ \gamma_{i+1}^{data})$ **then**
      $\gamma_i \Leftarrow \gamma_i^{voice}$ *or* γ_i^{data}
      select MCS_i for γ_i
      $C_k \Leftarrow \frac{1}{L}\ R_{c_i}\ log_2(1 + \frac{1.5\ G_{c_i}}{-ln(5\ P_b)} \gamma_i \times H_{k,l^*} \times noise)$
    end if
    $C_k \Leftarrow 0$
  end if
  $\lambda_{l^*} \Leftarrow 1,\ \rho_{k,l^*} \Leftarrow 1$
end for
while $\sum_i \lambda_i < L$ **do**
  $R(max) \Leftarrow max(\frac{U_k}{S_k} C_k) \qquad \forall k \in [1, 2, \ldots, K]$
  if $\gamma_{avg} < \gamma_1^{data}$ **then**
    $K \Leftarrow K_{voice}$
  end if
  for $k = 1 \rightarrow K$ **do**
    if $\frac{U_k}{S_k} C_k < R(max)$ **then**
      $R(max) \Leftarrow \frac{U_k}{S_k} C_k$, $\hat{k} \Leftarrow k$
    end if
  end for
  $\bar{H}(init) \Leftarrow 0$
  for $l = 1 \rightarrow L$ **do**
    if $\lambda_l = 0$ & $H_{\hat{k},l} > \bar{H}(init)$ **then**
      $\bar{H}(init) \Leftarrow H_{\hat{k},l}$, $\hat{l} \Leftarrow l$
    end if
  end for
  $k \Leftarrow \hat{k}$ & *SELECT BER*
  if $p/noise \in [\gamma_i^{voice},\ \gamma_{i+1}^{voice})$ *or* $\in [\gamma_i^{data},\ \gamma_{i+1}^{data})$ **then**
    $\gamma_i \Leftarrow \gamma_i^{voice}$ *or* γ_i^{data}
    select MCS_i for γ_i
    $C_k \Leftarrow \frac{1}{L}\ R_{c_i}\ log_2(1 + \frac{1.5\ G_{c_i}}{-ln(5\ P_b)} \gamma_i \times H_{k,l^*} \times noise)$
  else
    $C_{\hat{k}} \Leftarrow 0$
  end if
end while

APPENDIX E: POWER ALLOCATOR

POWER ALLOCATOR

Power Allocator for LTE

The power allocator algorithm for LTE based on waterfilling approach as discussed in Section 3.4.2 is shown in Algorithm 11 and the terms used in this algorithm are defined as follows: $sort(X)$ sorts the elements of vector X in an ascending order and $EFE(X)$ represents a vector that includes all the elements except for the first element of vector X.

The working of Algorithm 11 is described in detail in the following. The power allocation algorithm takes the parameters total power allocated to a particular user $P_{k(tot)}$, which is the sum of power allocated to every PRBs (Π_k) that are specific to a particular user, channel gain to noise ratio of the PRBs that has been allocated to a particular user $H_{k,l}$ and the parameters like $K,\ L,\ B,\ P_b$ as an input when invoked in the resource allocator algorithm. First the channel gain vector is ordered in ascending order. Then the channel gain vector is updated by eliminating the PRBs with the lowest channel gain to noise ratio until $\sum_{m} \frac{H_{k,m} - H_1}{H_{k,m} * H_1} \leq P_{k(tot)}$. The set of PRBs allocated to users is then updated to Π_k^* such that it contains only the PRBs that are considered while meeting the above specified power condition. The eliminated are allocated zero power and then the power given by $P = P_k - \sum_{m^*} \frac{\vec{H}_{k,m^*} - \vec{H}_{k,1}}{\vec{H}_{k,m^*} * \vec{H}_{k,1}}$, where $m^* \in \{1, 2, \ldots, |\Pi_k^*|\}$ is equally divided among the remaining PRBs and is denoted as p. Subsequently, capacity is calculated using the power p and the channel gain as obtained in the previous step. Finally the total capacity corresponding to each user C_k is evaluated.

Algorithm 11 Power Allocator (Waterfilling approach) for LTE

Input: $K,\ L,\ B,\ \Pi_k,\ P_{k(tot)},\ H_{k,m},\ P_b$, where $m \in \{1, 2, \ldots, |\Pi_k|\}$
$H_{k,m} \Leftarrow sort(H_{k,m})$
while $\sum_n \frac{H_{k,m} - H_{k,1}}{H_{k,m} * H_{k,1}} > P_{k(tot)}$ **do**
    $H_{k,m^*} \Leftarrow EFE(H_{k,m})$
end while
$P \Leftarrow P_k - \sum_n \frac{H_{k,m^*} - H_{k,1}}{H_{k,m^*} * H_{k,1}}$
$p \Leftarrow \frac{P}{|H_{k,m^*}|}$
$\mu \Leftarrow p + 1/H_{k,1}$
$C_k \Leftarrow \sum_{m^*} log_2(\mu \times H_{k,m^*})$
return: // C_k

Power Allocator for AMC implementation

The power allocator algorithm based on waterfilling approach as discussed in Section 4.3.2 is shown in Algorithm 12 and the terms used in this algorithm are defined as follows: $sort(X)$ sorts the elements of vector X in an ascending order and $EFE(X)$ represents a vector that includes all the elements except for the first element of vector X.

The working of Algorithm 12 is described in detail in the following. The power allocation algorithm takes the parameters total power allocated to a particular user $P_{k(tot)}$, which is the sum of power allocated to every subchannels (Π_k) that are specific to a particular user, noise adjusted channel gain of the subchannels that has been allocated to a particular user $H_{k,l}$ and the parameters like $K,\ L,\ B,\ P_b,\ G_{c_i},\ \gamma_i^{voice},\ \gamma_i^{data},\ \gamma_{avg}$ as an input when invoked in the resource allocator algorithm. First the channel gain vector is ordered in ascending order. Then the channel gain vector is updated by eliminating the subchannels with the lowest SNR until $\sum_m \frac{-ln(5\ P_b)}{1.5\ G_{c_i}} \frac{H_{k,i} - H_{k,1}}{H_{k,i} \times H_{k,1}} \leq P_{k(tot)}$. The set of subchannels allocated to users is then updated to Π_k^* such that it contains only the subchannels that are considered while meeting the above specified power condition. The eliminated subchannels are allocated zero power and then the power given by $P = P_{k(tot)} - \sum_{m^*} \frac{-ln(5\ P_b)}{1.5\ G_c} \frac{H_{k,m^*} - H_{k,1}}{H_{k,m^*} \times H_{k,1}}$, where $m^* \in \{1, 2, \ldots, |\Pi_k^*|\}$, is equally divided among the remaining subchannels and is denoted as p. Based on the QoS requirement of the user being served, the required BER is selected.

Once the power corresponding to a subchannel is calculated, BER corresponding to the user being scheduled and average system SNR are then considered to pick the best γ_i value and the corresponding MCS_i. Finally the total capacity corresponding to each user C_k is evaluated.

Algorithm 12 Power Allocator (Waterfilling approach) for AMC implementation

Input: K, L, B, Π_k, $P_{k(tot)}$, $H_{k,m}$, γ_i^{voice}, γ_i^{data}, γ_{avg}, P_b, G_{c_i}, where $m \in \{1, 2, \ldots, |\Pi_k|\}$
$H_{k,m} \Leftarrow sort(H_{k,m})$
while $\sum_m \frac{-ln(5\ P_b}{1.5\ G_{c_i}} \frac{H_{k,m} - H_{k,1}}{H_{k,m} \times H_{k,1}} > P_{k(tot)}$ **do**
    $H_{k,m^*} \Leftarrow EFE(H_{k,m})$
end while
$P \Leftarrow P_{k(tot)} - \sum_{m^*} \frac{-ln(5\ P_b)}{1.5\ G_c} \frac{H_{k,m^*} - H_{k,1}}{H_{k,m^*} \times H_{k,1}}$
$p \Leftarrow \frac{P}{|H_{k,m^*}|}$
if $(\gamma_{avg} \leq \gamma_i^{data}\ and\ P_b = 10^{-3})$ *or* $(\gamma_{avg} > \gamma_i^{data})$ **then**
    if $p/noise \in [\gamma_i^{voice},\ \gamma_{i+1}^{voice})$ *or* $\in [\gamma_i^{data},\ \gamma_{i+1}^{data})$ **then**
        $\gamma_i \Leftarrow \gamma_i^{voice}$ *or* γ_i^{data}
        select MCS_i for γ_i
        $C_k \Leftarrow \sum_{m^*} \frac{1}{L}\ R_{c_i}\ log_2 \left(\frac{H_{k,m^*}}{H_{k,1}} + \frac{1.5\ G_c}{-ln(5\ P_b)} \gamma_i \times H_{k,m^*} \times noise \right)$
    end if
else
    $C_k \Leftarrow 0$
end if

APPENDIX F: PRC ALLOCATOR

PRC POWER ALLOCATOR ALGORITHM

In PRC power allocator algorithm, the result of subchannel/PRB allocator as discussed in Appendix 5 is used and the power is allocated to every subchannel/PRB that is specific to a user as explained in [5]. The PRC power allocator algorithm is depicted in Algorithm 13 and the terms used in this algorithm are defined as follows: N_k is the total number of subchannel/PRB associated with k^{th} user, Θ is a set storing channel gain values corresponding to the used subchannel/PRB by a user, Φ represents a null set, h'_{min} is the minimum value of gain in set Θ, w_k is the ratio as defined in (C-28), c_k is the ratio as defined in (C-29), α is a flag used to determine whether or not any of c_k is ∞ and $P_{k(tot)}$ is the total power allocated to k^{th} user.

The working of Algorithm 13 is described in detail in the following. First the algorithm takes the parameters K, L, B, P_{tot}, N_0, $h_{k,l}$, $\rho_{k,l}$, and $\Gamma_{SF^x(k)}$ as inputs when invoked in the resource allocator algorithm. Subchannels/PRBs assigned to a particular user are then arranged in ascending order. Then parameters w_k, c_k, d_k as derived in [5] are evaluated as follows

$$w_k = \prod_{l=2}^{N_k} \frac{H_{k,l}}{H_{k,1}} \tag{C-28}$$

$$c_k = \begin{cases} 1 & \text{if } k = 1 \\ \dfrac{(\frac{H(1,1)w(1)^{1/N_1}}{N_1})^{\frac{N_1\Gamma_{SF^x(k)}}{N_k\Gamma_{SF^x(1)}}}}{\frac{H_{k,1}w_k^{1/N_k}}{N_k}} & \text{if } k = 2,3,\ldots,K \end{cases} \tag{C-29}$$

$$d_k = \begin{cases} 1 & \text{if } k = 1 \\ \frac{N_1\Gamma_{SF^x(k)}}{N_k\Gamma_{SF^x(1)}} & \text{if } k = 2,3,\ldots,K \end{cases} \tag{C-30}$$

Then it is required to solve the following equation so as to find the value of $P_{1,tot}$

$$\sum_{k=1}^{K} c_k (P_{1(tot)})^{d_k} - P_{tot} = 0 \tag{C-31}$$

Newton's root finding method as depicted in Algorithm 14, and explained in the following paragraph, is then used to find the root of (C-31), i.e, the value of $P_{1(tot)}$. If $P_{1(tot)}$ is an indeterminate number then the power for all the users $P_{k(tot)}$ is allocated as $P_{k(tot)} = \frac{P_{tot}}{L} \times N_k$ where $k = 1, 2, 3, \ldots, K$ else the following equation is used to find $P_{k(tot)}$:

$$P_{k(tot)} = c_k (P_{1(tot)})^{d_k} \tag{C-32}$$

where, $k = 2, 3, \ldots, K$. Then finally, $P_{k(tot)}$ is returned to the resource allocator algorithm.

Newton's Root finding method: The algorithm to implement Newton's root finding method that has been used to find the roots of (C-31) is depicted in Algorithm 14 and the terms used in this algorithm are defined as follows: P stores the guess value of total power allocated to first user i.e, $P_{1(tot)}$, i indicates the total number of iterations, i_{min} is the iteration number corresponding to minimum value of P, $f(y)$ is the function of y, $root(f(y), x)$ finds the root of a function $f(y)$ using the initial guess value of x and ¿ is the indeterminate number.

The working of Algorithm 14 is described in detail in the following. First a guess value ($P^* = 0.01 \times i \times P_{tot}$) is chosen and $\sum_{k=2}^{K} c_k (P^*)^{d_k}$ is calculated where i is the number of iterations. Then the root of the equation is found near the value P^* for which $\sum_{k=2}^{K} c_k (P^*)^{d_k}$ is minimum. The root so found is the required $P_{1,tot}$. If the root is indeterminate, then the power is equally divided among each subchannel else the power corresponding to k^{th} user is found using (C-32).

Algorithm 13 PRC Power Allocator

Input: $L,\ K,\ H_{k,l},\ \rho_{k,l},\ P_{tot},\ N_o,\ \gamma_{SF^x}(k)$ where $k \in \{1, 2, \ldots, K\},\ l \in \{1, 2, \ldots, L\}$
for $k = 1 \rightarrow K$ **do**
  $N_k \Leftarrow \sum_i \rho_{k,l}$
end for
for $k = 1 \rightarrow K$ **do**
  $\Theta \Leftarrow \Phi$
  for $l = 1 \rightarrow L$ **do**
    if $\rho_{k,l} > 0$ **then**
      $\Theta \Leftarrow [\Theta \quad h_{k,l}]$
    end if
  end for
  for $l = 1 \rightarrow |\Theta|$ **do**
    $h_{min} \Leftarrow min(\Theta)$
    $\Theta \Leftarrow \Theta - [h_{min}]$
    $\vec{h}_{k,l} \Leftarrow h_{min}$
  end for
  $H_{k,l} \Leftarrow h_{k,l}/N_0 \frac{B}{L}$ & $w^* \Leftarrow 1$
  for $l = 2 \rightarrow |\Theta|$ **do**
    $w^* \Leftarrow w^* \frac{H_{k,l}}{H_{k,1}}$
  end for
  $w_k \Leftarrow w^*$
end for
$c_1 \Leftarrow 1$, $d_1 \Leftarrow 1$
for $k = 2 \rightarrow K$ **do**
  $c_k \Leftarrow \frac{(\frac{H(1,1)w(1)^{1/N_1}}{N_1})^{\frac{N_1\gamma_{SF^x(k)}}{N_k\gamma_{SF^x(1)}}}}{\frac{H_{k,1}w_k^{1/N_k}}{N_k}}$ & $d_k \Leftarrow \frac{N_1\gamma_{SF^x(k)}}{N_k\gamma_{SF^x(1)}}$
end for
$\alpha \Leftarrow 0$
for $k = 1 \rightarrow K$ **do**
  if $c_k = \infty$ **then**
    $\alpha \Leftarrow 1$
  end if
end for
if $\alpha = 1$ **then**
  for $k = 1 \rightarrow K$ **do**
    $P_{k(tot)} \Leftarrow \frac{P_{tot}}{L \times N_k}$
  end for
else
  invoke: Newton Root finding method $\rightarrow$ **returns** $P_{k(tot)}$
end if
return: $P_{k(tot)}$

Algorithm 14 Newton Root finding method

Input: $K,\ L,\ c_k,\ d_k,\ P_{tot},\ N_k$, where $k \in \{1, 2, \ldots, K\},\ l \in \{1, 2, \ldots, L\}$
for $i = 1 \rightarrow 100$ **do**
  $P^* \Leftarrow 0.01 \times i \times P_{tot}$
  $P_i \Leftarrow P^*$
  for $k = 2 \rightarrow K$ **do**
    $P_i \Leftarrow P_i + c_k(P^*)^{d_k}$
  end for
  $P_i \Leftarrow P_i - P_{tot}$
end for
$P_{min} \Leftarrow min(P_i)$
$i_{min} \Leftarrow [min(P_i)]_{arg}$
$f(P_{min}) \Leftarrow \sum_{k=1}^{K} c_k(P_{min})^{d_k} - P_{tot}$
$x \Leftarrow 0.01 \times i_{min} \times P_{tot}$
$P_{1,tot} \Leftarrow root(f(P_{min}), x)$
if $P_{1,tot} =$ ¿ **then**
  for $k = 1 \rightarrow K$ **do**
    $P_{k,tot} \Leftarrow \frac{P_{tot}}{L} \times N_k$
  end for
else
  for $k = 2 \rightarrow K$ **do**
    $P_{k,tot} \Leftarrow c_k(P_{1,tot})^{d_k}$
  end for
end if
return: $P_{k,tot}$

VITA

VITA

Born in Kathmandu, Nepal in 1987, Bimal Paudel received his B.S. degree in electronics and communication engineering from the Pulchowk Campus, Institute of Engineering, Tribhuwan University, Nepal in May 2009. He is currently working towards the M.S. degree in electrical engineering at the University of Mississippi (Olemiss), University, MS. He worked as a telecom engineer in ZTE Corporation, Nepal from May 2009 to Dec. 2011. He had worked in the Electrical Engineering Department as a graduate teaching assistant and also worked as a graduate research assistant to Dr. Mustafa M. Matalgah. He is currently working as a platform engineer in Nexius Insight Inc., Redmond, WA.