UNIVERSITY OF KWAZULU-NATAL

Cross-layer Design for support of delay bound Quality of Service Guarantees over Fading Channels

Suvania Moodliar

EXAMINER'S COPY

2011

Cross-layer Design for support of delay bound Quality of Service guarantees over Fading Channels

By Suvania Moodliar

Submitted in fulfilment of the academic requirements
for the degree of M.Sc Engineering
in the Faculty of Engineering
at the University of KwaZulu-Natal, Durban, South Africa

January 2011

As the candidate's supervisor I have approved this dissertation for submission.

Signed: _____

Name: Professor HongJun Xu

Date: 26 January 2011

Declaration

I, Suvania Moodliar, declare that

i. The research reported in this dissertation, except where otherwise indicated, is my original work.
ii. This dissertation has not been submitted for any degree or examination at any other university.
iii. This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
iv. This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
    a. their words have been re-written but the general information attributed to them has been referenced;
    b. where their exact words have been used, their writing has been placed inside quotation marks, and referenced.
v. Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.
vi. This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed: _____

Acknowledgements

I would like to thank my supervisor, Professor HongJun Xu, for the endless support, dedication and immense guidance he has shown me in these past 2 years. He is an extra-ordinary person who goes beyond the call of duty to assist his students, and is someone I greatly admire. He has always pushed me to give off my very best at all times, based on his philosophy of hard work and perseverance. His continuous belief in me has been overwhelming, and I have been extremely privileged and honoured to be his student.

I would like thank my family for being a source of encouragement, and strength through the difficult times. Specifically, I want to thank my parents for all the sacrifices they have made, in leaving no stone unturned to see me reach my goals and dreams. Their love and guidance show no bounds, and I am the person I am today, because of them.

To all my friends, namely the 203s, 204s, 205s and 206s, I would like to thank you for making my university career the most epic and legendary time of my life. These past couple of years have been awesome, and I will always cherish them.

Finally, many thanks go to Telkom SA for their support of this dissertation.

Abstract

Quality-of-service (QoS) guarantees have become critically important for the transmission of real-time multimedia traffic in next generation mobile wireless networks. The aim of this dissertation is to investigate the cross-layer design for support of delay bound QoS guarantees over fading channels.

Providing diverse QoS guarantees presents a challenge due to the time-varying fading nature of wireless channels. Existing physical layer modelling is inadequate in supporting real time QoS metrics such as delay, hence adaptive techniques need to be extended to the upper-protocol layers. The first objective of this dissertation is to introduce a cross-layer design framework which investigates the impact of the physical layer on the data link delay bound QoS performance. At the physical layer, adaptive modulation and coding (AMC) is utilized for transmission over block fading channels. At the data link layer, the effective capacity approach is used to model the delay bound QoS performance subject to physical layer variations. The effects of varying physical layer parameters, such as average signal-to-noise ratio (SNR), the fading parameter for the Nakagami-m model, and target packet-error rate (PER), on the analytical delay bound performance are investigated and then validated by the simulation of a queuing system.

Due to the cross-layer design framework, the system's throughput has a significant impact on bounded delay at the data link layer. The switching levels of the conventional AMC scheme used in the first objective were fixed, subject to a target PER. However, fixed switching levels results in the system's throughput limiting the delay bound performance. The second objective of this dissertation is to optimize the switching levels of the AMC scheme employed at the physical layer, by maximizing the average throughput, while maintaining the target PER constraint. The analytical and simulated results show that by optimizing the switching levels, a superior delay bound performance is achieved, when compared to the deterministic switching levels of the conventional AMC scheme used in the first objective.

Contents

List of Figures

List of Tables

List of Acronyms

AWGN	Additive white Gaussian noise
AMC	Adaptive modulation and coding
ARQ	Automatic repeat request
BER	Bit error rate
CDMA	Code division multiple access
CSI	Channel state information
EDF	Earliest deadline first
FDM	Frequency division mode
FIFO	First-in-first-out
FSMC	Finite state Markov chain
GE	Gärtner-Ellis
MAC	Medium access control
MGF	Moment generating function
MIMO	Multiple-input-multiple output
OFDM	Orthogonal frequency division multiplexing
OSI	Open systems interconnections

PER	Packet error rate
QAM	Quadrature amplitude modulation
QoS	Quality of service
SISO	Single input single output
SNR	Signal-to-noise ratio
TDM	Time division multiplexing
WFQ	Weighted fair queue

Chapter 1

Introduction

Next generation mobile networks have allowed modern societies to communicate with flexibility, freedom, and efficiency. Today, the dependence on mobile services extends beyond communication, to entertainment, education, health services and location placement, just to name a few. Real time multimedia applications used for such services encompass different traffic types including video, voice and data. Quality of Service (QoS) guarantees are hence critical for meeting user set requirements for multimedia wireless transmission. QoS is defined as a set of service requirements that must be met by the network while transmitting a packet stream from source to destination [1]. If these requirements are not met, the applications or services are rendered unusable. Some of the QoS guarantees required to be supported by the network include bandwidth, bounded delay, probability of packet loss, and jitter. Guaranteeing QoS requirements however, is a complex problem. Unlike its wired counterpart, wireless networks pose a far greater challenge in guaranteeing a reliable service given that it has limited resources. Some of the factors contributing to the difficulties in supporting QoS guarantees in wireless networks are listed below.

- In a wireless network, the transmission from a mobile user not only uses up local bandwidth, but also consumes the bandwidth of neighbouring users. The network bandwidth is hence in contention among mobile users, and thus limits the number of users accessing the wireless network.
- The mobility and low power capabilities of wireless networks lead to links between mobile users being broken, hence resulting in service disruption and security problems. This contributes to the difficulties in providing timeous delivery of multimedia traffic such as voice and video.
- Wireless channels are susceptible to fading, interference and high bit/packet error rates, which leads to less reliable services.

There has been extensive research in overcoming the challenges presented in providing support of QoS guarantees for real time multimedia services. One of the main techniques that have emerged in recent years is cross-layer design, which is discussed below.

1.1 Cross-Layer Design

Traditionally the approach adopted for overcoming the challenges of supporting QoS guarantees, has been to adapt conventional wired protocols for multimedia transmission, to the varying, limited resources of the wireless network [2]. The layered architecture of the OSI (open systems interconnection) structure was considered, where protocols at different layers are designed independently. In the OSI structure communication only occurs between neighbouring adjacent layers, with no interaction between non-adjacent layers. The optimal protocol design of each layer however, does not incorporate the diverse QoS requirements of multimedia traffic. Hence this type of "layered" optimization results in a suboptimal QoS provisioning performance for multimedia transmission.

In recent years researchers have focused on absolving the use of the layered structure in wireless networks, by viewing the layers of the protocol stack as being interdependent and not isolated entities. The exploitation of these interdependencies allows information to be shared among different layers, which leads to enhanced protocol design and consequently allows for improved multimedia transmission. This departure from the layered architecture was denoted as cross-layer design. Cross-layer design generally refers to protocol design done by actively exploiting the dependence between protocol layers to obtain performance gains [3].

In order to further understand the principles that govern proposed cross-layer designs in current literature, the following classifications were presented in [2]:

- **Top-down Approach**: A higher layered protocol has its parameters and strategies optimized at the next lower adjacent layer.
- **Bottom-up Approach**: In this approach lower layers try to prevent higher layers from incurring losses and bandwidth variations. This methodology however is not always the optimal solution for transmission of multimedia traffic as it may result in large delays and poor throughput.
- **Application-centric approach**: The application layer, based on its requirements, optimizes lower layer parameters individually in a top down, or alternatively a bottom up fashion. This approach is not always feasible given that the application layer operates at a slower rate than that of lower layers, hence instantaneous adaption is not achieved and thus lowers performance.
- **MAC-centric approach**: The medium access control (MAC) layer dictates the application and physical layer parameters based on their specific requirements

which include traffic and channel information respectively. However the main disadvantage of this approach is that the MAC layer lacks the ability to perform adaptive source channel coding, which is required to cope with the effects of the fading channel, and supporting QoS guarantees.

- **Integrated approach**: The establishment of joint strategies are considered however the high complexity in attempting all possible combinations of strategies and theory parameters can make this approach infeasible.

The aforementioned cross-layer design approaches have advantages and disadvantages, but the ideal, optimal solution of multimedia transmission depends solely on the application's power requirements, complexity and used protocols.

1.2 QoS with Cross-Layer Design

The support of QoS requirements has fast become the focus of research in the area of mobile wireless networks, given the necessity to produce reliable services. The challenges discussed in supporting QoS guarantees in wireless networks have been surveyed in [2][4-5]. The works [2][5] provide excellent reviews on the principles and new paradigms used in guaranteeing QoS requirements based on the cross-layer design approach.

Numerous techniques, such as resource allocation and scheduling algorithms, have been researched in accordance with cross-layer design to provide QoS guarantees. Resource allocation algorithms are based on adaptively reserving adequate resources for mobile users, in order to transmit real time traffic with stringent delay bounds. Providing QoS guarantees using cross-layer design resource allocation algorithms can be seen in [6-8]. In [6], the cross-layer design architecture of a 1xEV-DV system was presented, which aimed to obtain high throughput while providing QoS provisioning. The set of cross-layer design modules included a resource scheduling algorithm which achieved an improved throughput and supported specific user QoS requirements, a priority admission control algorithm which admitted users according to the importance of their QoS requirements, and finally a resource allocation module which was designed to assign the optimal arrangement of system parameters. The work in [7] integrated adaptive modulation and coding (AMC) and truncated automatic repeat request (ARQ) at the physical layer, which influenced queuing behaviour. Based on the queuing process modelled using Markov chains, the proposed scheme was adapted in order to support QoS guarantees. In [8], a cross-layer design dynamic resource allocation scheme for a multiple-input multiple-output (MIMO)/ Orthogonal frequency division multiplexing (OFDM) system, which specifically focused on the physical

and MAC layers, was presented and aimed to support diverse QoS requirements. In particular, the dynamic resource scheme at the physical layer increased spectral efficiency by adaptively allocating transmit antennas and modulation levels to the users, subject to a set BER QoS requirement. Parallel to the physical layer functionality, the operation of the MAC layer accomplished two things, namely dynamically assigning time slots for real time users while in a time division multiplexing (TDM) mode, and maximising throughput by adaptively allocating subcarriers while being in a (frequency division multiplexing mode) FDM mode subject to maintaining fairness among users.

Scheduling algorithms are used to commit system resources, such as channel capacity, to mobile users. The utilization of scheduling techniques in the cross-layer design approach was seen in [9-13]. An adaptive resource reservation scheme which used a scheduling algorithm was presented in [9], in order to accommodate multimedia traffic. The novelty behind the algorithm was that channel allocations were based on actual traffic buffered from mobile stations, and not estimations. In [10], a cross-layer framework is proposed for efficient packet scheduling in wireless code division multiple access (CDMA) networks. At the MAC layer the next state of the wireless channel was estimated, given the physical layer characteristics. Consequently, based on the resultant estimation, users were selected to be serviced in the next frame, by the packet scheduler. Liu et al. proposed a cross-layer design scheduler in [11], to address multimedia transmission, where the MAC layer was combined with the AMC scheme employed at the physical layer. The admission and scheduling policy viewed, and analyzed the QoS of the wireless link in terms of throughput, packet loss rate and average delay, subject to a reserved bandwidth. The low complexity scheduler provided support for set user QoS requirements and accomplished effective bandwidth utilization. A multiple connection cross-layer scheduler developed at the MAC layer was seen in [12]. Each admitted connection was associated with a priority, which was continuously updated with the connection's channel quality, QoS requirement and service priority. The connection with highest priority was scheduled first. The cross-layer scheduler used in [13] minimized a prescribed cost function based on the channel quality and queuing packet delay. It was shown through simulations that the proposed scheduler outperformed weighted fair queue (WFQ) and earliest deadline first (EDF) schedulers in terms of packet delay and throughput.

1.3 Effective Capacity

In real time multimedia transmission, the support of QoS bounded delay is crucial. However based on existing channel modelling, the extraction of metrics like delay is difficult. To

investigate the delay bound performance, queuing analysis is required. In order to develop a platform in which delay bound violation can be evaluated, an approach which encompasses the characteristics of both physical and data link layer needs to be established.

The concept of effective capacity was proposed in [14], and happened to be the dual concept of effective bandwidth. The theory of effective bandwidth has been extensively studied in recent years as seen in [15-18], in order to statistically model the arrival traffic and buffer behaviour. Based on the asymptotic results showed in [15], Wu and Negi in [14] established that the effective bandwidth theory can be adapted to service curve characterization. In [23-26], the effective bandwidth and effective capacity functions are used to characterize arrival and service processes from a QoS point of view. The exploitation of this unique relationship consequently allowed the analysis of statistical QoS and buffer overflow probabilities.

The evaluation of the effective bandwidth and effective capacity functions requires the modelling of processes at the data link and physical layer respectively. The fashion in which packets enter the data link layer, is denoted as the arrival process, where as the adaptive transmission scheme used at the physical layer is known as the service process. In service process modelling, unlike non adaptive schemes, AMC is not deterministic, and hence presents a challenge in characterising its behaviour subject to fading wireless channels. In [19], the use of finite state Markov chains (FSMC) was proposed in order to model the adaptive modulation scheme given variations in the wireless channel. This type of service process modelling has been applied in numerous scenarios in order to investigate the QoS provisioning performance, as seen in [20-26]. In [20], MIMO transmit/receive diversity was integrated with AMC. Based on the FSMC modelling developed at the physical layer, the effective capacity function was used to evaluate the QoS performance at the data link layer. Two-state Markov process modelling was seen in [21], where the arrival process characterised by the ON-OFF model, and service process by the Gilbert-Elliot model, was integrated in the cross-layer design algorithm. The resultant analysis on the delay bound violation probability was a superb example of how to translate cross-layer design into a platform to investigate QoS provisioning. However the framework established was not feasible to accommodate more realistic scenarios. In particular, the Gilbert-Elliot model explicitly looked at the channel being in two states, where each state was associated with a BER. However the variations in the wireless channel, such as fading, is much more complex to characterise, and cannot be limited to a two state Markov chain. The work of [22] analyzed the combined effect of the applied AMC scheme and finite length queuing. The FSMC developed used state pairs which contained both queue and server states. The joint

cross-layer design over the data link and physical layers in [22] allowed the evaluation of the QoS metric, average delay. This however is not sufficient in supporting delay sensitive transmission where the main QoS metric of interest is bounded delay.

The novelty behind the effective capacity concept has led to the investigations of different optimization techniques in order to enhance the service process and subsequent QoS provisioning performance, as seen in [23-26]. A QoS driven rate and power adaption scheme was proposed in [23], where the system's throughput was maximized subject to a QoS delay constraint. It was shown that by applying the effective capacity concept, the optimal power control policy tended towards the water-filling algorithm as the QoS constraint became loser, and converged to total channel inversion as the QoS constraint became stricter. An extension of [23] was presented by the same authors, Zhang and Tang in [24], which concentrated on the QoS provisioning performance for multichannel communications over wireless channels. The multichannel transmission had a significant advantage over the single channel transmission shown in [23], given that it achieved high throughputs and jointly guaranteed stringent QoS requirements. Unlike [23], which maintained a target BER requirement, the optimal adaption scheme used in [25] had to meet an instantaneous target PER constraint. The continuous power, discrete rate adaption scheme proposed, optimized QoS by minimizing the packet loss rate subject to PER and average power constraints. Results showed that the QoS driven, optimized power and rate scheme, produced far a superior QoS delay bound performance than that of constant power AMC schemes. A cross-layer design based adapting resource allocation scheme was developed in [26]. In particular the scheme, through the effective capacity concept, adaptively allocated timeslots and power levels to mobile users, such that the required QoS delay bound was met. It was shown that the simulated delay bound probability was lower than that of the prescribed delay bound probability. It was also observed that conventional power control schemes had a similar QoS delay bound violation performance to that of the proposed scheme. However to achieve this comparable performance, a larger amount of resources had to be used as compared to the proposed scheme.

1.4 Motivation and Contributions

Many of aforementioned cross-layer design frameworks, which applied the effective capacity concept, focused on optimizing the service process of the channel model, in order to improve the QoS provisioning performance. This notion is attributed to the significant impact that the effective capacity has on the QoS performance and hence motivates

techniques that aim to enhance the service process. In [27] optimum mode switching levels were derived for a constant power AMC scheme by maximising the average throughput while maintaining an average target BER. To the best of the author's knowledge, there have been no publications to date reporting the extension of this proposed scheme by determining its effective capacity, subject to an average PER requirement. The first contribution of this dissertation is to adapt the optimized switching level technique used in [27] to meet an average PER constraint. Packets is the reference unit used in this dissertation, and hence motivates the use of PER rather than BER. The effective capacity of the resultant optimized AMC scheme, is evaluated and compared to an AMC scheme with deterministic switching levels.

Given that the effective capacity of the optimized switching level scheme used in [27], has not been evaluated, consequently the QoS delay bound violation performance through the effective capacity concept has not been analyzed. The second contribution of this dissertation is to further extend the optimized switching level AMC scheme by integrating it into a cross-layer design framework, where its subsequent effect on the QoS delay bound performance at the data link layer can be investigated. The QoS metric of interest in this dissertation is bounded delay, given that it is a critical requirement of real time multimedia services. The effective bandwidth and effective capacity theories will be used to determine the analytical delay bound violation probability of the optimized scheme. This will further be validated by the delay bound violation probability of a simulated M/M/1 FIFO queuing system, using the optimized AMC scheme as the service process. To contextualise the substantial improvement in the QoS delay bound performance, due to the applied optimized AMC scheme, the delay bound violation performance of a fixed switching level AMC is also evaluated, as a source of comparison.

1.5 Outline of dissertation

The remainder of the dissertation is organized as follows:

Chapter 2 introduces the cross-layer design system model, which exploits the relationships between the physical and data link layers. This is followed by a detailed description of the system's constituent functional blocks, namely the AMC scheme, M/M/1 queuing model, and source traffic modelling. The notion of statistical QoS guarantees is presented where the approach for determining the delay bound violation probability using the concept of effective bandwidth and effective capacity, is discussed. The relationship between the effective

bandwidth and effective capacity functions is investigated, more specifically focusing on their role in providing QoS support.

Chapter 3 discusses the service process modelling using theory of effective capacity for block fading and correlated fading channels. Numerical results presented show the behaviour of the effective capacity function subject to variations in physical layer parameters.

Chapter 4 presents the optimization of switching levels of the underlying AMC scheme, by maximising the average throughput while maintaining an average target PER constraint. This optimized technique is then extended to incorporate the effective capacity function. The effective capacity performances of the optimized AMC scheme, and a conventional AMC scheme with fixed switching levels, are compared through numerical results.

Chapter 5 introduces a cross-layer design algorithm that is used to determine the delay bound violation probability. The optimized AMC scheme is integrated into the cross-layer design framework, and its delay bound violation performance compared to the fixed bound AMC scheme is shown, and validated through simulation and theoretical results.

Chapter 6 concludes the dissertation.

Chapter 2

Background

The purpose of this chapter is to introduce a system model which incorporates cross-layer design, and provide a detailed description of the integrated layers and their subsequent role in QoS provisioning.

This chapter is outlined as follows: In Section 2.1 the cross-layer design system model is presented, where the sub blocks, namely adaptive modulation and coding applied at the physical layer; queuing model based at the data link layer; and traffic source modelling of upper layer packets streaming into the data link layer, are discussed in Sections 2.2, 2.3 and 2.4, respectively. The notion of using statistical QoS guarantees is considered in Section 2.5, and a packet delay bound expression is presented. The concept of effective capacity which is based on the effective bandwidth theory is presented in Section 2.6, where its profound role in the cross-layer design framework and QoS provisioning performance, is analysed. Finally a chapter summary is given in Section 2.7.

2.1 System Model

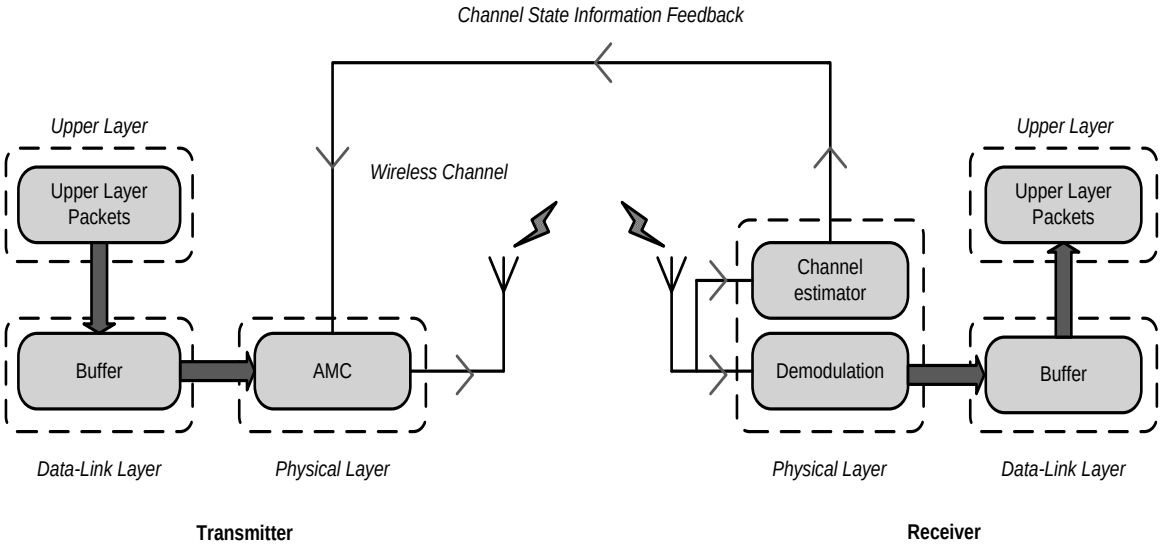

Figure 2-1 System model adapted from [26]

A single-input-single-output (SISO) antenna system model, adapted from [26], is depicted in Figure 2-1. In particular, we concentrate on the AMC scheme employed at the physical layer, and packet queuing at the data link layer. At the transmitter, upper protocol layer packets first stream into the transmit buffer at the data link layer. This behaviour is denoted as the arrival process. Packets that are stored in the buffer, await service from the AMC module at the physical layer, which is known as the service process. The buffer represents a first-in-first-out (FIFO) queue, which means that the first packet to arrive into the queue is serviced first. It is assumed that the queue length is infinite. The arrival process deals with the unit packets, however the applied AMC scheme uses the unit frames. Each frame contains a variable amount of packets, which is determined by the mode selected by the AMC scheme, based on the channel state information (CSI) feedback from the receiver. It is assumed that the CSI estimation and feedback is accurate and reliable over the fading channels. At the receive side it is observed that the reverse operation occurs, where the decoded bits are placed in a packet structure, and are fed into the upper protocol layer via the buffer at the data link layer. The size of the packet structure is denoted as $\boldsymbol{N_p}$ and contains 1080 bits [22][28].

The wireless channel can be represented by two distinct impairments, namely large scale fading and small scale fading. Large scale fading is primarily caused by signal path loss, and results in the reduction of signal power over a period of time. Small scale fading is attributed to multipath signal propagation, where several versions of the transmitted signal are added by the receiver producing a distorted signal. In this dissertation, small scale fading is assumed, hence the wireless channel is considered to be frequency flat. The discrete-time channel fading process is assumed to be stationary and ergodic, and is considered to follow the Nakagami-m distribution. Given that the wireless channel is frequency flat, the channel fading process is invariant for the duration of a frame, but varies on a frame by frame basis according to the channel fading distribution [22]. The Nakagami-m channel model, where m is the Nakagami fading parameter ($m \geq 1/2$), is very general, and applies to a large range of fading channels. These include one-sided Gaussian channel ($m = 1/2$), Rayleigh fading channel ($m = 1$), approximation of Rician fading channel ($m > 1$), and Gaussian channel with no fading ($m = \infty$).

Let the channel gain, which is assumed to be stationary and ergodic time-varying, be denoted by $\{\alpha(n), n = 1,2, \ldots\}$ where n is the frame index. The additive white Gaussian noise (AWGN) is given by N_0 with zero mean and variance σ^2. From Figure 2-1, the transmitter sends data over the wireless channel using an average transmit power denoted by $\bar{S}$. Given

the CSI, the pre-adaption received SNR in the n^{th} frame is defined as $\gamma(n) = \bar{S}\left(\alpha(n)\right)^2/\sigma^2$, where $\left(\alpha(n)\right)^2$ is the instantenous channel power gain. Based on the stationary assumption, the distribution of γ is independent of n and is denoted by $p_\gamma(\gamma)$. To simplify notation, the frame index n is omitted in SNR γ. The received SNR γ encapsulates the channel quality [22], and subject to channel variations per frame, the Nakagami-m distribution is adopted to obtain $p_\gamma(\gamma)$ which is expressed as [22, Eq. (1)],

$$p_\gamma(\gamma) = \frac{\gamma^{m-1}}{\Gamma(m)}\left(\frac{m}{\bar{\gamma}}\right)^m exp\left(-\frac{m}{\bar{\gamma}}\gamma\right), \qquad \gamma \geq 0 \tag{2-1}$$

where m is the Nakagami fading parameter ($m \geq 1/2$), $\bar{\gamma} = E(\gamma)$ denotes the average SNR where $E(\cdot)$ is the given expectation, and $\Gamma(m) = \int_0^\infty t^{m-1}exp(-t)\,dt$ is defined as the gamma function.

2.2 Adaptive Modulation and Coding

Various adaptive schemes employed at the physical layer, have been researched as viable solutions for overcoming the effects of the fading wireless channel. Adaptive modulation is a popular technique that aims to improve spectral efficiency while adhering to a target error performance over wireless channels [22][28][29]. In adaptive transmission, the modulation and coding parameters are adapted accordingly, based on the received SNR. The general parameters that are adapted include the transmission rate, transmit power, coding, error probability, and a combination of these aforementioned parameters. There are several adaptive techniques specified in [30] that are associated with the variations of these parameters relative to the received SNR. These are detailed as follows,

- Variable-rate techniques: Variable-rate modulation involves varying the data rate, relative to the received SNR. This can be achieved by setting the symbol rate of the modulation and using multiple modulation schemes, or fixing the modulation mode and changing the symbol rate.
- Variable power techniques: The basis of adapting the transmit power is to compensate for the variations in SNR due to fading. The objective is to satisfy a fixed bit error probability.
- Variable error probability: The instantaneous BER is adapted subject to a average BER constraint. Due to the instantaneous error probability varying as the received

SNR varies, typically another form of adaption is included in this technique, namely constellation or modulation type.

- Variable-coding techniques: In this technique different channel codes are used to produce variable amounts of channel gain to the transmitted bits. This approach however requires the channel to remain invariant for the duration of one block length, or constraint length of the code.
- Hybrid techniques: These techniques adapt several parameters of the transmission scheme, which can include combinations of rate, power, coding and instantaneous error probability. The optimization techniques are used to meet a set performance requirement. For example the combination of rate and power adaption is used to maximise spectral efficiency

In this dissertation constant power AMC schemes, which use M_k-ary rectangular or square quadrature amplitude modulation (QAM), are applied at the physical layer where the variable-rate technique is applied. AMC is a popular technique, given that it superimposes trellis coding on top of adaptive modulation, where the channel coding gain is independent of the modulation [30]. In M_k-ary rectangular or square QAM, the M denotes the number of points in the signal constellation [31]. We restrict the constellation sizes to 2, 4, 16, and 64, which represent BPSK, QPSK, 16 QAM, and 64QAM modulation schemes used in this dissertation. The specific use of these modulation schemes is attributed to their employment in the works of [20][25], which serve as replication sources. The spectral efficiency (bits/symbol) associated with each mode is a coded rate, given that the adaptive modulation is applied in conjunction with convolutional coding. In [28], adaptive modulation and convolutional coding was investigated, where a target packet error rate (PER) was met, given that the length of the packet was defined as $N_p = 1080$ bits. In particular, in [22], which is based on [28], the transmission modes, boundary points and code rates used for the adaptive modulation with convolutional coding are detailed in Table 2-1, and illustrated in Figure 2-2.

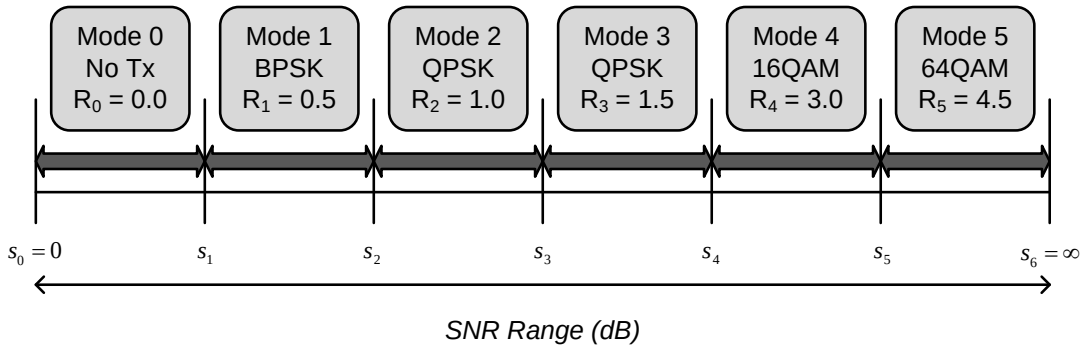

Figure 2-2 Partitioning of SNR range based on AMC parameters

Figure 2-2 illustrates the functionality of the AMC scheme. It is observed that the SNR range is partitioned into sections or bins. The total number of bins is denoted as K, hence producing $K+1$ boundary points. These boundary points are defined as $\{s_k\}_{k=0}^{K}$, where $s_0 < s_1 < \ldots s_K$, given that $s_0 = 0$ and $s_K = \infty$. Each bin is associated with a transmission mode. Hence given that the received SNR γ falls into the range $s_k \leq \gamma < s_{k+1}$, transmission mode k is selected. It is seen that, as the received SNR γ decreases, the system throughput decreases due to the diminishing channel quality. The worst case scenario is when transmission ceases in order to avoid deep fades in the channel. This occurs when γ is below the boundary point s_1, seen in Figure 2-2.

The unit of interest in the given system model is namely packets, hence the target PER that the applied AMC scheme must maintain is determined by [22, Eq. (5)]

$$\mathrm{PER}_k(\gamma) = \begin{cases} 1, & if\ 0 < \gamma < \gamma_k \\ a_k exp(-g_k \gamma), & if\ \gamma \geq \gamma_k \end{cases} \tag{2-2}$$

where a_k, g_k, and γ_k are mode-based parameters specified in Table 2-1. These parameters were obtained by fitting (2-2) to the exact PER in [22]. Unless otherwise stated; the parameters seen in Table 2-1 will be used by the AMC scheme applied at the physical layer, for remainder of this dissertation. Note that the parameter γ_k indicates the limiting threshold for mode k, where if the case of $0 < \gamma < \gamma_k$ occurs, the resultant PER would be 1.

Table 2-1 Adaptive modulation with convolutional coding mode parameters [22]

	Mode 1	**Mode 2**	**Mode 3**	**Mode 4**	**Mode 5**
Modulation	BPSK	QPSK	QPSK	16QAM	64QAM
Coding Rate R_c	$1/2$	$1/2$	$3/4$	$3/4$	$3/4$
R_k (bits/sym.)	0.5	1.00	1.50	3.00	4.50
a_k	274.7229	90.2514	67.6181	53.3987	35.3508
g_k	7.9932	3.4998	1.6883	0.3756	0.0900
γ_k(dB)	-1.5331	1.0942	3.9722	10.2488	15.9784

Based on (2-2), the boundary points of the underlying AMC scheme can be determined for a given target PER requirement. Using these thresholds, the probability π_k that the received SNR falls into a particular bin is given as,

$$\pi_k = \int_{s_k}^{s_{k+1}} p_\gamma(\gamma)\, d\gamma = \frac{1}{\Gamma(m)}\left[\Gamma\left(m, \frac{ms_{k+1}}{\bar{\gamma}}\right) - \Gamma\left(m, \frac{ms_k}{\bar{\gamma}}\right)\right] \tag{2-3}$$

where s_k is the boundary point of mode k, $\Gamma(m) = \int_0^\infty t^{m-1} exp(-t)\, dt$ is defined as the gamma function, m is the Nakagami fading parameter, $\bar{\gamma}$ denotes the average SNR, and $\Gamma(m,x) = \int_x^\infty t^{m-1} exp(-t)\, dt$ is the gamma incomplete function where x refers to $((ms_{k+1})/\bar{\gamma})$ or $((ms_k)/\bar{\gamma})$ as seen in (2-3).

2.3 Queuing model

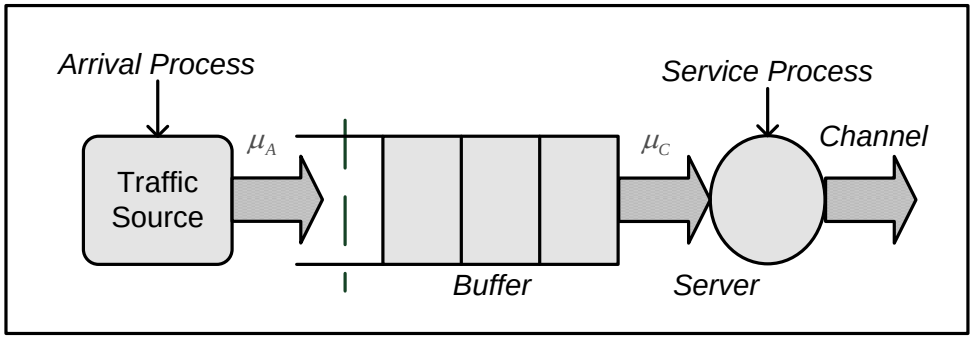

Figure 2-3 Queuing model

Based on the cross-layer design system model in Figure 2-1, the queuing model is situated at the data link layer, where the arrival and service processes form part of the upper protocol layer and physical layer respectively. A queuing model consists of an arrangement of what we shall call *items*, from an arrival process. These items are queued into an area of temporary storage, where they receive service, and then proceed to exit the system [32]. A simple, single server, queuing model is depicted in Figure 2-3, where μ_A represents the mean arrival rate, and μ_C the mean service rate of packets, per frame.

The queuing system is composed of a server and a storage buffer. If the server is busy, packets from the source traffic model are placed in the buffer to await service. However if the server is idle, packets are serviced immediately. The server accepts new packets when all previous packets waiting in the queue have completed their service or has their service time pre-empted by a packet of higher priority, according to a specific priority rule. Since our

main focus is not on scheduling algorithms, the simplest queuing priority rule was used, namely First in First out (FIFO). This implies that the first packet that enters the queue will be serviced first. The queuing model adopted follows an approximated M/M/1 type queue on a frame by frame basis, where M denotes a Markovian process which indicates that both arrival and service times are memory-less. There is one server and as mentioned, it is assumed that there is an infinite amount of space in the buffer. From Figure 2-1, the way in which upper layer packet stream into the data link layer is known as the *arrival process*, while the AMC scheme employed at the physical layer is denoted as the *service process* of the queuing model.

In real time multimedia traffic, the main QoS metric of interest is bounded delay and is hence the adopted key QoS requirement in this dissertation. Thus the queuing delay of each packet is measured, and is compared to the delay requirement set by the user. Since we are investigating the QoS provisioning performance over a range of bounded delay requirements, as seen in later in Chapter 5, we assume no packets are dropped from the queuing system. This can be reasoned if we consider a scenario where the QoS performance is evaluated at set requirements of 50ms and 100ms bounded delay. In this case, if packets were dropped when their queuing delay violated 50ms, the QoS performance at 100ms cannot be determined. Hence, for this reason, it becomes practical to ensure packets are not dropped from the queuing system.

2.4 Traffic Source Modelling

In Figure 2-1 and 2-3, the manner in which the upper layer protocol packets arrive into the buffer at the data link layer is denoted as the arrival process, as aforementioned. The traffic source of a given queuing model, can take many forms depending on the specified real time application. In [26], ON-OFF modelling and autoregressive models were used to represent audio and video services respectively. Several traffic sources were accommodated in [33], where the overall additive arrival process was modelled as a Markov modulated Poisson process, which had more than two states. It was stressed that each of the individual traffic sources had to be based on the same statistically parameters of the particular queuing model. In this dissertation the real time application of audio services, which sources bursty traffic, is considered where the arrival process is characterized using ON-OFF modelling. During the ON period, the source generates packets at some mean rate, while during the OFF period the source remains silent, by not generating any packets.

Let X_t denote the amount of packets generated by the source at time t. At the data link layer, the packets are queued in the buffer, awaiting service according to the specified priority rule, namely FIFO. In order to investigate the behaviour of the ON-OFF arrival process, a Markov modulated model is considered, to characterize the bursty nature of the source traffic. A discrete time, Markov modulated, two state (namely ON and OFF) source is depicted in Figure 2-4, where the self transition probabilities and transition probabilities between states are denoted as q_{ii} and q_{ij} respectively, where $1 \leq i, j \leq 2$. When the system is in state i at time t, as shown in Figure 2-4, the probability of transitioning to state j at time $t+1$ is given by q_{ij}. It is assumed that transitions only occur at discrete time intervals [33]. It is also seen that when the system is in state i, the amount of packets arriving in one time interval is given by $a_i(t)$, given that $1 \leq i \leq 2$.

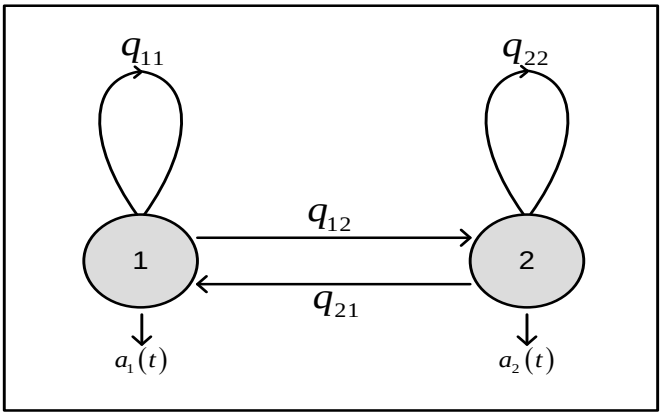

Figure 2-4 Discrete time, two state Markov modulated source model

The QoS provisioning performance investigated is of a statistical nature and depends on the log generating moments, or otherwise known as the Gärtner-Ellis (GE) [15] limit, of both arrival and service processes [17], as discussed in the next section. The Gärtner-Ellis theorem establishes the principle of large deviations, where the cumulant generating function converges towards a finite limit [15]. This resultant limit allows the traffic characterization of random, discrete-time, arrival and service processes.

Let $A(t)$ be a random variable representing the cumulative number of arrivals at time t and $a(t)$ be the number of arrivals at time t, where $t = 1,2,\ldots$ is specified. Then the GE limit of the arrival process $A(t)$ is determined by [15],

$$\Lambda_{\mathrm{B}}(\theta) \triangleq \lim_{t\to\infty} \frac{1}{t} \log\left[\mathrm{E}\left(\exp\left(\theta A(t)\right)\right)\right] \tag{2-4}$$

where $\Lambda_B(\theta)$ is the GE limit, $E(\cdot)$ denotes expectation; and the parameter θ is real, and positive [15]. The function given by (2-4) is validated, and verified in [15], to be convex, positive and increasing for $\theta > 0$. The parameter θ is denoted as the QoS exponent, and plays an important role in the cross-layer design framework, which will be discussed in subsequent sections.

Using the GE limit $\Lambda_B(\theta)$, the quantity of interest, namely the effective bandwidth function is determined by [34].

$$E_B(\theta) \triangleq \frac{\Lambda_B(\theta)}{\theta} = \lim_{t\to\infty} \frac{1}{\theta}\frac{1}{t}\log\left[E\left(\exp\left(\theta A(t)\right)\right)\right] \tag{2-5}$$

The effective bandwidth function models the source traffic from a QoS point of view, and is used in conjunction with the service process modelling, to form a cross-layer platform in which the QoS provisioning performance can be analyzed, as seen later in chapter 5. In [15], Based on the assumption that the arrivals at time t, namely $a(t)$, are random and independent of each other, Chang showed that the effective bandwidth expression seen in (2-5), has a closed form expression given by,

$$E_B(\theta) \triangleq \frac{\Lambda_B(\theta)}{\theta} = \frac{1}{\theta}\log[sp\{\mathbf{P}\boldsymbol{\Psi}(\theta)\}] \tag{2-6}$$

where $sp\{\cdot\}$ is the spectral radius, $\mathbf{P}$ is the probability transition matrix and $\boldsymbol{\Psi}$ is a diagonal matrix containing the moment generating functions (MGF) of $a_i(t)$. The probability matrix $\mathbf{P}$ is defined as,

$$\mathbf{P} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \tag{2-7}$$

where state 1 is denoted as the OFF state, and correspondingly state 2 is the ON state. Note that $q_{11} + q_{12} = 1$, and $q_{21} + q_{22} = 1$. The diagonal matrix $\boldsymbol{\Psi}(\theta)$ can be expressed as,

$$\boldsymbol{\Psi}(\theta) = \text{diag}[\Psi_1(\theta), \Psi_2(\theta)] \tag{2-8}$$

with

$$\Psi_i(\theta) = E\left(\exp(a_i(t)\theta)\right) \tag{2-9}$$

If the source traffic modelling is considered as an ON-OFF Markov modulated Poisson Process, then (2-8) becomes [25] (refer to Appendix A.1),

$$\boldsymbol{\Psi}(\theta) = \begin{bmatrix} 1 & 0 \\ 0 & \exp\left(v\left(\exp(\theta N_p) - 1\right)\right) \end{bmatrix} \tag{2-10}$$

where v is the mean arrival rate, and N_p is the amount of bits in a packet. It must be highlighted than the parameter μ_A denotes the average arrival rate over the entire ON-OFF period, and is obtained when $\lim_{\theta \to 0} \mathrm{E_B}(\theta)$. The parameter v denotes the mean packet arrival rate, specifically in the ON period, and is set by the user at the beginning of a simulation. The effective bandwidth function converges to v, also known as the peak rate [15], when $\theta \to \infty$.

Note that in [25] X_t is given in bits however for this dissertation, where the unit of interest is packets, X_t is in terms of packets. Thus for convenient discussion, and further comparisons with the results seen in [25], the MGF seen in (2-9) – (2-10) is expressed in bits, by introducing the term N_p. In [15], given the case of an ON-OFF model, the effective bandwidth function seen in (2-6) reduces to,

$$\mathrm{E_B}(\theta) = \frac{1}{\theta}\log\left(\frac{1}{2}\left(\mathrm{z}(\theta) + \sqrt{\left(\mathrm{z}(\theta)\right)^2 + 4\Psi_1(\theta)\Psi_2(\theta)(1 - \mathrm{q}_{11} - \mathrm{q}_{22})}\right)\right) \tag{2-11}$$

where $\mathrm{z}(\theta) = \mathrm{q}_{11}\Psi_1(\theta) + \mathrm{q}_{22}\Psi_2(\theta)$. Using (2-7) – (2-10) the effective bandwidth function seen in (2-11) can be determined (refer to Appendix A.2).

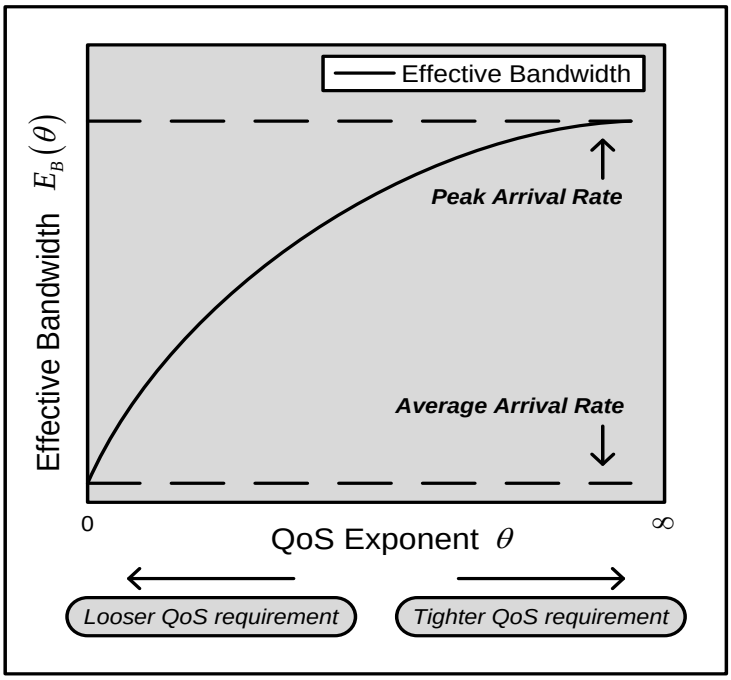

Figure 2-5 Effective Bandwidth Properties

A general representation of the effective bandwidth curve can be seen in Figure 2-5. It is shown that the effective bandwidth function converges to the average arrival rate and peak arrival rate when θ tends to zero and infinity, respectively.

2.5 Statistical QoS guarantees

In today's mobile wireless networks, the transmission of delay-sensitive multimedia traffic, such as voice and video, requires the support of bounded delay QoS guarantees. Existing physical layer channel models, given their time varying nature, present not only the challenge in providing QoS but make deterministic guarantees unrealistic to support. Thus it is more practical to investigate statistical QoS guarantees; in particular supporting delay bound requirements with a minor delay bound violation probability.

Another challenge presented by existing physical layer modelling, is the characterization of the fading channel in terms of delay, since conventionally this metric has been associated with queuing analysis. Thus alternative solutions, such as the aforementioned cross-layer design is considered, where the channel model is moved up to the data link layer. This notion allows the impact of the fading channel on the data link layer queuing behaviour, to be investigated. To provide such a framework, the arrival and service processes need to be modelled in terms of QoS metric, delay. In the previous section the concept of effective bandwidth was introduced as a technique to model arrival processes, and forms the basis of the cross-layer design framework.

The theory of effective bandwidth was used in [15-18] to model the statistical behaviour of arrival traffic accessing the buffer at the data link layer. The effective bandwidth of a given arrival process can be defined as the minimum service rate required in order to guarantee a QoS requirement [14]. From [15], the asymptotic results based on the effective bandwidth theory showed the probability that the queue length Q, exceeded a fixed threshold B_Q, can be expressed as

$$\Pr\{Q > B_Q\} = \exp(-\theta B_Q) \tag{2-12}$$

where θ was denoted as the QoS exponent, and B_Q was assumed to be very large. From (2-12) it can be seen that as B_Q increases, the probability that Q is greater than B_Q decreases.

For small values of B_Q, a more accurate approximation is given as [14, Eq. (9)],

$$\Pr\{Q > B_Q\} = \varepsilon \exp(-\theta B_Q) \tag{2-13}$$

where ε is defined as the probability of having a non-empty buffer. Using [15, Eq. (9.184)] ε can be approximated as,

$$\varepsilon \approx \frac{\mu_A}{\mu_C} \tag{2-14}$$

where μ_A is the average arrival rate shown in [15], and μ_C is the average service rate validated in [20]. The parameter is also commonly referred to as service efficiency.

As mentioned, the QoS metric of interest is bounded delay. Hence the probability of the queuing delay experienced by a packet, violating a specific delay constraint at time t can be expressed as follows [14, Eq. (10)],

$$\Pr\{D > D_{max}\} = \varepsilon \exp(-\theta^* \delta D_{max}) \tag{2-15}$$

where D, is the packet delay at time t, D_{max} is the delay constraint, δ is a parameter that is determined by arrival and service processes, and θ^* is the unique QoS exponent that describes the arrival rate that service process can support in order to meet the set QoS requirement. From (2-12), (2-13) and (2-15), it is observed that the QoS exponent θ plays a significant role in the degree of violation of QoS requirements. The QoS requirements in question are namely the buffer length, or more specifically for our discussion, bounded delay. As θ increases, the probability of violation decreases, hence a tighter or more *stringent* QoS requirement can be guaranteed. However as θ decreases, the probability of violation increases, thus a *looser* QoS requirement can only be supported.

2.6 Effective Capacity and Cross-Layer Design

In [14], Wu and Negi established that it was possible to adapt the effective bandwidth theory to accommodate the modelling of the service process of the wireless channel. The characterisation of the service process was called the effective capacity and was noted as the dual concept of the effective bandwidth theory. The effective capacity can be defined as maximum arrival rate that the channel can support, in order to fulfil the prescribed QoS requirements [14]. This meant that by applying the effective capacity theory, the service process of the channel model can be linked to the QoS performance at the data link layer, given the fact that both the effective bandwidth and capacity functions have a mutual

association with the QoS exponent θ. The effective capacity function, based on the GE limit $\Lambda_C(\theta)$ of the service process $C(0,\tau)$, can be expressed as,

$$\mathrm{E_C}(\theta) \triangleq -\frac{\Lambda_C(-\theta)}{\theta} = \lim_{t\to\infty}\frac{1}{\theta}\frac{1}{t}\log\left[\mathrm{E}\left(\exp\left(-\theta C(t)\right)\right)\right] \tag{2-16}$$

where $\Lambda_C(-\theta)$ is denoted as the delay exponent. From (2-16) it can be seen that the effective capacity function is the dual concept of the effective bandwidth function, expressed in (2-5). The effective capacity function in (2-16), and the delay exponent $\Lambda_C(-\theta)$, will be discussed in detail in the next chapter.

The properties of the effective capacity function $\mathrm{E_c}(\theta)$, are illustrated in Figure 2-6. It is observed that as the QoS exponent increases, the effective capacity correspondingly decreases. This behaviour is the result of the channel supporting reduced arrival traffic rates in order to support the more stringent QoS requirement. It is also shown that the effective capacity function converges to the average service rate and minimum service rate when θ tends to zero and infinity, respectively.

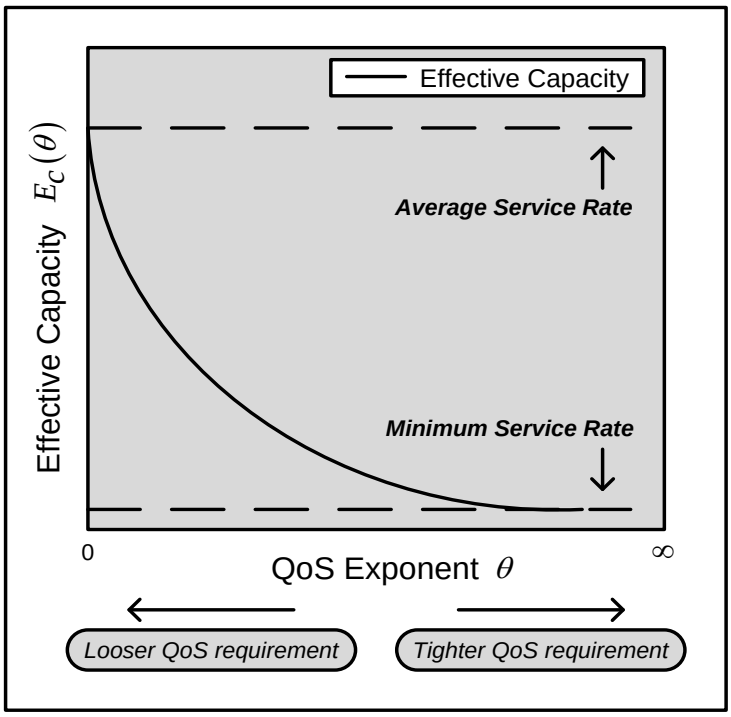

Figure 2-6 Effective Capacity Properties

Figure 2-7shows the relationship between the effective bandwidth and capacity curves, using Cases I, II and III, to help elaborate on the principles of these theories.

Case I: Average arrival rate is greater than average service rate

The scenario of where the average arrival rate is greater than the average service rate is depicted in case I. From Figure 2-7, it is observed that there are no intersections between arrival process A and service process B, since the channel can not support the given arrival rate. From queuing theory this results in a build up of packets in the storage buffer, where eventually the queuing size will reach infinity.

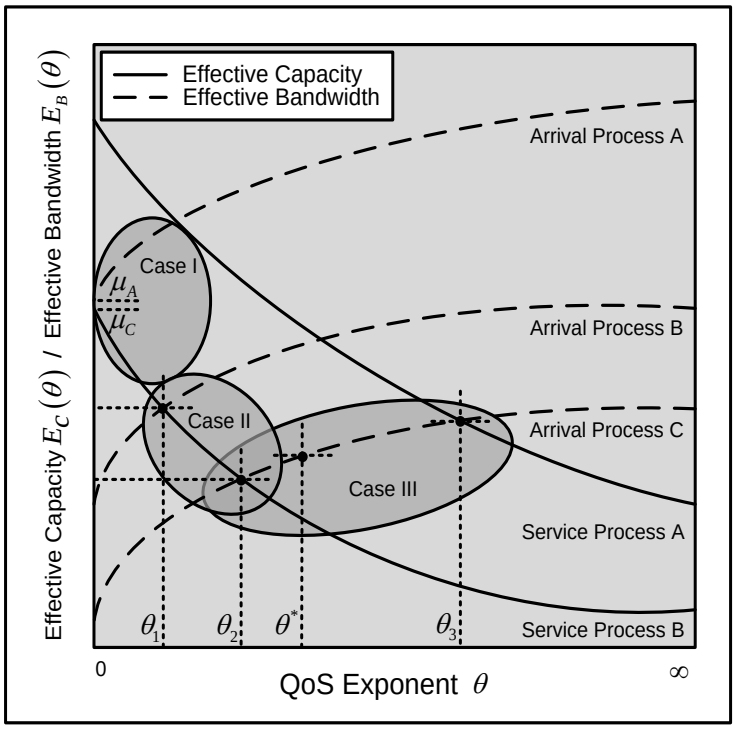

Figure 2-7 Relationships between Effective Capacity and Effective Bandwidth, adapted from [35]

Case II: Fixed service rate, while varying the arrival rate

For Case II, we investigate the fixed service process B, subject to arrival process B, with a higher bandwidth, and arrival process C with a lower bandwidth. From Figure 2-7 it is seen that arrival processes B and C, cut service process B at θ_1 and θ_2 respectively, where $\theta_1 < \theta_2$. As mentioned previously, a lower θ implies that a looser QoS requirement can be guaranteed, which in this scenario is θ_1. This is expected given the fact that as the arrival rate increases for a fixed service rate; the queuing delay correspondingly increases thus increasing delay bound violation.

Case III: Fixed arrival rate, while varying the service rate subject to a user specific QoS requirement θ^*

In Case III arrival process C is kept fixed while subjected to service process A and B, where the user QoS requirement θ^* must be met. From Figure 2-7, it is observed that the lower service process B cuts arrival process A at θ_2, whereas the higher service process A cuts at θ_3. It is clearly seen that $\theta_2 < \theta^* < \theta_3$. Thus the higher service process A can guarantee the user specified QoS requirement, while service process B is unable to support the QoS guarantee given its higher delay bound violation probability.

2.7 Chapter Summary

In chapter 2, the cross-layer design system model was introduced where detailed descriptions of the AMC, M/M/1 queuing model, and source traffic modelling, were given. The concept of effective bandwidth and effective capacity theories was presented and their subsequent role in supporting statistical QoS guarantees was discussed. The aim of the next chapter is to investigate service process modelling of the applied AMC scheme at the physical layer, using the effective capacity concept, in block fading and correlated fading channels.

Chapter 3

Effective Capacity

In chapter 2, the concept of effective capacity theory was discussed as a tool, to be used in cross-layer based design. As specified in [14], the effective capacity is defined as the constant arrival rate that the channel can support in order to guarantee a QoS requirement. The aim of this chapter is to introduce service process modelling over block fading and correlated channels using effective capacity, and to investigate its behaviour subject to varying physical layer parameters.

This chapter is outlined as follows: Section 3.1 presents the system model, particularly focusing on the area of interest in this chapter. Section 3.2 introduces the formulation of the delay exponent and effective capacity expressions over block fading channels. Section 3.3 discusses the numerical results of these expressions, given variations in the average SNR, target PER, and fading parameter m. In Section 3.4 FSMC modelling in correlated fading channels is discussed, and the numerical evaluation of the effective capacity subjected to varying fading parameter m, and Doppler frequency is shown in Section 3.5. Finally in Section 3.6, a chapter summary is given.

3.1 System model

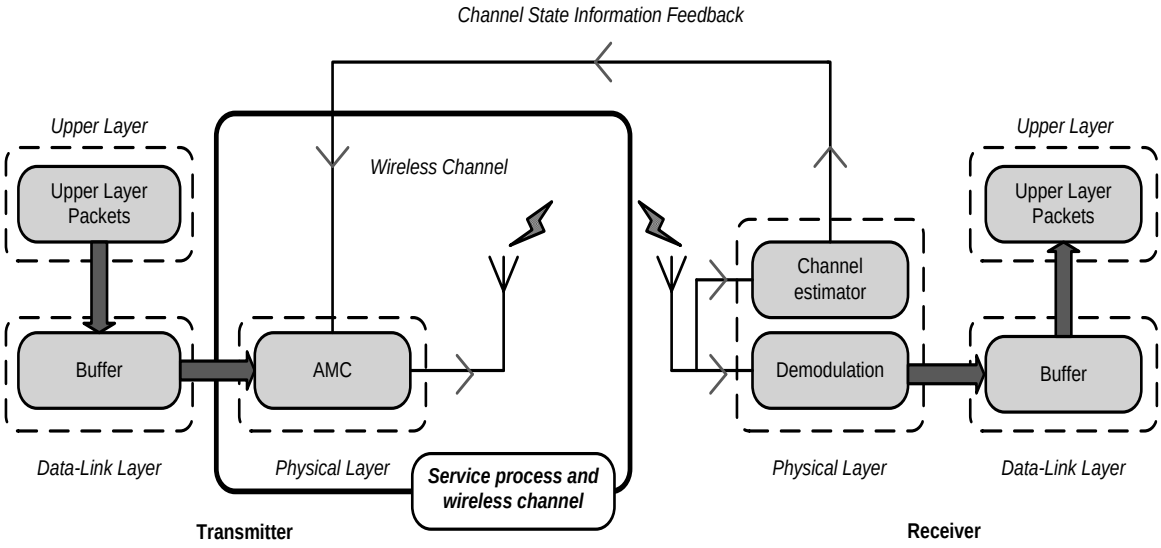

Figure 3-1 System model adapted from [26]

The cross-layer design framework introduced in chapter 2, involved the computation of the effective capacity and effective bandwidth functions, which correspondingly required the modelling of arrival and service processes. The arrival process, specifically the behaviour in which upper layer packets stream into the buffer at the data link layer, was modelled in Section 2.4, where a general expression for the effective bandwidth function of an ON-OFF process was provided (2-11). Thus based on the cross-layer design system model as seen in Figure 3-1, the white block depicts the restricted interest of this chapter, which is to investigate the different types of service modelling and consequent effective capacity, in both block fading and correlated fading channels.

3.2 Block fading channel Model

In a block fading model, the channel between the transmitter and receiver (refer to Figure 3-1) is frequency flat, and remains invariant per frame, but is allowed to vary from frame to frame [22]. As a result, the AMC is adjusted on a frame by frame basis hence the service process of each frame is uncorrelated.

Let the sequence $\{c_k(n),\ n = 1,2,\dots\}$ denote a discrete-time stationary and ergodic service process, where the time index of the frame is n and k denotes the transmission mode. The partial sum of the service process over the time sequence $n = 1,2,\dots t$ is determined by $C(t) = \sum_{n=1}^{t} c_k(n)$ and correspondingly its GE limit is expressed, as defined in [16] as,

$$\Lambda_{\mathrm{C}}(\theta) \triangleq \lim_{t\to\infty} \frac{1}{t} \log\left[\mathrm{E}\left(\exp\left(\theta C(t)\right)\right)\right] \tag{3-1}$$

where θ is the QoS exponent, and E{·} denotes the expectation. The expression seen in (3-1) is a convex function differentiable for all real θ [16].

It follows that the effective capacity of the service process, based on (3-1), is evaluated as given in [14, Eq. (12)],

$$\mathrm{E}_{\mathrm{C}}(\theta) \triangleq -\frac{\Lambda_{\mathrm{C}}(-\theta)}{\theta} = -\lim_{t\to\infty} \frac{1}{\theta t} \log\left[\mathrm{E}\left(\exp\left(-\theta C(t)\right)\right)\right] \tag{3-2}$$

If the sequence $\{c_k(n),\ n = 1,2,\dots\}$ is considered to be an uncorrelated process, then (3-2) reduces to [23, Eq. (4)],

$$\mathrm{E}_{\mathrm{C}}(\theta) = -\frac{1}{\theta} \log\left[\mathrm{E}\left(\exp\left(-\theta c_k(n)\right)\right)\right] \tag{3-3}$$

From (3-2), let $\Lambda_C(-\theta)$ be denoted as the delay exponent. Hence for the uncorrelated sequence $\{c_k(n),\ n = 1,2,\dots\}$, as seen in (3-3), $\Lambda_C(-\theta)$ can be determined by,

$$\Lambda_C(-\theta) = \log\left[\mathrm{E}\left(\exp\left(-\theta c_k(n)\right)\right)\right] \tag{3-4}$$

The service process $c_k(n)$ is defined as the amount packets that can be accommodated in frame n. Based on [20, Eq. (19)], $c_k(n)$ is determined by $c_k(n) = T_f W R_k$, where T_f is the frame duration, W is the spectral bandwidth, and R_k is the transmission rate for mode k. For simplicity we omit the frame index n in service process c_k.

By evaluating the MGF of the uncorrelated service process c_k, given by $\mathrm{E}\{\exp(-\theta c_k)\}$, (3-3) and (3-4) reduce to [25],

$$\mathrm{E}_C(\theta) \triangleq -\frac{\Lambda_C(-\theta)}{\theta} = -\frac{1}{\theta}\log\left[\sum_{k=0}^{K-1}\pi_k \exp\left(-\theta N_p c_k\right)\right] \tag{3-5}$$

and,

$$\Lambda_C(-\theta) = \log\left[\sum_{k=0}^{K-1}\pi_k \exp\left(-\theta N_p c_k\right)\right] \tag{3-6}$$

respectively, where N_p is the amount of bits in a packet, k denotes the transmission mode selected by the AMC scheme for the given frame, and π_k is the probability that the received SNR falls into particular bin, which is detailed in Section 2.2, (2-3).

3.3 Numerical Results of block fading channel

In this section the behaviour of the delay exponent $\Lambda_C(-\theta)$, and effective capacity $\mathrm{E}_C(\theta)$ of the applied AMC scheme, are evaluated by numerical solutions under different physical layer parameters. Unless otherwise stated, the reference system parameters are set as follows: upper-layer packet size $N_p = 1080$ bits; frame duration $T_f = 2$ms; fading parameter $m = 1$, average SNR $\bar{\gamma} = 10$dB and PER target $P_{th} = 10^{-3}$. The mode dependent parameters a_k, b_k and R_k, of the AMC scheme are detailed in Table 2-1. In [25], the service process c_k is defined as $c_k = bR_k/R_1$, where b is defined as the bandwidth parameter. It was specified that $b = 2$, hence the service process reduced to $c_k = 4R_k$, given $R_1 = 0.5$. For purposes of replicating the delay exponent results seen in [25] over block fading channels, we define the service process c_k as $c_k = T_f W R_k = 4R_k$. Thus the spectral bandwidth W is

set to $W = 2$KHz. The conventional, constant power AMC scheme applied in this section, over block fading channels, is denoted as AMC scheme 1 for the remainder of this dissertation.

3.3.1 Delay Exponent

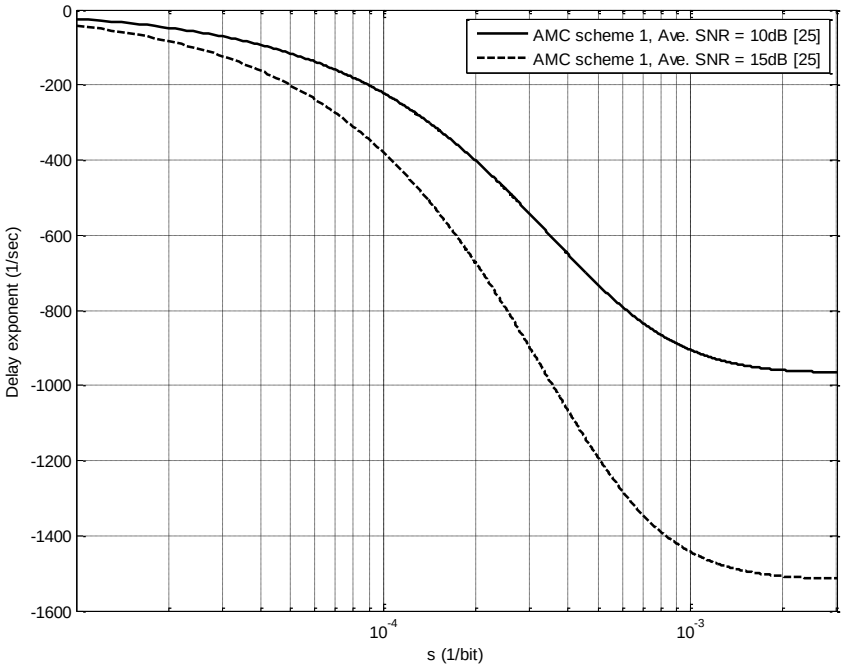

Figure 3-2 Delay exponent function over block fading channels varying average SNR $\overline{\gamma}$, given $m = 1$ and $P_{th} = 10^{-3}$

The delay exponent $\Lambda_C(-\theta)$ results shown in Figure 3-2, were replicated from [25], under the effects of varying average SNR. The unit of the delay exponent used in [25], was specified as 1/sec. The delay exponent seen in (3-6) however, has the unit 1/frames. Thus, in order to replicate the delay exponent results of [25], (3-6) is simply divided by T_f to subsequently produce the unit 1/sec as seen in as seen in [25, Eq. (9)],

$$\Lambda_C(-\theta) = \frac{1}{T_f}\log\left[\sum_{k=0}^{K-1}\pi_k \exp\left(-\theta N_p c_k\right)\right] \tag{3-7}$$

It is shown that, as the average SNR increases, the delay exponent decreases accordingly. The behaviour of the delay exponent $\Lambda_C(-\theta)$ function can be explained by examining its

relationship with the analytical delay bound violation probability expression seen in (2-15). If (3-2) is rewritten such that the delay exponent is the subject of the expression, it follows that,

$$\Lambda_C(-\theta) = -\theta E_C(\theta) \tag{3-8}$$

Thus for a unique QoS exponent θ^*, by substituting the delay exponent seen in (3-8) into (2-15), the analytical delay bound violation probability is given by,

$$Pr\{D > D_{max}\} = \varepsilon exp(\Lambda_C(-\theta^*)D_{max}) \tag{3-9}$$

where D is the packet delay, D_{max} denotes the set delay bound, and ε the probability the buffer is not empty. Note that the parameter δ, from the original analytical delay bound violation probability expression in (2-15), is the y axis co-ordinate when the effective capacity curve $E_C(\theta)$ intersects the effective bandwidth curve $E_B(\theta)$ at θ^*, as will be discussed later in chapter 5. Thus δ can be represented by either $E_C(\theta^*)$ or $E_B(\theta^*)$, but for convenient discussion, given that the subject of this chapter is effective capacity, $E_C(\theta^*)$ is considered. By analyzing (3-9) it can be seen as the delay exponent decreases in value, or in other words gets more negative, correspondingly the delay bound violation probability reduces. The converse also applies where as the delay exponent increases, correspondingly the delay bound violation probability increases. Thus there exists a direct relationship between the delay exponent and delay bound violation probability.

In Figure 3-2, it is seen that the larger average SNR value, produces the smaller delay exponent. The improved average SNR, or equivalently, increased transmission power, results in an increased mean service rate. This correspondingly increases the amount of packets serviced from the queuing system hence producing a smaller queuing delay. Thus for a fixed delay bound, a smaller delay bound violation probability, which translates to a smaller or more negative delay exponent, is achieved by the higher average SNR value, which in this case is $\bar{\gamma} = 15dB$.

Note that the delay exponent function is evaluated in terms of bits by introducing the parameter N_p, which denotes the amount of bits in a packet. This is attributed to the delay exponent results of the given AMC scheme being replicated from [25], where the unit of bits was adhered to.

3.3.2 Effective Capacity

The effective capacity $\mathrm{E_C}(\theta)$, as seen in (3-2), is based on simply dividing the delay exponent $\Lambda_\mathrm{C}(-\theta)$, by the QoS exponent θ. Thus, the numerical methods applied in obtaining the replicated delay exponent results, as seen in Figure 3-2, can be used to validate the computation of the effective capacity function. The effects of varying physical layer parameters such as: target PER P_{th}, fading parameter of Nakagami-m model, and average SNR on the effective capacity function can be seen in Figure 3-3, 3-4 and 3-5 respectively. In these plots the effective capacity is normalized to ease the comparison with spectral efficiency, given by $\mathrm{E_C}(\theta)/T_f/W/N_p$ (packets/sec/Hz).

Note that range of smaller QoS exponent values shown in these figures, in particular from 10^{-3} to 10^{-1}, is of a greater accuracy compared to that of the general effective capacity illustration seen in Figure 2-6 in Section 2.6. The reasoning behind this is to depict the levelling of the effective capacity function as $\lim_{\theta\to 0}\mathrm{E_C}(\theta)$. This will allow an easy form of comparison between spectral efficiency, and the average service rate.

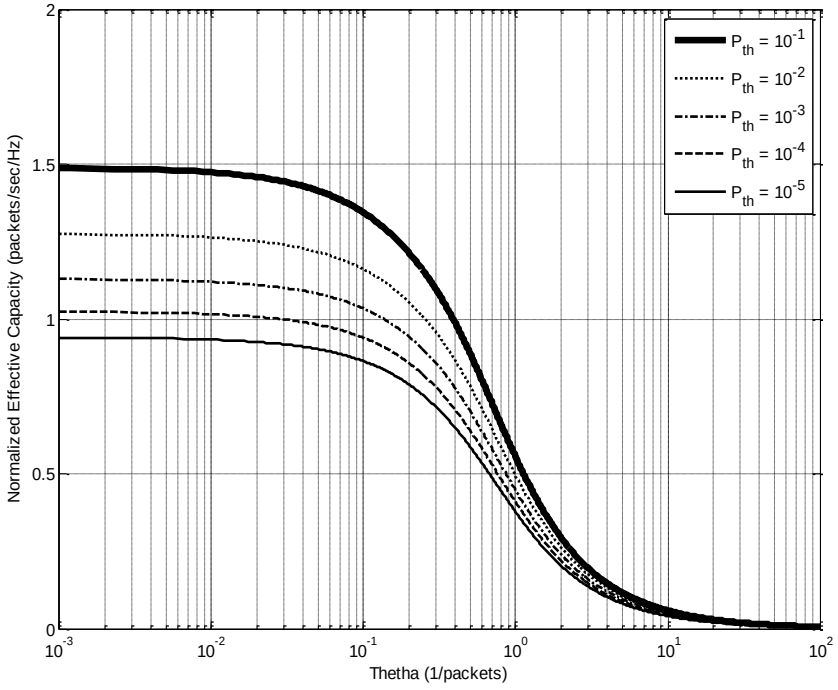

Figure 3-3 Effective Capacity over block fading channels varying target PER P_{th}, given $\bar{\gamma} = 10$dB and $m = 1$

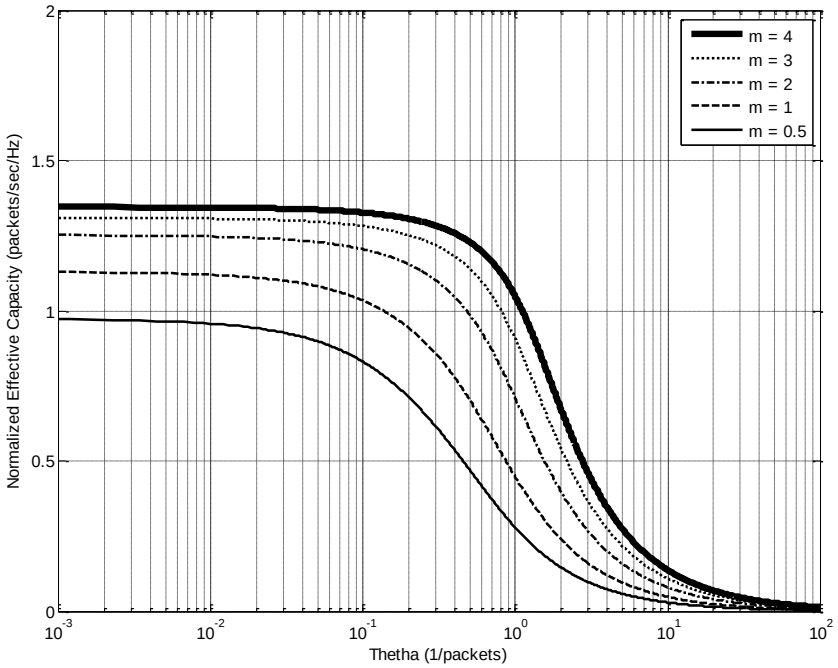

Figure 3-4 Effective Capacity over block fading channels varying fading parameter m, given $P_{th} = 10^{-3}$, and $\overline{\gamma} = 10$dB

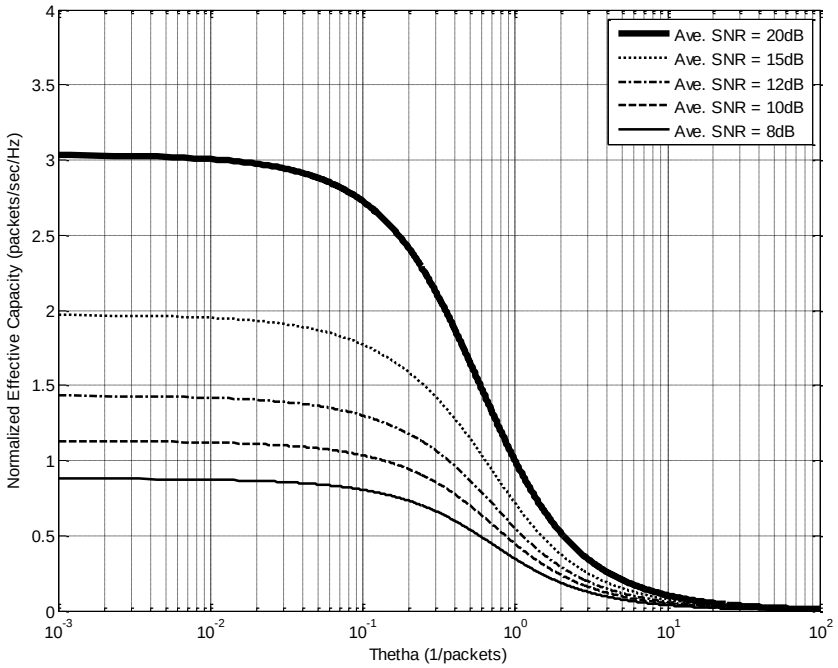

Figure 3-5 Effective Capacity over block fading channels varying average SNR $\overline{\gamma}$, given $P_{th} = 10^{-3}$, and $m = 1$

The numerical results obtained in Figure 3-3, 3-4 and 3-5 match the expected behaviour properties of the effective capacity as discussed in Section 2.6. These properties include: $E_C(\theta)$ decreases as the QoS exponent θ increases, $\lim_{\theta \to 0} E_C(\theta)$ approaches the average service rate and $\lim_{\theta \to \infty} E_C(\theta)$ approaches the minimum service rate.

In Figure 3-3, it is shown that a lower effective capacity guarantees a more stringent, user reliability-QoS requirement of PER. Conversely it is observed that a higher effective capacity secures a looser PER QoS guarantee. Based on (2-2) a larger PER constraint, say 10^{-1}, will result in smaller set of threshold points than that of a smaller PER constraint, namely 10^{-5}. Given the received SNR, a larger PER constraint will hence allow the AMC scheme to select a higher transmission mode, resulting in a larger throughput, and correspondingly a better effective capacity. Figure 3-4 shows that as the fading parameter m increases, the effective capacity increases respectively. This is expected given that a better channel quality can guarantee a more stringent QoS. From Figure 3-5, it is observed that a higher average SNR $\bar{\gamma}$ improves the effective capacity. Following discussions on Figure 3-2, the increased effective capacity reflects the improved system's throughput, given an increase in transmission power. When $\bar{\gamma}$ tends to infinity, the effective capacity will tend to the highest spectral efficiency of the applied AMC scheme.

3.4 Correlated Fading Channels

Unlike block fading channels discussed in Section 3.2, the correlation between subsequent channel path gains in correlated fading channels, is more complex to model. Although it was assumed that block fading channels would be adopted in the given system model, it is still interesting to investigate the characterisation of correlated fading channels, due to it being a more practical fading scenario. In the early 1960's the work of Gilbert in [36] and Elliot in [37], looked at modelling the wireless channel in two states, which described channel quality as being either noisy or noiseless. However this type of modelling is inadequate in characterizing the dynamic variations of channel quality. A simple solution would be to extend the channel model to include more than two states. This concept, motivated the works of [19], where instead of two states, an FSMC, was used to describe the wireless channel behaviour. The received SNR was partitioned into intervals, where the probability of being in a particular interval was denoted as a state in the FSMC model. This type of modelling has been extremely popular [8][20][22-26] in characterising the behaviour of the AMC scheme, which as described in Section 2.2, divides the SNR range into a finite amount of bins or

intervals. Hence in this section, the FSMC is used to model the dynamic and statistical behaviour of the AMC based, wireless channel service process.

In the FSMC, each state corresponds to a mode of the applied AMC scheme which has $K+1$ transmission modes. Let the transition probability matrix of the FSMC be denoted as $\mathbf{P} = [\rho_{ij}]_{(K+1)\times(K+1)}$, where ρ_{ij} represents the transition probability that the FSMC moves from state i, at time t, to state j, at time $t+1$. A slow fading channel is assumed; hence the transition activities are restricted to only neighbouring states [19]. The adjacent entries of **P** are calculated as [19]:

$$p_{k,k+1} \approx \frac{\mathrm{N_s}(s_{k+1})T_f}{\pi_k} \qquad \text{, where } k = 0,1 \ldots., K-2 \tag{3-10}$$

and

$$p_{k,k-1} \approx \frac{\mathrm{N_s}(s_k)T_f}{\pi_k} \qquad \text{, where } k = 1,2 \ldots., K-1 \tag{3-11}$$

where π_k is the probability that the received SNR falls into mode k, and the level crossing rate, denoted as $\mathrm{N_s}(\gamma)$, is given as [19],

$$\mathrm{N_s}(\gamma) \approx \frac{\sqrt{2\pi} f_d}{\Gamma(m)} \left(\frac{m\gamma}{\bar{\gamma}}\right)^{m-\frac{1}{2}} \exp\left(-\frac{m\gamma}{\bar{\gamma}}\right) \tag{3-12}$$

where f_d is the maximum Doppler frequency of the channel fading process, γ is the received SNR, $\bar{\gamma}$ is denoted as the average SNR, and $\Gamma(\cdot)$ is the gamma function. The remainder of the transition probabilities of matrix **P** are obtained using (3-10) and (3-11)

$$p_{0,0} = 1 - p_{0,1} \tag{3-13}$$

$$p_{K-1,K-1} = 1 - p_{K-1,K-2} \tag{3-14}$$

$$p_{k,k} = 1 - p_{k,k-1} - p_{k,k+1} \tag{3-15}$$

The boundary points $\mathrm{s_k}$, used in determining the entries of **P**, can be obtained using [20, Eq. (5)]. Finally, based on the FSMC based modelling, the effective capacity of the queuing service process is determined by

$$\mathrm{E_C}(\theta) \triangleq -\frac{\Lambda_\mathrm{C}(-\theta)}{\theta} = -\frac{1}{\theta}\log\left[\mathrm{sp}\left\{\mathbf{P}\left(\exp(-\theta\mathbf{\Phi})\right)\right\}\right] \tag{3-16}$$

where sp{.} is the spectral radius of the matrix, and $\boldsymbol{\Phi}$ is determined by $\boldsymbol{\Phi} = \text{diag}(\mu_0, \ldots, \mu_{K-1})$ which has the units bits per frame. The transmission rate μ_k is defined as $\mu_k = R_k T_f W$, where R_k is the spectral efficiency of mode k, and W denotes the spectral-bandwidth [20].

3.5 Numerical Results of correlated fading channels

3.5.1 Effective Capacity

The evaluation of the effective capacity function of the employed AMC scheme over correlated channels, subject to varying the fading parameter m, and Doppler frequency f_d is shown in Figure 3-6 and Figure 3-7 respectively. The results seen in Figure 3-6 were replicated from [20] given the following reference parameters: upper-layer packet size $N_p = 1080$ bits; frame duration $T_f = 2$ms, fading parameter $m = 1$, average SNR $\bar{\gamma} =$ 10dB, PER target $P_{th} = 10^{-3}$, Doppler frequency $f_d = 5$Hz and system spectral bandwidth $W = 100$KHz. Unless otherwise stated in the figure legend, parameters are defaulted to their reference values. The mode dependent parameters, namely a_k, b_k and R_k, for the AMC scheme applied in this section are given in [28, Table II], which are also used for the numerical evaluations in [20].

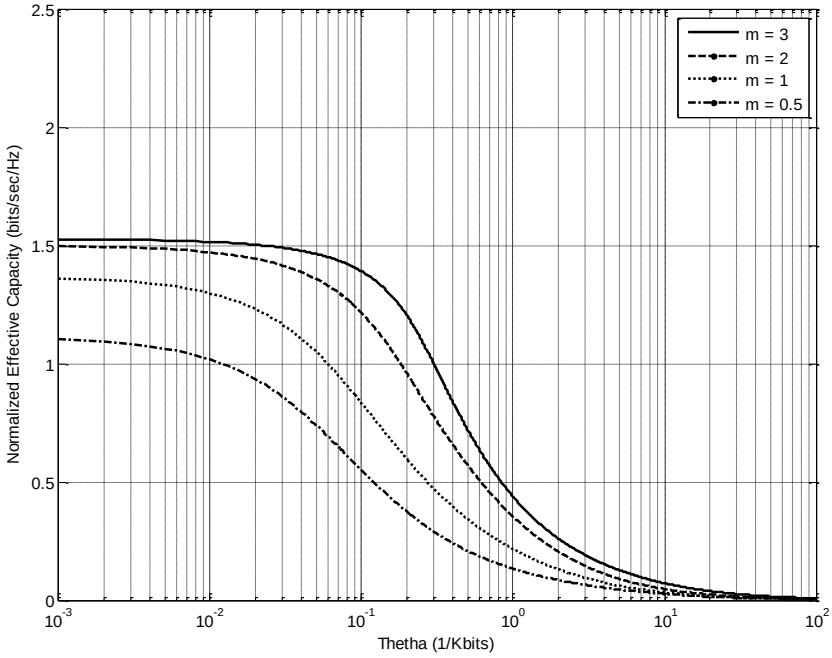

Figure 3-6 Effective Capacity over correlated fading channels varying fading parameter m, given $P_{th} = 10^{-3}$, $\bar{\gamma} = 10$dB, and $f_d = 5$Hz

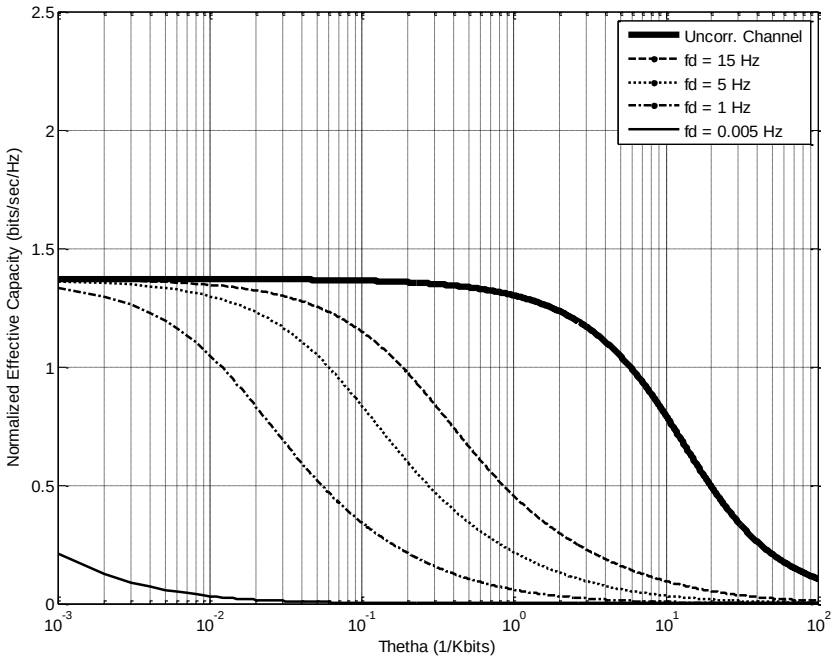

Figure 3-7 Effective Capacity over correlated fading channels varying Doppler frequency f_d, given $P_{th} = 10^{-3}$, $\bar{\gamma} = 10$dB, and $m = 1$

The replication of Figure 3-6, from [20], validates the method in obtaining Figure 3-7. It must be noted that the frame units set in block fading and correlated fading channels followed the respective system parameters of the papers replicated. Thus in the block fading channel scenario, the frame units was set in terms of packets per frame, whereas in correlated fading channels it was set to bits per frame. The units of normalized effective capacity for correlated fading channels is hence bits/sec/Hz, given by $\mathrm{E_C}(\theta)/T_f/W$.

In Figure 3-6, it is observed that the behaviour of the effective capacity follows the same trends as in the case of block fading channels. As the fading parameter m, or equivalently channel quality, improves, the effective capacity increases. The effects of varying the degree of channel correlation on the effective capacity function are shown in Figure 3-7. As the Doppler frequency increases, the correlation between fading coefficients decreases due to the corresponding increase in time diversity. On the other extreme, as time diversity decreases (Doppler frequency decreases), the channel becomes more correlated. Thus the two channel limits of varying the Doppler frequency are an uncorrelated channel (0% correlation), and a fully correlated channel (100% correlation). Based on the proposition, $\mathrm{E_{c_1}}(\theta) = \mathrm{E_{c_2}}\left((f_{d_2}\theta)/f_{d_1}\right)$ in [20], it is expected that the effective capacity increases as the Doppler

frequency increases, which is seen in Figure 3-7. If $(f_{d_2}/f_{d_1}) > 1$, then $E_{c_1}(\theta)$ is the horizontally left-shifted form of $E_{c_2}(\theta)$ [20]. The direct translation of this is that if the Doppler frequency increases from time t to time t + 1, the effective capacity will increase correspondingly. The converse also applies, where if $(f_{d_2}/f_{d_1}) < 1$, then $E_{c_1}(\theta)$ is the horizontally shifted right form of $E_{c_2}(\theta)$ [20].

3.6 Chapter Summary

This chapter investigated the use of effective capacity in modelling service processes over block fading and correlated fading channels. Based on the numerical results, it was observed that variations at the physical layer had a significant impact on the effective capacity function. Thus by characterizing the behaviour of the physical layer, a simple platform is created where the effective capacity and QoS provisioning performance are linked. The next chapter aims to optimize the switching levels of the AMC scheme, applied over block fading channels, which will increase the effective capacity, and hence improve the delay-bound QoS performance.

Chapter 4

Optimum mode switching Adaptive Modulation Coding

In the previous chapter, service process modelling was investigated, where it was shown how the effective capacity quantified the achievable service rate of the AMC scheme, employed at the physical layer. The switching levels adopted by the applied AMC schemes in chapter 3, were fixed given a target PER constraint, as seen in [20]. However, by maintaining deterministic thresholds, the opportunity for further improvement on the system's average throughput, and hence enhancement on the effective capacity, is lost. The works in [23] presented a power and rate adaption scheme, which aimed to maximise the system's throughput subject to meeting a QoS constraint. The optimized policy achieved a superior effective capacity when compared to water filling, constant power, and channel inversion schemes. The system model however, was constrained by a BER QoS requirement. In this dissertation, the system model is based on the unit of packets, thus motivating the use of a PER constraint. This avenue was considered in [25], where the proposed power adaptive scheme optimized packet delay performance at the data link layer, while maintaining the average power and instantaneous target PER constraints. Although the aspect of meeting an instantaneous target PER constraint was addressed, no relationship between the average PER of the proposed scheme, and target PER was available. In [27] the objective of the optimized scheme was to maximise the achievable bits-per-symbol (average throughput) while meeting a target average BER. The work in [27] however, did not investigate the effective capacity or QoS performance of the system, and as in [23], did not incorporate PER constraints.

The aim of this chapter is to extend the optimized scheme in [27] to accommodate an average PER constraint, while maximising the system's average throughput, and to investigate the effective capacity over block fading channels. The significance of this extension is attributed to the impact that the effective capacity has on the delay-bound QoS provisioning performance. By maximising the system's average throughput, the effective capacity will hereby increase, and will be able to support a more stringent delay-bound QoS requirement.

This chapter is outlined as follows: Section 4.1 presents the system model, where the area of interest of this chapter is highlighted. Section 4.2 presents the optimized switching level AMC scheme. In Section 4.3 the delay exponent and effective capacity of the optimized scheme and AMC scheme 1 are evaluated over block fading channels. AMC scheme 1 refers to the conventional AMC scheme with fixed switching levels used in Section 3.3. The performance of the optimized scheme is compared to AMC scheme 1 in terms of the delay exponent and effective capacity, in order to validate the use of the optimization technique in enhancing the service process. Lastly, a chapter summary is given in Section 4.4.

4.1 System model

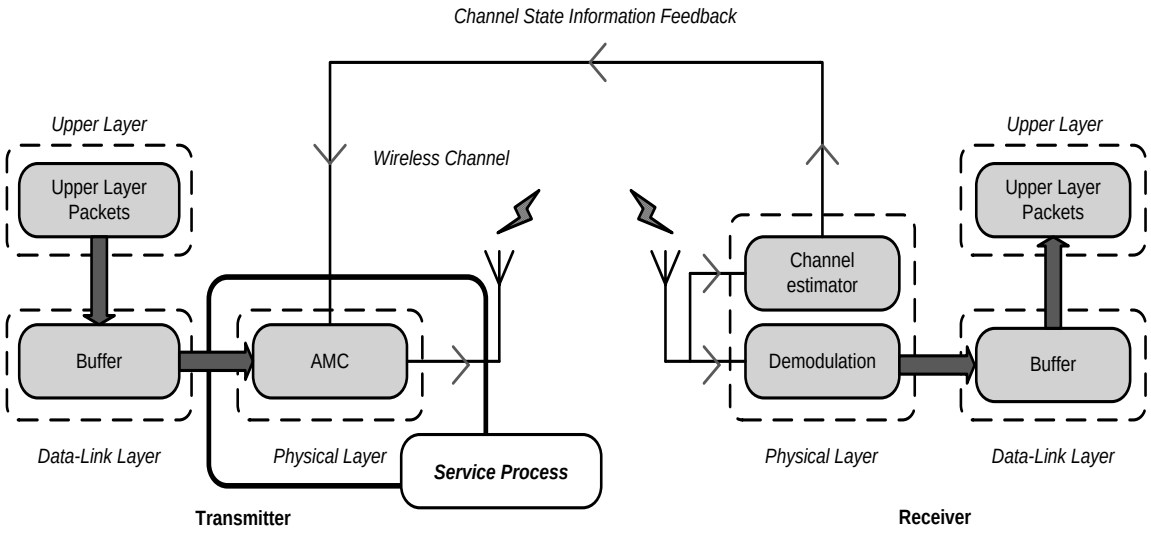

Figure 4-1 System model adapted from [26].

Many of the referenced cross-layer design methodologies [23-26] that used the effective capacity concept, proposed optimization techniques in order to enhance the underlying service process. The system model presented in Section 2.1 is still adopted, as seen in Figure 4-1, where the AMC scheme applied at the physical layer is the highlighted area of interest in this chapter. The optimization of the switching levels of the given AMC scheme, aims to improve the service process, which will affect the delay bound performance at the data link layer.

4.2 Optimum mode switching levels scheme

In [27], the Lagrangian optimization technique was used to derive the globally optimized modulation switching levels, while maintaining a target average BER. The concept of this

optimization technique was conceded through the realisation that the performance of adaptive modulation schemes, are predetermined by the value of the switching levels employed. In current literature such as [38] BER curves were used to locate the SNR values in which each modulation mode satisfied the target BER requirement. The resultant instantaneous BER constraint however, produced conservative switching levels. Thus the use of an average BER constraint was considered in [27], which insinuates that the each modulation mode can produce a variable instantaneous BER as long as the BER across all modes (average) meets the average BER requirement. As previously mentioned, the metric of interest in this dissertation is packets, and not bits, hence a PER constraint will be enforced. As defined in [39], the Lagrangian provides an algorithm for optimizing a function subject to constraints. Therefore the Lagrangian optimization method used in [27] is just a mathematical tool, and is not limited to the type of error rate constraint or underlying channel scenario. Thus in this section the scheme in [27] is adapted to optimize the switching levels of the given AMC scheme, by satisfying a target average PER, while maximising average throughput.

4.2.1 Average PER

The PER when using the *k*th AMC mode (*k* = 2, 3,...*K)* is given in Section 2.2 as

$$\mathrm{PER}_k(\gamma) = \begin{cases} 1, & if\ 0 < \gamma < \gamma_k \\ a_k exp(-g_k \gamma), & if\ \gamma \geq \gamma_k \end{cases} \tag{4-1}$$

where the parameters a_k, and g_k are mode dependent parameters that can be found in Table 2-1. The received SNR is given by γ, and its corresponding pdf $p_\gamma(\gamma)$ is defined in (2-1).

The average throughput *B* given in [27, Eq. (4)] is defined as,

$$B = \sum_{k=0}^{K-1} R_k \pi_k \tag{4-2}$$

where π_k denotes the probability that the SNR falls into mode k given in Section 2.2 (2-3), and R_k denotes the spectral efficiency of mode k.

Using (4-1), the mode specific average PER is determined by,

$$P_k = \int_{s_k}^{s_{k+1}} \mathrm{PER}_k p_\gamma(\gamma)\, d\gamma = \left(\frac{m}{w_k}\right)^m \frac{a_k}{\Gamma(m)} \left[\Gamma\left(m, \frac{w_k s_{k+1}}{\bar{\gamma}}\right) - \Gamma\left(m, \frac{w_k s_k}{\bar{\gamma}}\right)\right] \quad (4\text{-}3)$$

where $w_k = g_k \bar{\gamma} + m$, $\Gamma(\cdot)$ is the gamma function, and $\Gamma(\cdot,\cdot)$ is the incomplete gamma function. Given (4-2) and (4-3) the average PER, denoted by P_{avg}, is obtained by evaluating [modification, 20, Eq. (5)]

$$P_{avg} = \frac{1}{B} \sum_{k=0}^{K-1} R_k P_k \quad (4\text{-}4)$$

4.2.2 Formulating Optimum Switching Levels

Based on the Lagrangian method introduced in [27], the main goal is to maximize an objective function under a constraint, in order to optimize a set of switching levels, denoted as $\boldsymbol{s}$, for an applied AMC scheme which contains $k = 1, 2, .. K-1$ switching levels.

The first step in obtaining the optimized switching levels is to redefine the average throughput, B and average PER, P_{avg} in terms of $\boldsymbol{s}$. Due to the optimum switching levels being dependent on the average SNR $\bar{\gamma}$ [27], B and P_{avg} are also functions of $\bar{\gamma}$. Thus the average throughput B, given in (4-2) can be rewritten as,

$$B(\bar{\gamma}; \boldsymbol{s}) = \sum_{k=0}^{K-1} R_k \pi_k \quad (4\text{-}5)$$

and the average PER, P_{avg} in (4-4) can be redefined as,

$$P_{avg}(\bar{\gamma}; \boldsymbol{s}) = \frac{1}{B(\bar{\gamma}; \boldsymbol{s})} \sum_{k=0}^{K-1} R_k P_k \quad (4\text{-}6)$$

where $\boldsymbol{s}$ is a vector containing switching levels $k = 1, 2, .. K-1$ and $\bar{\gamma}$ is the average SNR.

Using the (4-5) and (4-6), the second step is to identify the objective function and constraint. Given that the optimized switching levels $\boldsymbol{s}$ are obtained by maximizing the *average throughput* under a constraint, let the objective function be denoted by $B(\bar{\gamma}; \boldsymbol{s})$ given in (4-5). The error rate constraint given in [27, Eq. (29)] can be defined in terms of packets, due to it being the unit of interest, and is hence expressed as,

$$P_{avg}(\bar{\gamma}; \boldsymbol{s}) = P_{th} \tag{4-7}$$

where P_{th} represents the target PER and $P_{avg}(\bar{\gamma}; \boldsymbol{s})$ is given by (4-6).

The constraint given in (4-7) can be rewritten in terms of the average throughput, $B(\bar{\gamma}; \boldsymbol{s})$ given in (4-5) as follows,

$$P_R(\bar{\gamma}; \boldsymbol{s}) = P_{th} B(\bar{\gamma}; \boldsymbol{s}) \tag{4-8}$$

where $P_R(\bar{\gamma}; \boldsymbol{s})$ is expressed as,

$$P_R(\bar{\gamma}; \boldsymbol{s}) = \sum_{k=0}^{K-1} R_k P_k \tag{4-9}$$

The final step in obtaining the optimized switching levels $\boldsymbol{s}$ is to use Lagrangian multiplier method to maximize the objective function $B(\bar{\gamma}; \boldsymbol{s})$ subject to the constraint given in (4-8). The *Lagrangian optimization problem formulation* is hence expressed as,

$$\max_{\boldsymbol{s}} \sum_{k=0}^{K-1} R_k \pi_k \tag{4-10}$$

$$\text{subject to} \quad \sum_{k=0}^{K-1} R_k P_k = P_{th} \sum_{k=0}^{K-1} R_k \pi_k \tag{4-11}$$

It must be highlighted that the Lagrangian multiplier method used in [27], is just a mathematical tool, which is used to find the optimal switching levels. Thus the PER constraint given in (4-8) can be used, given that the constraint unit does not factor into the overall optimization technique. Hence the relationship between the optimum switching levels can be computed as [27, Eq. (35)].

$$y_k(s_k) = y_1(s_1), \qquad for\ k = 2,3, \ldots . K-1 \tag{4-12}$$

where based on the PER constraint, $y(s_k)$ in [27, Eq. (36)] can be expressed as

$$y_k(s_k) \triangleq \frac{1}{v_k}\{R_k \mathrm{PER}_k(s_k) - R_{k-1}\mathrm{PER}_{k-1}(s_k)\} \tag{4-13}$$

where $v_k = R_k - R_{k-1}$.

The relationship between the optimum switching levels s_k, where $k = 2, 3, \ldots K-1$, and s_1 as seen in (4-12) and (4-13), is significant given the dependence on only AMC related parameters, such as b_k, v_k, and PER_k.

The AMC parameters, as given in Table 2-1, are substituted in (4-12) and (4-13) which results in the following:

$$y_1(s_1) = \mathrm{PER}_1(s_1) \tag{4-14}$$

$$y_2(s_2) = 2\mathrm{PER}_2(s_2) - \mathrm{PER}_1(s_2) \tag{4-15}$$

$$y_3(s_3) = 3\mathrm{PER}_3(s_3) - 2\mathrm{PER}_2(s_3) \tag{4-16}$$

$$y_4(s_4) = 2\mathrm{PER}_4(s_4) - \mathrm{PER}_3(s_4) \tag{4-17}$$

$$y_5(s_5) = 3\mathrm{PER}_5(s_5) - 2\mathrm{PER}_4(s_5) \tag{4-18}$$

4.2.3 Determining Optimum switching levels

It has been established that the optimum switching levels, s_2, s_3, s_4 and s_5, can be obtained based on s_1, using numerical methods to solve (4-12) and (4-13). Hence it is critical to determine the optimum value of s_1 given the P_{th}; the fading channel parameter m; and the average SNR, $\bar{\gamma}$. In order to obtain s_1 the constraint function in [27, Eq. (41)] must be evaluated, which is defined as,

$$Y\big(\bar{\gamma}; \boldsymbol{s}(s_1)\big) \triangleq P_R\big(\bar{\gamma}; \boldsymbol{s}(s_1)\big) - P_{th}B\big(\bar{\gamma}; \boldsymbol{s}(s_1)\big) \tag{4-19}$$

The optimum value of s_1 is obtained when $Y\big(\bar{\gamma}; \boldsymbol{s}(s_1)\big) = 0$, which follows the aim of maximising the average throughput while satisfying the constraint given in (4-8).

The impact of varying P_{th} and fading parameter m on the optimum switching level s_1, is depicted in Figure 4-2 and 4-3 respectively. Figure 4-2 depicts the constraint function $Y(s_1)$ for varying P_{th} values given, $\bar{\gamma} = 10dB$ and $m = 1$. It is seen that the solution to $Y(s_1) = 0$ unique and exists for all P_{th} values. As the PER target decreases, the optimum switching level, s_1, increases. To meet a stringent PER requirement, the channel can only support a

small amount of transmitted packets. Hence the transmission rate of the given AMC needs to be reduced, which is achieved by increasing the values of the switching levels.

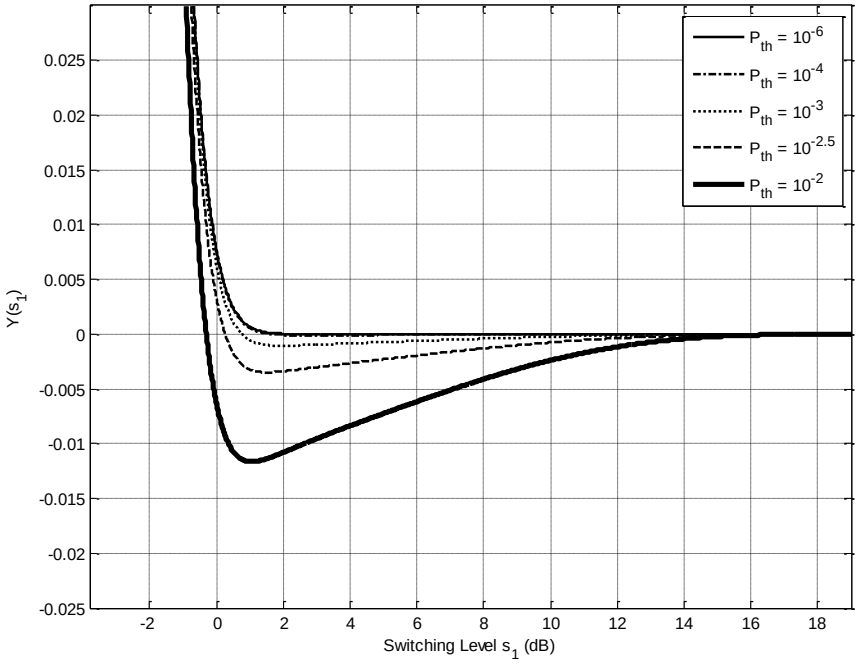

Figure 4-2 The optimum constraint function subject to varying target PER P_{th}, given $m = 1$, and $\overline{\gamma} = 10$dB

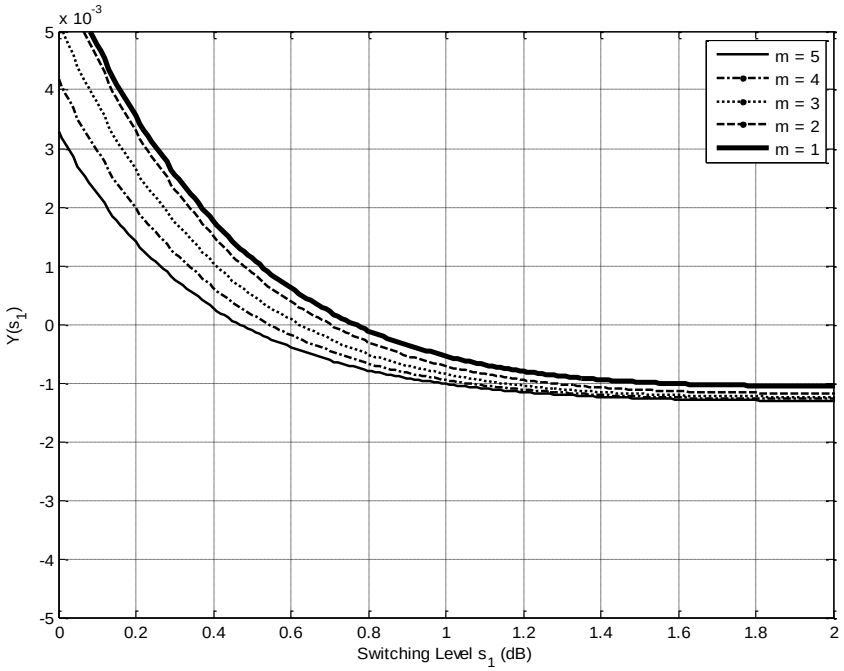

Figure 4-3 The optimum constraint function subject to varying *m*, given $P_{th} = 10^{-3}$ and $\overline{\gamma} = 10$dB

In Figure 4-3, given $P_{th} = 10^{-3}$ and $\bar{\gamma} = 10dB$, it is observed that as the fading parameter m increases, s_1 correspondingly decreases. This is expected given that as the channel quality gets better, transmission can occur at lower switching values.

The optimized switching levels obtained for varying $\bar{\gamma}$, are detailed in Table 4-1, where $P_{th} = 10^{-3}$ and $m = 1$. It is seen that an increase in the average SNR, or equivalently transmission power, results in the lowering of the optimized switching levels. All bounds obtained satisfy the constraint, $\gamma \geq \gamma_k$ stated in (4-1), where γ_k is the mode dependent parameter given in Table 2-1.

Table 4-1 Optimum Switching Levels (dB) for varying average SNR

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
	BPSK	QPSK	QPSK	16QAM	64QAM
$\bar{\gamma} = 10dB$	0.77	4.17	7.39	13.60	19.80
$\bar{\gamma} = 12dB$	0.74	4.13	7.36	13.57	19.77
$\bar{\gamma} = 15dB$	0.68	4.07	7.29	13.50	19.70

4.3 Performance analysis of Optimum AMC switching levels

Now that the optimized switching levels of the applied of AMC scheme have been determined, the focus will be on investigating its consequent effect on the service process performance. This will be done by evaluating the delay exponent, $\Lambda_C(\theta)$ and effective capacity, $E_C(\theta)$ of the optimized AMC scheme and comparing it to that of AMC scheme 1, with deterministic switching levels used in Section 3.3.2. For the remainder of this dissertation, let the optimized AMC scheme be denoted as AMC scheme 2. The delay exponent and effective capacity of both schemes are evaluated under different physical layer parameters over a block fading channel. The reference parameters set, as in Section 3.3.2, are: packet size $N_p = 1080$ bits; frame duration $T_f = 2$ms; fading parameter $m = 1$, average SNR $\bar{\gamma} = 10$dB, and PER target $P_{th} = 10^{-3}$. As discussed in Section 3.3, the delay

exponent results in [25], used the service process c_k, defined as $c_k = bR_k/R_1$, where b is defined as the bandwidth parameter. It was specified that $b = 2$, hence the service process reduced to $c_k = 4R_k$, given $R_1 = 0.5$. The delay exponent of AMC scheme 1 is a replication of the delay exponent result seen in [25], and serves as a source of comparison for the delay exponent of AMC scheme 2. Thus we define the service process c_k as $c_k = T_f W R_k = 4R_k$, as seen in Section 3.3. It follows that the spectral bandwidth W is set to $W = 2$KHz. Unless otherwise stated parameters are defaulted to their reference values.

4.3.1 The Delay Exponent

Given the optimized vector $\boldsymbol{s}$, switching levels $k = 1, 2, .. K-1$ can be used to evaluate the delay exponent $\Lambda_C(-\theta)$. Based on the GE limit of an uncorrelated service process defined in Section 3.2 (3-1), the delay exponent $\Lambda_C(-\theta)$ from Section 3.2 (3-6), is given by,

$$\Lambda_C(-\theta) = \log\left[\sum_{k=0}^{K-1} \pi_k \exp\left(-\theta N_p c_k\right)\right] \tag{4-20}$$

where the switching level at $k = 0$, which indicates no transmission, is simply 0dB. As seen in Section 3.3, the delay exponent of AMC scheme 1 was replicated from [25], where the delay exponent has the unit of 1/sec. Given that AMC scheme 1 is being used as a source of replication, (4-20), which has the unit 1/frame, is amended by dividing $\Lambda_C(-\theta)$ by T_f to obtain the unit 1/sec for the delay exponent. This is given, as in Section 3.3.1 (3-7) by,

$$\Lambda_C(-\theta) = \frac{1}{T_f}\log\left[\sum_{k=0}^{K-1} \pi_k \exp\left(-\theta N_p c_k\right)\right] \tag{4-21}$$

Figure 4-4 depicts the evaluated delay exponent (4-21), of AMC scheme 1 and 2, as a function of the QoS exponent, θ. Given that the delay exponent of AMC scheme 1 was replicated from [25], shown in Section 3.3.1, it thus serves as validated source of comparison in order to gauge the performance of AMC scheme 2. There are two limiting cases of the QoS exponent that can be analyzed in order to investigate the performance of both schemes in terms of the delay exponent. These are namely when a more stringent QoS requirement is required given by $\theta \to \infty$, and when a looser QoS requirement is required, in other words $\theta \to 0$. It is observed that as the QoS constraint becomes more stringent, AMC scheme 2 has lower, or more negative delay exponent than that of AMC scheme 1 for both average SNR scenarios. On the other extreme, if a looser QoS constraint is required, AMC scheme 1 has a

similar performance to that of AMC scheme 2. In Section 3.3.1 the direct relationship between the delay exponent and analytical delay bound violation probability was discussed, where as the delay exponent increased or decreased in value, correspondingly the probability of violating a given delay bound increased or decreased.

Based on the optimization technique for AMC scheme 2, a higher average throughput was achieved. This subsequently results in packets being serviced faster from the queuing system, reducing the queue length, and hence reducing the waiting period in the queue. Thus a smaller delay bound violation probability, and therefore a smaller or more negative delay exponent is produced by AMC scheme 2, as compared to AMC scheme 1. When the system can only tolerate small delay bound violations (stringent QoS requirement), the enhanced average throughput of AMC scheme 2 becomes more apparent given that it can support a higher arrival rate than that of AMC scheme 1. Thus AMC scheme 2 has more negative delay exponent for both average SNR scenarios. However when looser QoS requirements are required, both AMC schemes are adequate to support the given arrival process due to large delays being acceptable. The behaviour of varying the average SNR follows the same trends as in Section 3.3.1, where as the average SNR increases, the increasing transmission power results in a more negative delay exponent. This is observed in Figure 4-4 where $\bar{\gamma} = 15dB$ has a smaller delay exponent than $\bar{\gamma} = 10dB$.

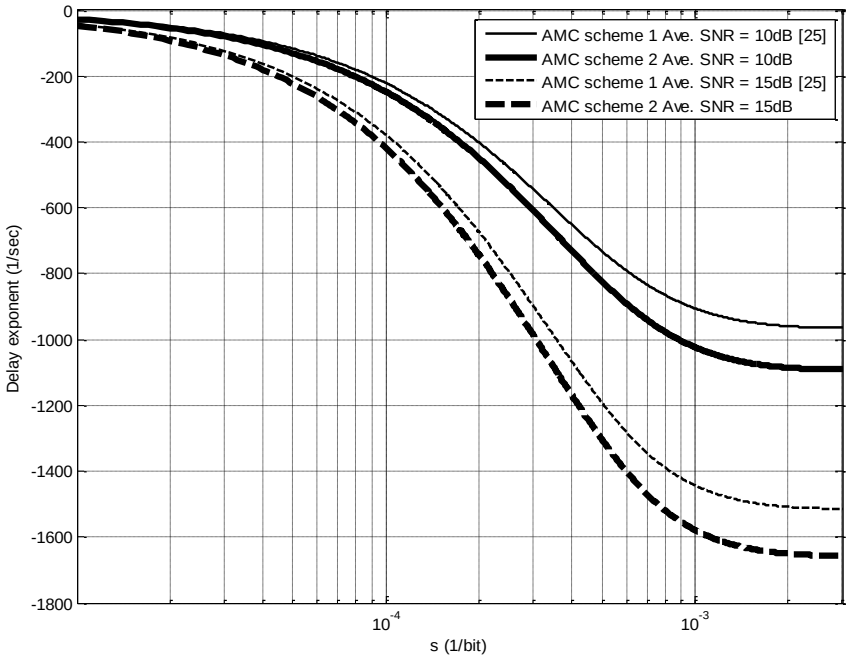

Figure 4-4 Delay Exponent of AMC scheme 1 and 2 varying average SNR $\bar{\gamma}$, given $P_{th} = 10^{-3}$ and $m = 1$

Note that the delay exponent function is evaluated in terms of bits by introducing the parameter N_p, which denotes the amount of bits in a packet. This is attributed to the delay exponent of AMC scheme 1 being replicated from [25], where the unit of bits was adhered to. Hence the conditions are applied to AMC scheme 2, for sake of comparison.

4.3.2 Effective Capacity

The effective capacity and delay exponent are related, as given in Section 3.2 (3-3),

$$\mathrm{E_C}(\theta) \triangleq -\frac{\Lambda_\mathrm{C}(-\theta)}{\theta} = -\frac{1}{\theta}\log\left[\sum_{k=0}^{K-1}\pi_k \exp\left(-\theta N_p c_k\right)\right] \tag{4-22}$$

Thus as with the evaluation of the delay exponent, the effective capacity, for the applied AMC scheme, can also be determined using the optimized switching level set $\boldsymbol{s}$. The performance of the effective capacity of AMC scheme 1 and 2, under different physical layers parameters, are shown in Figure 4-5, 4-6 and 4-7. Note that the effective capacity is normalized in order to ease comparison with spectral efficiency. This normalization is given by $\mathrm{E_C}(\theta)/T_f/W/N_p$ (packets/sec/Hz).

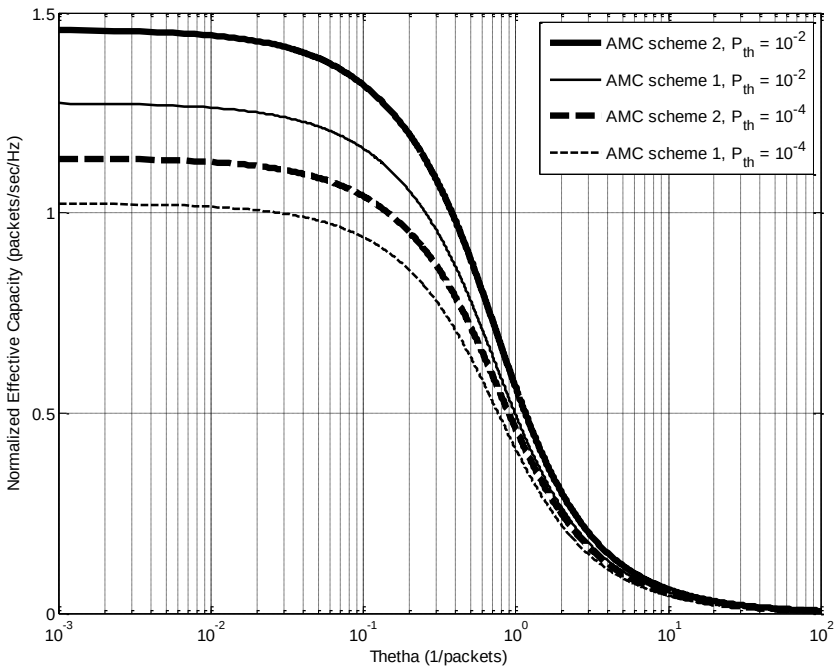

Figure 4-5 Effective Capacity of AMC scheme 1 and 2 varying target PER P_{th}, given $\bar{\gamma} = 10$dB, and $m = 1$

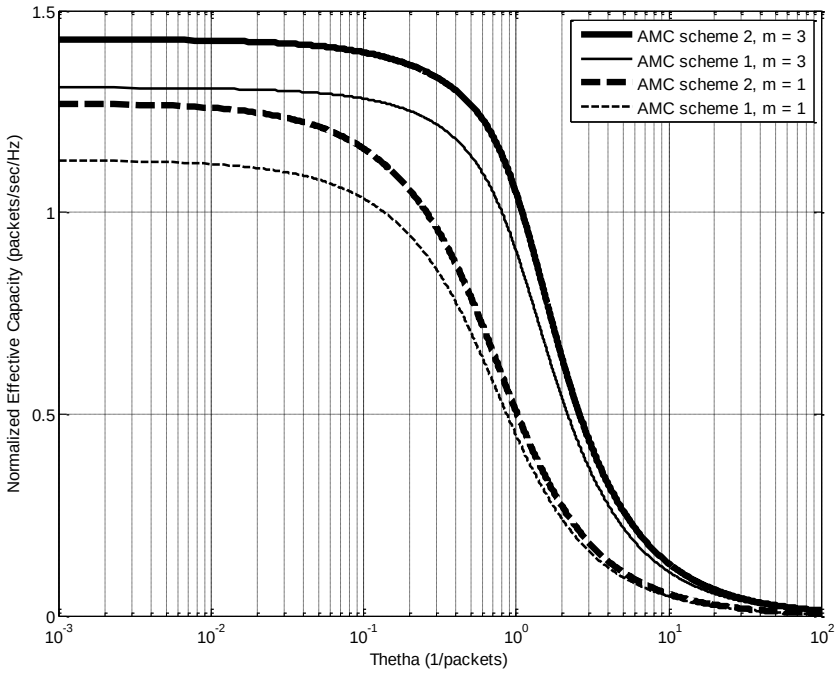

Figure 4-6 Effective Capacity of AMC scheme 1 and 2 varying fading parameter *m*, given $P_{th} = 10^{-3}$ and $\overline{\gamma} = 10$dB

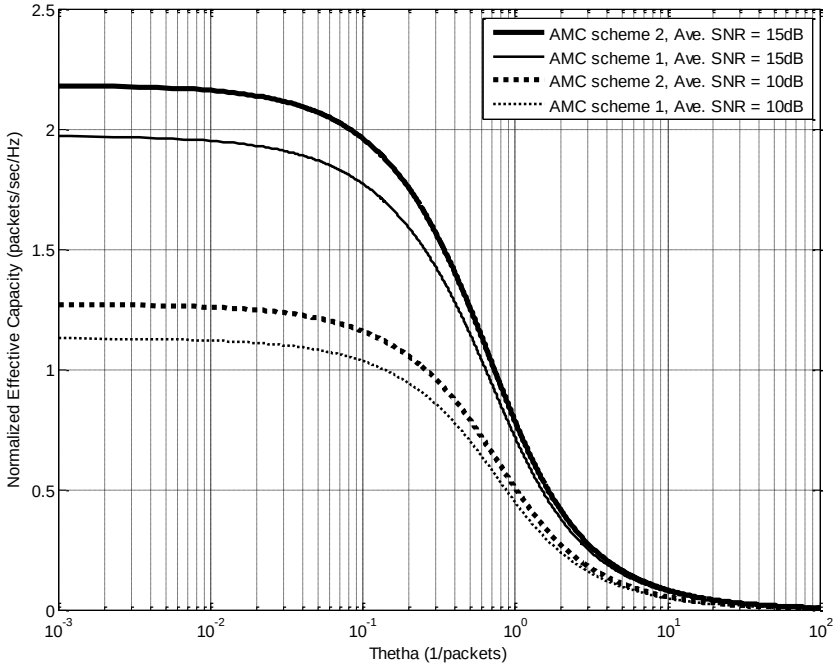

Figure 4-7 Effective Capacity of AMC scheme 1 and 2 varying average SNR $\overline{\gamma}$, given $P_{th} = 10^{-3}$ and $m = 1$

The behaviour of the effective capacity of AMC scheme 2 (optimized) subjected to varying PER constraint, fading parameter m, and average SNR, follows the same trend as the results of AMC scheme 1, discussed in Section 3.3.2. However it is observed that AMC scheme 2 increases the effective capacity in all varying, physical layer scenarios in respect to AMC scheme 1.

Table 4-2 Average service rates of AMC schemes 1 and 2

	PER constraint P_{th}		**Fading Parameter m**		**Ave. SNR $\bar{\gamma}$**	
	$P_{th} = 10^{-4}$	$P_{th} = 10^{-2}$	$m = 3$	$m = 1$	$\bar{\gamma} = 15dB$	$\bar{\gamma} = 10dB$
AMC Scheme 1 (packets/sec/Hz)	1.023	1.2735	1.3089	1.1287	1.9683	1.1288
AMC Scheme 2 (packets/sec/Hz)	1.1349	1.4565	1.4288	1.2691	2.118	1.2692
Gain	10.94%	14.37%	9.16%	12.42%	10.75%	12.44%

In particular, if we focus on the average service rate, the gains achieved by AMC scheme 2 in respect to AMC scheme 1, given the variations in the underlying physical layer structure, are shown in Table 4-2. The average performance gain of 11.68% by AMC scheme 2, is attributed to maximising average spectral efficiency under the PER constraint, (4-8). Hence a higher average spectral efficiency translates to a better effective capacity.

4.4 Chapter Summary

In this chapter, an optimized switching level AMC scheme, which maximized average throughput while maintaining a target average PER constraint, was presented. The optimized scheme, when compared to AMC scheme 1, in terms of the delay exponent and effective capacity, had a far superior performance. More specifically, AMC scheme 2's average service rate showed an average gain of 12.66%, 10.79% and 11.6%, given variations of the PER constraint, fading parameter m and average SNR, respectively. The aim of the next chapter is to investigate the cross-layer design approach and to analyse the delay-bound QoS performance of both AMC schemes.

Chapter 5

Delay Bound Violation Probability

Cross-layer design was proposed to overcome the challenges in providing QoS guarantees, particularly bounded delay. The development of the cross-layer design framework was based on the modelling of the arrival and service processes from a QoS point of view. The effective bandwidth and effective capacity functions, discussed in chapter 2 and 3 respectively, were able to model the arrival and service processes in terms of the QoS exponent θ. Thus using these functions as tools in the specified cross-layer design methodology, a relationship between the QoS delay bound and arrival and service processes can be established. The delay bound violation performances for audio and video traffic were presented in [35] for a MIMO system, which enforced the aforementioned cross-layer design technique. It was shown that AMC and MIMO diversity had a significant effect on the delay bound performance. The objective of [35] was to validate the relationship between the effective bandwidth and capacity functions in the cross-layer design framework by simulations, which was achieved. Optimized power and rate adaption policies derived in [25] were applied in the cross-layer design framework and it was shown that the proposed schemes produced a far superior QoS performance than that of conventional AMC schemes. The works of [35] and [25], hence validate the adopted cross-layer design technique, in providing support for statistical QoS guarantees. The integration of the given cross-layer design framework with the optimized switching bounds technique discussed in chapter 4, has yet to be investigated. Thus the resultant QoS delay bound violation performance has not been accounted for, which hence serves as motivation for this chapter.

The aim of this chapter is to analyze the QoS provisioning performance subject to variations in the arrival and service processes, in terms of delay bound violation. The optimized switching bounds of AMC scheme 2 will also be incorporated into the cross-layer design framework, which allows the investigation of the scheme's impact on the delay bound QoS performance. This will provide a further extension on AMC scheme 2's optimized switching level technique used to maximise average throughput while meeting an average PER constraint.

This chapter is outlined as follows: In Section 5.1, a system model is depicted which emphasizes the cross-layer design framework. Section 5.2 introduces a simple algorithm for

formulating the cross-layer design model and determining the delay bound violation probability. Simulated and analytical results of this framework is presented in Section 5.3, where delay bound QoS performance of AMC scheme 1 and 2 are evaluated subject to variations in both arrival and service processes. A chapter summary is provided in Section 5.4.

5.1 System Model

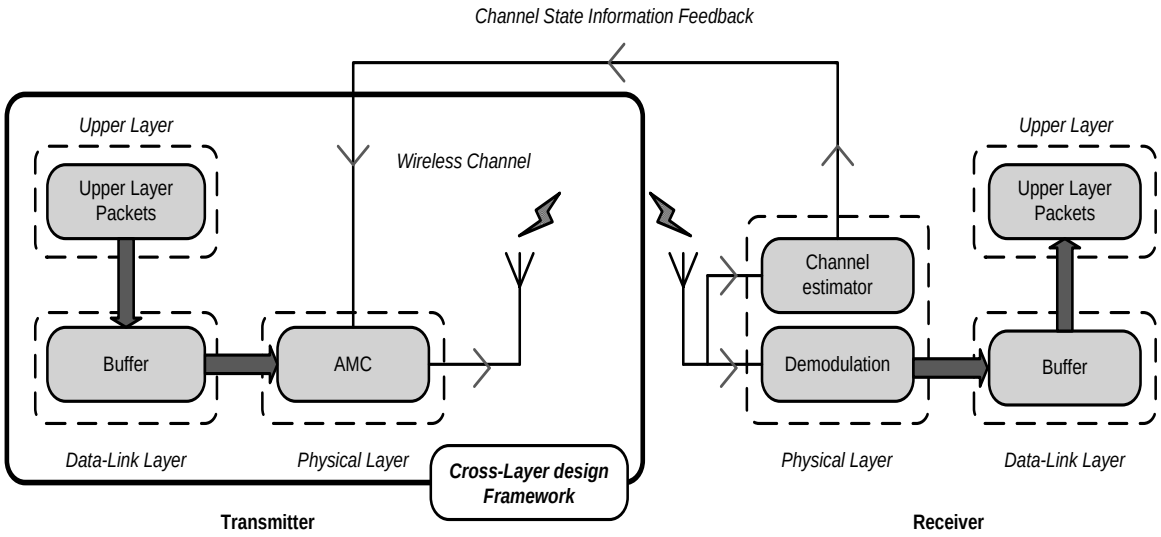

Figure 5-1 System model adapted from [26]

In chapter 2 and 3, the individual modelling of the arrival and service processes were discussed respectively. In particular, the system model used in Section 2.1 and seen in Figure 5-1, detailed the streaming of upper layer packets into the buffer at the data link layer (arrival process), and AMC variations at the physical layer (service process). Now that the functional blocks of the cross-layer design framework have been analyzed, the focus shifts to investigating the effects the service process has on the subsequent data link layer QoS delay bound performance. This specific focal point is highlighted in Figure 5-1 by the white block. The tools that can be used to achieve this, as discussed in chapter 2, are the effective bandwidth and effective capacity functions, which characterise the arrival and service process models in terms of QoS.

5.2 Cross-Layer Design and Delay Bound QoS guarantees

The effective capacity and effective bandwidth functions are linked through the common association with the QoS exponent. Exploiting this relationship, a unique θ^* can be obtained,

which represents the maximum arrival rate that the channel can support. Given the required QoS exponent θ^*, the analytical delay bound violation probability, given in Section 2.5 (2-15) is expressed as,

$$\mathrm{P_r}\{D > D_{max}\} \approx \varepsilon \exp(-\delta\theta^* D_{max}) \tag{5-1}$$

where D is the packet delay, D_{max} denotes the user set delay bound requirement, ε is defined as the probability that the buffer is not empty, and δ is a parameter that is jointly determined by the effective capacity and effective bandwidth functions. The buffer non-empty probability or otherwise known as service efficiency ε, can be approximated as given in Section 2.5 (2-14),

$$\varepsilon = \frac{\mu_A}{\mu_C} = \frac{\lim\limits_{\theta \to 0} \mathrm{E_B}(\theta)}{\lim\limits_{\theta \to 0} \mathrm{E_C}(\theta)} \tag{5-2}$$

where μ_A and μ_C are the average arrival and service rates.

Figure 5-2 shows the relationship between the effective capacity and effective bandwidth functions, where the delay bound violation probability given in (5-1) can be easily calculated.

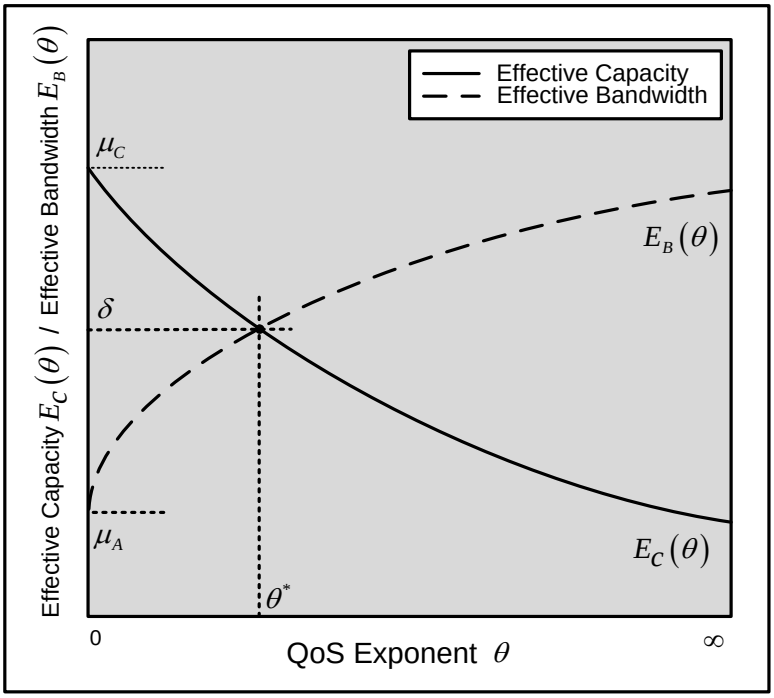

Figure 5-2 Effective Capacity and Effective Bandwidth functions

It can be seen that the effective bandwidth and effective capacity curves intersect at a unique QoS exponent value, namely θ^*, which correspondingly results in the parameter δ being obtained. A simple algorithm that can be used in determining the analytical delay bound violation probability is given as follows:

Step 1: Model the given arrival and service processes using the effective bandwidth and effective capacity theories respectively. Once $E_B(\theta)$ and $E_C(\theta)$ functions have been obtained, numerical search for the unique $\{\theta^*, \delta\}$ pair such that,

$$E_B(\theta^*) = E_C(\theta^*) = \delta \tag{5-3}$$

Step 2: Using the average arrival rate μ_A, and the average service rate μ_C, solve for the probability of a non empty buffer, ε, given by (5-2).

Step 3: The triplet $\{\theta^*, \delta, \varepsilon\}$ obtained in Step 1 and 2, can now be used to determine the delay violation probability expressed in (5-1), given the user set delay bound requirement D_{max}

5.3 Delay bound violation performance

From the cross-layer design algorithm, it can be seen that the delay bound violation probability is significantly impacted by the effective capacity function. Thus the effective capacity of the optimized AMC scheme 2 can further be extended by evaluating its consequent effect on the delay bound violation performance. It is hence important that delay bound violation probability in (5-1) is verified. This will ensure the correctness of optimized AMC scheme 2, which influences the effective capacity.

The approach taken to validate the analytical expression seen in (5-1) is to simulate a queuing system for a block fading channel. The source traffic is modelled as an ON-OFF Markov modulated Poisson process, and the AMC scheme applied at the physical layer is adopted as the service process. Based on the block fading assumption, the service rate will change on a frame by frame basis, due to the channel variation. Thus, there is no distinct distribution of service times for the overall simulation period. However for a frame time duration T_f, when the channel is assumed to be invariant, the service rate based on the given AMC transmission mode, is fixed. Hence for the period T_f, the simulated queuing model follows an approximation of an M/M/1 queuing model. The simulations will be run for both AMC scheme 1 and 2, as service processes, where AMC scheme 1 will serve as a validated source of comparison.

The reference parameters used in simulations are set as follows: packet size $N_p = 1080$ bits; frame duration $T_f = 2$ms; fading parameter $m = 1$, average SNR $\bar{\gamma} = 10$dB, PER target $P_{th} = 10^{-3}$, source parameters $q_{11} = 0.99$, $q_{22} = 0.9$ [25], mean source rate $v = 20$ packets/frame, and user set delay bound requirement $Dmax = 30$ frames. A block faded channel is assumed. As mentioned in Section 3.3 and 4.3, the delay exponent results in [25], used the service process c_k, which reduced to $c_k = 4R_k$. Given that [25] serves as source of validation for AMC scheme 1 we define the service process c_k as $c_k = T_f W R_k = 4R_k$. It follows that the spectral bandwidth W is set to $W = 2$KHz. Unless otherwise stated in the figure legend, parameters are defaulted to their reference values. The delay bound violation probability for varying physical layer parameters; PER target P_{th}; fading parameter m; and average SNR are shown in Figure 5-3, 5-4 and 5-5 respectively. Figure 5-6 and 5-7 depict varying source parameters namely, mean source rate v, and mean T_{ON} period correspondingly. It must be highlighted that both analytical and simulated results are depicted in the figures. The analytical results are based on (5-1), where the effective bandwidth and effective capacity functions that model the arrival and service processes of the queuing model respectively, are used to obtain the statistical delay bound performance. The effective bandwidth is used to model the ON-OFF source traffic, which was discussed in Section 2.4, where as the effective capacity characterises the respective AMC scheme over block fading channels, as shown in Section 3.2.

It is observed that the analytical and simulated results closely agree for both AMC scheme 1 and 2. It is also shown that the optimized AMC scheme 2 performs better than the conventional AMC scheme 1 for all variations of underlying parameters in both arrival and service processes. As previously discussed in Section 4.3.2, the optimized switching levels and maximized average throughput of AMC scheme 2 consequently results in an improved effective capacity, in comparison to the deterministic switching levels of AMC scheme 1. The enhanced effective capacity of AMC scheme 2 can be directly translated to packets being serviced faster, due to the better average service rate. This results in AMC scheme 2 having a lower delay bound violation probability, when compared to AMC scheme 1. From an analytical point of view, if a larger effective capacity is achieved, consequently the effective bandwidth and effective capacity functions will intersect at a larger or more stringent QoS exponent. From (5-1), for a stationary arrival process, a larger QoS exponent will result in a smaller delay bound violation probability, which is obtained by AMC scheme 2, as compared to AMC scheme 1, for fixed delay bound.

5.3.1 QoS delay bound performance with varying physical layer parameters

In Figure 5-3, it can be seen that as P_{th} decreases (becomes more stringent); the delay bound violation probability increases, which is expected. In order for the more stringent QoS requirement of PER ($P_{th} = 10^{-5}$) to be guaranteed, less packets are transmitted. Hence the mean service rate adaptively decreases, thus leading to an increase in packets waiting in the queue. From Little's Law [32] there exists a direct relationship between the queue size, and queuing delay, where as the queuing size increases, the queuing delay correspondingly increases. Thus the lengthening of the queue size translates to a larger delay bound violation probability.

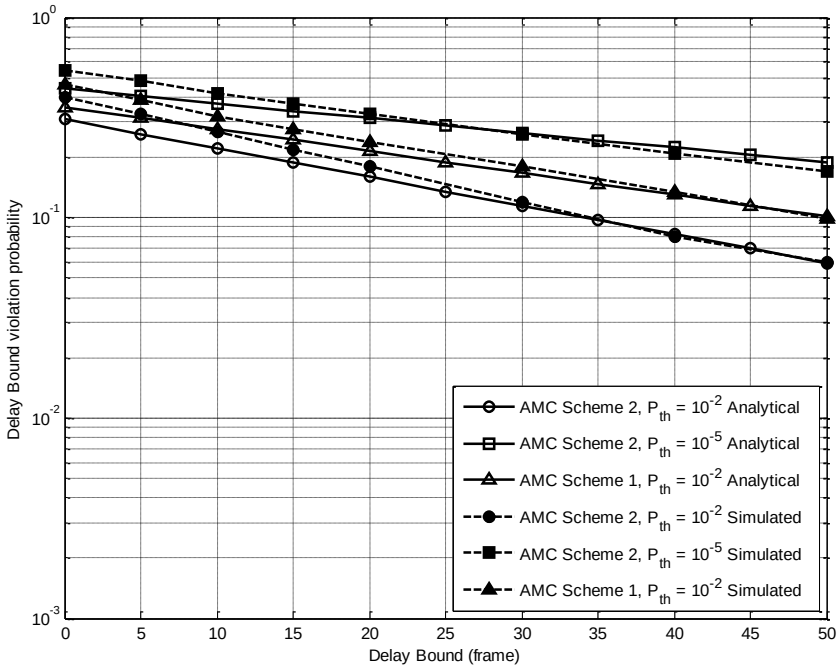

Figure 5-3 Delay bound violation probability varying target PER P_{th}, given $\overline{\gamma} = 10$dB and $m = 1$

It is shown in Figure 5-4 that as fading parameter m decreases in value, the service efficiency ε or $\Pr(D > 0)$ and delay bound violation probability increase. The value of the fading parameter m, indicates the channel quality. Hence if the channel quality is good (m is large), more packets will be transmitted, since there will be less transmission errors. However if the channel quality is bad (m is small), the service rate will be adaptively lowered to accommodate the increased error rates, hence reducing the amount of transmitted packets. This will result in a larger service efficiency, which consequently leads to an

accumulation of packets in the storage buffer. The lengthening of the queuing size hence produces a higher delay bound violation probability. Note that a larger service efficiency ε, is defined as the average service rate μ_C, converging towards the average arrival rate μ_A, which is based on the definition of ε seen in (5-2). By increasing the service efficiency, the server will end up continuously servicing packets without a chance of a free moment, and hence will lead to an increase in queuing delay. The extreme case of μ_A being greater than μ_C, will result in the server no longer being able to sustain the arrival traffic, and will hence lead to a build up in the queue.

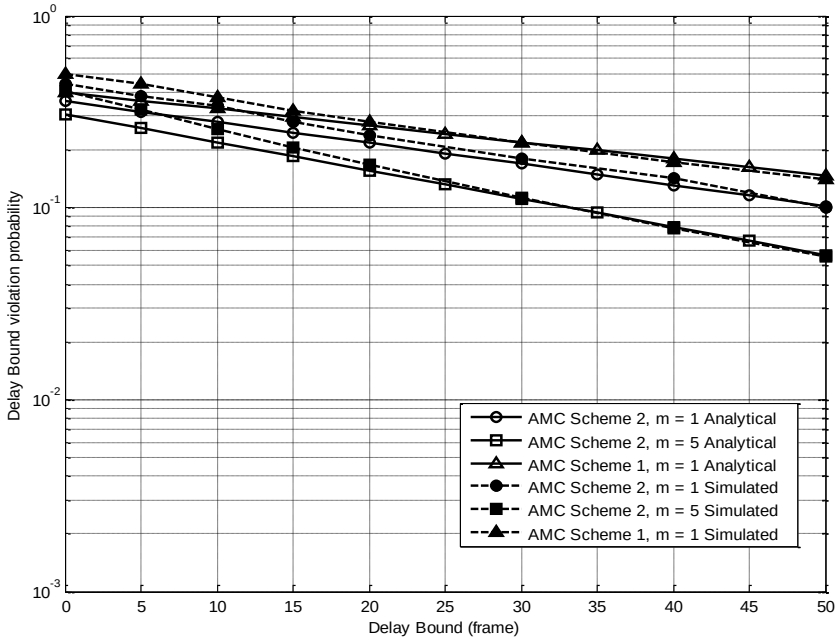

Figure 5-4 Delay bound violation probability varying fading parameter *m*, given $P_{th} = 10^{-3}$ and $\bar{\gamma} = 10$dB

Figure 5-5 shows that as the average SNR, $\bar{\gamma}$, increases, the delay bound violation probability decreases. An improved $\bar{\gamma}$ (improved transmission power), results in an increased average service rate, hence a decrease in service efficiency. This constitutes to packets being serviced faster, hence lowering the overall queuing delay, and correspondingly the delay bound violation probability. It can also be seen that as the delay bound increases, the delay bound violation probability decreases.

In Figure 5-3, 5-4, 5-5, the arrival process parameters were fixed, while the service process changed depending on the variation in the physical layer parameters. It was observed that a lower service efficiency ε, given that μ_C was substantially larger than μ_A, resulted in a lower

delay bound violation. This notion validates the motivation of threshold optimization of the applied AMC, as given in chapter 4. By improving the system's average throughput, AMC scheme 2 enabled packets to be serviced faster than AMC scheme 1, resulting in a lower delay bound violation, for a set delay bound.

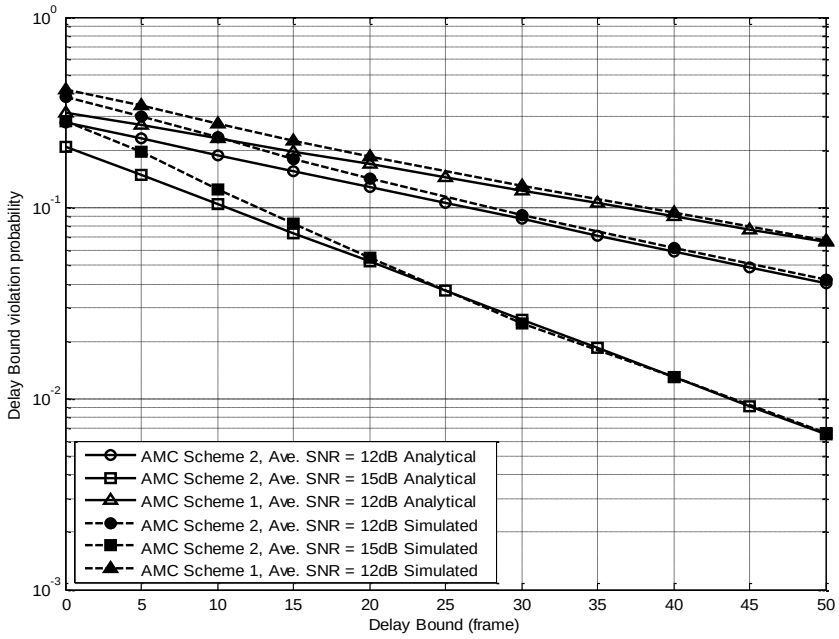

Figure 5-5 Delay bound violation probability varying average SNR $\overline{\gamma}$, given $P_{th} = 10^{-3}$, and $m = 1$

5.3.2 QoS delay bound performance with varying source traffic parameters

In Figure 5-6 and Figure 5-7, the arrival process parameters were varied, while the service process parameters were set. The effect of arrival process variations on the QoS delay bound performance was investigated, since it was interesting to analyse the data link layer element of the cross-layer design framework.

Figure 5-6 shows the delay bound violation probability for different traffic loads, namely setting the mean source rate to $v = 20$ packets/frame and $v = 25$ packets/frame, given $\bar{\gamma} = 15dB$. Based on Little's Law, if a significant amount of packets are injected into the buffer, the queue size increases and consequently increases the queuing delay. Thus compared to the reduced mean arrival rate $v = 20$ packets/frame, $v = 25$ packets/frame produced a higher delay bound violation probability for a set delay bound, as seen in Figure 5-6, given that the higher arrival traffic rate resulted in a larger service efficiency (μ_C is tends

towards μ_A). As aforementioned, AMC scheme 2's better effective capacity results in a lower delay bound probability, when compared to AMC scheme 1.

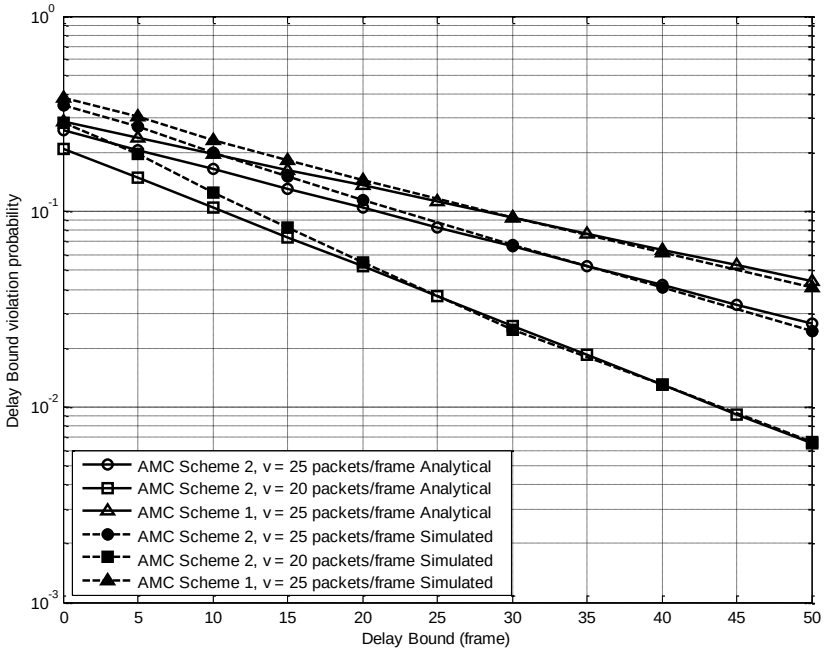

Figure 5-6 Delay bound violation probabilities varying mean source rate, given average SNR $\bar{\gamma} = 15$dB, $P_{th} = 10^{-3}$ and $m = 1$

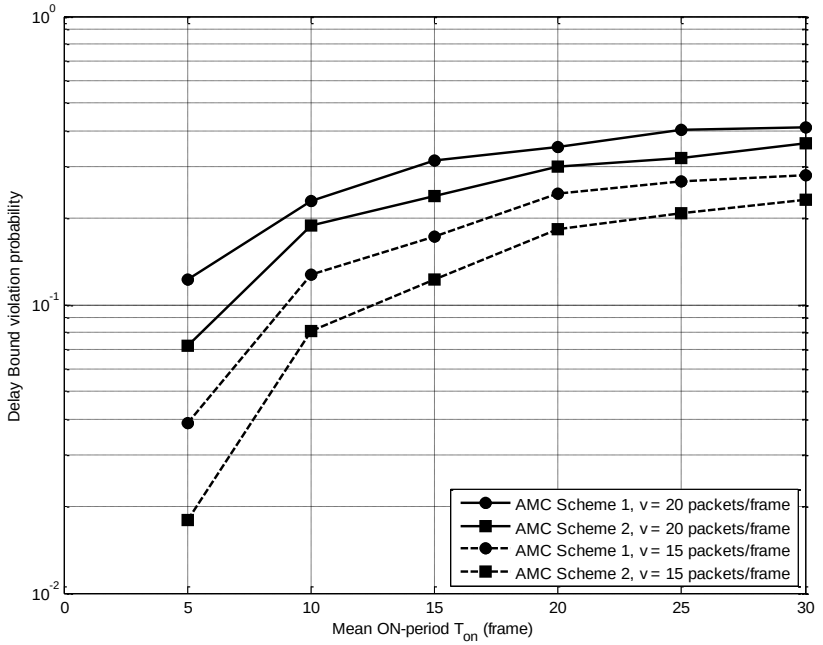

Figure 5-7 Delay bound violation probabilities varying mean ON period, given $P_{th} = 10^{-3}$, $\bar{\gamma} = 10$dB, and $m = 1$

Figure 5-7 demonstrates the influence of source burstiness on the delay bound violation performance of AMC scheme 1 and 2 under different traffic loads. The mean ON period ($T_{ON} = 1/(1 - q_{11})$), is varied while maintaining the fixed ratio of T_{ON}/T_{OFF}. The variation of T_{ON} reflects on the burstiness of the ON-OFF traffic source. The results seen in Figure 5-7 can be explained as follows. Let the time period in which packets are in the queuing system be denoted as an observation window. In the ON-OFF model, packets only arrive in the ON-period. The start of the ON period is marked as the start of the observation window. Packet servicing occurs during the ON-period and stops only when the buffer is empty in the OFF period. At this point the observation window will end. If given a small observation window, say $T_{on} = 10$ms, there exists a reduced time period for packets to build up in the queue. Hence, based on Little's Law, a small queue size will result in a small queuing delay that will in effect produce a small delay bound violation probability. Conversely, when a large observation window, for instance $T_{on} = 50$ms, is considered, the sizeable time period allows the queuing size to increase substantially, resulting in large queuing delays, and hence creating a large delay bound violation probability. This behaviour is clearly illustrated in Figure 5-7, where as T_{ON} increases, correspondingly, the delay bound violation probability increases. It is observed that the optimized AMC scheme 2 outperforms AMC scheme 1 in terms of providing a smaller delay bound violation probability, for a given delay bound requirement.

Table 5-1 Average service rates of AMC schemes 1 and 2

	$v = 15$ packets/frame	$v = 20$ packets/frame
AMC Scheme 1 $P_r\{D > D_{max}\}$	0.1269	0.2288
AMC Scheme 2 $P_r\{D > D_{max}\}$	0.0811	0.1891
Gain	56.47%	20.99%

Table 5-1 outlines the specific performances of both schemes for a T_{ON} of 10 frames, in terms of delay bound violation probability for fixed delay bound $D_{max} = 30$ frames. In Table 5-1, AMC scheme 2 showed a lower delay bound violation performance by 56.47% and 20.99% over AMC scheme 1, given mean arrival rates $v = 15$ packets/frame and

$v = 20$ packets/frame respectively. AMC scheme 2's smaller delay bound violation results is attributed to its optimized switching levels, and maximised average throughput. This consequently led to an enhanced effective capacity and correspondingly a lower service efficiency. Packets hence exited the queuing system faster, and thus produced the lower delay bound violation probability.

5.4 Chapter Summary

A cross-layer design approach for providing delay bound QoS guarantees was presented in this chapter. Simulations of the cross-layer system model demonstrated that as the channel quality and transmission power improved, and target PER requirement loosened, the service efficiency decreased which consequently decreased the delay bound violation probability. It was also shown that simulated and analytical results validated the enhanced performance of AMC scheme 2 when compared to AMC scheme 1. Its superior average service rate ensured reduced service efficiency (μ_C is far greater than μ_A), thus giving AMC scheme 2 its lower delay bound violation probability subject to a set delay bound. The effects of varying the source traffic parameters showed that an increased average arrival rate, increased the service efficiency (μ_C is tends towards μ_A), which hereby resulted in an increased the delay bound violation. However given the modifications in the arrival process, in particular the variation of the mean ON period, AMC scheme 2 still had a lower delay bound violation performance by 56.47% and 20.99% for corresponding mean arrival rates $v = 15$ and $v = 20$ packets/frame, given a $T_{on} = 10$ frames and $D_{max} = 30$ frames.

Thus subject to variations in both arrival and service processes, AMC schemes 2's optimized switching levels, and consequently better effective capacity hence resulted in a superior QoS delay bound performance in comparison to AMC scheme 1. The implication of this is a more stringent QoS delay bound can be guaranteed by AMC scheme 2. This result hereby further validates and motivates the optimization of the switching bounds of the applied AMC scheme at the physical layer by maximising the system's average throughput while maintaining an average PER target constraint.

Chapter 6

Conclusion

The support of QoS guarantees is critical for the transmission of real time multimedia traffic over wireless networks. The challenge of providing QoS guarantees over fading wireless channels were discussed, where the layered structure was introduced as a possible solution. However this approach was mainly suited for wired networks, whereas wireless networks opposed the stringent layered architecture given its limited resources and unstable links. Thus the novel approach, namely cross-layer design, was introduced which allowed information to be shared among different protocol layers, in order to enhance performance.

A cross-layer design system model was presented, that integrated the functionality of the data link layer and physical layer. The arrival and service processes, namely the bursty source traffic and employed AMC scheme, were characterised using ON-OFF modelling and FSMC. Based on these models, the theory of effective bandwidth and effective capacity were used to represent the statistical behaviour of the arrival and service processes respectively.

The analytical delay bound violation expression presented, showed that the effective capacity had a significant impact on the delay bound violation performance. Thus service process modelling and its subsequent effective capacity was investigated for block fading and correlated fading channels, subject to physical layer parameter variations. It observed a more stringent QoS guarantee was provided when the following separate cases occurred: PER constraint loosened, average SNR increased and fading parameter m increased. The converse cases occurred when a loosened QoS requirement was supported: PER constraint tightened, average SNR decreased and fading parameter m decreased. Hence it was concluded that modifications of physical layer parameters had a considerable effect on the service process and consequently the effective capacity in both block fading and correlated fading channels.

Many of the referenced effective capacity approaches used optimization techniques on the service process in order to enhance the delay bound performance. A set of optimized switching levels was introduced for a constant power AMC scheme which maximized the average throughput given an average PER constraint. Based on FSMC modelling the delay exponent and effective capacity of optimized AMC scheme 2 was obtained and compared to

that of AMC scheme 1, which had deterministic switching levels. It was observed that AMC scheme 2 had a far superior performance than that of AMC scheme 1. In particular AMC scheme 2's average service rate showed an average gain of 12.66%, 10.79% and 11.6%, given variations of the PER constraint, fading parameter *m* and average SNR, respectively.

The optimized technique used by AMC scheme 2 was further extended by investigating its effect on the delay bound violation performance. The simulated results of an M/M/1 FIFO queuing model, validated by the analytical delay bound violation expression, showed that AMC scheme 2 had a lower delay bound violation probability than AMC scheme 1, subject to a set delay bound requirement. Variations in both arrival and service processes were applied, and it was observed that the superior effective capacity of AMC scheme 2 consistently produced a lower delay bound violation when compared to AMC scheme 1. This hereby confirms that optimization techniques on service processing can support more stringent QoS delay bound requirement.

Appendix

A.1 Proof of (2-10)

For the ON-OFF model, where state 1 denotes the OFF state, and state 2 denotes the ON state, the diagonal matrix $\boldsymbol{\Psi}(\theta)$, is given by (2-8) as,

$$\boldsymbol{\Psi}(\theta) = \mathrm{diag}[\Psi_1(\theta), \Psi_2(\theta)] \tag{A-1}$$

with

$$\Psi_i(\theta) = \mathrm{E}\big(\exp(a_i(t)\theta)\big) \qquad , \text{where } i = 1, 2 \tag{A-2}$$

where $\Psi_i(\theta)$ is the moment generating function of the arrival process $a_i(t)$ at time t.

Given that the source traffic is considered as an ON-OFF Markov modulation Poisson Process, let X_t denote the number packets arriving into queue, in the interval $[0, t]$ with an average rate of v during the ON state. Thus $\Psi_2(\theta)$ can be determined by,

$$\Psi_2(\theta) = \mathrm{E}\left(\exp\big(X_t N_p \theta\big)\right) \tag{A-3}$$

where (A-3) is expressed in bits by introducing the parameter N_p, which denotes the amount of bits in a packet. Given that X_t is a Poisson process, it follows,

$$\begin{aligned}
\Psi_2(\theta) = \quad & \mathrm{E}\left(\exp\left(\big(N_p\theta\big)X_t\right)\right) \\
= \quad & \sum_{j=0}^{\infty} \exp\big(\theta N_p j\big)\left(\frac{\exp(-v)v^j}{j!}\right) \\
= \quad & \exp(-v)\sum_{j=0}^{\infty}\left(\frac{\big(\exp\big(\theta N_p\big)v\big)^j}{j!}\right) \\
= \quad & \exp(-v)\exp\big(\exp\big(\theta N_p\big)v\big) \\
= \quad & \exp\left(v\big(\exp\big(\theta N_p\big) - 1\big)\right)
\end{aligned} \tag{A-4}$$

During the OFF state, no packets arrive into the queue, hence,

$$\Psi_2(\theta) = \mathrm{E}\big(\exp(0)\big)$$

$$= 1 \tag{A-5}$$

Thus substituting (A-4) and (A-5) into (A-1) we have,

$$\boldsymbol{\Psi}(\theta) = \begin{bmatrix} 1 & 0 \\ 0 & \exp\Big(\mathrm{v}\big(\exp\big(\theta \mathrm{N_p}\big) - 1\big)\Big) \end{bmatrix} \tag{A-6}$$

A.2 Proof of (2-11)

The effective bandwidth close form expression as given in (2-6) is detailed as,

$$\mathrm{E_B}(\theta) = \frac{1}{\theta}\log[\mathrm{sp}\{\mathbf{P}\boldsymbol{\Psi}(\theta)\}] \tag{A-7}$$

By substituting the probability matrix **P**, given by (2-7) and diagonal matrix $\boldsymbol{\Psi}(\theta)$ containing the MGFs of $\mathrm{a_i}(\mathrm{t})$detailed in (2-10), the result is,

$$\mathrm{E_B}(\theta) = \frac{1}{\theta}\log\left[\mathrm{sp}\left\{\begin{bmatrix} \mathrm{q_{11}} & \mathrm{q_{12}} \\ \mathrm{q_{21}} & \mathrm{q_{22}} \end{bmatrix}\begin{bmatrix} \Psi_1(\theta) & 0 \\ 0 & \Psi_2(\theta) \end{bmatrix}\right\}\right]$$

$$= \frac{1}{\theta}\log\left[\mathrm{sp}\left\{\begin{bmatrix} \mathrm{q_{11}}\Psi_1(\theta) & \mathrm{q_{12}}\Psi_2(\theta) \\ \mathrm{q_{21}}\Psi_1(\theta) & \mathrm{q_{22}}\Psi_2(\theta) \end{bmatrix}\right\}\right] \tag{A-8}$$

For the ease of calculations, the general form of $\boldsymbol{\Psi}(\theta)$ is used as opposed to using (2-10). The spectral radius sp{·} seen in (A-8) is evaluated by computing the magnitude of the largest eigenvalue assiociated with the resultant matrix of $\mathbf{P}\boldsymbol{\Psi}(\theta)$. We first solve $det(\mathbf{P}\boldsymbol{\Psi}(\theta) - \lambda\mathbf{I}) = 0$, where $det(\cdot)$ is defined as the determinant, **I** is the identity matrix, and λ represents the possible set of eigenvalues. Evaluating $det(\mathbf{P}\boldsymbol{\Psi}(\theta) - \lambda\mathbf{I})$ we obtain,

$$det(\mathbf{P}\boldsymbol{\Psi}(\theta) - \lambda\mathbf{I}) = det\left(\begin{bmatrix} \mathrm{q_{11}}\Psi_1(\theta) - \lambda & \mathrm{q_{12}}\Psi_2(\theta) \\ \mathrm{q_{21}}\Psi_1(\theta) & \mathrm{q_{22}}\Psi_2(\theta) - \lambda \end{bmatrix}\right)$$

$$= (\mathrm{q_{11}}\Psi_1(\theta) - \lambda)(\mathrm{q_{22}}\Psi_2(\theta) - \lambda) - \big(\mathrm{q_{21}}\Psi_1(\theta)\big)\big(\mathrm{q_{12}}\Psi_2(\theta)\big)$$

$$
\begin{aligned}
&= q_{11}\Psi_1(\theta)q_{22}\Psi_2(\theta) - \lambda q_{22}\Psi_2(\theta) - \lambda q_{11}\Psi_1(\theta) - \lambda^2 \\
&\quad -q_{21}\Psi_1(\theta)q_{12}\Psi_2(\theta) \\
&= \lambda^2 - \lambda\big(q_{11}\Psi_1(\theta) + q_{22}\Psi_2(\theta)\big) + \\
&\quad \big(q_{11}\Psi_1(\theta)q_{22}\Psi_2(\theta) - q_{21}\Psi_1(\theta)q_{12}\Psi_2(\theta)\big)
\end{aligned} \tag{A-9}
$$

Substituting $q_{12} = 1 - q_{11}$, and $q_{21} = 1 - q_{22}$ into (A-9) we have,

$$
\begin{aligned}
det(\mathbf{P\Psi}(\theta) - \lambda\mathbf{I}) = &\ \lambda^2 - \lambda\big(q_{11}\Psi_1(\theta) + q_{22}\Psi_2(\theta)\big) + q_{11}\Psi_1(\theta)q_{22}\Psi_2(\theta) - \\
&\ (1 - q_{11})(1 - q_{22})\Psi_1(\theta)\Psi_2(\theta) \\
= &\ \lambda^2 - \lambda\big(q_{11}\Psi_1(\theta) + q_{22}\Psi_2(\theta)\big) + q_{11}\Psi_1(\theta)q_{22}\Psi_2(\theta) - \\
&\ (1 - q_{11} - q_{22} + q_{11}q_{22})\Psi_1(\theta)\Psi_2(\theta) \\
&\ \lambda^2 - \lambda\big(q_{11}\Psi_1(\theta) + q_{22}\Psi_2(\theta)\big) + q_{11}\Psi_1(\theta)q_{22}\Psi_2(\theta) - \\
&\ q_{11}q_{22}\Psi_1(\theta)\Psi_2(\theta) + (q_{11} + q_{22} - 1)\Psi_1(\theta)\Psi_2(\theta) \\
&\ \lambda^2 - \lambda\big(q_{11}\Psi_1(\theta) + q_{22}\Psi_2(\theta)\big) + \\
&\ (q_{11} + q_{22} - 1)\Psi_1(\theta)\Psi_2(\theta)
\end{aligned} \tag{A-10}
$$

From (A-10), based on $det(\mathbf{P\Psi}(\theta) - \boldsymbol{\lambda}\mathbf{I}) = 0$, the eigenvalues can be solved using the quadratic formula,

$$
\begin{aligned}
\lambda = &\ \frac{q_{11}\Psi_1(\theta) + q_{22}\Psi_2(\theta)}{2} \pm \\
&\ \frac{\sqrt{\big(q_{11}\Psi_1(\theta) + q_{22}\Psi_2(\theta)\big)^2 + 4(1 - q_{11} - q_{22})\Psi_1(\theta)\Psi_2(\theta)}}{2}
\end{aligned} \tag{A-11}
$$

Using the magnitude of the largest eigenvalue in (A-11), the spectral radius is determined by,

$$\mathrm{sp}\{\mathbf{P\Psi}(\theta)\} = \frac{q_{11}\Psi_1(\theta) + q_{22}\Psi_2(\theta)}{2} + \frac{\sqrt{\left(q_{11}\Psi_1(\theta) + q_{22}\Psi_2(\theta)\right)^2 + 4(1 - q_{11} - q_{22})\Psi_1(\theta)\Psi_2(\theta)}}{2} \quad \text{(A-12)}$$

Substituting (A-12) into (A-8), the effective bandwidth function reduces to,

$$E_B(\theta) = \frac{1}{\theta}\log\left(\frac{z(\theta) + \sqrt{\left(z(\theta)\right)^2 + 4\Psi_1(\theta)\Psi_2(\theta)(1 - q_{11} - q_{22})}}{2}\right) \quad \text{(A-13)}$$

where $z(\theta) = q_{11}\Psi_1(\theta) + q_{22}\Psi_2(\theta)$

References

[1] N. Sarma and S. Nandi, "QoS support in Mobile Ad hoc Networks," in the proceedings of *IFIP International Conference on Wireless and Optical Communications Networks*, 2006.

[2] M. van Der Schaar and N. Sai Shankar, "Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms," *IEEE Wireless Communications*, vol. 12, no. 4, pp. 50-58, August 2005.

[3] V. Srivastava and M. Motani, "Cross-Layer Design: A Survey and the Road Ahead," *IEEE Communications Magazine*, pp. 112-119, December 2005.

[4] Q. Zhang and Y.-Q. Zhang, "Cross-Layer Design for QoS support in Multihop Wireless Networks," *Proceedings of the IEEE*, vol. 96, no. 1, pp. 64-76, January 2008.

.

[5] J. K. Choi, J. S. Park, J. H. Lee, and K. S. Ryu, "Review on QoS issues in IEEE 802.11 W-LAN," in the proceedings *International Conference on Advanced Communication Technology*, pp. 2109-2113, February 2006

[6] M. Bourouha, S. Ci, G. B. Brahim and M. Guizani, "A cross-layer design for QoS support in the 3GPP2 wireless systems," *IEEE Global Telecommunications Conference Workshops*, 2004. GlobeCom Workshops 2004. pp. 56 – 61, December 2004.

[7] X. Wang, Q. Liu and G. B. Giannakis, "Analyzing and Optimizing Adaptive Modulation Coding Jointly With ARQ for QoS-Guaranteed Traffic," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 2, pp. 710 - 720, March 2007.

[8] J. Tang and X. Zhang, "Cross-layer design of dynamic resource allocation with diverse QoS guarantees for MIMO-OFDM wireless networks," *Sixth IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks*, 2005. WoWMoM 2005. pp. 205-212, June 2005.

[9] D. Skyrianoglou, N. Passas and A. Salkintzis, “Traffic scheduling for multimedia QoS over wireless LANs,” *IEEE International Conference on Communications*, 2005. ICC 2005. vol. 2, Page(s):1266 - 1270, May 2005.

[10] S. M. Abd El-atty, “Efficient Packet Scheduling with Pre-defined QoS using Cross-Layer Technique in Wireless Networks,” *11th IEEE Symposium on Computers and Communications*, 2006. ISCC '06. pp. 820 - 826, June 2006.

[11] Q. Liu, S. Zhou, and G. B. Giannakis, “A Cross-Layer Scheduling With Prescribed QoS Guarantees in Adaptive Wireless Networks”, *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 5, pp. 1056-1065, May 2005.

[12] Q. Liu, X. Wang, and G. B. Giannakis, “A Cross-Layer Scheduling Algorithm With QoS Support in Wireless Networks”, *IEEE Transactions on Vehicular Technology*, vol. 55, no. 3, pp. 839-847, May 2006.

[13] K. B. Johnsson and D. C. Cox, “An adaptive cross-layer scheduler for improved QoS support of multiclass data services on wireless systems”, *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 334 – 343, February. 2005.

[14] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 630-643, July 2003.

[15] C.-S. Chang, *Performance Guarantees in Communication Networks*. Berlin, Germany: Springer-Verlag, 2000.

[16] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Transactions on Automatic Control*, vol. 39, no. 5, pp. 913-931, May 1994.

[17] G. Kesidis, J. Walrand, and C.-S. Chang, "Effective bandwidths for multiclass Markov fluids and other ATM sources," *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 424-428, August 1993.

[18] A. I. Elwalid and D. Mitra, "Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks," *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pp. 329-343, June 1993.

[19] H. S. Wang and N. Moayeri, "Finite-state Markov channel — a useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163-171, February 1995.

[20] J. Tang and X. Zhang, "Cross-Layer modeling for Quality of Service guarantees over Wireless Links," *IEEE Transactions on Wireless Communications*, vol. 6, no. 12, pp. 4504-4512, December 2007.

[21] M. M. Krunz and J. G. Kim, "Fluid analysis of delay and packet discard performance for QoS support in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 2, pp. 384–394, February 2001

[22] Q. Liu, G. B. Giannakis and S. Zhou, "Queuing with adaptive modulation and coding over wireless link: cross-layer analysis and design," *IEEE Transactions on Wireless Communications* , vol. 4, no. 3, pp. 1142-1153, May 2005

[23] J. Tang and X. Zhang, "Quality-of-Service Driven Power and Rate Adaption over Wireless Links," *IEEE Transactions on Wireless Communcations*, vol. 6, no. 8, pp. 3058-3068, August 2007.

[24] J. Tang and X. Zhang, "Quality-of-Service Driven Power and Rate Adaption for Multichannel Communications over Wireless Links," *IEEE Transactions on Wireless Communications*, vol. 6, no. 12, pp. 4349-4360, December 2007.

[25] J. S. Harsini and F. Lahouti, "Adaptive Transmission Policy Design from Delay-Sensitive and Bursty Packet Traffic over Wireless Fading Channels," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 776-786, February 2009.

[26] J. Tang and X. Zhang, "Cross-Layer-Model Based Adaptive Resource Allocation for Statistical QoS Guarantees in Mobile Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 12, pp. 1-11, December 2007.

[27] B. Choi and L. Hanzo, "Optimum Mode-Switching-Assisted Constant-Power Single- and Multicarrier Adaptive Modulation," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 3, pp. 536-560, May 2003.

[28] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer Combining of Adaptive Modulation and Coding with Truncated ARQ Over Wireless Links," *IEEE Transactions on Wireless Communications* , vol. 3, no. 5, pp. 1746-1755, September 2004.

[29] A. J. Goldsmith, S.-G. Chua, "Adaptive coded modulation for fading channels", *IEEE Transactions on Communications*, vol 46, no 5, pp 595-602, May 1997.

[30] A. Goldsmith, *Wireless Communications*. New York, United States of America: Cambridge University Press, 2005.

[31] A. J. Goldsmith, S.-G. Chua, "Adaptive coded modulation for fading channels", *IEEE Transactions on Communications*, vol 46, no 5, pp 595-602, May 1997.

[32] G. R. Dattatreya, *Performance Analysis of Queuing and Computer Networks*. USA: Chapman & Hall/CRC Computer & Information Science Series, 2008.

[33] M. Schwartz, *Broadband Integrated Networks*. New Jersey: Prentice Hall PTR, 1996.

[34] C.-S. Chang and J. A. Thomas, "Effective bandwidth in high-speed digital networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 1091-1100, August 1995.

[35] J. Tang and X. Zhang, H.-H. Chen, C. Song, and M. Guizani, "Cross-Layer-Based Modeling for Quality of Service Guarantees in Mobile Wireless Networks," *IEEE Communications Magazine*, pp. 100-106, January 2006.

[36] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell System Technical Journal*, vol. 39, pp. 1253-1265, September 1960.

[37] E. O. Elliot, "Estimates of error rates for codes on burst-noise channels," *Bell System Technical Journal*, vol. 42, pp. 1977-1997, September 1963.

[38] W. T. Webb and R. Steele, "Variable rate QAM for mobile radio," *IEEE Transaction Communications*, vol. 43, pp. 2223-2230, July 1995.

[39] G. S. G. Beveridge and R. S. Schechter, *Optimization:Theory and Practice*. New York: McGraw-Hill, 1970.