Cross-Layer QoE Improvement with Dynamic Spectrum Allocation in OFDM-Based Cognitive Radio.

Zhong, Bo

Cross-Layer QoE Improvement with Dynamic Spectrum Allocation in OFDM-Based Cognitive Radio

Bo Zhong

Supervised by Dr John A. Schormans

Department of EECS

Queen Mary University of London

Thesis Submitted for the degree of

Doctor of Philosophy

March 2014

To my family

Acknowledgements

My deepest and sincerest gratitude goes to my supervisor, Dr John Schormans, for his persistent guidance, suggestions, advices, patience and caring that supported me through my PhD study. Many thanks to Dr Eliane Bodanese, Dr Chris Philips and Professor Steve Uhlig for the valuable and enlightening suggestions and advices. Thanks to all my friends in the Networks Research Group, for the encouragements and valuable suggestions.

Special thanks to my good friends Zelun, Ran, Xin and Shiwen, for accompanying me and made my life in UK colourful.

Finally, many thanks to my beloved wife Xiajing, for the always lasting supports, understanding and trust.

Abstract

Rapid development of devices and applications results in dramatic growth of wireless traffic, which leads to increasing demand on wireless spectrum resources. Current spectrum resource allocation policy causes low efficiency in licensed spectrum bands. Cognitive Radio techniques are a promising solution to the problem of spectrum scarcity and low spectrum utilisation. Especially, OFDM based Cognitive Radio has received much research interest due to its flexibility in enabling dynamic resource allocation. Extensive research has shown how to optimise Cognitive Radio networks in many ways, but there has been little consideration of the real-time packet level performance of the network. In such a situation, the Quality of Service metrics of the Secondary Network are difficult to guarantee due to fluctuating resource availability; nevertheless QoS metric evaluation is actually a very important factor for the success of Cognitive Radio. Quality of Experience is also gaining interest due to its focus on the users' perceived quality, and this opens up a new perspective on evaluating and improving wireless networks performance. The main contributions of this thesis include: it focuses on the real-time packet level QoS (packet delay and loss) performance of Cognitive Radio networks, and evaluates the effects on QoS of several typical non-configurable factors including secondary user service types, primary user activity patterns and user distance from base station. Furthermore, the evaluation results are unified and represented using QoE through existing mapping techniques. Based on the QoE evaluation, a novel cross layer RA scheme is proposed to dynamically compensate user experience, and this is shown to significantly improve QoE in scenarios where traditional RA schemes fail to provide good user experience.

Author's Publications

B. Zhong, J. Schormans, and E. Bodanese, Evaluating QoE in Cognitive Radio Networks for Improved Network and User Performance, Communications Letters, IEEE, vol. 17, no. 12, pp. 23762379, December 2013.

B. Zhong, J. Schormans, and E. Bodanese, QoE Aided Performance Compensation for Cognitive Radio Networks, Transaction on Wireless Communications, IEEE, submitted and under review.

Contents

List of Figures

Nomenclature

BER Bit Error Rate

CDF Cumulative Distribution Function

CR Cognitive Radio

DSA Dynamic Spectrum Access

IWT Inverse Wavelet Transform

LRD Long Range Dependence

MOS Mean Opinion Score

OFDM Orthogonal Frequency Division Multiplex

PESQ Perceptual Evaluation of Speech Quality

PU Primary Users

QoE Quality of Experience

QoS Quality of Service

RA Resource Allocation

SINR Signal to Interference and Noise Ratio

SN Secondary Network

SNBS Secondary Network Base Station

SNR Signal to Noise Ratio

SU Secondary Users

VoIP Voice over IP

WT Wavelet Transform

Mathematical Expressions

M	Number of Subchannels
N	Number of Secondary Users
N_p	Number of Primary Users
ω	Critical Interference Threshold
N_0	Gaussian Noise Power
$\overline{P_j}$	Power Constraints Vector of Channel j
P_{max}	Maximum Transmission Power of SNBS
L	Number of Time Slots in One Transmission Frame
F	Number of Time Slots in A Sub-Transmission Frame
g_{ij}	Channel Gain of User i in Channel j
h_{ij}	Ricean Fading Gain of User i in Channel j
d_0	Far-Field Crossover Distance
d_i	Distance From User i to the SNBS
η	Path Loss Exponent
R_z	Packet Transmission Rate of User in Mode z
γ_z	SNR Requirement in Mode z
r_1	Maximum Distance From SU to SNBS
r_2	Maximum Distance From PU to SNBS
ΔP_{ij}	Power Efficiency of Updating Transmission Mode (Ratio Between Power Difference and Transmission Rate Difference)
λ_{RT}	Packet Arrival Intensity for Real-Time Secondary Users
λ_{NRT}	Packet Arrival Intensity for Non-Real-Time Secondary Users
P_{On-Off}	State Transition Probability from On to Off for Secondary Users
P_{Off-On}	State Transition Probability from Off to On for Secondary Users
$P_{primary-I-A}$	State Transition Probability from Inactive to Active for Primary Users

$P_{primary-A-I}$	State Transition Probability from Active to Inactive for Primary Users
δt	Time Slot Length in ms
$\phi_j^m(t)$	Haar Mother Wavelet
d_j^m	m^{th} Wavelet Coefficient at Time Scale j
ϕ	Combined Priority Index
α	QoE Compensation Index
I_{QoE}	Evaluated Average QoE Index in Previous Sliding Window

Chapter 1

Introduction

With the rapid development of devices and applications such as VoIP, video streaming and wireless online games, the demand on wireless system capacity is continuously increasing. The wireless bandwidth, which determines the capacity of a mobile network system, has become valuable and scarce. However, a report from the Spectrum Efficiency Working Group of Federal Communications Commission (FCC) pointed out that the shortage of spectrum is actually a spectrum access problem. The utilization of the spectrum is limited by traditional fixed spectrum allocation policies [3]. In order to tackle this problem, the Cognitive Radio (CR) technique was introduced by Joseph Mitola III in 1998 [4]. With CR, unlicensed users (secondary users, SU) in a secondary network (SN) could have access to the licensed spectrum under the condition that primary users (PU) of the licensed spectrum are not interfered with. Since then, a large number of papers have been published on CR, and CR technology has experienced significant development.

There has been increasing research interest in CR based on Orthogonal Frequency Division Multiplex (OFDM) systems in recent years. The reason for that is not only because OFDM systems have the ability to achieve a high data rate in wideband digital communication and cope with severe channel conditions, especially frequency-selective fading, but also for its flexibility in dynamic radio resource allocation (RA), which makes it very attractive for CR. As stated in [5], OFDM modulation is very convenient for a spectrum rental system as it is possible to leave a set of subcarriers unused. Due to the advantages of implementing

CR using OFDM modulation systems, great efforts have been recently made on OFDM-based CR research [2, 6, 7, 8, 9, 10, 11]. All of these papers aim at efficiently allocating resources to improve certain parameters of any OFDM-based CR system. For example, [8] tries to analyze the total bit rate of SUs, [9] tries to maximize the spectral efficiency of the frequency band occupied, while in [2] a max-min rate of all the SUs is used as a measurement of system performance, and a model is proposed to analyze it.

However, few of these takes into account the real-time quality of service (QoS) of higher layer applications, which is actually to some extent a more important aspect for the satisfaction of users. In current research, although secondary networks have been optimised in many ways, most of the current research focuses on the network performance in a particular short time frame [1]. It is true that when we observe the network at a particular time, certain network parameters (user throughput, spectral efficiency, power efficiency etc.) can be optimised. However, this does not guarantee the QoS for individual users. An evaluation of QoS for CR users is critical for a continuous time evolving scenario, where the average QoS of a user over a longer period can be studied. In reality, QoS provisioning in secondary networks is a very complex problem, and there are few existing papers addressing this problem[2]. In this research, in order to achieve QoS provisioning in secondary networks, the first step we take is to identify the factors that may have impact on QoS of SUs, and how much each of these factors affects the QoS parameters. The following factors are of most interest since these are the typical non-configurable factors in practical scenarios: *SU traffic characteristics, PU activity patterns, distance of SU from base station.*

Due to the variety of applications, the properties of the traffic that are being transmitted in a cognitive network could be very different. One typical example is the difference between the traffic of a VoIP service and the traffic of a web browsing session: VoIP traffic has a lower load and does not have much variation, while traffic for web browsing has high load bursts. The heterogeneous traffic that is being transmitted on a CR network can affect the effectiveness of the spectrum allocation algorithm used, but this is seldom considered in existing papers. In [2],

[1] The details of such simulation is introduced in Section 3.2.2

[2] details of these papers are introduced in Chapter 2

the buffer backlog of each SU is used as the criterion in the proposed RA scheme, but the proposed algorithm does not consider the variation of backlog over time at all. Thus the proposed RA scheme might be ineffective if heterogeneous traffic is fed into the buffers of SUs.

On the other hand, the heterogeneity of arriving traffic could lead to the need to support very different QoS requirements. Different traffic arrival processes may lead to very different queue backlog conditions, which will also affect the QoS of users. For example, some traffic is sensitive to delay but has higher packet loss tolerance such as VoIP traffic, while some traffic has tolerance for delay but is sensitive for packet loss and error. Satisfying the QoS of heterogeneous service types is difficult yet very important. Because of these problems of different service types and different QoS requirements, QoE becomes a more promising way of evaluating network performance. QoE unifies QoS requirements of different services and wraps them up into a single, comparable metric. One significant novelty of the compensation scheme for RA of CR proposed in this thesis, is the use of QoE as the improvement target.

One significant characteristic that differentiates CR networks from common wireless systems is that the available physical resource to the networks is not preallocated and is thus fluctuating through time. The fluctuation of available resource is obviously going to affect the QoS of the SUs; but how much is yet to be discovered. The available resource to secondary networks is determined by how PUs access their licensed spectrum. So the PUs' activity pattern is likely to be a key factor affecting QoS of SUs. There are a variety of models that can be applied to user activity patterns, and recent research [12] indicates that a human inter-activity model is better approached by using a power law distribution. This change in the fundamental modelling of human activity pattern could affect many existing results related to human activity. For example, if a CR network shares the spectrum band with PUs whose activities are driven by human activities (e.g. cellular network, LAN), the modelling of PU activity pattern needs to be changed. It's very important to examine how different PU activity patterns could result in different secondary network performance. It's also very helpful as a reference to future design of secondary networks, so that different measurements can be applied in secondary networks sharing resource with different PUs.

Individual users are usually not concerned about the system performance as a whole; rather, users only care about their own perceived experience when they use wireless applications. This kind of user experience, more commonly referred in the literature as the quality of experience (QoE), is directly related to the QoS parameters. QoE itself is not directly measurable in real-time, and the traditional way to gain an accurate QoE score is to statistically analyze user experience feedback. However QoS is measurable in real-time and if we know the mapping from QoS to QoE, we can measure user QoE indirectly in real-time, and further, try to improve user experience by optimising QoE. There are many existing papers on QoS to QoE mapping for different services [13, 14, 15, 16]. The incentive of using QoE as a metric instead of QoS such as throughput (one of the most focused optimisation target for CR in literature is individual user throughput) is that providing good throughput does not necessarily result in good user experience, and providing good user experience does not necessarily require a high throughput. The key evidence supporting this is shown later in this thesis.

In this thesis, I focus on the resource allocation problem of the downlink transmission in an OFDM based CR network. In this scenario, SUs in the CR network share the same band of OFDM sub-channels with a number of PUs. I consider the underlay CR mode so that an SU can still make use of the sub-channel when a PU is transmitting in the same sub-channel, but the transmission power from the base station needs to be limited to protect the transmission of the PU. I assume that both the co-channel interference of the SUs and the interference caused by PU transmission can be neglected.

The main contributions of this research includes the following:

- The evaluation of QoS and QoE of an SN is enabled, and the influences of several contributing factors are evaluated.
- We show QoE of SUs degrades linearly to the activity burstiness of PUs.
- We propose a novel spectrum resource allocation scheme to improve the QoE of the SUs

The steps taken to achieve these goals is described as follow. Firstly we have

extended existing research on CR, which is limited to short frame time Physical/MAC metric optimization problems, to be able to analyse network layer QoS of individual users inside secondary networks in a dynamic time evolving manner. Based on the continuous time simulation tool built, the effect of different QoS contributing factors and their combined effects are analysed and quantified. Furthermore, application of the evaluation results are represented as QoE to gain a more accurate understanding of user experience over CR. And finally based on these results, a novel QoE based RA scheme is proposed, and is shown to outperform traditional RA schemes in terms of providing better user experience.

This thesis is organised as follow: Chapter 2 is a literature review of recent related development on CR and relating techniques. Chapter 3 introduces a repetition of experiments of an existing resource allocation algorithm, and the steps for building the simulation platform for numerical evaluation in a short frame time. In Chapter 4, the process of extending the platform in Chapter 3 is introduced. Real time experimental results show the significant impact of packet level traffic arrival process of SUs on the network performance. Chapter 5 introduces experiments for evaluating the effect of PU activity patterns on the SN performance. An end-to-end video transmission model is introduced in Chapter 6 for the evaluation of QoS across CR networks. In Chapter 7, the novel QoE based RA scheme is introduced, and it's performance is evaluated, including a sensitivity analysis of the configurable parameters in the compensation scheme. Chapter 8 concludes the contributions of this research and indicates possible future works.

Chapter 2

Literature Review

2.1 Cognitive Radio

With the rapid development of wireless devices and wireless applications, the demand for wireless radio spectrum resources has seen dramatic increase in the last decades. On the other hand, the current resource assignment policy by regulatory institutions have restricted spectrum usage to licensed users only, resulting in severe under-utilization of the spectrum resources. A survey illustrates the current usage of spectrum [1]. See Figure 2.1

As can be observed from the figure, only part of the spectrum experiences heavy usage, while many, if not most, parts of the spectrum are severely under-utilized. The traditional spectrum allocation policy strictly limits the usage of spectrum to licensed users only. This worked fine when the demand on spectrum resource wasn't so heavy. But nowadays, the increasing demand [17] for wireless radio resource is forcing regulators and researchers to reconsider the existing spectrum allocation policy. Dynamic spectrum access technology which enables dynamic use of spectrum for higher utilization rate has gained much research interest.

CR [4] is regarded as the key technology for dynamic spectrum access. The basic idea of the CR technique is to allow users to access the same spectrum band as the licensed user without interfering with their activities. Thus CR networks should have "cognitive ability" so that they can gain knowledge of the

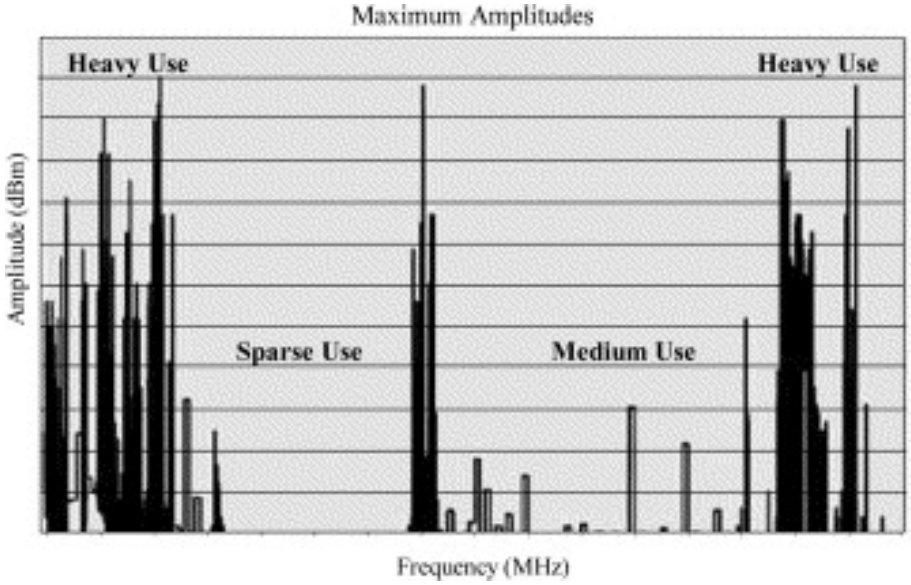

Figure 2.1: Spectrum Usage at different frequencies [1]

activity of the licensed users. The CR networks also need to be able to reconfigure operational parameters to adapt to dynamic changes in licensed users' activities.

The key functions that a CR network should support include [1]:

- Spectrum sensing and analysis
- Spectrum management and hand off
- Spectrum allocation and sharing

There are efforts at standardization for CR networks. IEEE 802.22 is a standard for Wireless Regional Area Network (WRAN) using white spaces in the TV frequency spectrum. It defines a CR network operating at analog and digital TV broadcasting white bands, without causing harmful interference. The IEEE DySPAN standard committee is working on a new standard series IEEE P1900 which includes:

- dynamic spectrum access radio systems and networks with the focus on improved use of spectrum,

- new techniques and methods of dynamic spectrum access including the management of radio transmission interference, and
- coordination of wireless technologies including network management and information sharing amongst networks deploying different wireless technologies.

Three types of mode for the CR networks are commonly studied in literature: overlay, underlay and interweave [18]. For the overlay mode [19, 20, 21], the SUs can transmit simultaneously with PUs, while the SUs helps to relay the PUs message to compensate for the interference caused to the PUs. For the underlay mode [22, 23, 24], the SUs can communicate at the same channel with the PU, as long as the interference caused by the communication of SU does not exceed certain threshold level for PUs. For the interweave mode [25, 26], SUs can access the spectrum resource only when the PU is completely idle, i.e. a sub-channel is not occupied any PU. The mode we assume the CR system to be working in is the underlay mode.

2.2 Spectrum Sensing

Spectrum sensing is one of the most important enabling technologies for CR Networks. By sensing the environment in which the CR network is deployed, CR can gain knowledge of the surrounding radio environment, including the activity of the licensed user, and the available spectrum resource for the secondary network. Only with this information can the CR network reconfigure itself to avoid interference to the licensed users, and achieve optimal performance. There is extensive research on spectrum sensing techniques. Many theoretical models and practical implementations of the spectrum sensing techniques are proposed in the literature [27, 28, 29, 30, 31]. The three different modes of CR requires different sensing information. In the underlay mode, SUs needs to know the channel strength of the PUs transmission. In the overlay mode, SUs requires the channel gains, codebooks and the messages of the PUs. For the interweave mode, SUs need to know the On/Off events of the PU activities.

Detailed sensing methods are not of interest for this research so they are not discussed here. In this research, perfect distributed sensing is assumed so CR networks always have perfect knowledge of the activity of licensed users.

2.3 Resource Allocation in CR

The resource allocation mechanism allows a CR network to reconfigure itself according to the dynamic changes of its surrounding radio environment. And this is where most of the research on CR networks has focused. A good resource allocation method should be able to effectively increase the spectrum utilization and at the same time guarantee limited interference to the licensed users. Most importantly, the resource allocation method should provide satisfactory network performance for the SUs.

Many papers on the resource allocation mechanism aim at optimizing certain performance parameters of the secondary networks, such as total transmission power, total system capacity, transmission delay and minimum transmission rate, etc. A variety of algorithms, heuristics and optimisation models have been proposed to improve secondary network performance. These are reviewed below.

In [6], the authors try to further increase the total transmission capacity of all SUs by loosening the PU protection threshold. SUs do not stop transmitting at once after PU activities are detected, or in some other methods, they strictly control the transmission power so that PUs are protected. Instead, the scheme allows SU transmission to cause a certain level of interference to PUs, as long as the BER threshold of PUs is not reached. The capacity of SU is seen to be increased using the proposed heuristics. However, in [6] the PU's activity is assumed to be static, and the BER of the PU is obtained only from the SINR caused by SUs. Fairness is not considered in [6], nor is QoE.

In [8], the author also proposed a heuristic to approach suboptimal total throughput of the secondary network, at the same time ensuring fairness among users in the sense that they try to allocate bits to users who have not received their fair share of service. Specifically, the author analyzed total throughput under different interference tolerance of PUs. The analytical optimisation problem is formed in this article, and the author approaches a suboptimal solution using

heuristics. However, only one PU, whose activity is modelled as constant, is considered together with several SUs. Although fairness is considered, the only aim is to increase TX rate of non-satisfied SUs as much as possible, rather than analyzing the QoS requirements and different classes of SUs. User QoE is not considered at all.

Like [6, 8], a large number of papers were published to solve the SN optimisation problem by dynamically allocating resources. However, like [6, 8], PU activities are usually modelled as constant. However in real world scenarios, PU activity could be highly dynamic, and the activity pattern could be very different from case to case. As a result, the available resource to the SN is also highly dynamic, which may lead to ineffectiveness of the proposed algorithms. On the other hand, in these papers SUs' activities are considered static, and homogeneous, and the traffic for SUs is assumed to be saturated where traffic modelling is considered at all. This is not the case in real world scenarios, where SUs could use different kinds of applications so that the traffic pattern can be very different, and the queue backlog for SUs can also vary through time. Furthermore, QoS and QoE of individual SUs is often neglected in these researches, while the authors focus mainly on the overall performance of the whole network.

There are a few papers that do take into account the QoS of SUs. For example in [32], a Markov chain queueing model is proposed to analyze the queue length variation in an SU network. The author applies a Discrete Time Markov Chain analysis for the queue model, and then uses a numerical simulation to evaluate the scenario. PU activity intensity is taken into account so that results are obtained under low, medium and high PU activity intensity. However, PU activity is only modelled as a two state Markov chain, with exponentially distributed sojourn time, and SU traffic is also assumed to be homogeneous. On the other hand, [32] also fails to take into account the heterogeneous characteristics of SU traffic.

Recent research [12] shows that activity driven by people is more accurately modelled by more bursty and heavy tailed processes such as power law distribution process. As pointed out by the author, models of human behavior are crucial for better resource allocation and pricing plans, trying to incorporate such characteristic to PU activity may lead to very interesting results. A later section of this thesis demonstrates the significant difference in the evaluated network perfor-

mance when PU activity model changes from exponentially distributed off time to Pareto (heavy tailed) distributed off time.

Research in [33] does consider the QoS of heterogeneous SUs. The author defines a priority factor to ensure that real-time user's QoS requirements are met. The average delay and total throughput of the secondary network are optimised using the algorithm proposed by the author. However, the simulation used to verify the result consider the optimisation of the metrics within a single transmission time slot, and PU activity in [33] is also assumed to be static.

As a summary, existing research focuses on optimizing secondary network performance using analytical optimisation and heuristics for suboptimal solutions, but mostly without QoS consideration for SUs. Those papers which take into account the QoS requirements either fail to consider the dynamic nature of PU activity, or fail to consider the heterogeneity of SUs.

2.4 PU activity analysis

One of the key factor that affects the resource availability to the CR networks is the PU activity. The CR network shares the spectrum band resource with the licensed PUs. As soon as the PU starts using the wireless resource, the activity of the PU needs to be immediately detected by CR network and CR activity at that spectrum band needs to stop at once. The continuously changing PU activity will result in fluctuation of the resource availability to secondary networks, and introduce unpredictable interruptions to CR transmissions.

The uncertainty in resource availability caused by PU activity may have a great impact on the performance of a CR network. It is important to have a thorough understanding of how PU activity will affect the CR network QoS. In order to evaluate the PU activity effects on the CR network, proper PU activity models are required. The choice on different models will affect the results obtained by numerical evaluations.

[34] provides an analysis of primary usage on real data collected inside a cellular network operator. In the paper, the observation results show that, when using a call based model, similar to an On/Off model, the distribution of the inter-call time and call holding time basically fits with the exponential distribution. When

the data is analyzed using an event-based random walk model, which ignores details of individual calls and instead models only the load, the results shows more heavy-tailed characteristics which cannot be accurately modelled using exponential distribution. As a conclusion, the author claims that an exponential call arrival model (coupled with a non-exponential distribution of call durations) is often adequate to model the primary usage process.

the research on PU activity modelling includes [35], which uses a two state On/Off Markov chain to model the activity of a single PU, where a state transition matrix is used to characterize the activity. The values in the transition matrix are determined by observing real network traffic over a IEEE 802.11 wireless AP. Furthermore, the aggregated PU activity of K users is modelled using the Fritchman-model [1]. The complementary CDF of the off time of aggregated PU activity is provided as a result, indicating the probability of resource being available for a certain amount of time from the aggregated PU activity. The model used in the research is quite simple, while the results are inspiring. However, there are several limitations: [35] assumes that all PUs are using the same frequency band, so as long as there is one or more PU occupying the resource, the frequency band is considered unavailable to the CR networks. This is not the case in certain network environment such as an OFDM network, where each user can occupy a subchannel at a different frequency, and different users can transmit simultaneously without interfering with each other. So the model used in the research is not applicable in such cases. On the other hand, the On/Off time in [35] is modelled as the traditional Exponential distribution. This is the simplest case but probably not necessarily the most accurate case. Recent research [12] suggest that activity driven by human actions are more precisely modelled by heavy tailed processes such as power law distributions. If PU activities are driven by human actions, which is the case for many primary networks, then using only exponential on/off modelling is not sufficient to accurately approach the performance in practical cases. The real traffic data in [35] is from a 802.11 wireless AP. But the CR network could operate on many other frequency bands, with various kind

[1]This model was originally proposed in [36] to characterize bit errors in binary channel. It uses a partitioned Markov chain to model error-free run and error cluster, which is then modified in [35] to model the aggregated use of channel by PUs.

of primary networks to share the resource. Different kinds of PU network could result in very different PU activity patterns and having different impact on the performance of CR networks.

[37] studies the activity of PUs and the effect of their locations on the opportunity for SUs to access the resource. The individual PU in [37] is characterized by a two state On/Off Markov chain model. The duration of the On/Off model, however, is still modelled using the exponential distribution with different average length of On time λ.

[38] proposes an optimisation method for CR transmission rate considering PU activity. The numerical results shows improvement in total transmission rate of the secondary network. The simulation over a single transmission frame is repeated many times to obtain an averaged result. For each simulation of a transmission frame, the PU activity is represented merely by a probability ϕ that it occupies the subchannel. The research in [38] fails to consider the real time QoE of the SUs. Most importantly, although the paper claims to consider the PU activity, it is actually just performing the proposed resource allocation method under different number of active PU in each simulation iteration, and averaging the final results. This is actually equivalent to a simulation under the average active PU number. The evolution over time of the PU activity cannot be taken into account by this model.

As a summary, PU activity is a very important contributing factor to CR networks' performance. Many papers have been published on the effect of PU activity, and the performance optimisation of CR network considering PU activities. However, most of the models used either cannot accurately characterize the activity pattern of the PU, or fail to characterize the real time evolution of PU activities. Real time QoS of the CR networks under different PU activity patterns needs to be studied.

2.5 Quality of Experience

User Quality of Experience is defined by the ITU in [39] as *The overall acceptability of an application or service, as perceived subjectively by the end-user.* And an important note for the definition is *Quality of Experience includes the complete*

end-to-end system effects (client, terminal, network, services infrastructure). The most important characteristic of QoE is that it focuses on the end users' subjective opinion of a service. And the evaluation of QoE of a service must be done with a complete end-to-end system.

This could be compared with the Quality of Service definition, which is extensively used in network performance literatures. The definition is also provided by ITU [40]: *the collective effect of service performances, which determine the degree of satisfaction of a user of the service.*

As pointed out in [41] : *Most publications, including many standards, use the term QoS but either do not define it, or else point to one of these other few definitions.*

While in terms of the IP related QoS research publications and standards, QoS is usually used to refer to packet level network performance parameters such as delay, loss and jitter.

These QoS parameters provide good measurements for the network operators to monitor their network speed, availability and reliability. They provide a quantitative metric for the network operators. However, the end users of the network don't necessarily benefit from the QoS that a network can provide. In other words, although QoS will to some extent determine the degree of satisfaction of the end user, satisfiable network QoS for a network operator does not guarantee satisfiable services for an end users. Thus there has been increasing research interest in QoE provisioning for users, aiming at providing better user satisfaction rather than looking at the network performance alone.

The evaluation of QoE is based on different services, because users have a different definition of satisfaction for different services. Take the VoIP service and the file downloading service as examples: users are satisfied when VoIP service could provide continuous and high quality voice signals; while users that are downloading files are satisfied when the data are loaded and in a shorter time. The natural measurement of QoE is of course by gathering subjective user feedback, since QoE is defined as the subjectively perceived service quality. For example, the most commonly used method for voice service QoE measurement is the Mean Opinion Score (MOS) method, in which testers are required to give a score to the audio perceived, ranging from 1-5 (bad to good). The arithmetic

mean of all user opinion scores are taken as the final result.

Subjective methods can provide accurate results for QoE. However, they are usually very time consuming, and require human activity, which is very expensive. Furthermore, the measurement cannot be carried out online, meaning that the resulting QoE cannot serve as real time feedback to the network, so the network configuration cannot be dynamically adjusted according to the QoE by a subjective method.

There has been extensive research on objective measurement approaches for QoE for services, and also QoE provisioning for different services. These are reviewed below.

The ITU E-model [42] for monitoring end-to-end voice quality takes into account a wide range of possible impairments and represent the voice quality using an R-factor, which can then be transformed to MOS. In [43], the authors try to simplify the E-model for VoIP service and find a relationship between transport layer factors and the R-factor. Several transport layer QoS factors including packet delay, packet loss and jitter are identified as relevant factors, and an expression of R-factor value by these factors are described in the paper. The R-factor expression proposed in this paper can be directly used to transform packet level QoS to the user perceived QoE.

[13] presents the exponential relationship between several packet layer QoS parameters and the perceived QoE represented by MOS. And it is shown that the proposed IQX hypothesis, which is a natural and generic exponential relationship between user-perceived QoE and network-caused QoS, can provide a better approximation than the former logarithmic relationships. The MOS functions presented in this paper can be directly applied to map from evaluated QoS parameters such as packet delay and loss to MOS.

In [44] an objective method for voice quality monitoring called Perceptual Evaluation of Speech Quality (PESQ) is proposed. The system used in this paper is trained by comparing the reference signal with the degraded signal after transmission over the networks. After a certain amount of training time, the PESQ method is able to provide very accurate estimation on the subjective quality score in real time, with over 74 percent of observed estimation value within absolute error range of 0.25. The PESQ method is extensively used in the evaluation of

VoIP over various networks [45, 46, 47].

The transmission capacity of wireless networks are continuously increasing, enabling mobile devices to implement services that requires higher transmission links. Mobile video streaming is one of the most popular services, but user experience is likely to suffer, because of the fluctuation in wireless transmission environments. Recently, researchers have become more interested in evaluating QoS and QoE of Video streaming over networks. Especially, since the dynamic spectrum access (DSA) techniques including CR has the potential to provide high speed wireless links, there are many papers on providing QoS and QoE over such networks [48, 49].

Some more details about ITU-T E-Model, and QoS to QoE mapping methods in [43] and [13] can be found in Appendix B.

2.6 Conclusions

CR techniques have been studied intensively in recent years, due to its potential to solve the spectrum resource scarcity problem. The major problems in the CR techniques include spectrum sensing, resource allocation, the analysis of PU activities and QoS, QoE provisioning for CR. All of these areas require more research. Especially, current resource allocation algorithms proposed for the CR systems usually lacks the capability to perform real time QoS evaluations. The heterogeneous user traffic and different requirements for QoS need also to be considered when designing and implementing resource allocation algorithms for CR. The effects of the activities of PUs needs to be evaluated using better models for the activity patterns of PUs. And quality of experience of users using specific services in CR networks requires more research. In summary, although extensive researches have been carried out on CR, many fundamental and challenging problems are yet to be solved.

Chapter 3

Resource Allocation Optimisation Problem in Short Transmission Frame

3.1 Introduction

Many different models and approaches to creating CR networks have been proposed in recent years. OFDM based CR networks are suggested as one of the most promising technology for CR networks [5]. The ultimate purpose of the research in this thesis is to evaluate and improve user perceived experience of an OFDM based CR network. The model used in this thesis is adopted and modified from packet level QoS related research on an OFDM based CR RA algorithm [2]. Although the model proposed in [2] fails to deal with continuous time performance, it can be extended and converted to a continuous time evolving model with some modifications. Thus it is reasonable to firstly build a simulation platform based on the existing optimisation model within a transmission frame, and make sure that the RA algorithm is correctly implemented. And then this simulation platform can be extended to further evaluate the real-time performance of the network. In this chapter, the basic scenario of the OFDM based CR system is described in detail, and the process of building a simulation platform based on the scenario is provided. In section 3.2, the system model focusing on an opti-

misation problem within a transmission frame is presented. In section 3.3, the experimental parameters and experimental process are introduced. Experimental results are shown and compared with results from [2] in section 3.4. A conclusion for the optimisation problem within a transmission frame is provided in section 3.5.

3.2 System Model of Optimisation Problem within A Transmission Frame

This section describes the system model of Max-Min optimisation for the OFDM based CR network. Section 3.2.1 gives a general overview of the scenario, including how users are distributed, and interaction between PUs and SUs. Section 3.2.2 describes the MAC layer frame structure. Section 3.2.3 introduces the wireless channel characteristic model used. Finally the modelling of SUs is provided in 3.2.4.

3.2.1 Scenario Overview

Consider an OFDM downlink RA problem. There are M subchannels, N SUs and N_p PUs. PUs are randomly located inside a circular area with radius of 60km while SUs are randomly located inside a circular area with radius of 33km. At the center of these two areas, a Secondary Network Base Station (SNBS) is located to provide downlink transmission for SUs. Figure 3.1 demonstrates the basic scenario. PUs and SUs share the same frequency band in this area. Each PU occupies one OFDM subchannel. We consider an underlay CR mode[22, 23, 24], so that when a subchannel is occupied by a PU, an SU could still use this subchannel as long as the largest possible signal power (from the SNBS) perceived by the PU in the subchannel the PU is licensed to occupy is lower than ωN_0, where N_0 is the Gaussian noise power and ω is the critical interference threshold used to adjust the maximum amount of interference that a PU can tolerate.

Perfect distributed sensing is assumed for all SUs in the system, so that PU activities are sensed by both the SUs and the SNBS. As a result, for each of the subchannels, a transmission power constraint from the SNBS can be calcu-

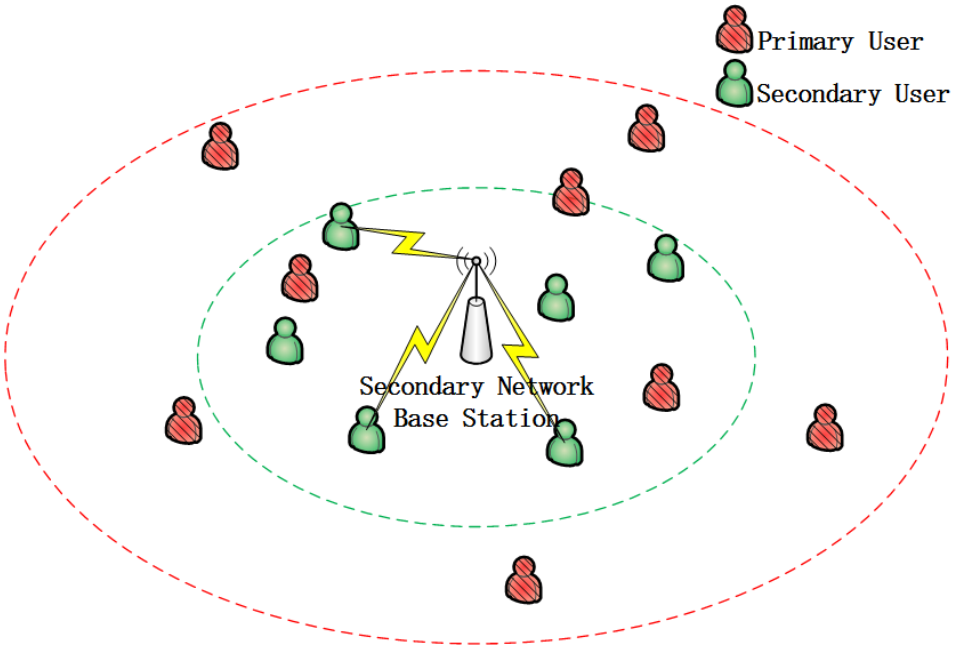

Figure 3.1: OFDM based Cognitive Radio System Model

lated according to the sensed information. The SNBS will limit signal power in subchannels according to the constraints in order to protect PU transmissions. All these constraints form a vector $\overline{P_j}$, in which each element corresponds to the power constraints in a subchannel. If there is no PU occupying subchannel x, then $\overline{P_j}(x) = \infty$. Meaning that the SNBS can transmit in this subchannel using as much power as necessary. The total transmission power of the SNBS is limited by P_{max}. The detail of how to calculate the power constraint vector is introduced in section 3.3.3.

The purpose of the RA algorithm is to allocate subchannel, power and transmission rate for SUs to maximize the minimum transmission rate among all SUs. In other words, it is an optimisation problem in which the target function is the minimum transmission rate among all SUs.

3.2.2 Resource Allocation Process - The Frame Structure

Time is assumed to be slotted and grouped into frames of L time slots. At the beginning of each frame, SUs send the results of sensing and channel gain

information to the SNBS. The SNBS runs the RA algorithm and then sends the allocation map to all the users so that they can tune their radio parameters to the correct subchannels on a time-slot basis. The frame structure is illustrated in Figure 3.2

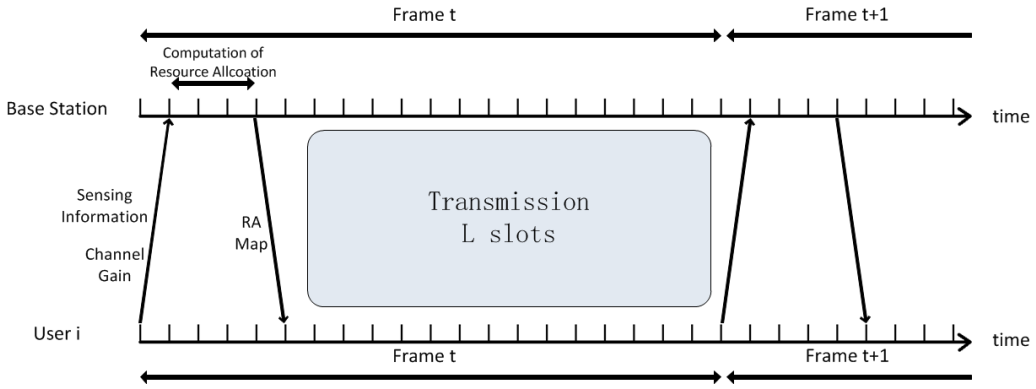

Figure 3.2: OFDM based Cognitive Radio Frame Structure

The computation of RA is based on:

1. Subchannel Power constraints
2. Channel gains for each SU in each subchannel
3. Current backlog of each SU

These three factors are clearly dynamic from one frame to the next frame in any real case. In order to simplify the scenario, frame time length L is assumed to be short enough so that all these parameters remain unchanged during a frame time, which makes it a quasi-static condition. This indicates that PUs activities are assumed to remain unchanged within one frame, so the power constraints are static, and channel gains for each SU remain unchanged. On the other hand, the queue backlog of each SU is affected by transmission and arrival of packets. Throughout the whole frame, for each of the time slots, the SNBS will transmit data packets to SUs according to the RA map calculated at the beginning of the frame. New packets will arrive at the SU buffers for each of the time slots, while packets in the buffer will also be transmitted to SUs for each of the time slots.

However, only the observation of the SU backlogs at the beginning of each frame will be taken as the input parameter of the RA algorithm.

It is important to notice that the computation of the RA map is a very time consuming process if all the L time slots in one frame are considered together. In other words, the optimisation problem becomes much more complicated if time slots 1 to L in one frame are optimised together, because this would introduce another degree of freedom into the optimisation problem. In practical applications, an SNBS needs to calculate the RA map in a very short frame time, which should be significantly smaller than the frame duration.

In order to help improve the responsiveness, firstly, heuristics are used to approximate the optimal solution, and secondly, the problem is optimised in a shorter range of time slots, say, F slots. And the same RA map will be applied $k = L/F$ times to the rest of the frame. Specifically, the cases where $F = 1$ and $F = 3$ are considered here. As an example, when $L = 30$ and $F = 3$, the heuristic tries to find the optimal solution by allocating subchannel time-slot pairs to SUs for 3 time slots. When the optimal solution is found, the SNBS allocates the resource in the first 3 time slots according to the solution and transmit data, and the same RA map is used for the next 3 time slots, and so on, for a total of 10 times. It is obvious that when F gets larger, the optimal target value can be improved, but however, the computational complexity will be increased significantly as a cost.

The detailed Heuristics used for the RA algorithm will be introduced in section 3.3.5.

3.2.3 Wireless Channels

The wireless channels for both PU and SU are characterized by path loss combined with fading effect [50]. Let g_{ij} denote the channel gain from the SNBS to user i (Primary or Secondary) in subchannel j. Then $g_{ij} = |h_{ij}|^2 (d_0/d_i)^\eta$, where h_{ij} is an independent Ricean fading gain characterized by K-factor, η is the path loss exponent, d_0 is the far-field crossover distance and d_i the distance from user i to the SNBS [50]. Wireless channels are considered to be quasi-static[51, 52], that is, the frame length is short enough so that the channel condition is assumed to

be stable during the frame time. The fading gain, which is Ricean distributed, is randomly generated only at the beginning of one frame, and remains as the same value during the frame. When the simulation starts for the next frame, the wireless channel condition for all users in all subchannels will be initialized again. The detailed calculation of path loss and fading gain is described in section 3.3.4.

3.2.4 Secondary Users

Recent research enables adaptive modulation and coding schemes according to different SNR threshold [53, 54, 55]. Assume that an SU downlink transmission could work in one of $\overline{z}$ different transmission modes with different transmission capacities, resulting from different modulation and coding schemes being used. R_z packets could be transmitted in one time slot if mode z is applied. Assume $0 < R_1 < R_2 < \cdots < R_{\overline{z}}$. Which mode an SU works in is decided by the Signal to Noise Ratio (SNR) in the subchannel the SU occupies. Mode z requires a minimum SNR of γ_z to meet the required error rate.

Let $f_{ij(z)}$ denote the minimum transmission power from the SNBS to SU i on subchannel j using transmission mode z. $f_{ij(z)}$ is a function of corresponding channel gain g_{ij}, SNR threshold γ_z, noise power at the receiver and the interference from PUs on subchannel j. For simplicity, take $f_{ij}(z) = \gamma_z N_0 / g_{ij}$.

There is a separate buffer for each of the SU in the network, to store the packets that are to be transmitted to the SU. The short frame time RA optimisation problem is limited to one frame time, so the backlogs for SUs do not evolve through time. But the RA algorithm considers the queue backlog of each of the SUs and decides priority among SUs to access the spectrum according to the queue backlog. So in the repeated experiment of the short frame time simulation, a fixed value of backlog is initialized at the beginning of each simulation iteration, to serve as input to the RA algorithm. This means that the actual packet arrival process is not considered at all here. A saturated buffer (though may have different saturated queue length) is assumed for each of the SUs.

3.3 Experiment Setup for Short Frame Time Optimisation Problem

A simulation platform is implemented in MATLAB, following the system scenario and the algorithm described in the previous section. Because not all the detailed simulation steps and attributes are clearly stated in [2], in order to further test the effectiveness of the algorithm, it is necessary to firstly repeat the experiments in [2] to verify that the algorithm was correctly implemented in the simulation platform.

In the following sections, the experiment process is introduced step by step in detail. A detailed Pseudocode of the introduced algorithm can be found in [56].

3.3.1 Initialization of Parameters

In the short frame time optimisation problem, after each optimisation in one frame time, all the parameters are re-initialized, including user backlog, channel conditions, and the position of PUs and SUs. Then the experiment is repeated to reduce the variance from random factors so that more precise result of the Max-Min downlink transmission rate is obtained.

Table 3.1 is a list of all the parameters used in the simulation platform.

The SNBS can transmit to an SU in downlink channel in one of the $\overline{z} = 5$ transmission modes, depending on the channel condition from SNBS to the SU. See Table 3.2 for the transmission rate R and the Signal to Interference plus Noise Ratio (SINR) requirement γ for the corresponding transmission mode.

3.3.2 Initialization of User Location Distribution

PUs are distributed inside a circular area with radius of 60km. SUs are distributed inside a circular area with radius of 33km, and an SNBS located at the center. Position is expressed in polar coordinates and the distance d and angle ϕ are generated using random functions 3.1, 3.2,

$$d = r \times \sqrt{\xi}, where\ \xi \sim U(0, 1) \tag{3.1}$$

Table 3.1: Parameter summary in simulation platform

Parameter	Value	Unit	Explanation
M	120	1	Total subchannel number
N	40	1	Total SU number
N_p	30	1	Total PU number
K	-10	dB	K factor of Ricean distribution
η	3	1	Path loss exponent
d_0	50	m	Far-field crossover distance
r_1	33	km	Max distance from SU to BS
r_2	60	km	Max distance from PU to BS
N_0	-100	dBW	Gaussian noise Power
ω	0	dB	Critical interference threshold
L	30	slot	Frame length
F	1, 3	slot	Sub-frame length
P_{max}	0-100	W	Maximum transmission Power from BS

Table 3.2: Transmission Rate and SINR Requirement

Mode	$R(Packets)$	γ (dB)
1	1	10
2	2	14.77
3	3	18.45
4	4	21.76
5	5	24.91

$$\phi = 2\pi \times \xi, where\ \xi \sim U(0,1) \tag{3.2}$$

By generating this pair of random variables, the positions of the users are uniformly distributed inside the circle. User distribution actually plays a very important part in system performance evaluation. Different distributions of users within a cell can affect the overall system capacity and there are works to analyse these effects [57, 58, 59]. We choose the uniform distribution of users within the cell because this is the basic model and is commonly used in literature, and also because we need to validate our simulation tool with the result from [2], which also uses uniform distribution model. An example distribution of users is shown in Figure 3.3

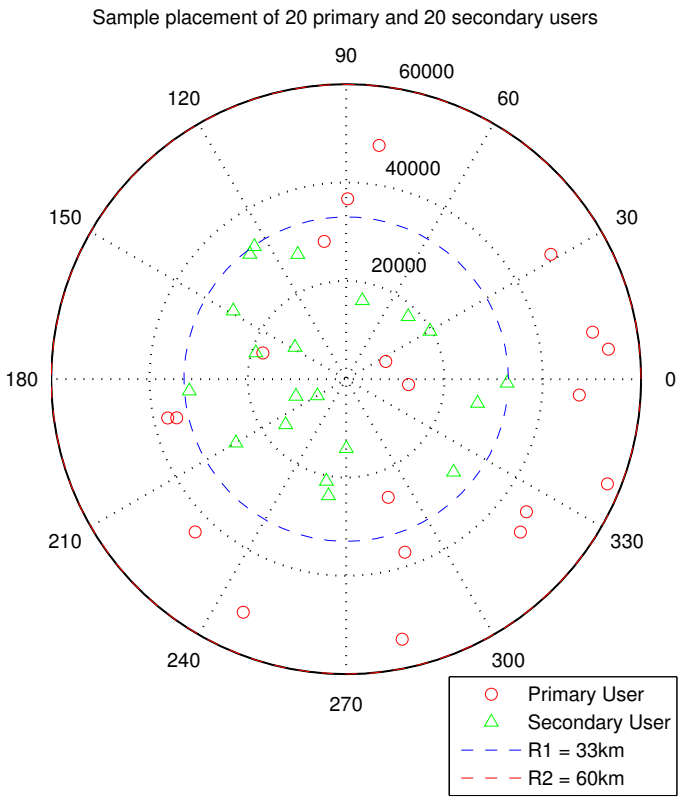

Figure 3.3: User Distribution Example

3.3.3 Calculating the Power Constraints Vector

Each of the PUs is randomly allocated one subchannel. All subchannels occupied by PUs form a set: $A = \{subchannels\ occupied\ by\ PU\}$. Transmission signal from the SNBS to an SU sharing the same subchannel with a PU is perceived by the PU as noise. It is necessary to limit the transmission power from the SNBS in the subchannel, so that the noise level perceived by any PU does not exceed a certain threshold. Here the power constraints vector $\overline{P_j}$ for all subchannels is calculated in the following way. Each element in $\overline{P_j}$ is taken to be the largest feasible value such that the received power from the SNBS to the PU is ωN_0. So if PU n occupies subchannel j, then the power constraints $\overline{P_j}$ is calculated by equation 3.5:

$$P_j \cdot (d_0/d_n)^\eta \leq \omega N_0 \tag{3.3}$$

$$P_j \leq \omega N_0 \, (d_n/d_0)^\eta \tag{3.4}$$

So,

$$\overline{P_j} = \omega N_0 \, (d_n/d_0)^\eta \tag{3.5}$$

Obviously the power constraint is closely related to the distance d_n from the SNBS to the PU. If a PU is far away from the SNBS, the power constraint on the subchannel which the PU currently occupies will be higher. In other words, the constraint is less restrict, meaning that an SU can transmit with higher power when sharing the subchannel with the PU. This is intuitively correct, because the further away the PU is from the SNBS, the less interference it will get from the SNBS.

if $M < N_p$, then some of the subchannels are not occupied by any PU. The power constraints in these subchannels should be set to infinity, indicating no power constraints.

$$\overline{P_{j'}} = +\infty, \quad j' \in A' \tag{3.6}$$

The power constraints for all subchannels are calculated one by one and recorded in vector $\overline{P_j}$.

3.3.4 Initialization of channel gain

The wireless channel gain is characterized by $g_{ij} = |h_{ij}|^2 (d_0/d_i)^\eta$, where h_{ij} is an independent Ricean fading gain characterized by K-factor, η is the path loss exponent, d_0 is the far-field crossover distance and d_i the distance from user i to the SNBS.

Firstly values of h_{ij} must be calculated. Considering only Ricean fading, the probability density function of the received signal power after Ricean fading is expressed as equation 3.7 [60]:

$$f_p(p) = \frac{(1+K)\, e^{-K}}{\overline{p}} exp\left(-\frac{1+K}{\overline{p}}p\right) I_0\left(\sqrt{4K(1+K)\frac{p}{\overline{p}}}\right) \tag{3.7}$$

Where $\overline{p}$ is the local mean power, I_0 is 0^{th} order modified Bessel function of the first kind. There is no Ricean random generator in MATLAB, so an easy way to generate a random variable using the probability density function is the Inversion Method [61].

Firstly the cumulative density function is numerically calculated from pdf, by equation 3.8:

$$F_p(p) = \int_0^p f_p(p)\, dp \tag{3.8}$$

Then a uniform random variable $\xi \sim U(0,1)$ is generated using the MATLAB random function. According to Inversion transform sampling, $p = F_p^{-1}(\xi)$ should have the same pdf as $f_p(p)$.

To test the random variable generation program, a set of experiments were done, with different values of K.

The red curve represents the PDF of equation 3.7, and the histogram is obtained from 100,000 generated random variables. Figure 3.4 shows that the simulation results generated from Inversion Method fits well with the original pdf. The Ricean fading gain h_{ij} is generated using this random variable generator and assigned to different user-subchannel pairs.

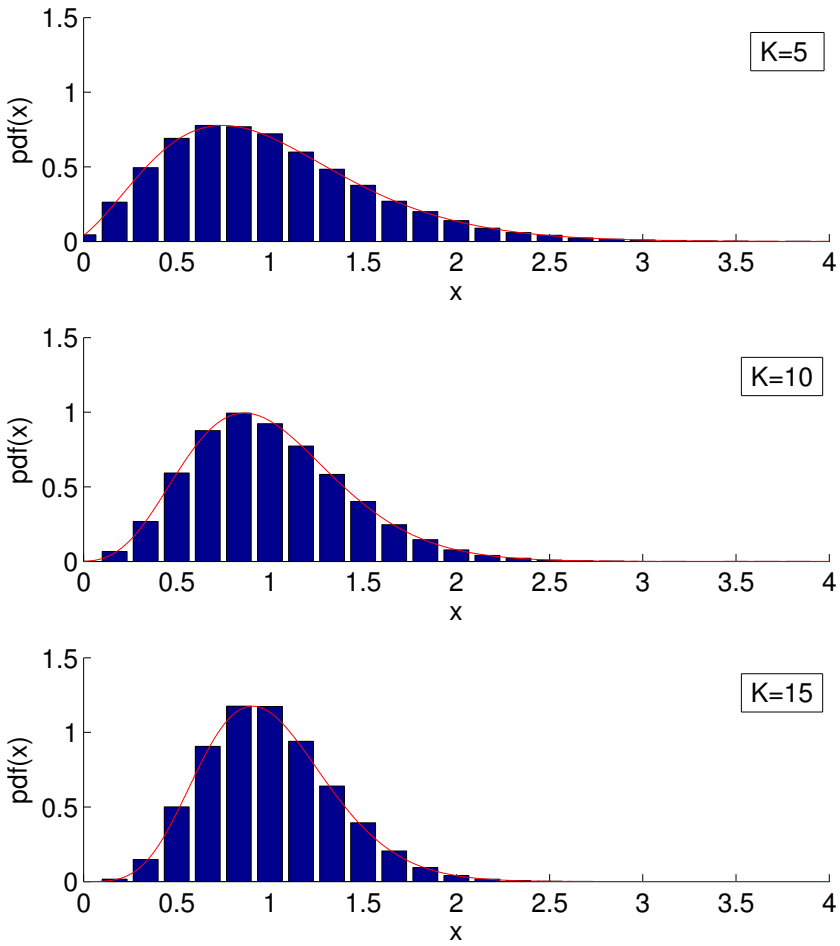

Figure 3.4: Ricean Gain Distribution

3.3.5 The Implementation of Max-Min Resource Allocation Algorithm

In this section, the detailed step by step implementation of the Max-Min optimisation Heuristic is introduced.

3.3.5.1 Step 1 - Power pre-allocation

The first step of the heuristic is to perform power allocation over subchannels by sharing P_{max} as uniformly as possible among all subchannels considering the power constraints in vector $\overline{P_j}$. The resulting power allocation P_j would be either the constraints power in $\overline{P_j}$ (subchannels with power constraints), or the same power as any subchannel which is not at its constraints limit. See Figure 3.5:

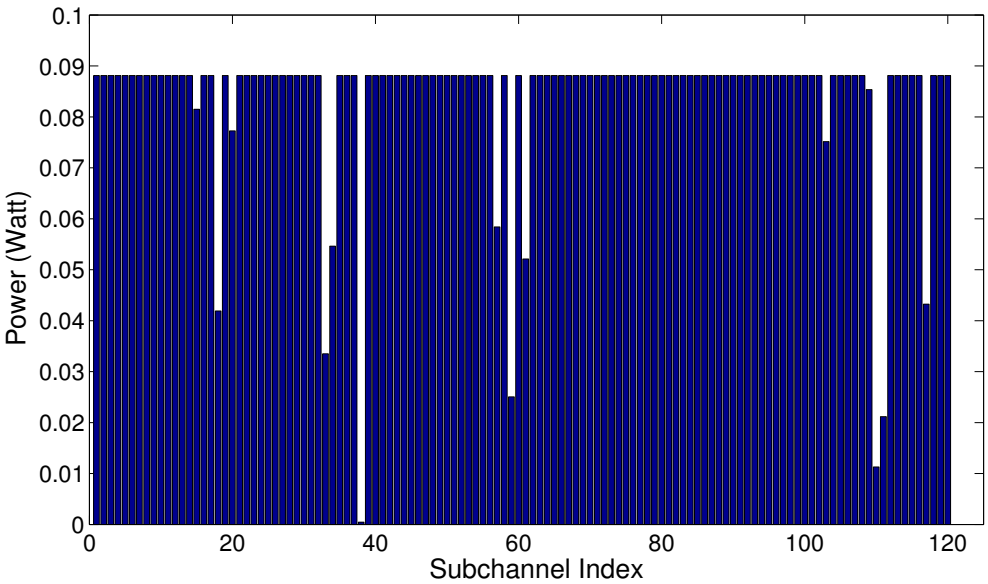

Figure 3.5: Power Allocation Map after Step 1

As can be seen from figure 3.5, subchannels such as $j = 38$, $j = 59$ are occupied by PUs and thus only allocated power same as the constraints $\overline{P_j}$. Other subchannels which are not occupied by PUs have equal allocated power. The total transmission power of the SNBS at different subchannels are:

$$\sum_j P_j = P_{max}, \qquad j = 1, 2, 3, \cdots M \tag{3.9}$$

With the transmission power allocated here, a maximum transmission rate for each of the SU in each of the subchannels can be obtained. Figure 3.6 and 3.7 shows the max transmission rate for SUs in all subchannels and the relationship between this rate and the distance of an SU from the SNBS. Note that the transmission rate appears to be unified along the sub-channels, this is because this result does not consider the fading effects yet, and the fading effect is included in later experiments. In that case, the transmission rate of different sub-channels for an SU are different, and an SU can choose the sub-channel with the highest transmission rate.

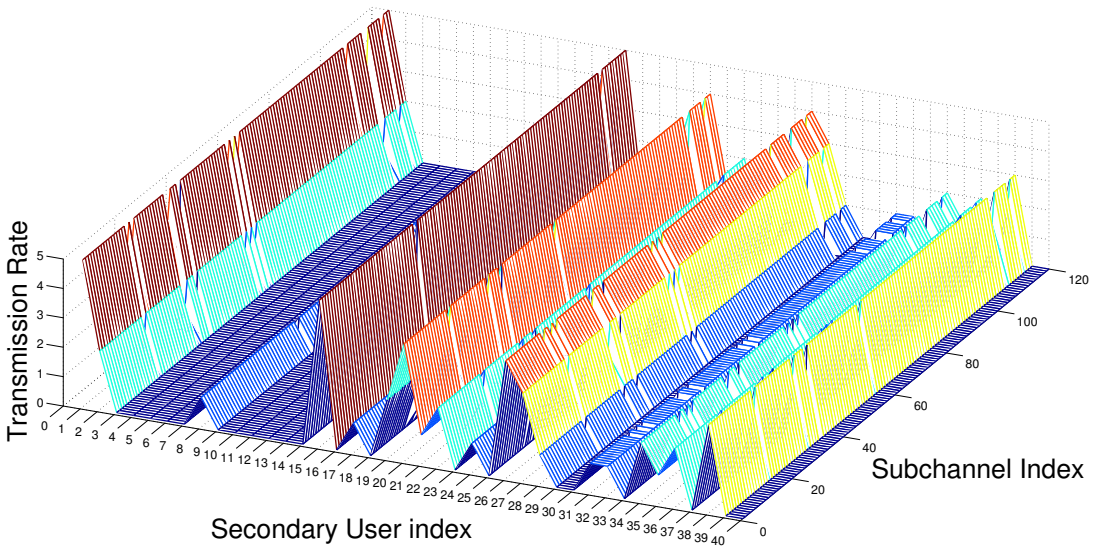

Figure 3.6: Transmission Rate Map under Power Constraints

We can see from Figure 3.7 that if an SU is nearer to the SNBS, it can achieve a higher transmission rate using almost all subchannels. However, in some subchannels which are occupied by PUs and thus with power constraints, all SUs only achieve very low transmission rate, see Figure 3.6.

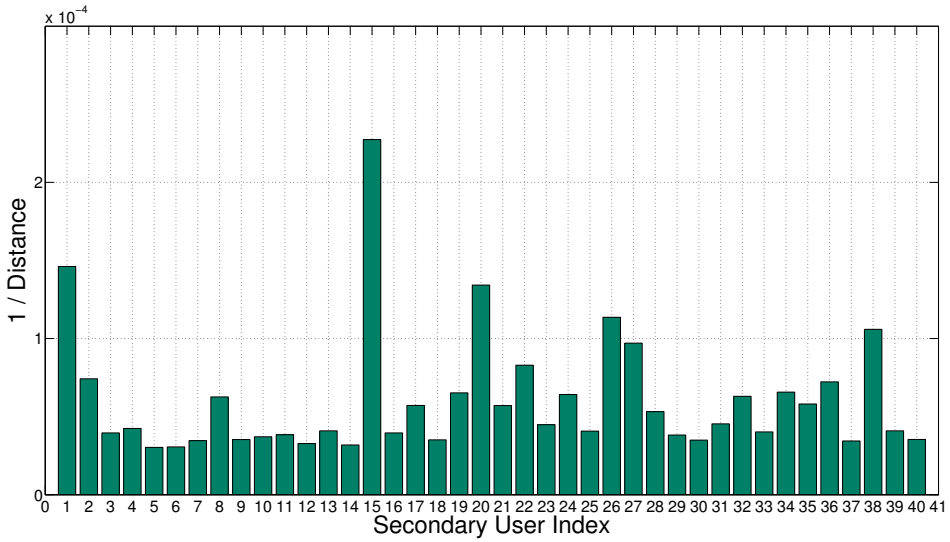

Figure 3.7: 1/Distance Relationship of Secondary Users

3.3.5.2 Step 2 - Allocation of subchannel-time slot pairs

In this step, subchannel-time slot pairs are allocated to SUs as resources. This is the initial creation process of the RA map. The resulting map will be a table with rows representing different subchannels and columns representing different time slots. Each element inside this table should be the SU number. See example table 3.3

As an example, SU3 in row 1 column 2 means that in this frame, at time slot 2, SU number 3 is allowed to download data from the SNBS using subchannel 1.

The subchannel-time slot pairs are allocated one after another iteratively. In each iteration, an SU with the lowest total transmission rate is allocated one subchannel-time-slot pair achieving the highest transmission rate subject to the subchannel power allocation in step 1. To clarify, in each iteration, the algorithm firstly chooses an SU with the lowest total transmission rate (in terms of the total transmission rate of this frame). Then for this SU, the achievable transmission rate in each subchannel-time slot pair is compared and the pair with highest transmission rate is selected and allocated to this SU. There may be ties in highest transmission rate for subchannel-time slot pairs, and this is broken in

Table 3.3: Example Resource Allocation Map

Time slot / Subchannel	1	2	3	⋯	L-1	L
1	SU5	SU3	⋯	⋯	⋯	⋯
2	SU7	⋱				
3	⋮		⋱			
⋮	⋮			⋱		
M-1	⋮				⋱	
M	⋮					⋱

favor of the highest channel gain. There may also be ties in the lowest total transmission rate for different SUs, and this is broken in favor of the smallest geometric mean $\overline{g_j} = \sqrt[|A|]{\prod_{j \in A} g_{ij}}$, where A is the set of remaining subchannels (not already allocated to an SU).

During this iterative allocation process, when SU i has received enough resources to satisfy its queue backlog, it is removed from the active SU list, and thus not allocated any resource in later iterations. In other words, these SUs are already satisfied and the queue backlog for these users are likely to be all transmitted in the next frame, so to continue allocating more resources to these users would be a waste, and would not further increase the target function which is the minimum transmission rate among SUs.

The iteration will continue allocating subchannel-time slot pairs, until in any iteration, the best subchannel cannot improve the total transmission rate of the SU under consideration. If this situation is reached, we say the algorithm reaches a saturation state, because the algorithm aims to maximize the minimum transmission rate of all SU, but when this state is reached, the total transmission rate of the SU under consideration (the one with lowest total transmission rate) cannot be further improved. Once the saturation state is reached, all the remaining available subchannel-time slot pairs are allocated to the SUs which are not satisfied in a round robin fashion in preparation for the next step.

Figure 3.8 shows an example of total transmission rate of different SUs after

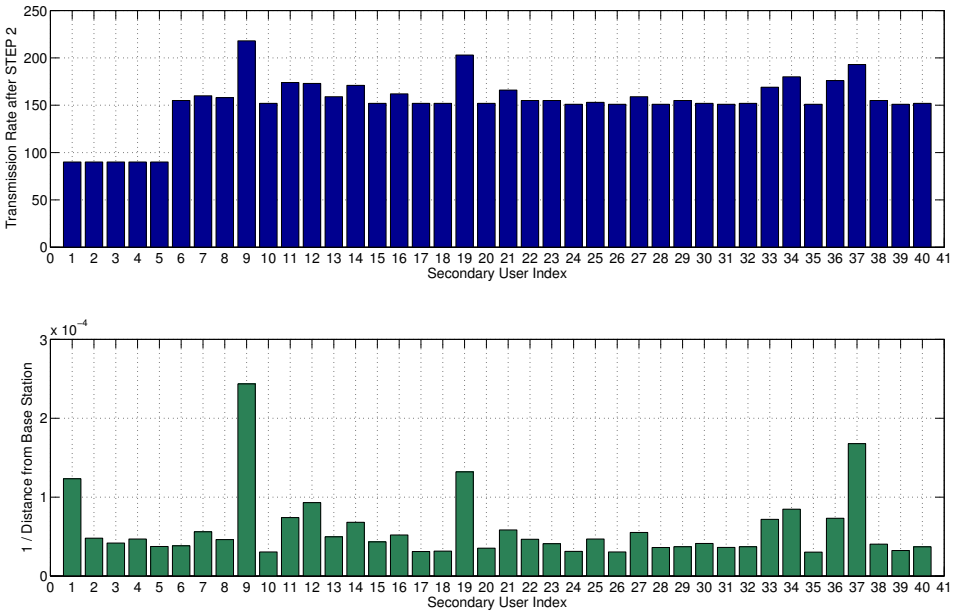

Figure 3.8: Transmission Rate of SU after step 2

Step 2. The fixed queue backlogs for SUs initialized at the beginning of each simulated frame are listed in Table 3.4

Table 3.4: Initial Backlog [2]

SU1 - SU5	90
SU6 - SU10	180
SU11 - SU15	270
SU16 - SU20	360
SU21 - SU25	900

We can see from figure 3.8 that for SU1 to SU10, their allocated transmission rate are clearly satisfied (enough for the transmission of remaining backlog for the next transmission frame) and thus not allocated more resource. However from SU11, all users are not satisfied. Comparing the rate of those unsatisfied users with their distance from the SNBS, we can see that users that are closer to the SNBS tend to get more resources than those far away from the SNBS. Here it is not very significant, but in the continuous simulation experiment introduced in later chapters, the effect of distance cumulate together through time and causes significant differentiation.

After each RA iteration when a subchannel-time-slot pair is allocated to an SU, the power allocated in step 1 is usually larger than the minimum transmission power required, which is $f_{ij}(z) = \gamma_z N_0 / g_{ij}$. For example if subchannel $j = 15$ is allocated to SU7, and the transmission mode in this pair for SU7 is $z = 4$, then $f_{7,15}(4) = \gamma_4 N_0 / g_{7,15}$ is likely to be larger than the power allocation scheme from step 1 in channel 15: $P_{j^*}, j^* = 15$. In this case, the remaining power $P_{15} - f_{7,15}(4)$ is allocated to the set of remaining subchannels in this time slot as evenly as possible considering the power limits due to the power constraints.

A comparison of the subchannel power allocation before Step 2 and after Step 2 is shown in figure 3.9

It can be observed that the previous evenly distributed power along different subchannels is now either decreased because the required transmission power is lower than the original allocated power, or increased because they are not allocated to any SUs and continuously receive residual power from other subchannels.

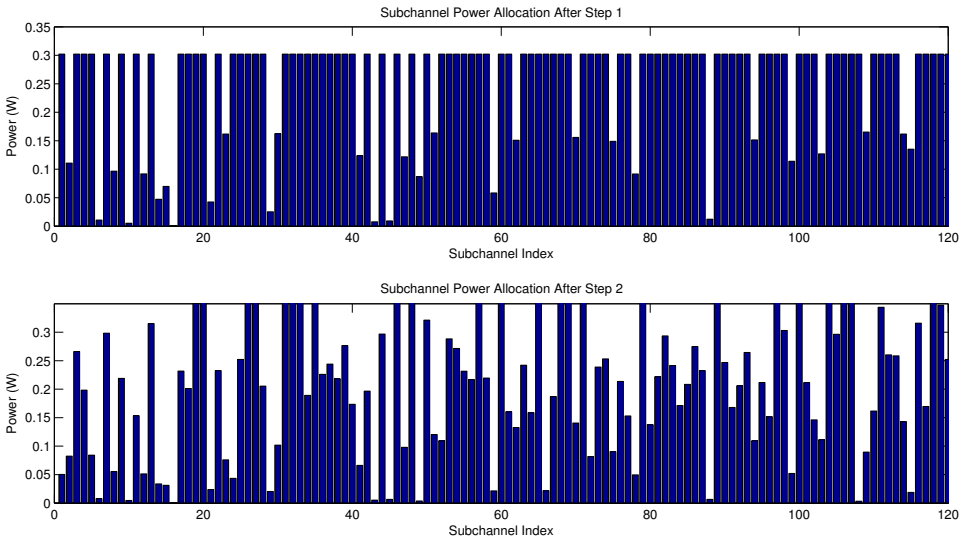

Figure 3.9: Power allocation of sub-channels after step 2

Note that power constraints still apply for subchannels occupied by PUs.

3.3.5.3 Step 3 - Further Improvement of Max-Min rate

With the subchannel-time slot pair allocation map from step 2, there is still potential Max-Min improvement by re-performing rate and power allocation. To do this, the algorithm sequentially increments the transmission mode of the SU with lowest transmission rate at the most power efficient subchannel-time slot pair that would not violate the primary protection constraints vector $\overline{P_j}$. The detailed steps are: firstly find out the unsatisfied SU with the lowest total transmission rate; then for all the subchannel-time slot pairs that are allocated to this user, find the one with the highest power efficiency ΔP_{ij}.

$$\Delta P_{ij} = \frac{[f_{ij}(z_{ij}+1) - f_{ij}(z_{ij})]}{R_{z_{ij}+1} - R_{z_{ij}}} \tag{3.10}$$

This is the extra power required to upgrade the transmission mode from z_{ij} to $z_{ij}+1$ over the transmission rate difference from z_{ij} to $z_{ij}+1$. In other words, this is the power efficiency (power difference over rate difference) of updating the

transmission mode by one.

In this updating process, whenever an SUs backlog is satisfied, this SU is removed from the update list and not considered any more. After several iterations of this updating process, it is likely that at some point all the resource pairs allocated to the unsatisfied SU with lowest total transmission rate cannot be further improved, and this is when step 3 terminates.

3.4 Experimental Results for Short Frame Time Optimisation Problem

Previously described steps in section 3.3 complete a simulation of the optimisation over one frame time.

In the simulation for short frame time optimisation problem, the RA map is calculated once, and the final allocated transmission rate results for all SUs are recorded. The randomness introduced to the experiments include position of PUs and SUs, and the channels conditions. To get a precise result (reduce the random effects of position distribution of users and channel conditions), for each of the P_{max} value, the simulation is repeated for 50 times and the minimum transmission rate is averaged. The simulation is also repeated for different subframe length F, with $F = 1$ and $F = 3$. The results of the averaged minimum transmission rate for SUs are plotted against total transmission power, and compared with the result from [2]. See Figure 3.10 and Figure 3.11.

Figure 3.10 is the simulation result of my simulation program and Figure 3.11 is the result from [2]. There are more results with Step 4 optimisation in Figure 3.11. In Step 4 limited perturbation on the sub-channel allocation is performed to improve the minimum rate among all SUs whose queue is not satisfied. Specifically, for each bottleneck SU (i.e., an SU which among all those whose queue is not satisfied, has a minimum rate), one subchannel-time slot pair is taken from a non-bottleneck SU, and reallocated to the bottleneck SU. In the research of this thesis, Step 4 is not implemented, because:

1. The extra benefit after Step 4 is not significant

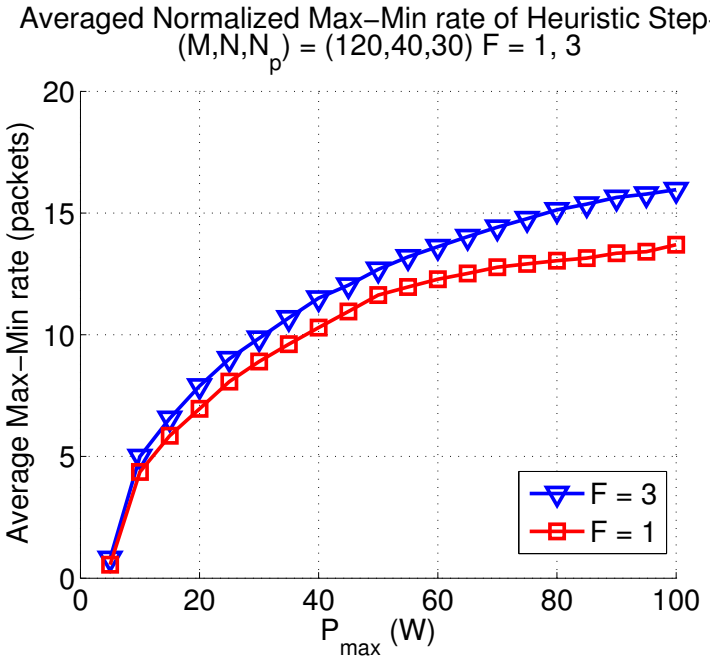

Figure 3.10: Simulation Result of Averaged Max-Min Rate

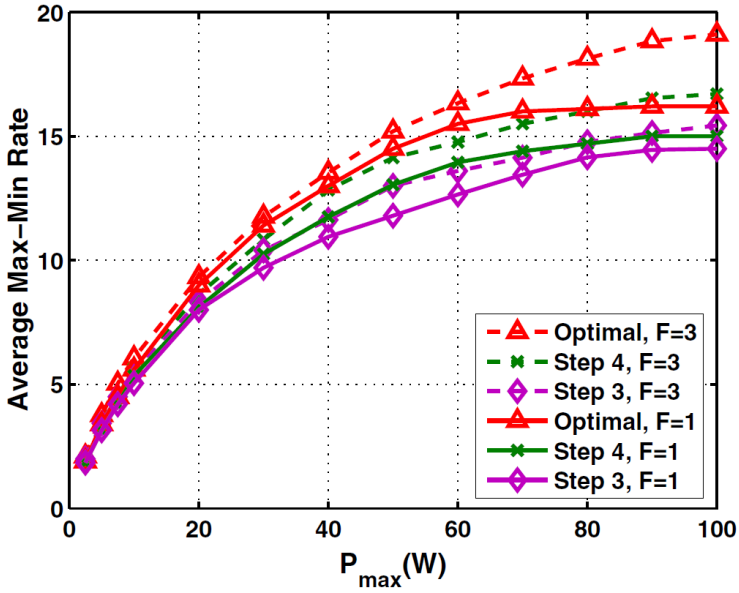

Figure 3.11: Simulation Result from [2] (unit in Packets)

2. Step 4 introduces extra computation complexity

3. The purpose of the research of this thesis is to evaluate and improve user experience rather than reaching optimal solution of RA for Max-Min rate

Comparing the two curves with the same parameter settings: Step 3, $F = 1$ and Step 3, $F = 3$, we can see that the results are very similar. This verifies that the simulation platform correctly implements the algorithm described in [2], and further extension and real time performance experiments can be done based on the simulation platform.

3.5 Conclusions

In this chapter, a optimisation problem over a transmission frame in OFDM based CR network is described. The scenario is adopted from the paper [2]. The reason for adopting this scenario is because it proposes a queue aware RA algorithm for short transmission frame time optimisation, and it can be extended to a continuous time evolving simulation. A simulation platform is built according to the scenario the described scenario in [2]. The numerical results obtained are compared with the that from paper to verify correct implementation of the scenario and the optimisation algorithm. This simulation platform provides a good basis for further experiments in the research of this thesis. The following chapters introduce extended model based on the short frame time optimisation problem and continuous time QoS and QoE evaluation based on the extended version of the simulation platform.

Chapter 4

Time Evolving Dynamic Simulation Platform for CR

4.1 Introduction

In most of the literature on RA for CR, the optimisation problems considered are limited in a short transmission frame. Extensive research has been carried out for such optimisation problems. In these scenarios, various algorithms and techniques are proposed to optimise the network performance parameters, mostly Physical/MAC layer metrics, such as system throughput, spectrum utilization, fairness and power consumption. In most cases, the proposed algorithms are designed to be executed repeatedly for different transmission frames to provide optimised network performance. However, optimisation at individual transmission frames does not guarantee the overall continuous QoS of users, such as average packet delay and loss. These packet layer metrics are not considered at all in most of these proposed RAs, but they are actually very important performance indicator of the network, and are vital to the success of the technique. Evaluation of real time packet level QoS of CR networks is urgently needed.

Some other factors, which cannot be taken into account with model of optimisation problem limited to a single transmission frame, may also have great impact on system performance. Users in the CR networks are likely to be using different types of service over CR, and this could result in very different traffic types being

transmitted over the network, which may have great impact on RA and network performance. CR networks share spectrum resources with PUs, whose activities are usually dynamic and can introduce fluctuations to the available resource to CR, impacting on user QoS. All these factors could not be taken into account with short transmission frame optimisation scenarios. It is necessary to extend the short transmission frame optimisation scenarios introduced in Chapter 3 to a time evolving dynamic model, so that the dynamic factors can be taken into account and their effect on user QoS and network performance can be evaluated.

In this chapter, the model for time evolving dynamic experiments and the extended simulation platform are introduced. Long term evaluation results including queue length, packet delay and loss are demonstrated.

4.2 Time Evolving Dynamic Model

Table 4.1 summarises a comparison between simulation of the traditional optimisation problems in literature (such as the one introduced in Chapter 3) and the time evolving dynamic simulation introduced in this chapter.

Table 4.1: Comparison of Simuations

	Traditional Optimisation Simulation	Time Evolving Dynamic Simulation
Simulation Period	Single Transmission Frame	Cascaded Transmission Frames
SU Backlog	Fixed, using the same value for every frame	Carried forward to next frame
Traffic Model	None. Saturated Backlog	Packet arrival process with different characteristics

4.2.1 Cascading Frames

In the scenario introduced in Chapter 3, the short frame time experiments are simply repeated with the same settings of parameters for an averaged result.

In order to evaluate packet level system performance in a continuous manner rather than just focusing on the optimisation in one frame time, the system model needs to be modified to test the resource allocation algorithm in consecutive frames. Instead of simply repeating experiments in one frame time, the frames need to be cascaded, and the parameters such as user backlog and user activities needs to be carried on to the next transmission frame. This would allow the conditions to evolve through time, as happens in practical CR networks.

Below is a list of situation comparison and changes made to extend the original short frame time simulation platform to adapt to time evolving dynamic evaluation.

Queue Backlog of SU

In short frame time simulation, each of the secondary users is assigned a fixed length of backlog at the beginning of the simulation. Different length of backlog indicates different needs for downlink transmission resource. In practical cases, the queue backlog evolves according to the traffic intensity and the downlink transmission rate allocated to the user. So in the time evolving dynamic simulation, instead of assigning fixed backlog, a packet buffer is created for each of the SUs. And packets are inserted to and flushed from the buffer in each time slot. The packets inside the buffer contains information of the time when it arrives at the buffer, so that the waiting time for a packet inside the buffer can be tracked. The queue backlog for the next transmission frame for user i can be expressed as [2]

$$q_{i,t+1} = max\left\{0, q_{i,t} - x_{i,t}\right\} + a_{i,t}, \tag{4.1}$$

where $a_{i,t}$ is the number of packets arriving in frame t for user i and $x_{i,t}$ is the number of packets transmitted to user i.

User Positions

In short frame time simulation, the positions of PUs and SUs are re-initialized at the beginning of each simulation iteration to eliminate effect of user position. In real time cases, proper mobility pattern of users should be considered. For simplicity, a set of positions for PUs and SUs are generated

uniformly distributed in the cell for once at the beginning of the simulation, and are fixed for the rest of the simulation. So all users stays at the same position during the whole simulation process[1].

Wireless Channel

The wireless channel conditions are still characterized by path loss and Ricean fading with quasi-static assumption. The wireless gains are re-initialized at the beginning of each simulation time frame, and remain unchanged during the frame time.

Resource Allocation Heuristics The heuristic for fast calculation of RA map is still based on one frame time. However, it now makes use of the inputs which evolve from frame to frame, such as queue backlog, wireless channel condition, and power constraints from PUs whose activities are also evolving through time.

4.2.2 Secondary User Activity Modelling

To evaluate system performance for users with different service types, SUs within a secondary network are classified into two types: Real-Time-Users SU_{RT} and Non-Real-time-users SU_{NRT}. In the time evolving dynamic simulation, packets are assumed to arrive at the beginning of each time slot, and be completely transmitted at the end of each time slot.

For SU_{RT}, traffic arrival pattern is characterized by a Poisson process with intensity λ_{RT}. So on average λ_{RT} packets will arrive at the beginning of each time slot to SNBS and buffered for SU_{RT}.

For SU_{NRT}, the traffic arrival pattern is characterized by an On-Off model, so that a bursty traffic pattern for non-real time users can be generated. To generate bursty traffic, the transition probability between On and Off states needs to be correctly set so that it stays in Off state for some time and jumps to On state for a short while and jumps back to Off state. Denote the transition probability from On state to Off state as P_{On-Off}, and from Off state to On state as P_{Off-On}. Figure 4.1.

[1]In here, a simple case with fixed position is considered to focus on the QoS evaluation

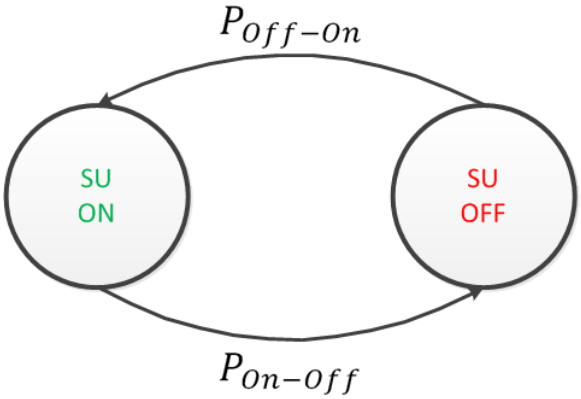

Figure 4.1: NRT Secondary User Markov Chain

When SU_{NRT} is in an On state, its packet arrival pattern is still characterized by a Poisson process, with intensity λ_{NRT}. In this case, the overall average packet arrival rate for SU_{NRT} will be

$$\overline{\lambda_{NRT}} = \frac{\lambda_{NRT} \times P_{Off-On}}{P_{Off-On} + P_{On-Off}} \tag{4.2}$$

An example of traffic arrivals for SU_{RT} and SU_{NRT} is illustrated in Figure 4.2

4.2.3 Primary User Activity Modelling

In most of the network optimisation scenarios proposed in literature, the modelling of PUs is static. In some cases, the interaction of SUs and PUs are simplified so that only one PU and one SU is considered. However, in real cases, multiple SUs share spectrum resource with multiple PUs. And the activity of multiple PUs combined together may have severe impact on resource available to secondary network. In this thesis, I would like to explore how the PU activity would affect the performance of secondary network, and thus the PU activities can no longer be assumed to be static.

There are many proposals for activity modelling for PUs, as introduced in the literature review section 2.4. The exact modelling of PU activities could be

itself. In later chapters, user mobility model is also taken into account

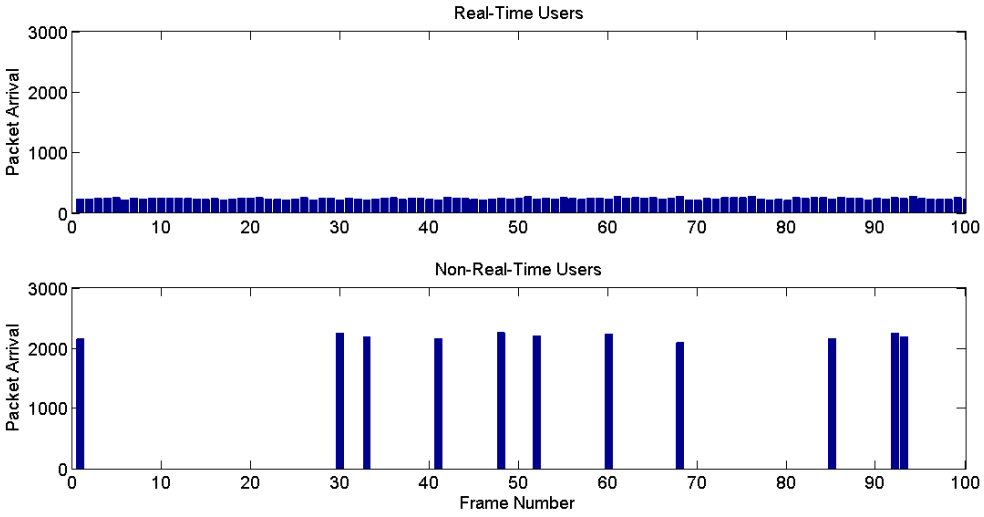

Figure 4.2: Comparison of Traffic Pattern for NT and NRT SU

very complicated and also depends on which spectrum band that CR operates in. The simplest model for PU activity is a two state Markov On Off model, with exponentially distributed dwelling time for each of the states. Figure 4.3. A PU still occupies one subchannel when it is in On state. But when it switches to the Off state, it releases the subchannel it occupies and other users could transmit in the subchannel without power constraints.

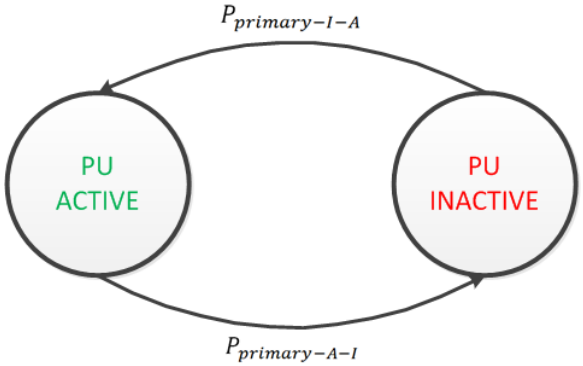

Figure 4.3: Primary User Markov Chain

The transition probability from active to inactive is $P_{primary-I-A}$, and from

inactive to active is $P_{primary-A-I}$. An example of activity of one PU is illustrated in Figure 4.4

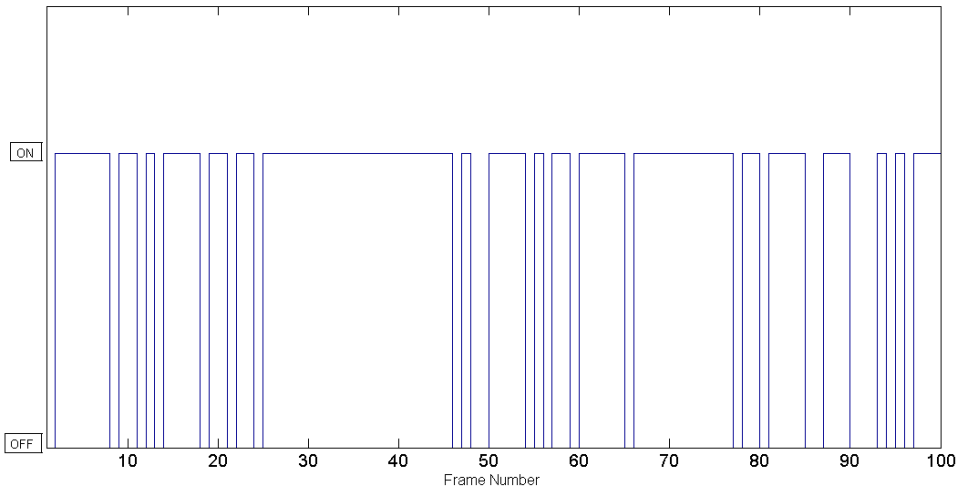

Figure 4.4: Example Activity of one Primary User

Suppose there are totally N_p PUs, then at a certain instance, the number of active PU is a random value, Figure 4.5, and the expected value can be calculated by

$$\overline{N_p} = N_p \times \frac{P_{primary-I-A}}{P_{primary-I-A} + P_{primary-A-I}} \tag{4.3}$$

The calculation of expected number of PU is very important because this decides the average resource available to the secondary network. And also, the simulation should start with number of active PUs equals to $\overline{N_p}$, in order to achieve steady state quicker.

The exponential distribution of dwelling time is the simplest, but probably not the most accurate model, for the modelling of PU activities. Though for now only exponential distribution is applied, the consideration of other heavy tailed distribution of inter-event times is introduced in Chapter 5.

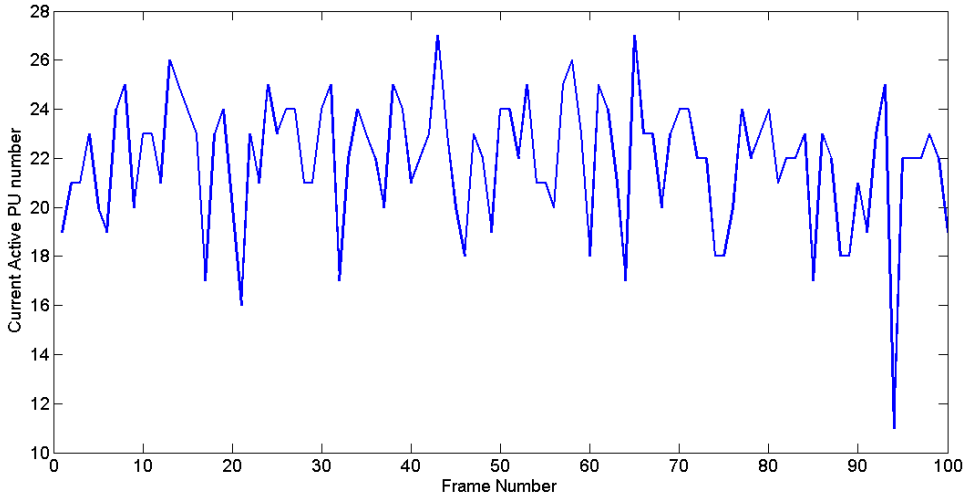

Figure 4.5: Fluctuation of Total Primary User Activities

4.2.4 Time Evolving Dynamic Experiment Process

Table 4.2 shows the parameters used in the time evolving dynamic experiments. Most of the parameters used are the same as those in the short frame time simulation, however, several new parameters are introduced to characterize evolving activities.

Figure 4.6 shows a flowchart illustrating the time evolving dynamic simulation process.

At the beginning of the simulation, the position of all users, including PUs and SUs are randomly initialized, using the same distribution equation (3.1, 3.2). The SUs are initialized with empty buffers. Then, channel conditions for both SUs in each of the subchannels are initialized as described in section 3.3.4. At one frame, the number of active PUs is a random value. The number of active PUs should be initialized as the expected value, which is calculated by equation 4.3. In this case $\overline{N_p} = 8.18$, so 8 PUs are initialized as active, while the rest 22 PUs are initialized as inactive.

At the beginning of each frame, the activity states of PUs and NRT SUs are decided according to the Markov chain transition probability. And the power constraints are re-calculated according to the activity states of PUs.

Table 4.2: Parameter summary for Time Evolving Dynamic Experiments

Parameter	Value	Unit	Explanation
M	120	1	Total subchannel number
N	40	1	Total SU number
N_p	30	1	Total PU number
K	-10	dB	K factor of Ricean distribution
η	3	1	Path loss exponent
d_0	50	m	Far-field crossover distance
r_1	33	km	Max distance from SU to BS
r_2	60	km	Max distance from PU to BS
N_0	-100	dBW	Gaussian noise Power
ω	0	dB	Critical interference threshold
L	30	slot	Frame length
F	1, 3	slot	Sub-frame length
Δt	30	ms	Time slot Length
P_{max}	30	W	Maximum transmission Power from BS
$P_{primary-A-I}$	0.3		Transmission probability from active to inactive
$P_{primary-I-A}$	0.8		Transmission probability from inactive to active
P_{On-Off}	0.8		SU transition probability from on to off
P_{Off-On}	0.1		SU transition probability from off to on
λ_{RT}	variable	Packets	Traffic Intensity of Real-Time SUs
λ_{NRT}	variable	Packets	Traffic Intensity of Non-Real-Time SUs

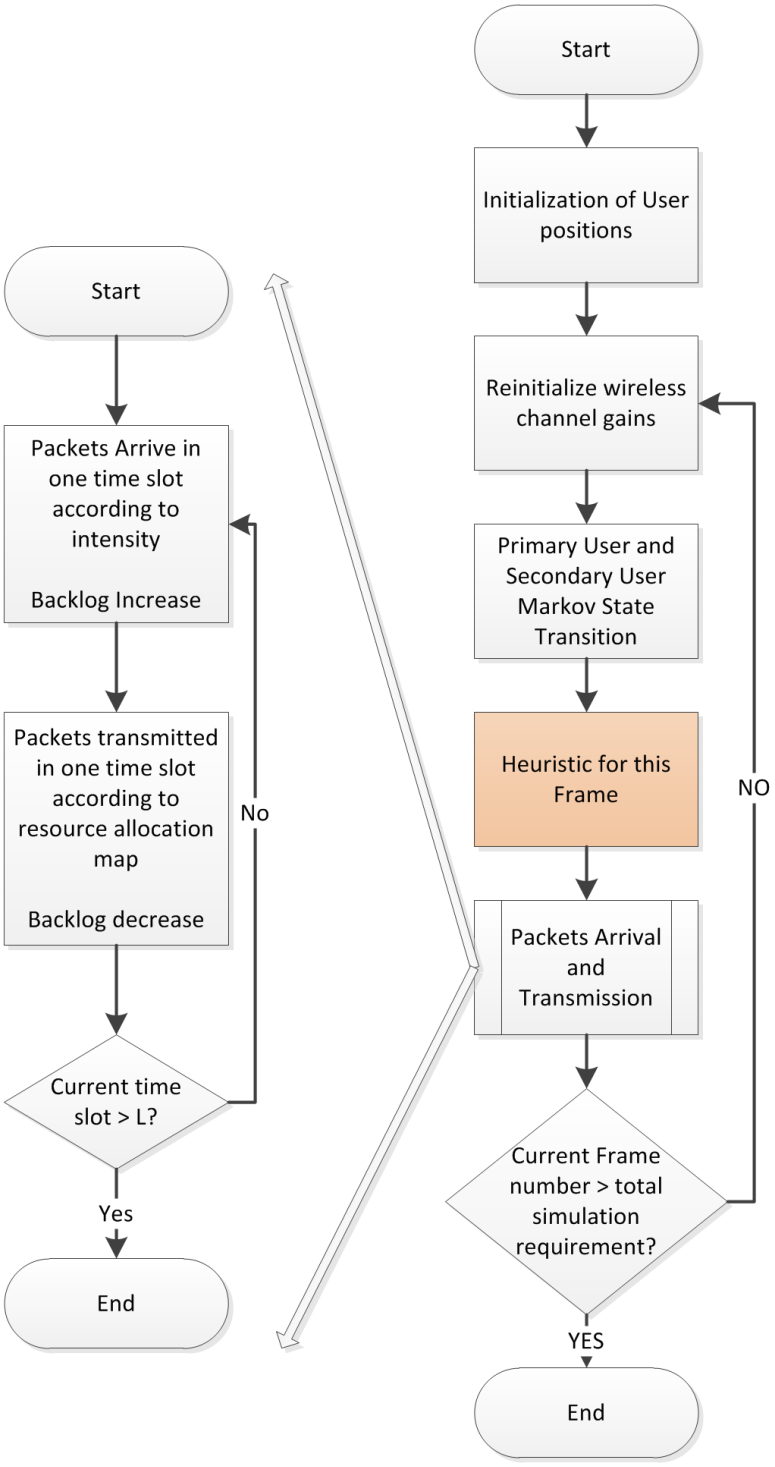

Figure 4.6: Flowchart of Time Evolving Dynamic Simulation Process

The RA heuristic is executed once, taking into account the evolving parameters including the backlog length of each of the SUs, the channel conditions, and the power constraints decided by PU activity situation. The RA map is produced as the result of the heuristic. According to the RA map, the downlink transmission rate (packet/slot) for each of the SUs can be calculated.

Next, the packets arrival and transmission process is simulated. For each of the SU, new packets are inserted at the end of the buffer queue.

$$q_{i,t+1} = max\left\{0, q_{i,t} - x_{i,t}\right\} + a_{i,t}, \tag{4.4}$$

,where $a_{i,t}$ is the number of newly arrived packets decided by the traffic intensity λ_{RT} and λ_{NRT}.

$$a_{i,t} = \begin{cases} \mu(\lambda_{RT}) & \text{when } i \text{ is a RT user} \\ \mu(\lambda_{NRT}) & \text{when } i \text{ is an NRT user and is in On state} \\ 0 & \text{when } i \text{ is an NRT user and is in Off state} \end{cases} \tag{4.5}$$

where $\mu(\lambda)$ is a Poisson random variable with average value of λ. The transmission rate for an SU is the sum of transmission rates on all subchannels for this user.

$$x_{i,t} = \sum_{j=1}^{M} r_{ij} \tag{4.6}$$

In order to keep track of the average waiting time for packets inside the buffer, when a packet n enters the buffer, the time instance t_{in}^{n} is recorded in the packet. When the packet gets transmitted from the buffer, the time instance t_{out}^{n} is also recorded. Suppose in total N_t packets are transmitted for a user, the expected waiting time for a packet in the buffer can then be calculated by

$$\frac{\sum_{n=1}^{N_t} t_{out}^{n} - t_{in}^{n}}{N_t} \tag{4.7}$$

The buffer size is originally set to be infinity to observe the trend of waiting time and queue length. In order to evaluate the loss probability of a packet due to buffer saturation, the buffer size is limited to a finite value. And when new

packets arrives while the buffer is full, these new packets are dropped and the number of lost packets are counted.

By this point, the simulation of a frame time is finished. The backlog length for all SUs are put into an array as the input for RA heuristic of the next frame. Other parameters such as packet loss and user activities are also recorded into files at this point for later analysis. The simulation of one frame time is then repeated for a large number of times for averaged results. In order to obtain accurate enough results, it is necessary to carry out experiments for enough number of times. Appendix B introduces research on simulation run length planning.

4.3 Experimental Results for Time Evolving Dynamic Evaluation

In this section, several numerical results are presented.

4.3.1 Distance and Average Queue Backlog

4.3.1.1 Real Time Users

The first interesting fact that is discovered is that the distance from an SU to the SNBS has a significant impact on QoS of SU in a cognitive radio network. The experiments were done separately for RT and NRT SUs.

Firstly the effect of user distance from the SNBS on RT users is evaluated. Considering one RT user, and let the distance of this RT user vary

$$D^*_{RT} \in \{3, 6, 9, \cdots 33\}\, km \tag{4.8}$$

while the position of all other SUs remains unchanged. For each of these distance points, the simulation is run for 3 hours, 12000 frames. The queue backlog length of each of the SU is observed once per frame, and are recorded and then averaged to obtain a final result.

The simulation is then repeated 3 times, each with different global traffic intensity. The traffic intensity values tested for SU_{RT} are $\lambda_{RT} = 8$, $\lambda_{RT} = 8.5$, $\lambda_{RT} = 9$, units are packets/time-slot.

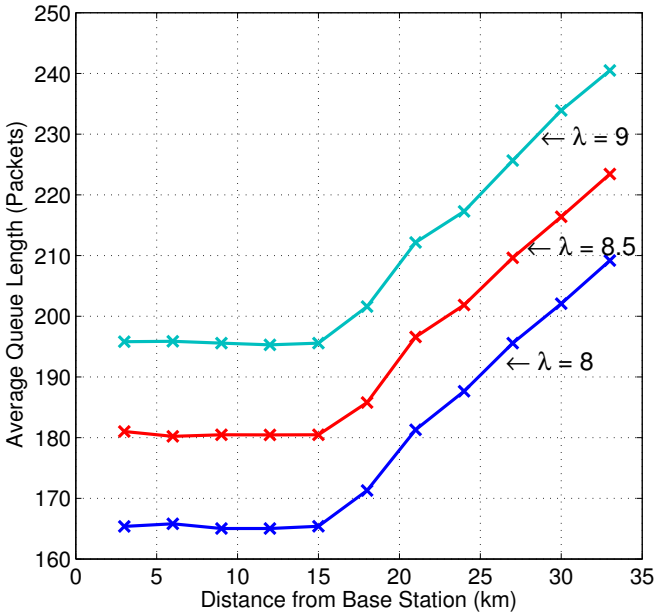

Figure 4.7: Average Queue length vs. Distance for RT SU

In figure 4.7, the queue backlog increases as distance from the SNBS increases. However, the increase is not very significant. Comparing a user at 33 km distance and a user at 15 km distance, the average backlog is increased by around 50 packets. The traffic intensity λ_{RT} also have impact on the queue backlog, and at the interest point $\lambda_{RT} = 8.5$, the relationship between queue backlog and λ_{RT} seems linear.

4.3.1.2 Non Real Time Users

The same simulation process is repeated for NRT users, with distance from SNBS,

$$D^*_{NRT} \in \{3, 6, 9, \cdots 33\}\, km \tag{4.9}$$

The simulation is also repeated 3 times, each with different global traffic intensity. The traffic intensity values tested for SU_{NRT} are listed in table 4.3 comparing with that of SU_{RT}. The traffic intensities are set as such so that the average traffic intensity for SU_{RT} and SU_{NRT} are the same, and the results are

comparable.

Table 4.3: Traffic Intensity

$\overline{\lambda}$	λ_{RT}	λ_{NRT}
8	8	72
8.5	8.5	76.5
9	9	81

In Figure 4.8 we can see that for NRT users, the average queue backlog increases significantly as distance from SNBS increases. The queue length difference between users at 33km distance and users at 15km distance can be up to nearly 1000 packets, which is a huge difference for user experience. And also, the relationship between queue backlog and λ seems to be non-linear at around $\lambda = 8.5$. The backlog increases much faster as λ becomes larger.

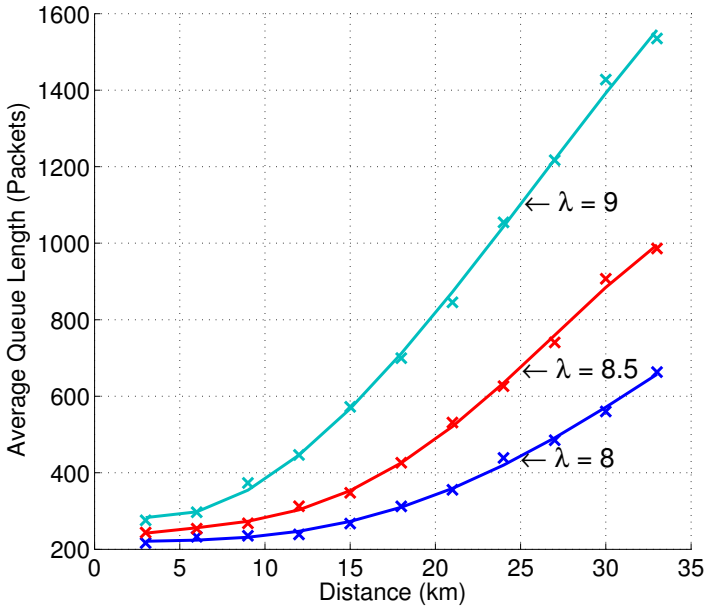

Figure 4.8: Average Queue length vs. Distance for NRT SU

4.3.1.3 Comparison of RT and NRT Users

Figure 4.9 shows a comparison between SU_{RT} and SU_{NRT}. There is a very significant discrimination between RT SUs and NRT SUs. Though queue backlogs for those near to SNBS are almost of the same length, users far away from the SNBS see a significant difference between RT users and NRT users. This would result in SUs with bursty traffic pattern to experience much longer delay time, and also a larger probability of packet loss.

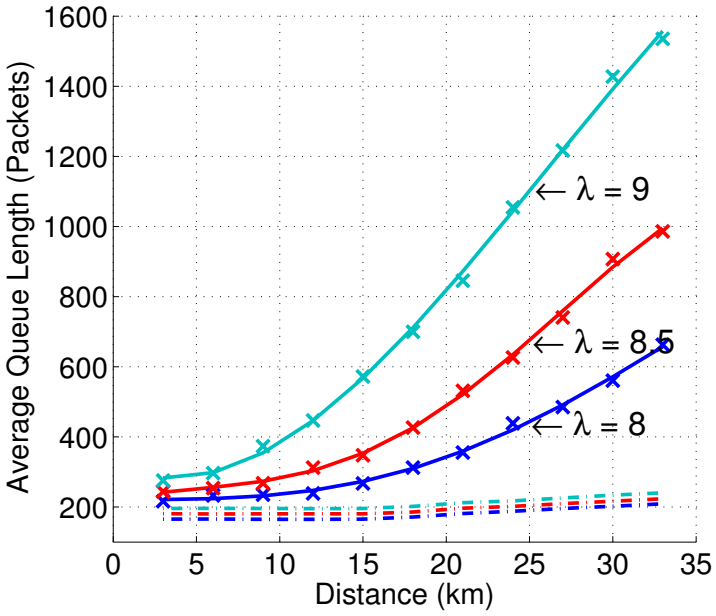

Figure 4.9: Average Queue length vs. Distance for NRT SU

4.3.2 Distance and Loss Probability

To evaluate the loss probability, the same parameters are used as for the queue length versus distance experiments. The difference is that the buffer size of all SUs are set at 5000 packets. So if the backlog length exceeds 5000, the SU buffer will overflow and newly arrived packets will be lost. The same experimental process is repeated as in evaluation of queue backlog, and the packet loss numbers are recorded.

For SU_{RT}, no packet loss is observed for the given traffic intensity, which means that during the simulation time, the buffers for SU_{RT} never overflowed.

For SU_{NRT}, queue backlog sometimes exceed buffer size and thus results in packet loss. The results of packets loss are plotted in Figure 4.10. Again, we can observe that the distance from SU to the SNBS has a great impact on packet loss probability. Secondary users that are further away from the base station are likely to experience very high packet loss and deteriorated QoS.

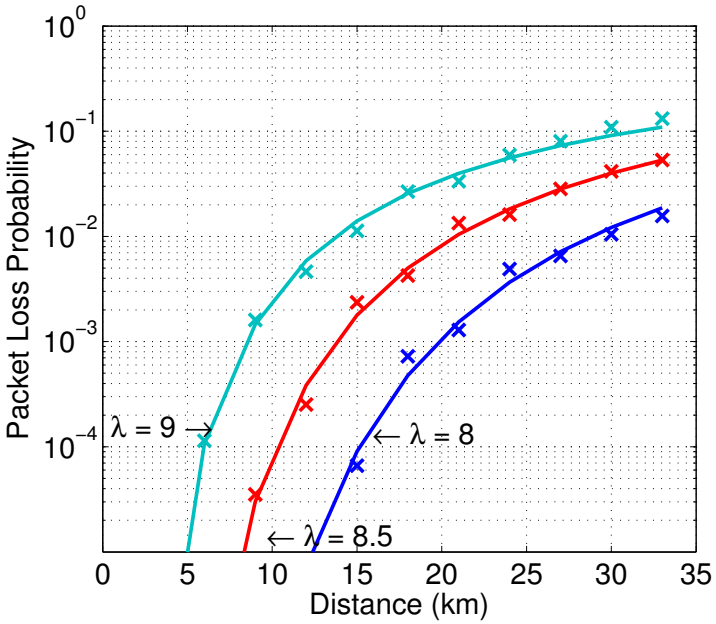

Figure 4.10: Average Queue length vs. Distance for NRT SU

4.3.3 Traffic Intensity and Average Queue Backlog

Another experiment is performed to evaluate the effect of global traffic intensity over average queue length. All SU's positions are fixed, and average traffic intensity λ_{RT} and $\overline{\lambda_{NRT}}$ varies from 7 to 11 packets per time slot. For each of the traffic intensity value, simulation runs for 10,000 frames to obtain the average result. Figure 4.11 shows the results of the simulation.

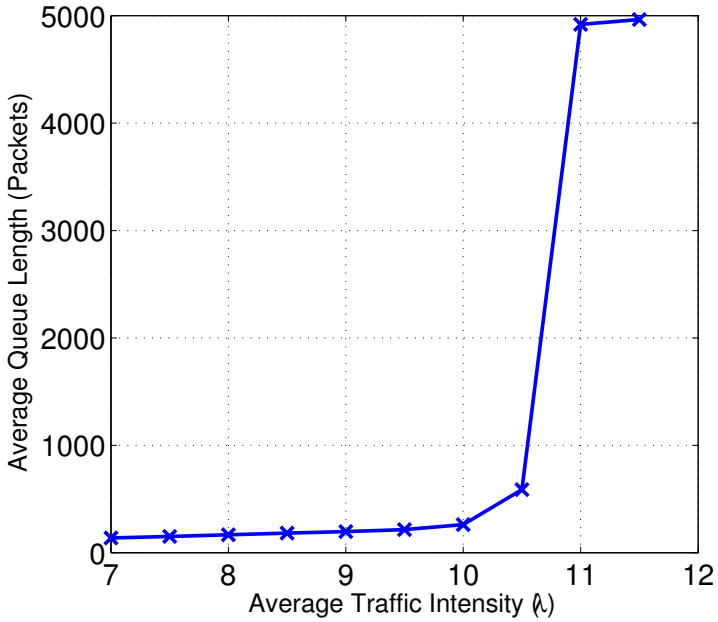

Figure 4.11: Average Queue Length vs. Global Average Traffic Intensity

The average queue backlog increases dramatically at around $\lambda = 10.5$, indicating that the global traffic intensity exceeds the SN system capacity.

4.4 Conclusions

The simulation platform is extended to be capable of real time evaluation of secondary network performance parameters. Experiments are designed to examine the real-time queueing performance of an existing RA scheme for downlink OFDM-Based CR network. Simulation results show that packet level QoS parameters for SUs in the system vary significantly with the distance from the SNBS. These results demonstrate the very significant impact of both packet level traffic arrival process, and user distance from the SNBS. The Max-Min RA scheme tested in this chapter discriminates against users with bursty traffic profiles and SUs at a large distance, meaning that the Max-Min RA scheme is ineffective in providing uniformly good user QoS. And it is reasonable to believe that other similar Physical/MAC layer focused RA schemes have similar problems. This

clearly shows that any future work on CR system design should also consider packet level network performances and packet level user QoS.

Chapter 5

Primary User Effect on Secondary User QoS

5.1 Introduction

In a CR network, SUs share spectrum resource with licensed users, and make use of the resource as long as no interference is caused to licensed users. This would require CR base stations and users to have the ability to sense the activities of the licensed users. There is extensive research on spectrum sensing techniques, and there is still research going on trying to find more efficient and accurate methods. In this research, perfect spectrum sensing is assumed for SNBS and SUs, so the SN always have knowledge of activities of all PUs within the same spectrum band.

One important characteristic of CR networks to distinct from other traditional networks is that the available resource to the users is not constant. Since SNs share spectrum resource with licensed users, and licensed users always have priority in spectrum accessing, the resource available to SN is likely to be highly fluctuating due to the spectrum usage of PUs. The characteristic of available spectrum resource to SN depends on the activity pattern of the PUs, with whom SUs share the spectrum. The available resource to SN may have great influence on both the system capacity, and the network performance. Consequently, the activity pattern of PUs is very likely to be a main contribution factor to the SN

performance and users' QoS and QoE.

The activity patterns of PUs are dependent on the characteristics of the primary network, and there is research on activity patterns on different spectrum bands [34, 35, 37, 38]. Designing and implementing CR network sharing spectrum with different types of primary network with different PU activity patterns can be difficult and time consuming. In this thesis, I want to focus on general characteristics of PU activity such as overall user activity intensity and user activity patterns, and also evaluate the difference when different PU activity models are employed. These factors can be related to specific practical primary networks, so that the evaluation results obtained in this research can be a reference for designing of CR networks in the future.

In this chapter, the evaluation of the effect of different PU activity, patterns and modelling on the performance of SN is introduced. Specifically, QoS metrics including packet level delay and loss for SUs are still the main concern.

5.2 PU Activity Pattern Modelling

The target of this chapter is to explore how the pattern and modelling of PU activity can affect SN performance and SU QoS. It is clear that the intensity of PU activity decides the average level of available resource to the SN, and there isn't much need to study the effect of average PU activity intensity because it can be treated equivalently as available bandwidth. On the other hand, the effect of PU activity pattern, specifically PU activity burstiness, on SN performance remains unknown, and yet is important to study.

The PU activity state model used in these experiments is the same as in section 4.2.3. The Markov On-Off model with exponentially distributed dwelling time for each of the states is the most basic way to model PU activity.

Both the PU activity intensity and the burstiness of activity can be decided by the transition probability. The activity intensity is decided by the average number of PUs in the active state. The average number of PU can be calculated by equation 4.3. The higher the value of average number of PU $\overline{N_p}$, the higher is the activity intensity.

The parameter of interest is the burstiness of PU activity. And one easy

way to approach different levels of burstiness is using different values for the transition probability between On-Off states. If the value of $P_{primary-I-A}$ and $P_{primary-A-I}$ are higher, then the frequency of transition between state is higher, resulting in more regular transitions and lower burstiness. And lower values of the transition probability will result in longer sojourn times in each state and thus higher burstiness.

To ensure that the activities with different burstiness are comparable, the activity intensity must be kept the same, but the activity intensity depends on both the transition probability values, so the two values are dependent. The relationship between the transition probability $P_{primary-A-I}$ and the probability of a PU being in On state P_{on} is shown in equation 5.1.

$$P_{primary-A-I} = \frac{(1 - P_{on})}{P_{on}} \cdot P_{primary-I-A} \tag{5.1}$$

Where P_{on} is the probability of a PU being in the On state.

Using equation 5.1, a set of transition probability value pairs can be calculated; each pair of probability values represents a different level of burstiness. Using this method, exponentially distributed Off state dwelling times for each of the PUs can be generated.

A recent paper suggests that exponentially distribution of inter-activity times may not be an accurate enough model for activities driven by humans[12]. Such activities are better approached by heavy tailed distributions, such as Power law distributions. Figure 5.1 shows an example comparison between Exponential distribution and Pareto distribution. The two distributions shown in the figure has the same mean value, but the Pareto distribution is much more heavy tailed (larger variance) compared to the exponential distribution. If the Off event time of a PU follows a Pareto distribution instead of an exponential distribution (even though with the same mean value), this change in the modelling of PU activity can affect SN performance[1]. The numerical evaluation of the effect of employing different PU activity model will be shown in section 5.3.5.

[1]There is study on the Power law characterized traffic patterns and queue patterns such as [62, 63], which has a similar concept, but is different from the scope here.

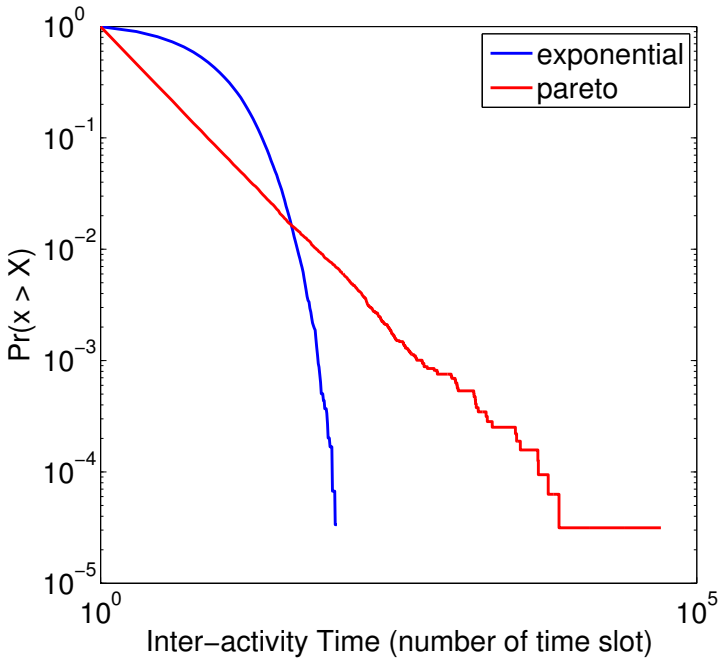

Figure 5.1: Comparison of Exponential and Pareto Distribution

5.3 Numerical Experiments

5.3.1 Experiment setup

Firstly, a set of transition probability pair values are calculated. Suppose the probability for a PU to be on $P_{on} = 0.7$, and a set of values are given for transition probability from Inactive state to Active state is $P_{primary-I-A}$, then the transition probability from Active state to Inactive state can be calculated by equation 5.1. A set of values used in the simulation are listed in Table 5.1.

Other parameters used in the simulation are listed in table 5.2.

5.3.2 Effect on total throughput

Different levels of burstiness of individual PU activities will result in different overall activity patterns, and consequently, the resource available to the CR network will have different characteristics. Figure 5.2 shows an example of the overall

Table 5.1: State Transition Probability Pairs

$P_{primary-I-A}$	$P_{primary-A-I}$
0.1	0.043
0.2	0.086
0.3	0.127
0.4	0.171
0.5	0.214
0.6	0.257
0.7	0.300
0.8	0.343
0.9	0.386

Table 5.2: Parameter summary for PU Activity Effect

Parameter	Value	Unit	Explanation
M	30	1	Total subchannel number
N	12	1	Total SU number
N_p	30	1	Total PU number
K	-10	dB	K factor of Ricean distribution
η	3	1	Path loss exponent
d_0	50	m	Far-field crossover distance
r_1	33	km	Max distance from SU to BS
r_2	60	km	Max distance from PU to BS
N_0	-100	dBW	Gaussian noise Power
ω	0	dB	Critical interference threshold
L	30	slot	Frame length
F	1	slot	Sub-frame length
Δt	30	ms	Time slot Length
P_{max}	30	W	Maximum transmission Power from BS
$P_{primary-A-I}$	variable		Transmission probability from active to inactive
$P_{primary-I-A}$	variable		Transmission probability from inactive to active
λ_{RT}	variable	Bytes	Traffic Intensity of Real-Time SUs

system throughput variation (through time) under different PU activity level of burstiness.

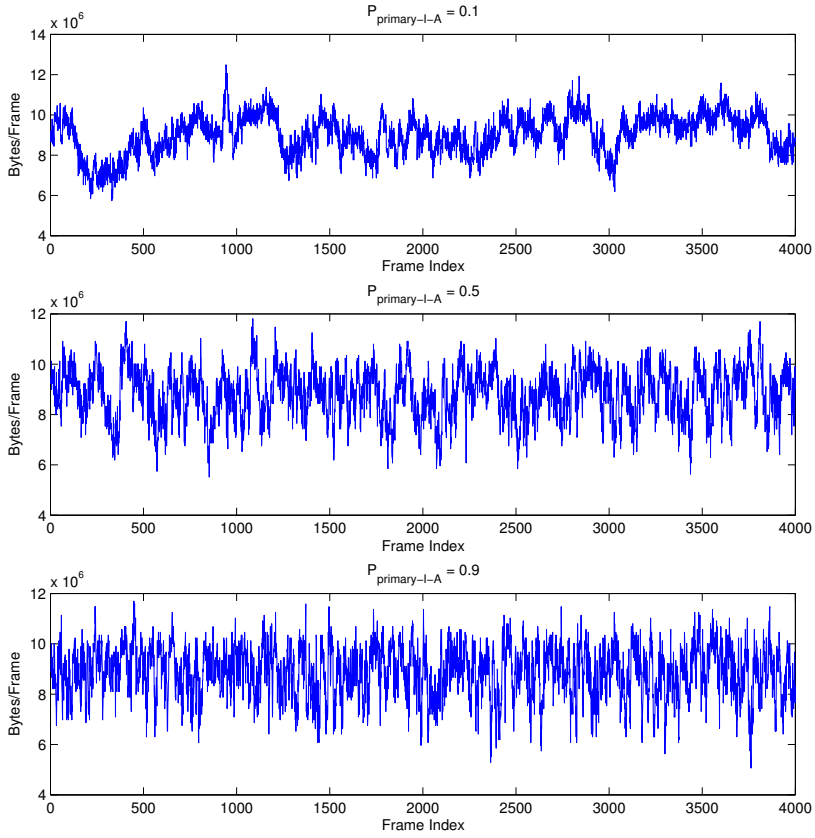

Figure 5.2: Burstiness Effect on Total Throughput

$P_{primary-I-A}$ has a value of 0.1, 0.5 and 0.9 for plots in Figure 5.2 from top to bottom. It can be observed that a higher burstiness in PU activity (top case with $P_{primary-I-A} = 0.1$) results in a slower fluctuation of SN throughput, while a lower burstiness in PU activity (lower case with $P_{primary-I-A} = 0.9$) results in a faster fluctuation of the SN throughput. Though the frequency of fluctuation of the total throughput is different, the average throughput is the same, since the average resource available to the CR network is the same.

5.3.3 Effect on queue backlog

The simulation is repeated for each of the state transition probability pairs, and the queue backlog length at each frame is recorded. The distribution for the queue length of an SU sharing spectrum with PUs with different levels of burstiness is illustrated in Figure 5.3. The figure shows the queue backlog distribution in the buffer for the same target SU, at distance from SNBS of 15 km (half the radius of the cell), and the only variable for different curves is the level of burstiness of the PUs. At $P_{primary-I-A} = 0.1$, the level of burstiness of PU activities is at maximum, and the average queue backlog length is the highest. As $P_{primary-I-A}$ increase, level of burstiness of PU activities decreases gradually, the average queue backlog length also becomes lower. The results indicate that level of burstiness of PUs do have a significant impact on the packet level performance of SUs in the CR network. Under the same resource availability, the QoS of SUs will deteriorate if the behaviour of the PUs in primary network is more bursty.

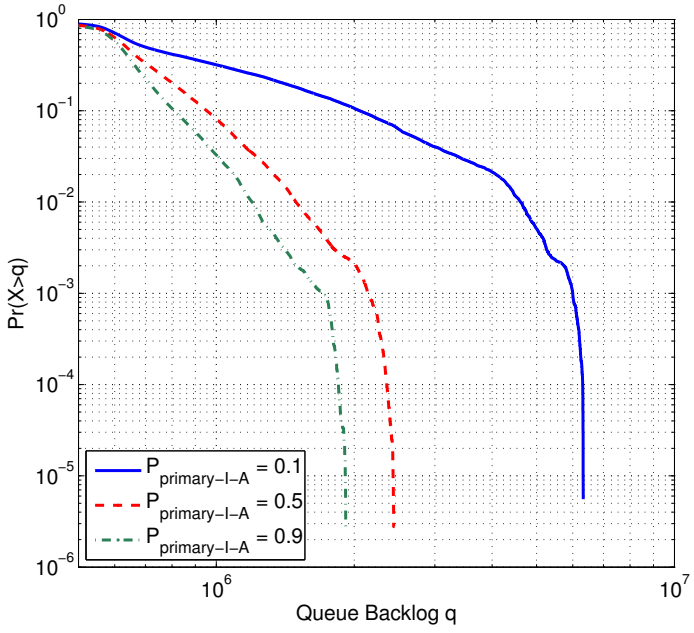

Figure 5.3: Burstiness Effect on Queue Distribution

Figure 5.4 illustrates the effect of PU activity burstiness on SUs at different

distances from the SNBS. For each of the SU, queue distribution under maximum burstiness and minimum burstiness are plotted for clear contrast. The user that is nearest to the SNBS is plotted as the blue curve. No significant effect of PU burstiness can be observed. But as the user gets further away from the SNBS, the effect of PU activity burstiness becomes more obvious. We can see that the users that are further away from the SNBS, and thus have worse channel conditions resulting in higher average queue backlog, are influenced more significantly by the burstiness effect of PUs. On the contrary, those nearer to the SNBS are almost not influenced by the activity burstiness effect of PU at all.

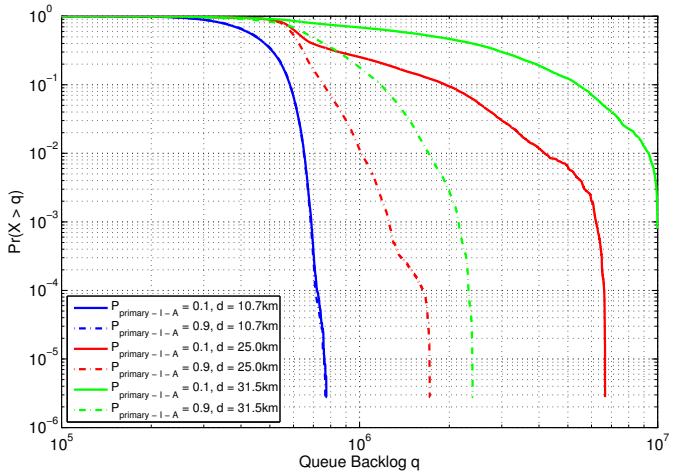

Figure 5.4: Burstiness Effect on Queue Distribution at Different Distances

5.3.4 Effect On Packet Delay

The effect of PU activity burstiness is more clearly shown in its effect on packet delay. Because the time a packet enters the buffer and leaves the buffer are recorded in the packet during the simulation, the delay time in buffer for each of the packet can be calculated and recorded in the simulation experiments.

The averaged delay for all SUs is plotted against different levels of burstiness of PU activities. Results are also compared for users at different distances from the SNBS. Figure 5.5.

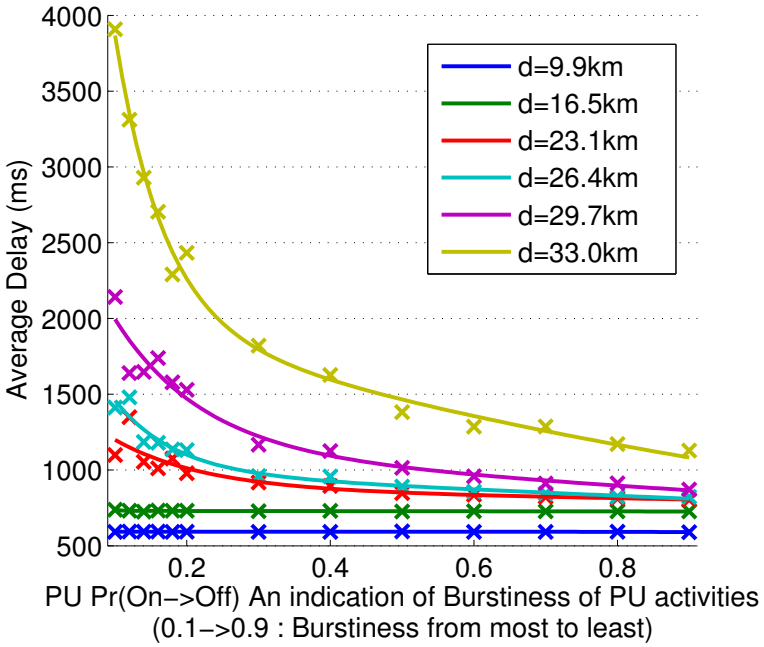

Figure 5.5: Average Delay vs PU Activity Level of Burstiness

Different PU activity patterns do have a significant impact on the QoS of secondary network though the activity intensities are the same. As the PU activity becomes more bursty, the performance of the secondary network tends to deteriorate. Under the current resource allocation algorithm used by the SNBS, the SUs that are further away from the base station experience more serious QoS deterioration because of PU activity burstiness, while those nearer to the base station barely have any significant change in QoS.

5.3.5 Heavy-Tailed PU Activity Modelling

The above experimental results all assumed an exponential distribution of Off state time for PU. In this section, another set of experiments are introduced to evaluate the effect of changing the modelling of the Off period from an exponential distribution to the Pareto distribution. The reason of this change of modelling is because recent research [12] suggests that inter-activity time of events driven by human activity is more accurately characterised by heavy tailed processes such

as power law distributions.

Power law distribution is a distribution whose density function has the form $p(x) \propto L(x)x^{-\alpha}$, where $\alpha > 1$ and $L(x)$ is a slow varying function. An example of power law distribution is the Pareto distribution. For a Pareto distribution, its probability density function is:

$$f(x) = \begin{cases} \frac{\alpha x_\alpha^m}{x^{\alpha+1}} & \text{if } x \geq x_m \\ 0 & \text{if } x < x_m \end{cases} \tag{5.2}$$

To generate example On-Off events with two different distributions, firstly a set of Pareto distributed random numbers are generated to serve as the Off event time for PU activity. Since heavy tailed distributions such as Pareto distribution have a very high variance and thus require a very large number of samples to converge to the mean value, it is more feasible to generate the random number series for Pareto distribution first, and then calculate the mean value of the generated series. Then generate the exponential distributed random number series as a control sample, according to the calculated mean value for the Pareto distributed random numbers. These two generated random series are then used to control the Off event sojourn time in the simulation.

Figure 5.6 shows a comparison of SU queue length distribution under these two different models. It can be observed that although these two different models should result in the same average available resource to the SUs, the resulting queue backlog distributions for these two cases are quite different. A Pareto distributed OFF time of PU can significantly deteriorate SU QoS because of its greater probability of creating larger queue lengths.

5.4 Conclusions

Current results reveal a complicated relationship between secondary network QoS and several contributing factors, including PU activity level of burstiness and distance from base station. The results show the general trend of how these factors affect QoS, and can be used as an indicator of how much each of these factors affects the performance of secondary networks. This could be used as an impor-

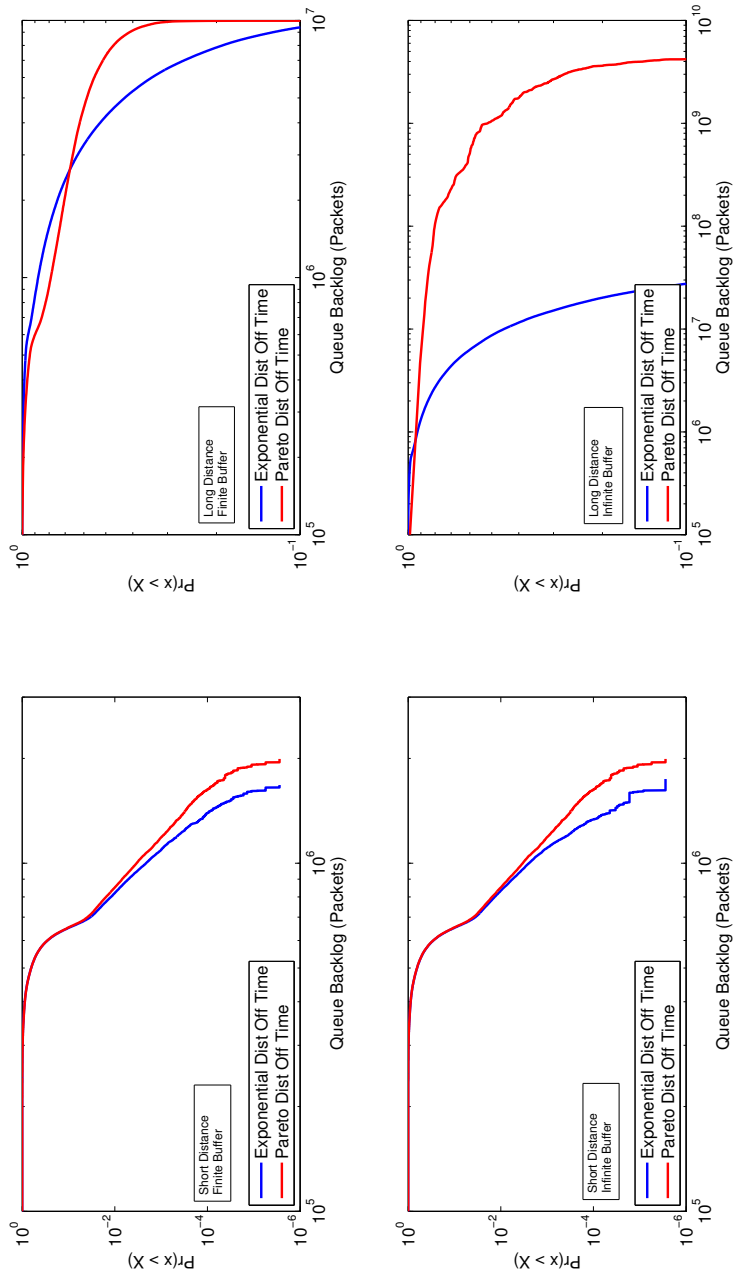

Figure 5.6: Queue Backlog Distribution Under Different PU Modelling

tant reference to design CR networks. The significant performance deterioration caused by PU activity pattern burstiness suggest that evaluation of PU activity pattern is very important before designing and deploying a CR network.

Chapter 6

Video Quality over Cognitive Radio Network

6.1 Introduction

The requirement for wireless transmission capacity has seen a dramatic increase in recent years. A report from Cisco [17] stated that mobile traffic has more than doubled each year over 2009-2011, and is predicted to double again in 2012. Part of the reason is that the wireless mobile devices are becoming more powerful and many of the mobile applications rely heavily on high speed wireless transmission. Among the mobile data traffic, video traffic contributes to more than half at the beginning of 2012, and is predicted to generate over 69 percent of mobile data traffic by 2018 [64]. Cognitive radio technique have the potential to provide very high speed wireless links for mobile devices, and thus may be a very good candidate for mobile video data carrier.

In this chapter, the process of evaluating QoS of video data transmission over CR network is introduced. IPTV is defined as "multimedia services such as television/video/ audio/text/graphics/data delivered over IP based networks managed to provide the required level of QoS/QoE, security, interactivity and reliability" [65]. And there is growing research interest in mobile IPTV implementation [66, 67, 68]. IPTV requires relatively high wireless transmission speed, especially for high quality video streams. It is interesting to see how different factors within

a CR network could affect the QoS of video service, whereas previous chapters considered voice and widely used Non-Real-Time data model.

6.2 System Modelling

6.2.1 IPTV over CR System Overview

In order to evaluate the QoS and QoE of IPTV over CR networks, it is necessary to model not only the CR network, but also all other elements in between the end to end transmission link, because video traffic has its own distinct characteristic, and there is no simplified model (like voice traffic) to evaluate video QoS and QoE. In previous experiments, only the transmission link section is modelled and simulated. In this section, additionally, the video sender (IPTV source) and the video receiver (secondary users) also need to be modelled.

Figure 6.1 illustrates an overview of the whole end to end system.

The key components included in the system are:

Sender

The sender is where video source are converted to consecutive packets containing the video information.

SNBS

The SNBS receives packets from the sender and put them into buffer of the SU. The SNBS allocate downlink resource to the SU according to the requirements, and transmit the video packets to the SU requiring IPTV service.

Primary Network

The primary network operates at the same spectrum band with the secondary network. The activity of PUs are sensed by the SNBS and may interrupt transmission of SUs when the activity intensity of PUs are high.

Secondary User

The secondary user which requires IPTV service receives the packets from the SNBS, process the packets and play the video for the user.

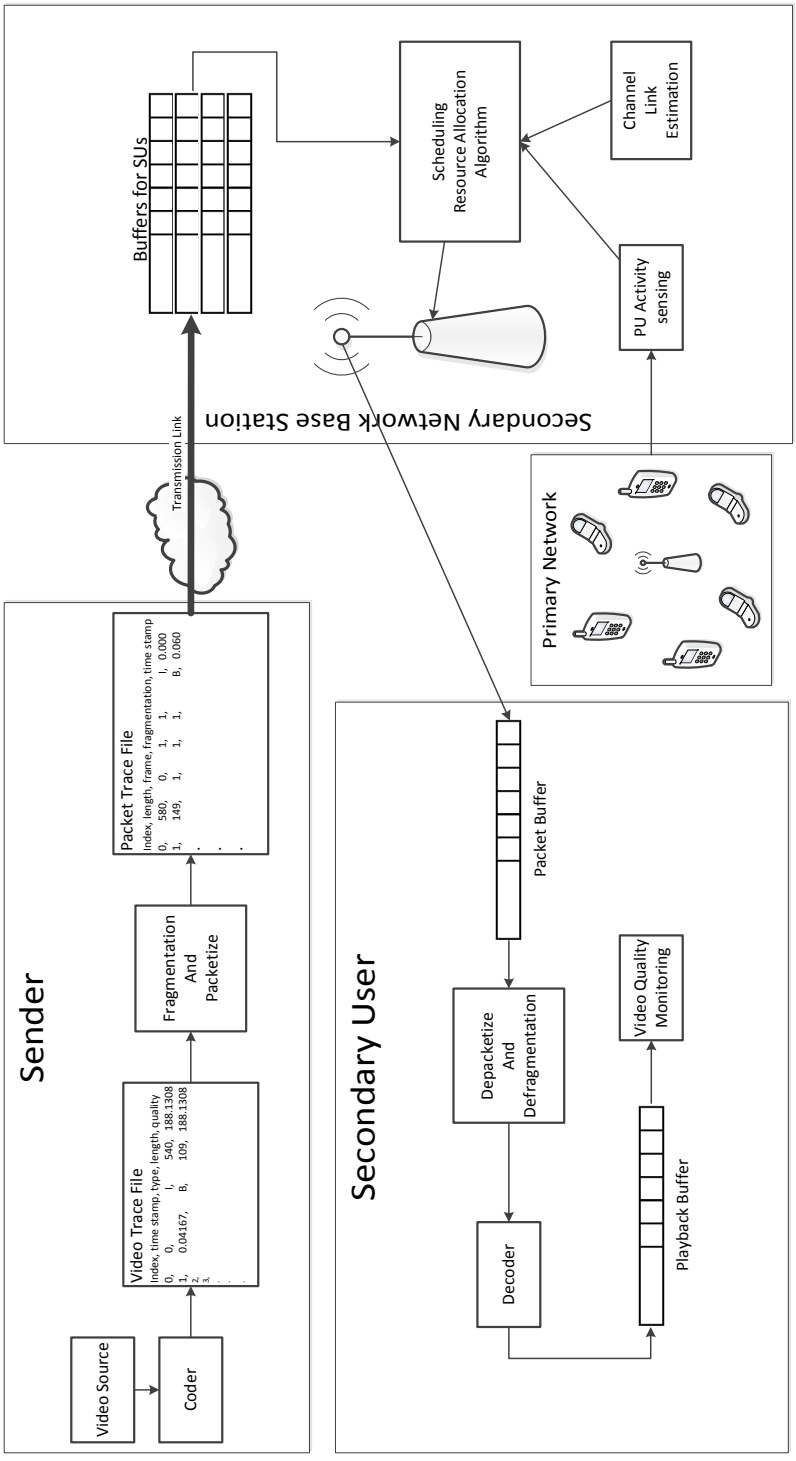

Figure 6.1: System Overview of IPTV over CR Networks

The SNBS and primary network part are modelled the same as introduced in previous chapters, and only slight modifications are made.

6.2.2 Sender

The modelling of the sender is a very important part in the video quality evaluation simulation experiment, because video traffic has its own distinct characteristics. In order to evaluate the network performance, the simulation usually requires long enough run time to obtain accurate results. Especially for CR network, the random effects and the analysis of heavy-tailed PU activity effects requires even longer simulation for accurate results.

However, common video sources are not long enough to satisfy the long time simulation. Long movies usually have a length of 2 to 3 hours, and still cannot satisfy the time requirement. In addition, current VBR (variable bit rate) video frame lengths see significant long range dependance (LRD) characteristic, and can be modelled by self similar processes [69]. So simply concatenating different video clips for a longer simulation time is not appropriate. On the other hand, using existing video sources (movie, TV channel) for research may lead to copyright problems. So I need to find a proper way to generate long video traffic source, while keeping the LRD characteristic of the video traffic.

Most recent researches [70, 71, 72] on video traffic generation suggest traffic modelling using Wavelets produces video traffic with the highest LRD precision, and it is also easy to implement. The detailed theory of generating video traffic pattern using Wavelet is introduced in section 6.3.

The variable bit rate video frames have different lengths, and the lengths of the frame depends on many factors, such as the characteristic of the pictures, motions and the coding method used. For most of the time, especially for high quality video, the size of the frame is too large to be transmitted in one packet, so the frame data needs to be fragmented and packetized. After the fragmentation and the packetizing, packets may have different lengths.

Though originally the CR network simulation tool can only process number of packets assuming all packets have the same packet length, the tool can be modified to be able to deal with packets with different sizes as described below.

The original transmission rate unit is packet/time-slot, and modified model's transmission unit is Byte/time-slot according to the SINR requirement listed in Table 6.1.

Table 6.1: Transmission Rate and SINR Requirement in Bytes

Mode	$R(Bytes)$	γ (dB)
1	3750	10
2	7500	14.77
3	11250	18.45
4	15000	21.76
5	18750	24.91

A list of packet trace file serve as the output of the sender. The information contained inside the list includes the packet index, packet length, which frame it belongs to, fragmentation index and scheduled transmission time.

6.2.3 Secondary User

The SU which requires IPTV service needs to be modelled properly so that video performance parameters can be measured. Firstly, the user needs to have a packet buffer to store the video packets received from the SNBS. After all the packets belonging to one video frame are received, the packets in the buffer can be de-packetized and de-fragmented. Then the frame data can be decoded and put into the playback buffer. A pointer in the playback buffer will be continuously moving forward indicating the current playing frame for the user. The physical and MAC modelling of the SU remains the same as in previous experiments.

6.3 Video Traffic using Wavelet Model

6.3.1 Wavelet Transform

The Wavelet Transform (WT) is very similar to the Fourier Transform. However, it overcomes some shortcomings of the Fourier Transform. A Wavelet Transform takes a signal in time domain as the input, and transforms the signal into

the Wavelet Domain. Instead of just only one series of frequency signals in the Fourier Transform, Wavelet Transforms results are a collection of time-frequency representations of the signal, all with different resolutions, and thus so called multiresolution analysis. Multiresolution analysis is the heart of Wavelet Analysis. It is the decomposition of a signal into subsignals of different size resolution levels.

Wavelets are complete orthonormal bases, which can be used to represent a signal as function of time. Discrete wavelets can be represented as

$$\phi_j^m(t) = 2^{-j/2}\phi\left(2^{-j}t - m\right) \tag{6.1}$$

Where j is the scale factor, or dilation index and m is the translation in time.

$\phi(t)$ is called the mother wavelet. The simplest mother wavelet is called Haar wavelet, where

$$\phi(t) = \begin{cases} 1 & \text{if } 0 \leq t < 1/2 \\ -1 & \text{if } 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{6.2}$$

Consider a discrete-time process $x(t), 0 \leq t \leq 2^K - 1$, the Wavelet Transform of $x(t)$ is given by:

$$d_j^m = \sum_{t=0}^{2^K-1} x(t)\phi_j^m(t) \tag{6.3}$$

where d_j^m is the m^{th} Wavelet coefficient at time scale j.

And $x(t)$ can be represented through its Inverse Wavelet Transform (IWT):

$$x(t) = \sum_{j=1}^{K} \sum_{m=0}^{2^K-1} d_j^m \phi_j^m(t) + \phi_0 \tag{6.4}$$

Equation 6.3 and 6.4 defines the Wavelet Transform and the Inverse Wavelet Transform respectively.

6.3.2 Video Traffic Synthesis using Wavelets

One very significant characteristic of Video Traffic is that its both SRD and LRD, which makes it very hard to generate using traditional random variable generation methods [72].

Consider frame length $x(t)$ of a video clip. If we perform wavelet transform on $x(t)$, we will get a set of wavelet coefficients d_j^m.

Taking wavelet coefficient d_j^m at a specific scale j, e.g. . $d_j^1, d_j^2, d_j^3 \cdot d_j^m$. Surprisingly, these coefficients are no longer LRD over m. Instead, they are i.i.d. Gaussian random variables with zero mean.

Inspired by this characteristic of traffic in wavelet domain, a method to generate complex (SRD+LRD) traffic using a simple model in wavelet domain is developed in [72].

- Perform wavelet transform on empirical video trace $\hat{x}(t)$ and obtain its wavelet coefficient $\hat{d_j^m}$.
- Analyze $\hat{d_j^m}$ at each time scale j. Obtain their mean, variance and cumulative distribution function.
- For each of the time scale j , generate zero mean Gaussian variables with variance evaluated from empirical traffic analysis. Then for all the random variables generated, transform the data according to the cumulative distribution functions of empirical data, so that the newly generated random variables fits into the empirical distribution. The resulting random variables are d_j^m.
- Perform an inverse wavelet transform on d_j^m to obtain the synthesized video traffic.

6.3.3 Efficient Method for Wavelet Transform

The purpose of generating arbitrarily long movie traces is because we need to provide long enough movie traffic for the simulation to obtain more accurate results. The Wavelet Transform and Inverse Wavelet Transform are required for video traffic generation method introduced in the last section. However, they

could require very high computational complexity, especially when the level of detail j has a higher value, or when a complex mother wavelet is chosen. To simplify the calculation, the simplest mother wavelet 'Haar Wavelet' is chosen as it is common in the literature [70, 73]. The Haar mother wavelet is shown in equation 6.2.

Suppose the Haar Wavelet is used for the Wavelet Transform, then the support (non-zero values) for the translated and dilated wavelets $\phi_j^m(t)$ defined in equation 6.1 can be represented by

$$\left[m2^j, \left(m+\frac{1}{2}\right)2^j - 1\right] \bigcup \left[\left(m+\frac{1}{2}\right)2^j, (m+1)\,2^j - 1\right] \tag{6.5}$$

So the Wavelet Transform in equation 6.3 can be simplified as:

$$d_j^m = 2^{\frac{-j}{2}} \left(\sum_{t=m2^j}^{(m+0.5)2^j-1} x(t) - \sum_{t=(m+0.5)2^j}^{(m+1)2^j-1} x(t) \right) \tag{6.6}$$

This form makes the wavelet coefficients very easy to calculate since it only involves summation of the time series $x(t)$.

The inverse Wavelet Transform is also easy to calculate for the Haar Wavelet. It is just a superposition of reconstructed signal at different time scales.

$$\begin{aligned} x(t) &= (d_1^1, -d_1^1, d_1^2, -d_1^2, \cdots d_1^m, -d_1^m) \\ &\quad + (d_2^1, d_2^1, -d_2^1, -d_2^1, \cdots) \\ &\quad \vdots \end{aligned} \tag{6.7}$$

By equations 6.6 and 6.7, both the WT and IWT can be calculated with minimum computational complexity.

6.4 Simulation Results and Analysis

6.4.1 Examining Empirical Traces

Video trace example files are available at an online video trace library [74]. Video traces are downloaded from the library and I wrote a MATLAB program to decode the data from the hex files downloaded.

The autocorrelation function is analyzed to show the LRD characteristic of the video traffic, and buffer loss probability are also analyzed for each of these empirical traces.

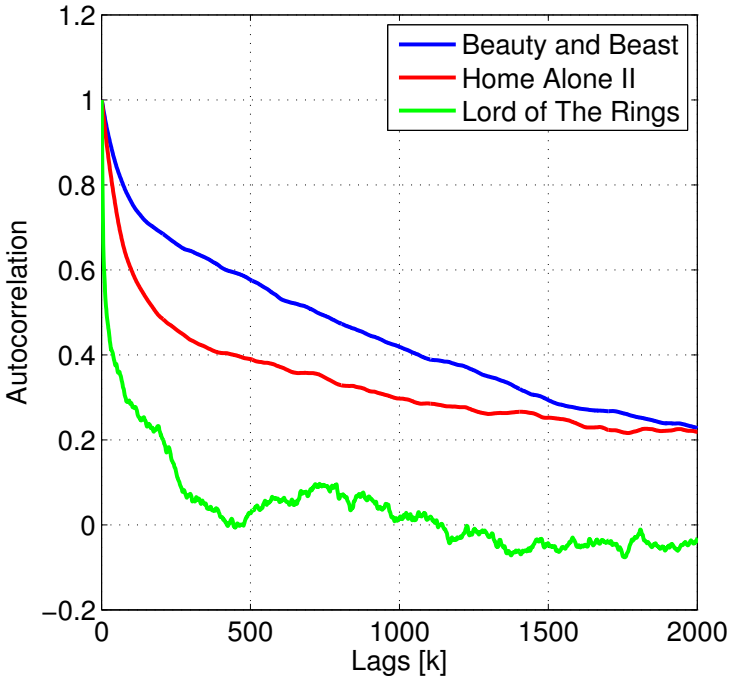

Figure 6.2: Autocorrelation of Empirical Traces

Figure 6.2 and 6.3 shows the autocorrelation and the buffer loss probability for the 3 movie traces downloaded from the library (Lord of The Rings, Home Alone II and Beauty and Beast).

For each of the video trace traffic, an infinite buffer is used to store the arrival data. At the beginning of each frame, the traffic is fed into the buffer, while the

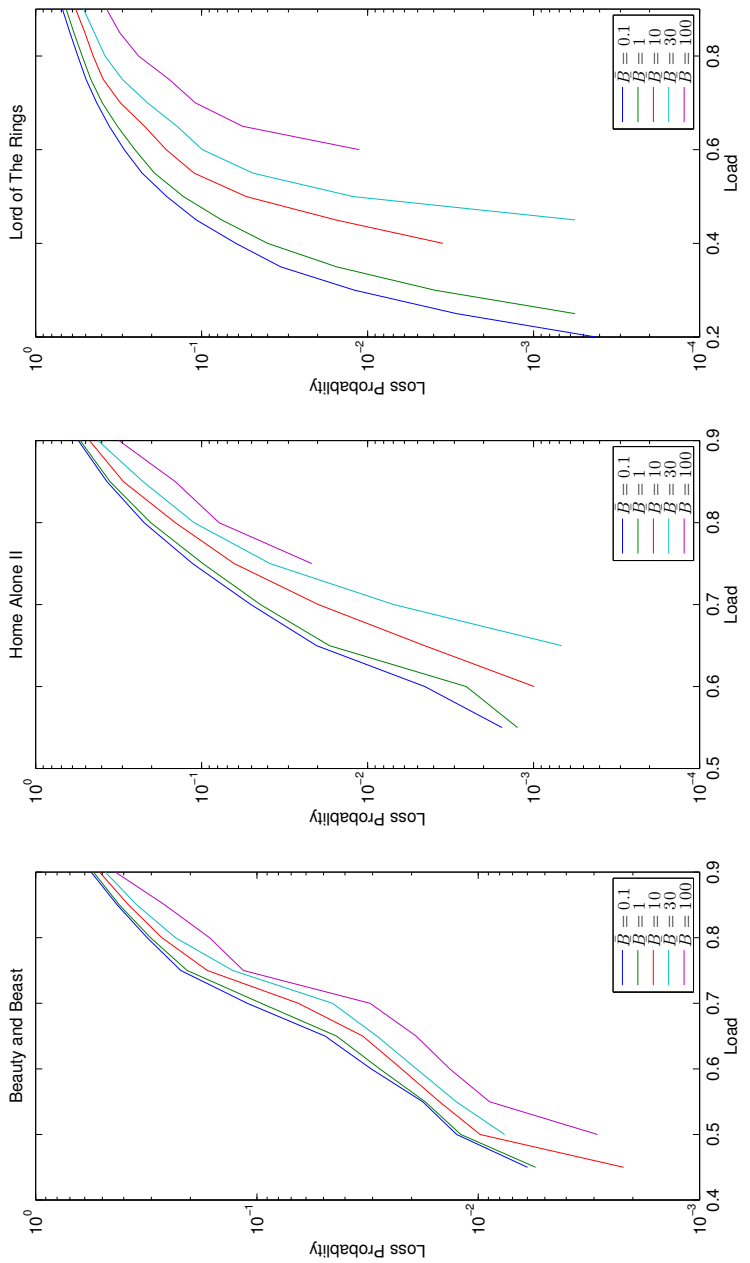

Figure 6.3: Buffer Loss Probability of Empirical Traces

server drains data from the buffer at a constant rate. We store the queue length variation $\hat{B}(t), 0 \leq t < 2^K$. The loss probability is defined as $P(x) = Pr(\hat{B}(t) > x)$.

Obviously, the frame size of the empirical traces of these three movies are different because they are completely different movies, resulting in different traffic intensity overall, where we define the average traffic intensity for an empirical trace as λ. So the buffer loss results are only comparable for these 3 movies if the load (defined as ρ) is kept the same for the 3 movies. This is done by applying different service capacities (service rate defined as C) to each of the movie traces.

On the other hand, the size of the buffer (defined as B) that is used to evaluate the loss probability also needs to be different for each of the 3 movies due to different traffic intensity, so that the packet loss probability results comparable. A normalized buffer size (defined as $\bar{B}$) is used and is defined as:

$$\bar{B} = \frac{B}{C} \tag{6.8}$$

Equation 6.9 and 6.10 is used to calculate the actual buffer size B and service capacity C with observed traffic intensity λ, and a predefined load ρ and normalised buffer size $\bar{B}$.

$$C = \frac{\lambda}{\rho} \tag{6.9}$$

$$B = \bar{B} \times C \tag{6.10}$$

The diagrams for loss rate against load in Figure 6.3 are plotted using different Normalised-Buffer-Size (0.1, 1, 10, 30, 100 from top down).

6.4.2 Wavelet Transform and Inverse Wavelet Transform on A Target Video Trace

A simulation tool was built to convert a discrete time series from time domain to wavelet domain using the simplified WT and IWT introduced in section 6.3.3. In order to test the correctness of the simulation tool, a simple way is to convert a

video trace file from the time domain to the Wavelet domain first, and obtain a set of wavelet coefficients d_j^m. And then, convert these coefficients in the Wavelet domain directly back to the time domain, and compare the results with the original time domain traffic. The two time domain traffic signal should be exactly the same if the simulation tool correctly implements the WT and the IWT.

We take one of the example video traces, which is "Home Alone II", and try to apply WT and IWT on the trace to test the correctness of the implemented algorithm. Figure 6.4 shows a comparison of the empirical trace and the traffic trace after WT and IWT. The comparison is plotted in time domain, and it can be observed that the traces are identical, indicating that the WT and IWT conversion tool works fine.

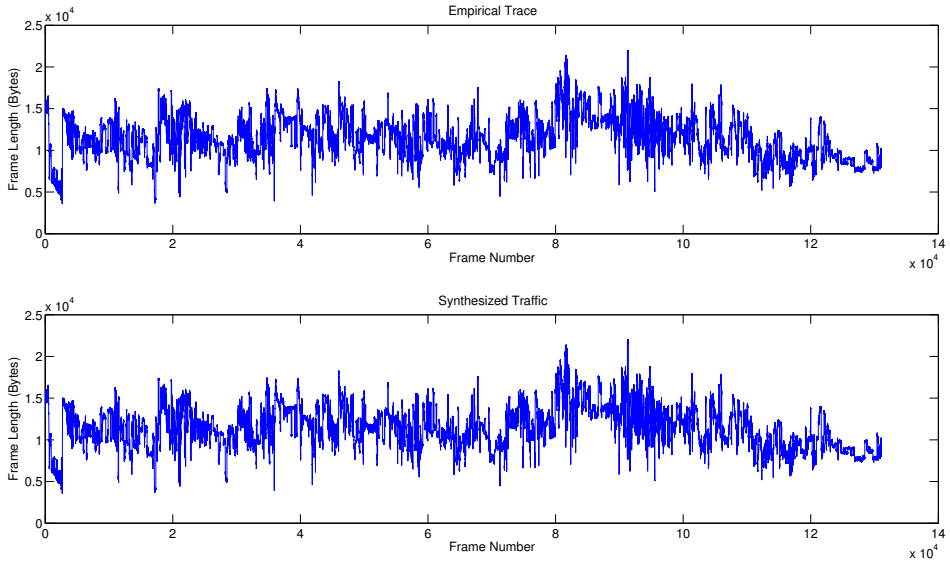

Figure 6.4: Comparison of Empirical Trace and Trace after WT and IWT (Home Alone II)

6.4.3 Generating New Traces From Empirical Traces

Using the algorithm introduced in section 6.3.2, I wrote a simulation tool to generate new traces according to the characteristics of the empirical video trace.

The example movie trace used in this section is the "Home Alone II" movie trace from the online movie trace library [74].

Firstly, Wavelet Transform is performed on the empirical video trace $\hat{x}(t)$ and the wavelet coefficients $\hat{d_j^m}$ are obtained.

The wavelet coefficients are analyzed at each time scale j for the mean and variance and cumulative distribution function of the coefficients.

Table 6.2 shows the mean and standard deviations of wavelet coefficients of the first 5 levels j for the example empirical video trace.

Table 6.2: Characteristic of Wavelet Coefficients of Empirical Video Trace

j	1	2	3	4	5
MEAN	-2.957146972	-0.688583374	0.034688542	0.11214447	-0.099043071
STD DEV	192.8995395	228.0501034	319.658761	426.0183178	603.3703875

Then a new set of random variables with the same length as the array of empirical wavelet coefficients are generated for each of the wavelet coefficients level j, denoted as d_j^m. These random variables should have the same marginal distribution as the empirical wavelet coefficients $\hat{d_j^m}$.

The Inverse Wavelet Transform is performed on the generated wavelet coefficients d_j^m to obtain the synthesized video trace in the time domain.

Figure 6.5 shows a comparison of the empirical trace and the synthesized video traffic in the time domain. And the two traces are also examined and compared with autocorrelation functions (Figure 6.6) and buffer loss probability (Figure 6.7).

We can see that both the autocorrelation and the buffer loss probability are very similar to the empirical video trace. It can be observed from Figure 6.7 that for the buffer loss probability curves of the empirical trace, as the loss probability decreases below 10^{-2}, it starts to show significant deviation from the original trend of the curve. However, for the probability curves of the synthesized traffic, they still keep a good "accuracy" because the sample size of the traffic is much larger. The ability to generate video traffic trace with arbitrary lengths enables the simulations to last for a long enough period for accurate result. Details about

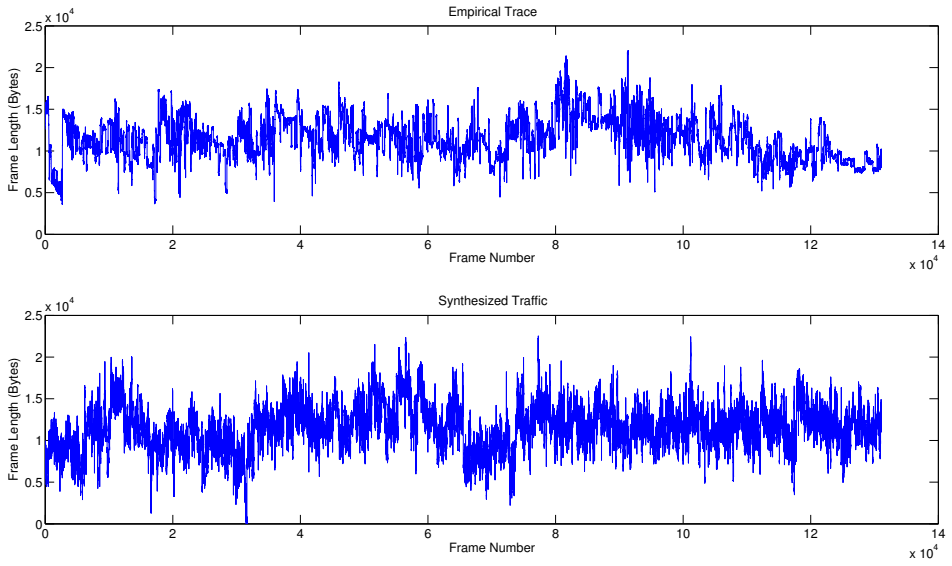

Figure 6.5: Comparison of Empirical Trace and Synthesized Video Traffic (Home Alone II)

simulation convergence can be found in Appendix C.

6.4.4 Video Traffic Transmission over CR Networks

Using the Wavelet method, we are already capable of generating arbitrarily long video trace traffic. Although the result may not be exactly the same as the original trace (as can be observed in Figure 6.6, there is still small deviations in the autocorrelation), the rough SRD and LRD characteristics of the traffic can still be generated. So now we can try to push the video traces traffic into the CR network and see how the network responds to it.

The experiment is designed as below:

There are 40 secondary users, sharing sub-channel resources with 80 primary users. For now the activity of Primary users are set to static. In the 40 SUs, 36 of them are RT users (arrival packets are Poisson distributed). 3 of them are NRT users (Markov On-Off model, with bursty ON-traffic) 1 of them uses the video trace traffic. We specifically set 1 of each type of users to an equal distance

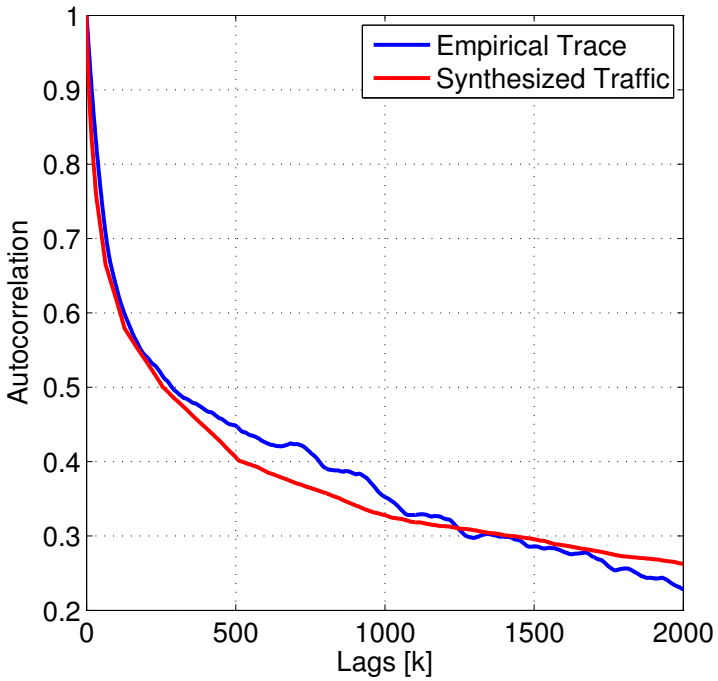

Figure 6.6: Autocorrelation Comparison of Empirical Trace and Synthesized Video Traffic (Home Alone II)

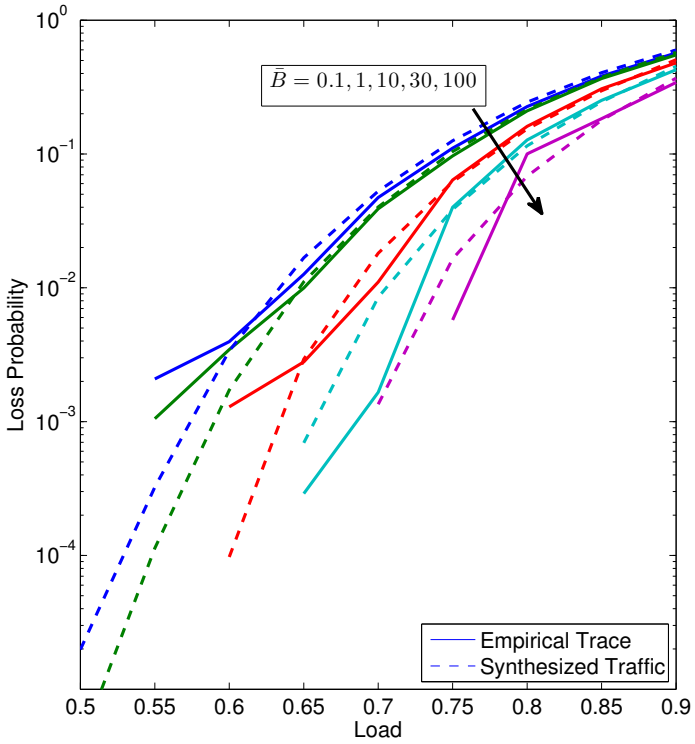

Figure 6.7: Packet Loss Probability Comparison of Empirical Trace and Synthesized Video Traffic (Home Alone II)

from the base station, so that the results are comparable. These 3 users are the experimental subjects.

The experiments were repeated using synthesized traffic from the 3 example video traces(Beauty and Beast, Home Alone II, Lord of The Rings). These video traffic traces are scaled so that the average traffic intensities are the same as that of the other RT and NRT users, and experiments are conducted with a number of frames = 50000.

Figure 6.8 shows a comparison of the autocorrelation functions for the three movie traces used.

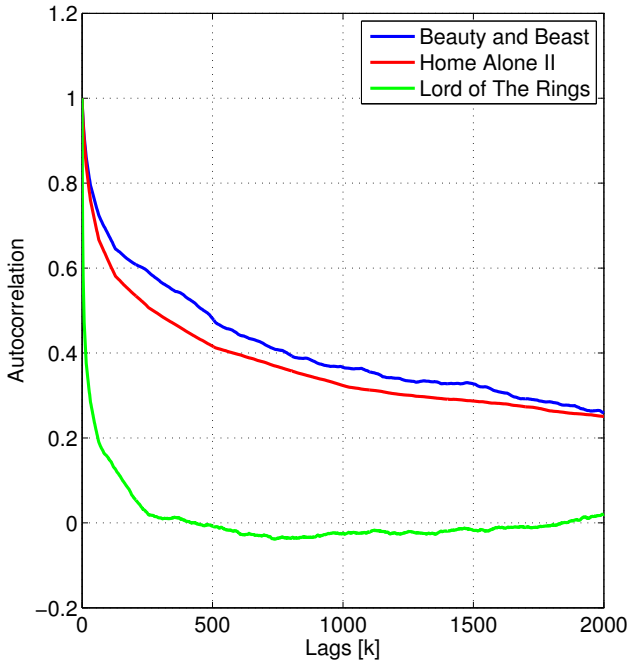

Figure 6.8: Comparison of the autocorrelation functions of three synthesized video traces

Lord of The Rings (green) is a synthesized trace from MPEG-4 empirical trace. Home Alone II (red) and Beauty and Beast (blue) are synthesized traces from M-JPEG frame level empirical traces. They exhibit significantly more LRD compared with the traffic of the Lord of The Rings.

The backlog decay function of the SU buffers are plotted in Figure 6.9.

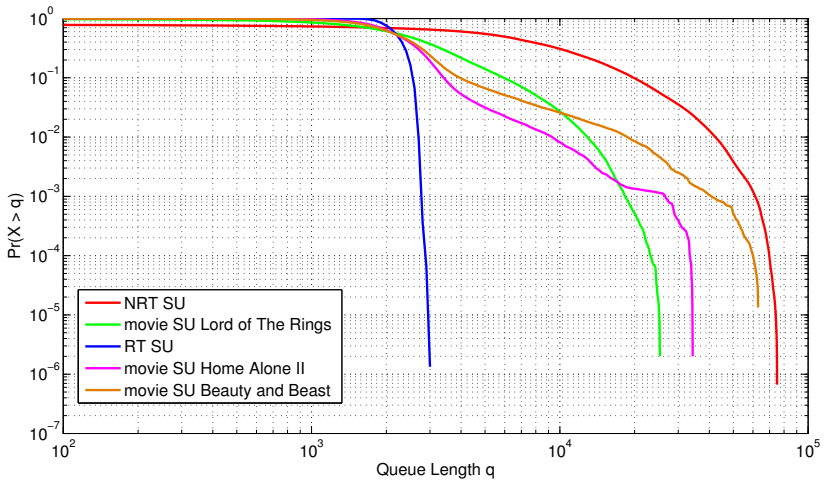

Figure 6.9: Queue Decay Characteristic After CR Transmission

It can be observed that the three Secondary Users receiving movie traffic have backlog decay function in between the RT users (Blue, e.g, VoIP users) and NRT users (Red, e.g, Web Browsing Users). This indicates that burstiness of the NRT traffic contributes to long backlog more than the LRD property of the video traffic. It also suggests that by correctly setting the parameter of SRD bursty traffic, it is possible for it to yield similar results to that of LRD video traffic.

6.5 Conclusions

Evaluating video quality over networks transmission is a very challenging problem. One of the challenges is to generate long movie traffic for simulation evaluation. In this chapter, a video traffic pattern generation method using Wavelet is introduced for network performance evaluation. The method is able to generate very long video traces for simulations that requires long time to reach steady state. The characteristic of the synthesized video trace traffic is very close to that of the empirical traces. When analyzing the video traces characteristics, it is noticed that even using the same coding method, different movie sources will have very different self-similar patterns. Thus the resulting traffic and buffer behaviours are vary significantly from one to another.

The significant LRD characteristic of movie traffic results in larger queue backlog than a SRD traffic. However, it is shown that the burstiness in traffic contributes to a even larger queue backlog than the LRD of the video traffic. Thus by adjusting the bustiness of a SRD traffic, it is possible to approximate the queue backlog of a LRD traffic.

Chapter 7

QoE Aided Performance Compensation for Cognitive Radio Networks

7.1 Introduction

The method of evaluating user QoS in a continuous time evolving manner was introduced in the previous chapters. In this chapter, we take a step forward from packet level QoS metric evaluation, to the evaluation of QoE for SUs. This would ultimately allow improvement of user QoE for CR networks by incorporating the QoE evaluating results into the RA schemes. In this chapter, we propose a novel QoE compensation scheme that increase the spectrum access priority for users with poor QoE, in order to achieve a uniformly good QoE among all users.

Similarly to most of the multiple access radio networks, CR also requires radio resource management in order to coordinate spectrum usage between users within the system, and to achieve better network performance (usually focusing on network throughput). Radio resource management for CR becomes even more challenging than that of the traditional radio networks because the radio resource available to a CR system is often highly dynamic. It is very likely that existing radio resource management methods and schemes for traditional radio networks will fail to yield consistently good performance for CR networks.

There are many papers proposing new radio resource management schemes for CR. Examples of these include [2, 6, 8, 9, 10, 11, 32, 33]. In [8], the author proposed a heuristic to approach suboptimal total throughput of the secondary network, at the same time ensuring fairness among users in the sense that they try to allocate bits to users who have not received their fair share of service. In [9], the spectral efficiency of the frequency band occupied by a CR network is used as an optimisation target function. In [6], the authors try to further increase the total transmission capacity of all secondary users by loosening the primary user protection threshold. In [32], a Markov chain queueing model is proposed to analyze the queue length variation in SU network. The author applies a Discrete Time Markov Chain analysis for the queue model, and then uses a numerical simulation to evaluate the scenario. [33] considers the QoS of heterogeneous secondary users. The author defines a priority factor to ensure that real-time user's QoS requirements are met. The average delay and total throughput of the secondary network are optimized using the algorithm proposed by the author. In [11], the authors proposed an algorithm with reduced complexity in the expense of total network throughput. In [10], the RA problem is solved using particle swarm optimisation, so that the the complexity of the algorithm is eased while maintaining a satisfactory throughput.

Most of these papers suffer from a few weakness which may render them inefficient when deployed in practical CR network scenarios. Firstly, almost all of these proposals consider the resource allocation problem within a short period of time, i.e. optimization of certain QoS parameters over a time slot[1]. The optimization of QoS over a time slot does not necessarily result in good user perceived quality, especially when the available resources to the system varies a lot over time. Secondly, licensed user (Primary User, PU) activities are usually not considered, or when considered, are modelled as static. However, dynamic resource availability is a key difference between CR and the traditional radio network, and thus must be properly modelled and integrated into the study in order to obtain more accurate results, and to test the effectiveness of resource

[1]The usual approach on RA for CR is to form an optimization problem over a short time slot for a MAC parameters such as user throughput, spectrum efficiency and power efficiency, while assuming everything within this short time slot remains constant, including saturated backlog, channel condition, etc.

allocation schemes. Last but not least, these papers do not consider the case that users may employe the CR network for different types of services, and their perceived experience can be very different.

In order to deal with the first two problems, it is necessary to consider the resource allocation problem in a continuously evolving manner. This work was introduced in Chapter 4 and 5. In this chapter we try to address the last problem, which is even more important, because the service quality that users perceive is crucial to the success of CR.

Little work has been done on analyzing packet level SN performance and SU QoE, partially due to the complex nature of the problem. In parallel, user QoE provisioning is gaining interest. In terms of the packet level QoS research publications and standards, QoS is usually used to refer to network parameters such as packet delay, packet loss and jitter. These provide a quantitative metric for the network operators and they are generally satisfied with them. However, the end users of the network don't necessarily benefit from the QoS that a network can provide. In other words, although QoS will to some extent determine the degree of satisfaction of the end user, good network QoS for a network operator does not guarantee satisfactory services for end users. QoE, on the other hand, can provide a more accurate evaluation in terms of user satisfaction. Another benefit of using QoE as a measurement method is that it unifies the network performance as a single metric, regardless of the types of service users' require, because users have a different definition of satisfaction for different types of service, and the QoS requirement on the network can be very different.

The main aim of this chapter is to show how user QoE in CR can be improved by informing the resource allocation algorithm with a continuously evaluated QoE index. The work is based on the packet level QoS evaluation results introduced in Chapter 4 and 5. These results are converted and unified as QoE, and the relationship between user QoE and the previously mentioned typical non-configurable factors are demonstrated. These factors include user distance from the base station, PU activity burstiness, and user service types. Based on the QoE evaluation results, we design a QoE feedback mechanism to inform the resource allocation algorithm with continuously monitored user QoE, so that the overall user experience can be improved.

7.2 Introduction to QoE

7.2.1 QoE Definition

User Quality of Experience is defined by the ITU in [39] as *The overall acceptability of an application or service, as perceived subjectively by the end-user.* And an important note for the definition is *Quality of Experience includes the complete end-to-end system effects (client, terminal, network, services infrastructure).* The most important characteristic of QoE is that it focuses on the end users' subjective opinion of a service.

In terms of the IP related QoS research publications and standards, QoS is usually used to refer to packet level network performance parameters such as delay, loss and jitter. These QoS parameters provide good measurements for the network operators to monitor their network speed, availability and reliability. However, the end users of the network don't necessarily benefit from the QoS that a network can provide. In other words, although QoS will to some extent determine the degree of satisfaction of the end user, satisfiable network QoS for a network operator does not guarantee satisfiable services for an end users. Thus there has been increasing research interest in QoE provisioning for users, aiming at providing better user satisfaction rather than looking at the network performance alone.

The evaluation of QoE is based on different services, because users have a different definition of satisfaction for different services. Take the VoIP service and the file downloading service as examples: users are satisfied when VoIP service could provide continuous and high quality voice signals; while users that are downloading files are satisfied when the data are loaded and in a shorter time. One of the most important advantage of using QoE as the unified metric of user experience is because it provides a single network performance quality metric that combines QoS metrics, and allows comparison of different user even when they are using different types of services.

7.2.2 QoS to QoE Mapping Techniques

According to the definition of QoE [39], it is a subjective measurement of a customer's experiences with a service. However, they are usually very time consuming, and require human activity, which is very expensive. Furthermore, the measurement cannot be carried out online, meaning that the resulting QoE cannot serve as real time feedback to the network, so the network configuration cannot be dynamically adjusted according to the QoE by a subjective method.

Recently, methods have been proposed to allow conversion from measurable QoS parameters to QoE [13, 43]. These use packet delay and loss, and by applying the mapping techniques, relate them to user QoE represented as a Mean Opinion Score (MOS).

The packet level evaluation results introduced in the previous chapters can be converted to QoE using a couple of QoS to QoE mapping techniques.

The ITU E-model [42] for monitoring end-to-end voice quality takes into account a wide range of possible impairments and represents the voice quality using an R-factor, which can then be transformed to MOS. In [43], the authors try to simplify the E-model for VoIP service and find a relationship between transport layer factors and the R-factor. Several transport layer QoS factors including packet delay, packet loss and jitter are identified as relevant factors, and an expression of R-factor value by these factors are described in the paper. The R-factor expression proposed in this paper can be directly used to transform packet level QoS to the user perceived QoE.

In this research we assume the RT users employ a VoIP service through the CR, and the corresponding packet delay and loss can be used to objectively measure user experience according to the mapping equation from [43]:

$$R \sim 94.2 - 0.024d - 0.11(d - 177.3)H(d - 177.3) - 11 - 40ln(1 + 10e) \quad (7.1)$$

,where d is packet delay in ms, e is the packet loss probability and H is the step function.

R can then be converted to user MOS by:

$$MOS = 1 + 0.035 \cdot R + 7 \cdot 10^{-6} \cdot R \cdot (R - 60) \cdot (100 - R) \quad (7.2)$$

[13] presents the exponential relationship between several packet layer QoS parameters and the perceived QoE represented by MOS. It is shown that the proposed IQX hypothesis, which is a natural and generic exponential relationship between user-perceived QoE and network-caused QoS, can provide a better approximation than the former logarithmic relationships. One result from [13] provides a delay to MOS mapping relationship for NRT services such as web page download. In this research we assume that the NRT users use CR to browse a web page, and that the web page loading time can be approximated by the packet delay time. The mapping equation for NRT users is

$$MOS(x) = 4.298 \cdot exp(-0.347 \cdot x) + 1.390 \tag{7.3}$$

, where x is the response delay in seconds. These mappings are applied to QoS results for RT and NRT users respectively.

Some more details about ITU-T E-Model, and QoS to QoE mapping methods in [43] and [13] can be found in Appendix B.

7.3 System Model and Simulation Platform

The system model used in this chapter for QoE evaluation of SUs in a CR network is the same as described in Chapter 4 and 5. However, the simulation platform is modified to implement QoS to QoE mapping to allow real time evaluation of QoE of individual users. Figure 7.1 shows the block diagram of the modified simulation platform. Note that the "QoE Evaluation by Mapping" block receives information from both the "SU Traffic Modelling" block and the "Packet Delay & Loss Evaluation" block in order obtain measured QoE of SUs.

7.4 QoE Evaluation for CR

The ultimate purpose of this chapter is to demonstrate a method to improve the overall long term experience of SUs in CR networks. However, the first step of improving it, is to be able to evaluate it. So in this section, we demonstrate experiments that focus on evaluating the effect of several parameters on SUs experience.

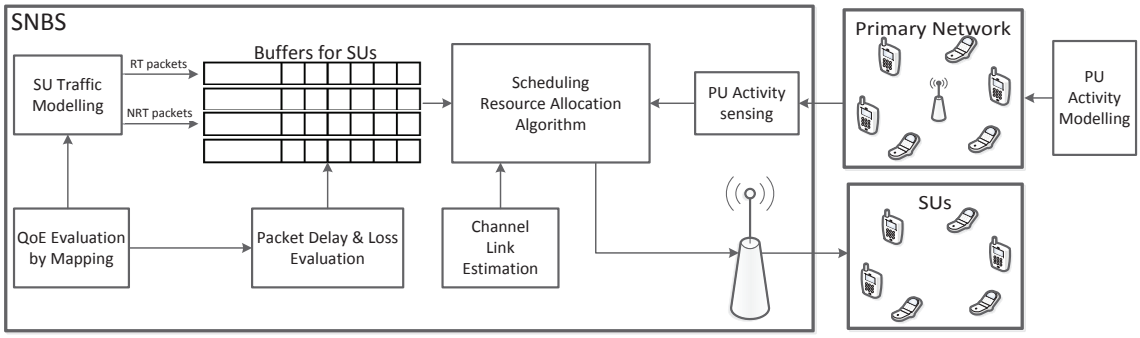

Figure 7.1: Packet Level Performance Evaluation Platform for CR

The parameters include typical non-configurable metrics that cannot be changed in practical scenarios. Using the simulation platform introduced in section 7.3, we have the capability to examine long term SU QoE. Firstly, the effect of user service types and distances from the SNBS is evaluated. Then we introduce the mapping method to convert the evaluated QoS metrics to QoE, so that the effects of service types and distance can be represented in the form of MOS. Finally we show the evaluation of PU activity effect on user QoE. These evaluation results demonstrate the significant impact of several contributing factors on user QoE, when the traditional throughput focused RA scheme is employed. In later sections, the QoE evaluation method introduced in this section is used provide input to the QoE based performance compensation scheme for RA. And the QoE results with QoE performance compensation are compared with the evaluated QoE results shown in this section to illustrate its effectiveness.

7.4.1 Service Type and Distance Effect

7.4.1.1 Modelling of SU Traffic

Here we model SUs as two types, real-time (RT) users and non-real-time (NRT) users. Each represents different service types, with different traffic characteristics.

For RT users, packet arrivals are modelled as a Poisson process, with arrival intensity of λ_{RT}. So on average λ_{RT} packets will arrive at the SU buffer in each time slot. The traffic arrival process for NRT users is characterized by an ON/OFF model to simulate bursty NRT traffic patterns. The transition prob-

Table 7.1: Parameters Setup

Variable	Explanation	Value
M	Total Subchannel Number	120
N	Total SU Number	40
N_p	Total PU number	30
K	Ricean Fading Factor	$-10db$
η	Path Loss Exponent	3
d_0	far-field crossover distance	$50m$
r_1	SUs Radius	$33km$
r_2	PUs Radius	$60km$
N_0	Gaussian Noise Power	$-100db$
ω	Critical Interference Threshold	$0db$
L	Number of Time Slot	30
δt	Time Slot Length	$30ms$
P_{max}	Total Transmission Power	$30W$
P_{On-Off}	NRT SU Transition Probability	0.8
P_{Off-On}	NRT SU Transition Probability	0.1

ability between the ON/OFF states, and the traffic intensity for NRT users in the On state, R_{NRT}, is set so that the overall average traffic intensity λ for RT and NRT users have the same value, i.e. $\lambda = \lambda_{RT} = Pr_{ON} \times R_{NRT}$. Pr_{ON} is the probability that an NRT user is in the On state.

7.4.1.2 Experimental Setup

The parameters used in the experiments are shown in Table 7.1. Each PU occupies one subchannel, and since the main concern in this set of experiments is on SU related parameters, PU activities are set to be static, though later, in the case where the PU activity patterns are considered, these are modelled dynamically. (We later show that PU activity pattern has a strong effect on the long term queue behaviour of the SUs) The SNBS will allocate the available subchannels to the SUs in order to maximize the minimum transmission rate among the SUs.

The experiment is repeated 3 times, each with different packet arrival intensity. For each of the experimental iteration set, one SU of each service type being observed is placed at a different location with different distances from the SBNS, so that the effect of user distance from the SBNS can be evaluated. We monitored the average packet delay and loss probability for the SUs.

7.4.1.3 Results and Analysis for Effect of Service Type and Distance

A comparison of the QoE of RT and NRT users for different traffic intensity λ is shown in Figure 7.2. Note that the solid lines for NRT user results are obtained by curve fitting. Figure 7.2 shows an NRT user's QoE degrades dramatically as its distance from the SNBS increases. This means that the users that are further away from the SNBS have a much worse experience when using NRT services than those nearer to the SNBS. For RT users, results for the same three λ values almost overlaps, indicating the limited effect of packet rate on QoE. Comparing the MOS of RT and NRT users, although both service types require similar capacity for mean packet rate, the difference in QoE for users at a large distance is significant.

These results demonstrate the very significant impact of both packet level traffic arrival process, and user distance from the SNBS. The Max-Min RA scheme discriminates against users with bursty traffic profiles and SUs at a large distance, meaning that the Max-Min RA scheme is ineffective in providing uniformly good user QoE.

7.4.2 Primary User Activity Pattern Effects

7.4.2.1 Modelling of PU Activity Patterns

As introduced in chapter 5, PU activity pattern has a significant effect on packet level SU QoS. In this section we try to research on the relationship between the PU activity pattern and user QoE.

It is obvious that the activity of PUs occupying the resource will directly affect the SN performance. In previous papers, PU activity is usually assumed to be static and the PU only affects an SN through the mean level of activity (mean

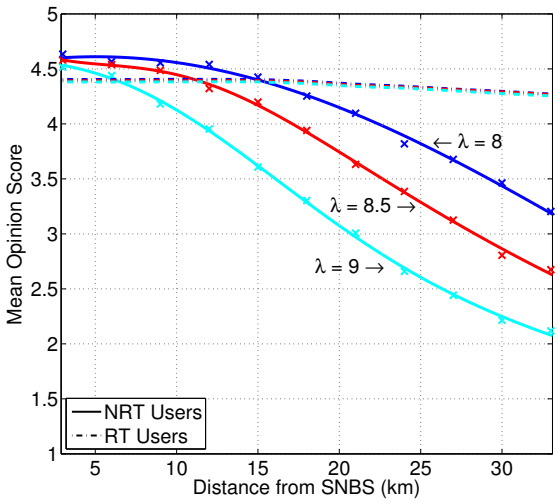

Figure 7.2: RT and NRT Mean Opinion Score over CR

number of active PUs). However, in reality, an SN could share spectrum with various types of primary networks. Even with the same type of primary network, the activity pattern of any PU could be different [34]. In our work, PU activities are dynamically modelled, and each PU is characterized by a Markov On/Off model. For a PU in an On state, an SU could still share the same sub-channel with the PU if the transmission power of the SNBS in this sub-channel does not reach the interference threshold, while for a PU in an Off state, the SNBS can transmit data without power limitation. We model PU activities with different transition probabilities, while keeping the average resource available to an SN constant, so that the effects of the PU activity pattern are evaluated.

For any PU, P_{I-A} indicates transition probability from inactive state to active state, while P_{A-I} indicates that from active state to inactive state. In our experiment, P_{I-A} takes a value from the set 0.1 to 0.9 with a 0.1 interval, while P_{A-I} is calculated correspondingly to ensure the average PU traffic intensities are kept constant so that the resource available to SNs is the same. We define PU activity burstiness using the Index of Dispersion of PU ON state periods $D_{ON} = \sigma^2/\mu$. Where σ^2 is the variance of PU ON state periods and μ the average value. A higher value of D_{ON} indicates a higher value of burstiness. As the value of P_{I-A} increases, the dwelling periods mean and variance will decrease, resulting in less

burstiness, i.e. more rapid fluctuations and shorter mean dwelling times.

7.4.2.2 Results and Analysis

The effect of PU activity burstiness on performance quality measured as QoE is shown in Figure 7.3. We can see that different PU activity patterns have a significant impact even though the PU activity intensities are kept constant. QoE deteriorates linearly to D_{ON} as PU activity burstiness increases. Results are also compared for users at different distances from the SNBS. Under the Max-Min RA scheme, the deterioration slope increases significantly as SUs move further away from the SNBS, while those nearer to the base station barely experience any significant change in QoE. This results in very poor QoE under bursty PU activity for users at large distances. The linear relationship between MOS and D_{ON} can be expressed as: $MOS = 5 - 4.5 \cdot 10^{-4} \cdot exp(0.16 \cdot d) \times D_{ON} - 0.0186 \cdot d$. The results may seem conter-intuitive: intuitively a rapid fluctuation on available resource means bad channel quality and would lead to low transmission quality; however, the results here show that more rapid fluctuations of PU activity leads to better SU quality. This is due to the combination of individual activity of all PUs affecting the overall resource availability being more stable when individual activity fluctuation is more rapid, while slow varying of individual PU activity results in higher variation in resources availability to SUs, which is known to be bad in queueing theory.

With these results, we can draw very important conclusions. When deploying CR networks, it is preferable to choose an environment or frequency bands in which the PU activities are less bursty. More importantly, the Max-Min RA scheme is ineffective in providing good user QoE in a highly dynamic spectrum environment, and the lack of equality among users in terms of QoE is very obvious, especially for users far away from the SNBS.

7.5 QoE Aided Performance Compensation

As illustrated in Section 7.4, the classical Max-Min resource allocation scheme(details introduced in Chapter 3) for CR results in poor user experience in certain sce-

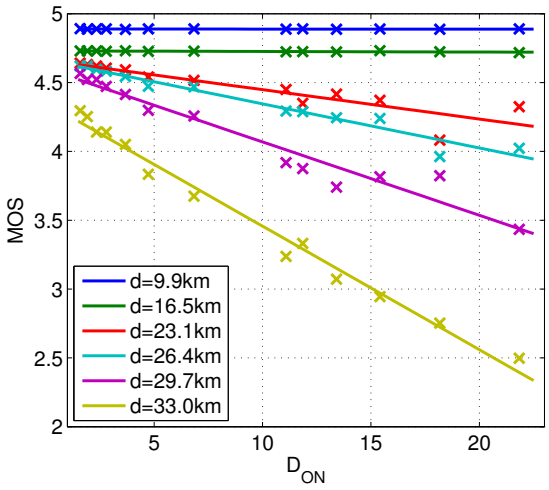

Figure 7.3: Average MOS vs Burstiness of PU Activities

narios. One of the reasons that classical radio resource management schemes fail to provide overall good user experience is because they are designed to optimize MAC/Physical layer parameters (e.g. user throughput, spectral efficiency, power efficiency, etc) under the assumption of saturated queue backlog, rather than looking at the user experience over a longer time scale. With the help of the simulation platform introduced in Section 7.3 and the QoE evaluation results illustrated in Section 7.4, it is now possible to provide the resource allocation algorithm with additional information on user QoE to aid resource allocation. In this way, our results show that it is possible to compensate those users with poor experience and significantly improve user QoE at low cost. The block diagram of the QoE compensation scheme is illustrated in Figure 7.4.

7.5.1 QoE Aided Performance Compensation Algorithm

Here we evaluate the effect of incorporating the compensation algorithm into the radio resource scheme for performance improvement. The algorithm is still based on the Max-Min scheme introduced in [2], but is modified to incorporate QoE compensation. In fact we expect that this compensation scheme can be generalised to any radio resource management scheme as long as it employs priority based scheduling.

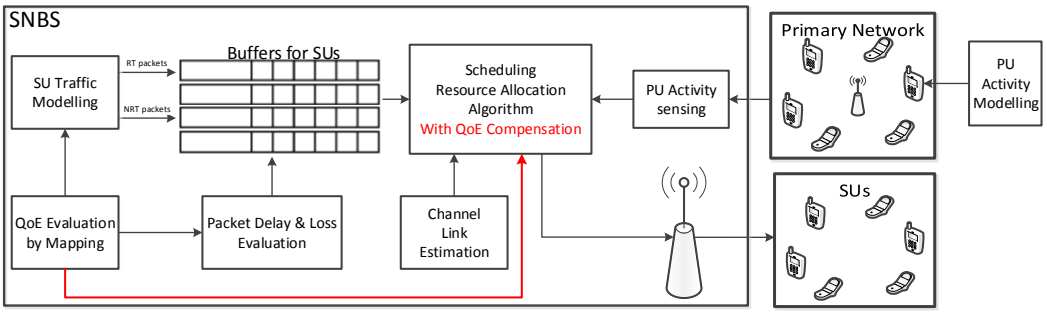

Figure 7.4: Block Diagram of QoE Compensation Scheme for CR

Algorithm 1: The Max-Min Algorithm

```
initialization;
while there is still active SU do
    if saturation = false then
        find SU with minimum allocated Tx rate;
    else
        find next unsatisfied SU;
    end
    find the subchannel with highest Tx rate;
    if target Tx rate > 0 then
        issue the subchannel to the SU;
        SU allocated Tx rate +=;
        if SU backlog satisfied then
            remove SU from active list;
        end
    else
        saturation = true;
        remove SU from active list;
    end
end
```

Firstly we introduce some details on the Max-Min algorithm proposed in [2]. The Pseudocode for the Max-Min algorithm is presented in Algorithm 1.

After initialisation, SUs are put into a list indicating that their requirement

for radio resources have not yet been satisfied. A flag variable named *saturation* is used to determine whether the SU with the minimum allocated transmission rate can be allocated more resources. In other words, *saturation* is used to determine whether the minimum transmission rate can be further improved or not. If saturation is false, the SU with minimum allocated Tx rate will be prioritised. Otherwise the rest of the resource will be allocated to the remaining SUs in a round robin fashion. Once an SU is selected, the next step is to find a subchannel that supports the highest Tx rate for the target SU. If this subchannel can be used to transmit data to the SU, then we allocate this subchannel to the target SU, and the SU allocated Tx rate is updated accordingly. Otherwise, if there is no available channel that can support data transmission for the target SU, the flag *saturation* is updated as true. During the allocation process, if a target SU receives enough radio resource to support transmission of its remaining backlog, it will be removed from the active list. On the other hand, if no more radio resource can be allocated to the target SU, it will also be removed from the list. The resource allocation iteration continues until the active SU list is empty. As we can see here, Algorithm 1 always prioritises the SU with lowest allocated Tx rate in order to maximize the minimum transmission rate among all SUs.

The fundamental idea of the QoE aided performance compensation algorithm is, instead of allocating subchannels to users with lowest transmission rate, the SNBS prioritises the user with the lowest evaluated QoE in previous transmission frames. The pseudocode for the proposed QoE aided compensation algorithm is presented in Algorithm 2. A combined index

$$\phi = R_{TX} + \alpha I_{QoE} \tag{7.4}$$

is used to determine the priority of allocating resource, where R_{TX} is the current allocated transmission rate, I_{QoE} is the MOS index evaluated using a sliding window of previous frames, and α is the weight coefficient to adjust the degree of QoE compensation. At the beginning of each transmission frame where the RA algorithm is required, I_{QoE} of individual users are calculated using a sliding window buffer containing the queue backlog condition of all users for the previous frames. In the original Max-Min RA scheme, priority will always be given to the

user with minimum R_{TX}. With QoE compensation, a user with lower QoE in previous frames will likely be allocated more resources. α is required because there is no defined mapping between transmission rate allocated to a user and its QoE, so we cannot formally optimise QoE. A too large value of α will cause an overall degradation of QoE for users with originally good experience, while a too small value of α will limit the improvement of the compensation algorithm. A sensitivity study on the parameter α is provided in a later section.

A new flag variable *QoE-saturation* is used here to indicate whether there is possible performance improvement for the SU with lowest ϕ value. If *QoE-saturation* is true, the RA won't stop immediately. It will downgrade to continue functioning as the original Max-Min algorithm. Since there is no defined relationship between R_{TX} and user QoE, the RA cannot be formed as an optimization problem. I_{QoE} here is a compensation index rather than an optimization function. Also note that I_{QoE} is calculated using different QoE functions for users with different service types, so compensation can be performed fairly regardless of there being different types of services.

7.5.2 Compensation for User Distance

Here we show how user experience can be compensated with the proposed algorithm using a set of compensation experiments. Informed by a set preliminary tests (See Appendix A), we set $\alpha = 6$ and a sliding window size of 100 frames. The experiment is repeated 50 times and the results are averaged, and the 95% confidence intervals are plotted. As shown in Figure 7.5, for users at a large distance from the SNBS, classical Max-Min optimization results in very low user QoE. With QoE incoporated in RA, the deterioration in QoE with distance is significantly compensated. Note that a slight QoE loss can also be observed where it was originally good, which is caused by part of the resource reallocated to compensate other users with low QoE, but compared with the significant gain, the loss is very small, and a loss in QoE at range > 4 would hardly make a user react at all.

Algorithm 2: The QoE Aided Compensation Algorithm

```
initialization;
while there is still active SU do
    if QoE-saturation = false then
        find SU with minimum φ = R_TX + αI_QoE;
    else
        if saturation = false then
            find SU with minimum allocated Tx rate;
        else
            find next unsatisfied SU;
        end
    end
    find the subchannel with highest Tx rate;
    if target Tx rate > 0 then
        issue the subchannel to the SU;
        SU allocated Tx rate +=;
        if SU backlog satisfied then
            remove SU from active list;
        end
    else
        if QoE-saturation = false then
            QoE-saturation = true;
        else
            saturation = true;
            remove SU from active list;
        end
    end
end
```

7.5.3 Compensation for User Mobility

Here we demonstrate the proposed algorithm's ability to compensate user QoE as the user moves within the cell. The user trajectory within the CR cell is predefined so that the SU under observation can move along the same trajectory for different experimental settings. When generating the trajectory [75], the velocity is predefined as 30 km/hour and the direction of movement is randomly decided; it turns back to a random direction as it reaches the boundary of the considered area. The example predefined trajectory used for the following experiments is

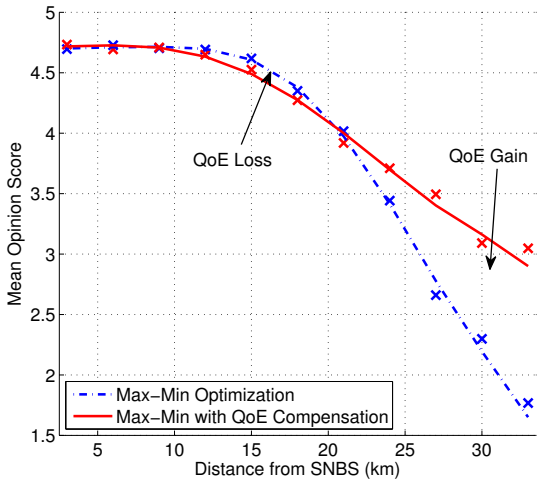

Figure 7.5: QoE Aided Compensation of MOS over Distance from SNBS

shown in Figure 7.6. User mobility is implemented by changing user location at the beginning of each frame, so that the channel link condition can be re-calculated accordingly. The rest of the SUs which are not observed remain still during the experiments. The experiments are repeated 2 times, and the only difference is the RA scheme being used by the SNBS, that is, one implementing the original Max-Min scheme, the other implementing the QoE aided compensation scheme. Figure 7.7 shows the QoE variation of one user moving inside the cell along the same trajectory for the two different cases. A significant QoE increase can be observed for periods with originally low QoE using the Max-Min scheme. With the Max-Min scheme, user MOS can degrade below 2 when the user moves to the edge of the cell; a MOS value less than 2 is considered poor for most users. On the other hand, the proposed QoE compensation scheme is able to maintain user MOS above 3 even in the worst cases, which is generally fair for most users [76]. Also notice the slight quality degradation when the user moves closer to the SNBS. This is caused by part of the radio resources diverted to other users which, at that period, have comparatively lower experience. Overall, the improvement is very significant, while the cost is very limited.

As can be observed from the results in Figure 7.5, distance of a user from the SNBS significantly affect its QoE. This is also related to the user mobility

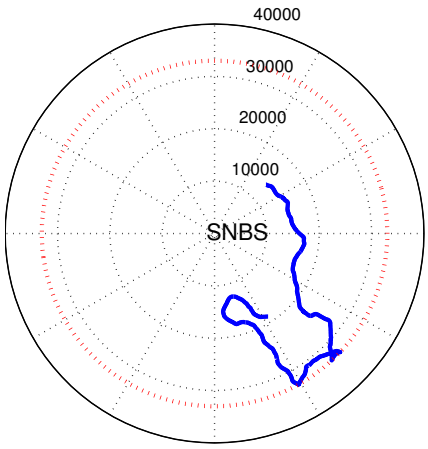

Figure 7.6: Sample Trajectory within CR Cell

model, because if a user moves faster within the cell, the distance from the user to the SNBS changes more rapidly, and would lead to more rapid changes in user QoE. Intuitively, in the case where user's QoE changes more rapidly, a smaller window size for QoE evaluation should be applied so that the changes can be captured more accurately, and the QoE compensation scheme can respond faster. The analysis of the sliding window size for QoE evaluation is not provided here, and can be a future work of this research.

7.5.4 Compensation for PU Activity Pattern

In Section 7.4.2 we show that patterns of PU activity have a strong effect on SU user experience. The degradation in user experience is especially strong when PU activities exhibit heavy burstiness characteristics, and when SUs are further away from the SNBS. Here we design experiments to show how the proposed algorithm outperforms the original Max-Min scheme under the influence of PU activity. The experiment setup is exactly the same as the evaluation experiment introduced in Section 7.4.2, with the exception that the RA scheme used here is the proposed QoE aided performance compensation algorithm. The results are

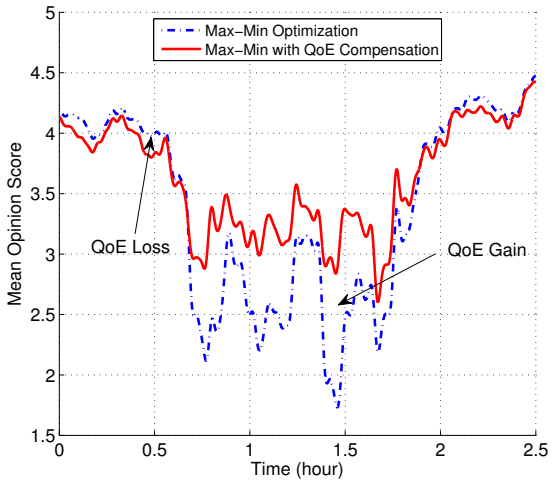

Figure 7.7: QoE Aided Compensation of MOS with Mobility

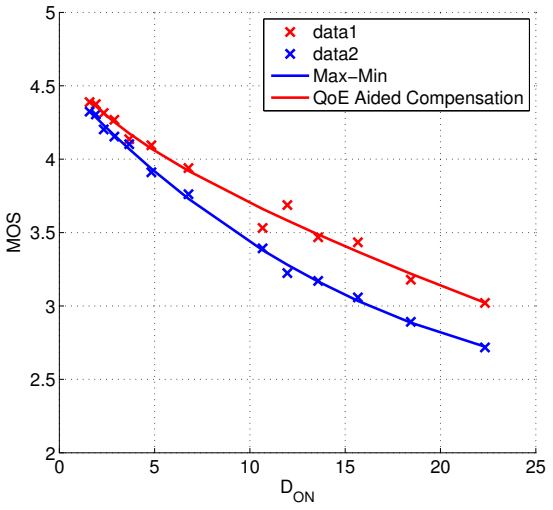

Figure 7.8: QoE Aided Compensation of MOS with PU Activity

shown in Figure 7.8.

7.5.5 Sensitivity Analysis for α

α is an important parameter in the proposed QoE compensation scheme. It is used in the proposed combined priority index for RA: $\phi = R_{TX} + \alpha I_{QoE}$. α works as a coefficient to adjust the weight of the evaluated I_{QoE}, so that the degree of

I_{QoE} affecting the priority index ϕ can be adjusted. The reason that α is required in the QoE compensation scheme is that there is no defined relationship between the allocated resources (transmission rate) to a user and its expected QoE. Of course with a much simpler model, e.g. a simple FIFO queue model, it is possible to predict packet delay and loss with the expected capacity (transmission rate), and then map these to QoE. However, the case we are looking here is a much more complicated case where finding a defined relationship between allocated resources (transmission rate) and the expected QoE is very difficult.

Although an RA's prioritising policy is affected by QoE in the compensation scheme, the fundamental purpose of an RA is still to allocate the transmission rate to users through subchannel allocation. And since we cannot directly translate R_{TX} to QoE, we cannot define an optimisation problem on I_{QoE} itself; instead, αI_{QoE} is employed as a compensation factor. In such a case, the value of α can affect the effect of the QoE compensation scheme. A too large value of α will make the users with poor perceived QoE always very "hungry" and always asking for more resource, and this may cause an overall degradation of QoE for users with originally good perceived QoE. On the other hand, a small value of α is likely to limit the improvement on QoE of the compensation algorithm.

Thus in this section we present a sensitivity analysis on the parameter α. The QoE of all the SUs within the CR network is evaluated while the values of α changes in the range [1, 30]. All the other parameters remain the same for the simulation. The results in Figure 7.9 illustrate the average QoE of each individual SU as α value varies. The solid lines represent the QoE of NRT users while the dotted lines are that of the RT users. As the value of α increases, the QoE of those users with originally poor perceived quality increases as a result of more resources being allocated to them to compensate for their QoE. However, the degradation in QoE for the users with originally good perceived quality can also be observed as α increases. The minimum value of QoE among all users is plotted separately as the value of α varies and the result is shown in Figure 7.10. As shown in the Figure 7.10, the compensation effect increases significantly as α increases from a smaller value. At a value over 15, the effect on QoE by α increment start to flat out. In both cases, $\alpha > 25$ results in an overall degradation of QoE; this is caused by the "saturation flag" being set too early because the

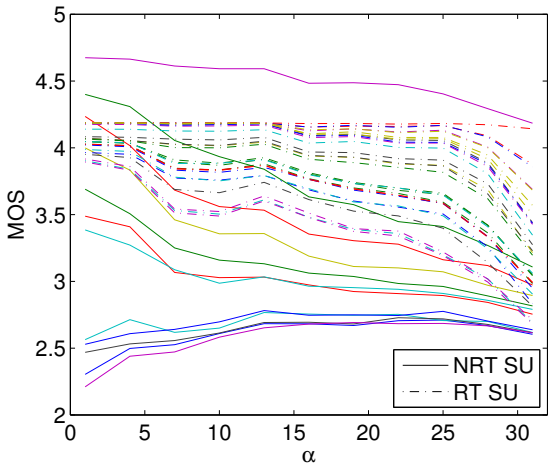

Figure 7.9: SU MOS compensation effect against α

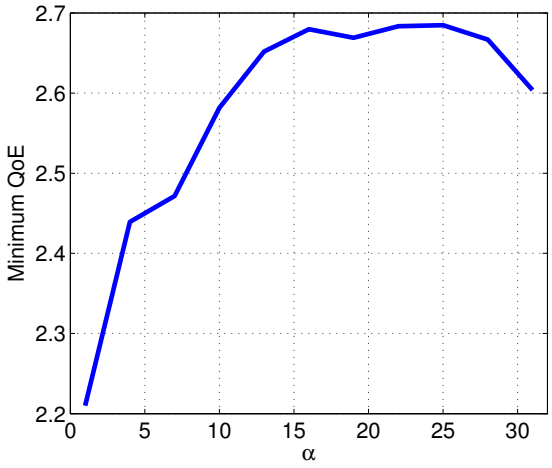

Figure 7.10: Minimum MOS against α

worst case SU is too hard to satisfy. By looking at Figure 7.10, we can see that the maximum value of minimum QoE can be achieved at around $\alpha = 15$, and further increasing the value of α will not increase the minimum QoE. However, by looking at Figure 7.9, at the range of $\alpha = [10, 15]$, the degradation of the QoE of other users are very significant. This is likely to cause a degradation of overall QoE. So, overall, a value of α in between 5 to 10 is recommended, so that QoE can be significantly compensated for SU with originally poor perceived quality, while causing a limited impact on the other users.

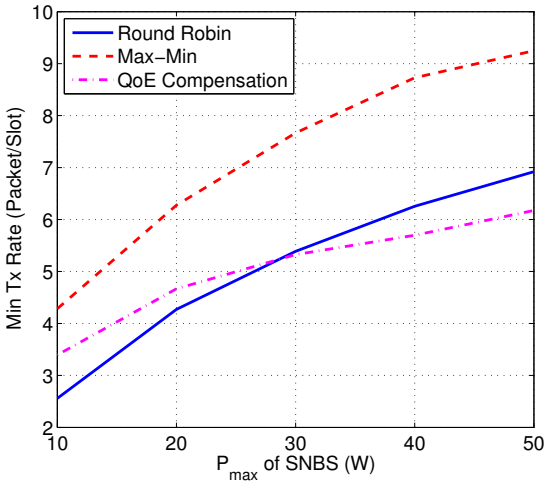

Figure 7.11: Throughput Comparison

7.6 Conclusions

Classical MAC/Physical parameter focused RA schemes are able to reach optimal or near optimal results for the targeted parameter (throughput, spectral efficiency etc). However, these parameters do not directly matter to the end users. What the end users are really concerned about is how good the experience is when they use the network. Here we provide comparisons in terms of throughput and QoE to emphasise this point. Figure 7.11 shows the minimum throughput among all users that can be achieved using different RA schemes. Comparatively, Figure 7.12 shows the minimum QoE that can be achieved with the same three schemes. Note that there is a slight dip in the result at the range of $P_{max} = [10, 20]$. This is caused by the unstable condition resulted from low P_{max}, thus the simulation run length is not long enough to obtain an accurate result (experiment time is already very long, typically more than a week, under current simulation configuration).

It is clear in Figure 7.11 that the Max-Min scheme achieves highest minimum throughput, while the QoE compensation scheme even results in lower minimum throughput compared with a simple round-robin scheme. However, if we look at the MOS under the same condition, a significant gain is achieved with the QoE compensation scheme, even though it does not result in a higher throughput.

Previous CR research has been limited to physical or MAC layer optimization

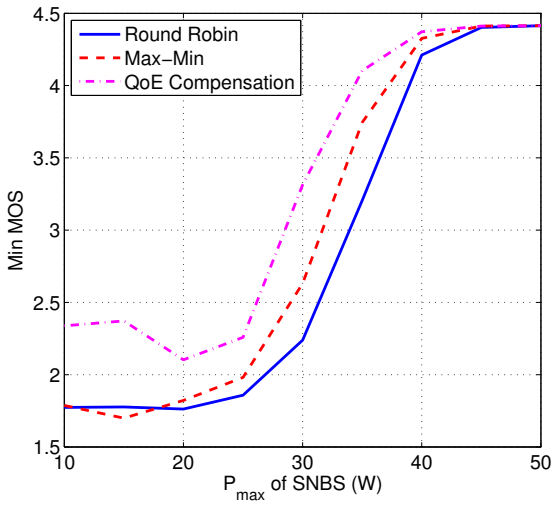

Figure 7.12: QoE Comparison

and lacks consideration of user satisfaction. Our results demonstrate that bursty activity patterns of PUs cause the MOS of users in CR networks to deteriorate far more than less bursty activity patterns. These effects are related to several other parameters such as SU traffic patterns and SU distances from the SNBS. These results expose limitations in existing RA schemes in terms of providing satisfiable QoE in CR. A classical RA algorithm is shown to result in poor performance in certain cases. Based on these results, we have proposed a new cross layer RA scheme, which was the QoE aided performance compensation algorithm. The newly proposed algorithm is tested and is shown to be effective in several scenarios. The user experience is significantly improved in cases where the classical RA scheme was unable to provide acceptable user experience.

Chapter 8

Conclusions and Future Work

In the first part of this chapter, the main contributions and conclusions introduced in this thesis is summarised. Then possible future future work related to my research is addressed in the second part of the chapter.

8.1 Contributions and Conclusions

The CR technique is recognized as a key technology for improving overall spectrum utilization, and providing high rate transmission links for mobile users. Although extensive research has covered different topics for CR networks, many fundamental problems remain unsolved. Especially, existing research on CR RA schemes lacks consideration of real time QoS provisioning for the SUs. The possible effect of activity patterns of the PUs on the coexisting SUs is also neglected in existing research. Most importantly, QoE evaluation and provisioning for SUs are also key factors to the success of the technology, and existing research wasn't able to take them into account. In this thesis, I have introduced my approach to address these problems.

Research on RA schemes in the literature is usually limited to optimisation problems within a short transmission frame time. A system level simulation platform is built in my research, which enables continuous time evaluation of packet level CR transmission parameters. The simulation platform allows continuous time evolution of queue backlog of individual SUs. Other parameters such as the

activity states of SUs and PUs are also allowed to evolve through time to simulate a more accurate practical scenario.

Using this simulator, the impact is evaluated for different SU traffic arrival patterns on the QoS parameters for the SUs. The results demonstrates a significant influence of the traffic arrival process on the queue length, delay and loss of users. The results suggests that packet level traffic arrival pattern needs to be taken into account when designing CR systems. Results also show that packet level QoS parameters for SUs in the system vary significantly with the distance from the SNBS. These results demonstrate the very significant impact of both packet level traffic arrival process, and user distance from the SNBS. The example RA scheme (Max-Min) under study discriminates against users with bursty traffic profiles and SUs at a large distance, meaning that the Max-Min RA scheme is ineffective in providing uniformly good user QoS.

With the help of the continuous time evolving CR Platform I built, I am able to model PU activity patterns dynamically, breaking the limitation in literature where PU activities are usually modelled as static or simplified as abstract probability. The impact of the PU activity patterns on the QoS parameters of the SUs is evaluated in my research. Results show that bursty activity patterns of PUs significantly deteriorate the performance of the CR networks. The bustiness effects are also related to several other parameters such as SU traffic patterns and SU distances from the SNBS. The results show a complicated relationship between all these parameters: SUs that are further away from the base station experience more serious QoS deterioration because of PU activity burstiness, while those nearer to the base station barely have any significant change in QoS. The effect of applying a different PU activity model on SU QoS is also evaluated. Results show that heavy-tail distributed Off period of PU activities results in a much larger SU queue backlogs, indicating that SUs may experience QoS degradation if PU Off activities are heavy-tail distributed, such as the case when PU activities are driven by human actions.

The CR system level simulation platform is then modified to enable simulation of video traffic transmission over the CR network. The method for generating arbitrary long video traffic source using wavelet is introduced, and the generated long video traffic is used to test video transmission over CR networks. Numerical

results show that the significant LRD characteristic of movie traffic results in larger queue backlog than a SRD traffic. However the burstiness in the generic data traffic model contributes to an even larger queue backlog than does the LRD of the video traffic. By adjusting the bustiness of a SRD traffic, it is possible to approximate the queue backlog of a LRD traffic.

I then take a step forward to map the QoS evaluation results obtained in my research to QoE, using existing QoS to QoE mapping techniques. MOS is used to represent user QoE, so that the user perceived quality can be compared for different types of services. The QoE evaluation results reveal limitation of existing RA schemes for CR in providing satisfiable QoE. Users that are further away from the SNBS experience much worse QoE than those close to the SNBS. It is also shown that QoE deteriorates linearly to PU activity burstiness (represented using Index of Dispersion D_{ON}).

Taking QoE into consideration, a novel QoE aided performance compensation RA scheme is proposed in this thesis. The compensation RA scheme prioritise users with poor QoE by increasing the probability of allocating radio resource to these users. A set of experiments were performed to test the effectiveness of the QoE compensation algorithm in different scenarios, including compensation for user distance, user mobility and PU activity patterns. Numerical results show a significant improvement, with limited cost, of QoE in cases where the classical RA scheme was unable to provide acceptable user experience.

8.2 Future Work

The main method used in this research is evaluation of parameters by simulation. The scenario my research focuses on includes so many different factors within a complex system that it is very difficult to formulate an analytical queue model to approximate the simulation results. Finding an analytical model is difficult, yet it is still possible, possibly with careful modelling and abstraction. So one possible future work is to find an analytical queue model for CR that can take into account different factors including traffic pattern and those affecting channel capacity.

The proposed QoE aided performance compensation scheme aims to increase

the probability of users with poor QoE to access radio resource. It is a compensation RA scheme instead of an optimisation RA scheme because the target function is a combined index of user QoE and allocated radio resource, instead of QoE itself. In order to formulate an optimisation problem, it is necessary to find a mapping relationship between allocated radio resource to a user and its perceived QoE. I anticipate that such a QoE optimisation RA scheme can further increase user QoE within a CR network.

The RA scheme for CR in this thesis is based on the Max-Min RA scheme, which is one of the typical classical RA schemes in the literature. However, there are many other existing RA schemes for CR trying to optimise different Physical/MAC layer metrics. I anticipate that the RA compensation scheme proposed in this thesis can be generalised to improve user QoE with any of these RA schemes, as long as they are priority based. In fact, any priority based scheme for wireless radio resource allocation, regardless of whether they are for CR or not, can benefit from the QoE compensation scheme. Future work can include generalisation of the QoE compensation scheme to other cases.

The RA compensation scheme introduced in my thesis has two configurable parameters: the sliding widow size for QoE evaluation, and the compensation coefficient α. A sensitivity analysis for α is already provided in the previous chapter. However, it is possible that changes in the model parameters (different user composition, change in traffic intensity etc.) can affect the optimal value of α. The choosing of value of α can be further analysed. On the other hand, the sliding window size for QoE evaluation is related to user mobility pattern. In the case where user's QoE changes more rapidly, a smaller window size for QoE evaluation should be applied so that the changes can be captured more accurately, and the QoE compensation scheme can respond faster. The analysis of the sliding window size for QoE evaluation is not provided in this thesis, and can also be future work.

8.3 Final Remarks

The cross layer RA scheme proposed in my research demonstrates the significant benefits of using QoE evaluation results to inform the MAC layer RA decision

making process. This opens up a whole new perspective for designing RA scheme in wireless communication systems: optimising MAC layer parameters can be ineffective in terms of satisfying users; taking into account the QoE of users, though not necessarily reaching optimised value for MAC layer parameter, can result in better user satisfaction. This type of cross-layer design for better QoE should be payed more attention to, because, it is the end user experience that matters most.

Appendix A - Preliminary Tests for α

Due to the lack of direct mapping between the resource allocated to an SU and the expected QoE of the user, we are unable to construct an optimisation problem using QoE as a target function. Instead, we take the approach to compensate for the QoE using the combined index:

$$\phi = R_{TX} + \alpha I_{QoE} \tag{1}$$

.

The compensation coefficient α is used to adjust the degree of effect of QoE compensation on the RA algorithm. A too large value of α will cause an overall degradation of QoE for users with originally good experience, while a too small value of α will limit the improvement of the compensation algorithm.

In order to choose an appropriate value of α for testing the effect of the compensation scheme, a set of preliminary tests are performed and are introduced in here. The tests are designed so that the only factors that varies between different tests is the value of α. The observed outcome of the tests is the minimum observed QoE among all the SUs in the CR network.

The minimum QoE among SUs is plotted against α to illustrate the effect of α. The results are shown in Figure 1.

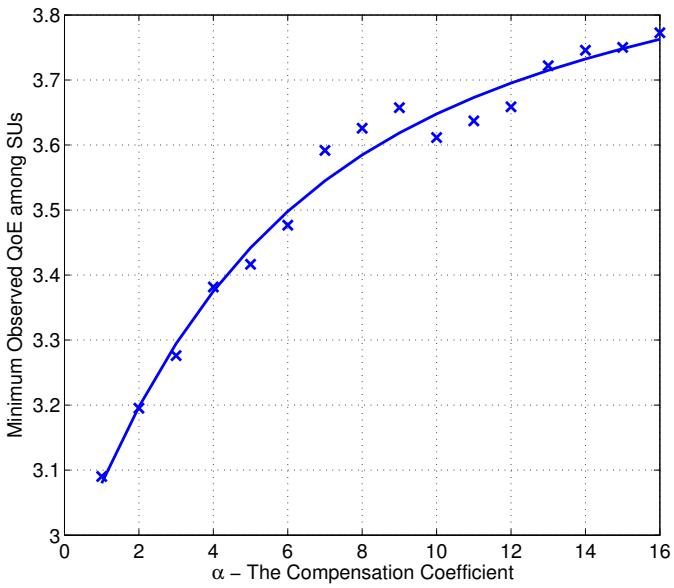

Figure 1: Minimum MOS against α

Appendix B - QoS to QoE Mapping Techniques

.1 Voice over IP Performance Monitoring

ITU-T Recommendation G.107 [42] describes a computational model, known as the E-model, that has proven useful as a transmission planning tool, for assessing the combined effects of variations in several transmission parameters that affect conversational quality of 3.1 kHz handset telephony. This computational model can be used, for example, by transmission planners to help ensure that users will be satisfied with end-to-end transmission performance whilst avoiding over-engineering of networks. The primary output from the model is the "Rating Factor" R but this can be transformed to give estimates of customer opinion. R is a scalar quality rating value and varies directly with the overall conversational quality.

[43] describes a method for monitoring Voice over IP (VolP) applications based upon a reduction of the ITU-T's E-Model to transport level, measurable quantities.

It focuses on the transport layer parameters including delay and packet loss, and rely upon the default values from the E-Model for the rest of the impairment factors. This simplifies the R-Factor from the E-Model as:

$$R = 94.2 - I_d - I_{ef} \tag{2}$$

, where I_d is the delay component and I_{ef} is the equipment impairment component, which includes the packet loss.

Each of these two components are given an analytical expression by estimates extracted from subjective measurements.

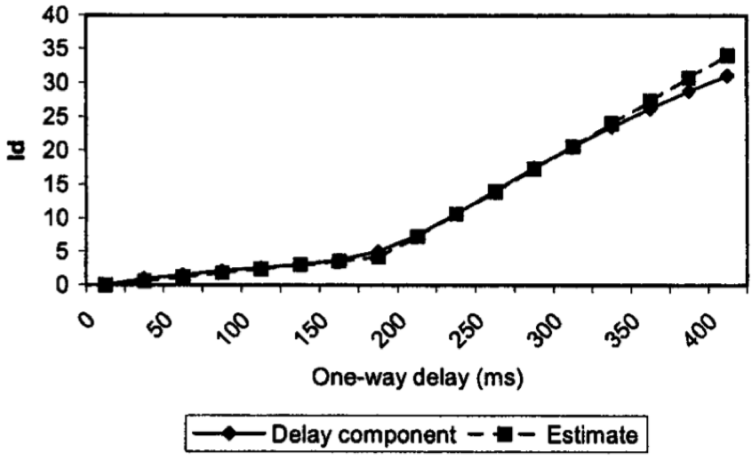

Figure 2: I_d as function of delay

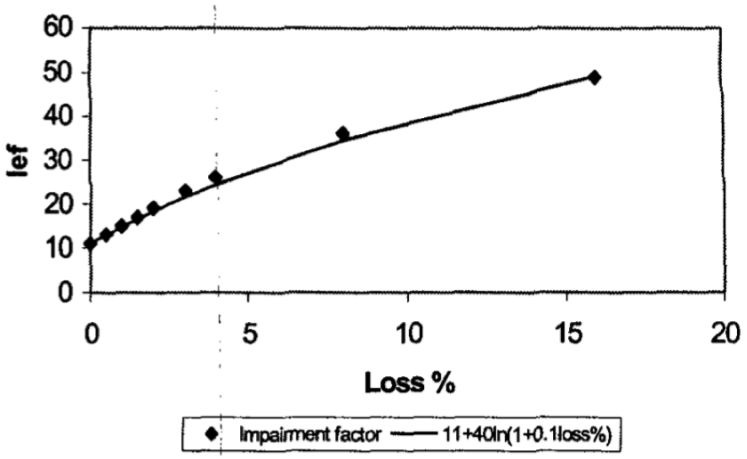

Figure 3: I_{ef} as function of loss

And finally the R factor can be expressed as an analytical function of packet delay and loss:

$$R \sim 94.2 - 0.24d - 0.11(d - 177.3)H(d - 177.3) - 11 - 40ln(1 + 10e) \qquad (3)$$

,where d is the packet delay in ms and e is the packet loss probability.

.2 Mapping of Weighted Session Time to Perceived Web Browsing Quality

[77] applies perceptual models to gauge user satisfaction (i.e., QoE) with the end-to-end performance (i.e., QoS) in terms of response and download times, measured in the network or calculated from the HTTP transaction times. Response and download times in a web session were manipulated, and users were asked to evaluate the perceived quality according to the five-point MOS scale In total, 49 experiments were conducted for each of the three network contexts with varied response and download times. Figure 4 shows the measurement results from [77]. Each point in the graph represents a single experiment with the weighted session time on the x-axis and the MOS on the y-axis.

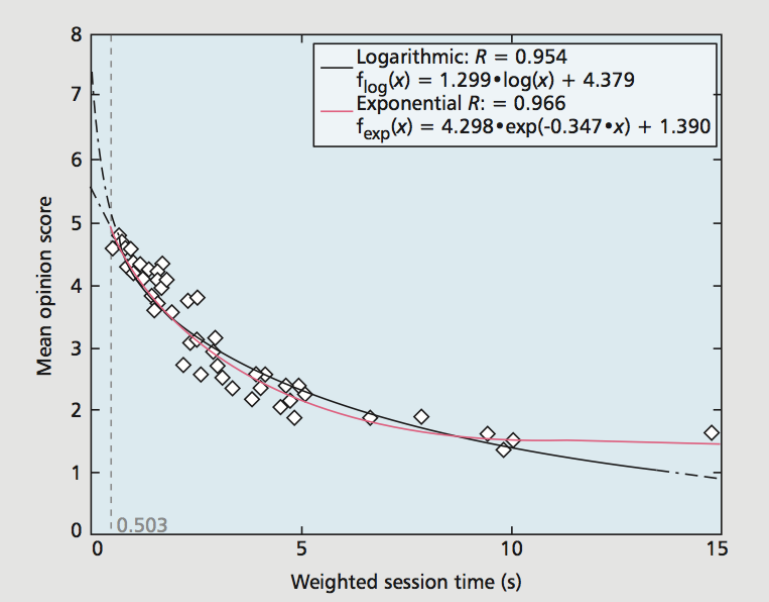

Figure 4: QoE of Web Browsing as Function of Weighted Session Time

[13] proposes the exponential approximation formula for the measurement

results, and is shown to fit quite accurately with the observations.

$$MOS(x) = 4.298 \cdot exp(-0.347 \cdot x) + 1.390 \tag{4}$$

Appendix C - Planning Simulation Time

The CR system model used in this research is very complex and consists interactions of many random variables, which results in a long time for the simulation results to reach their steady states. On the other hand, it is difficult to predict how long the simulations should last in order to obtain accurate enough results. However, there are works that attempt to address the simulation length planning problem. This appendix introduces work from Dr Ling Xu [78], who worked at the same research group with me in her PhD studying.

In [78], Ling claims that simulating large-scale networks can be very time and resource consuming. It can take several days to run one long replication of simulation experiment, which may be unaffordable. Large-scale network often have complex topology, combine various traffic pattern, multiple protocols, and applications, etc. Also, dimensioning and of large-scale network can be difficult, particularly in the presence of disparate traffic mixes. Under this circumstance, traffic partitioning is of great importance, since this will improve corresponding network performance and QoS. Also, the new age of 4G technology is coming, which provides greater bandwidth, higher data rates, efficient spectrum use, etc. Dimensioning 4G mobile will be even more challenging.

The packet multiplexing model that Ling uses in her research is already more complex than the ones in existing literature. Figure 5 illustrates the packet multiplex model in Ling's work. N homogenous ON/OFF VoIP packet sources are multiplexed into a FIFO queue, with service rate C (in pps). Each ON/OFF packet source generates packets with rate h (in pps) when active (the ON state), and sends no packets when it is idle (the OFF state).

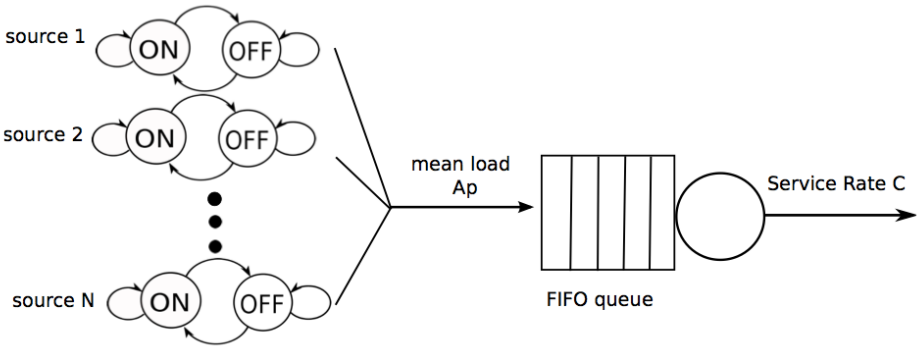

Figure 5: Packet Multiplexing Model

The required simulation run length is provided in Ling's work as:

$$N_r(WT) = 4 \cdot \frac{z^2_{1-\beta/2}}{\varepsilon_r^2} \cdot c^2(WT) \tag{5}$$

, where ε_r is predefined relative width and $z^2_{1-\beta/2}$ describing the confidence interval. And the required simulation time δT can be calculated using $N_r(WT)$ and sample interval.

Figure 6 shows the predicted required simulation time using Ling's formula.

Ling's simulation time prediction formula is shown to work fine with the simplified model using a FIFO queue and constant service rate. However, it is not directly applicable to the CR scenario introduced in this thesis, because the randomness introduced by wireless channel and interruptions from the PUs makes the model much more complex. Thus the strategy used in our research is to allow simulation to run as long as possible.

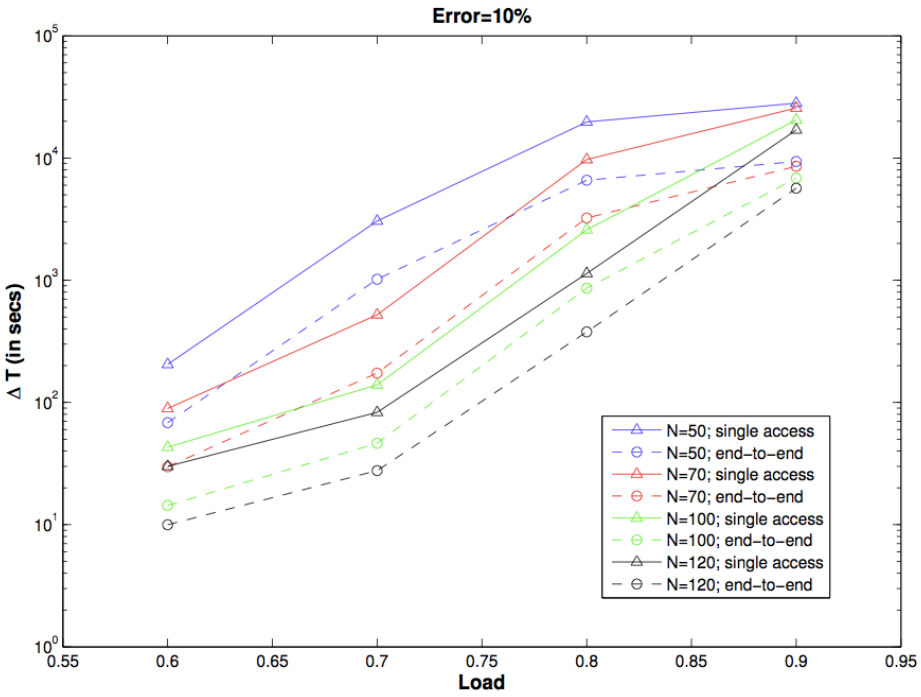

Figure 6: Simulation Time Prediction

References

[1] I.F. Akyildiz, W.Y. Lee, M.C. Vuran, and S. Mohanty. Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey. *Computer Networks*, 50(13):2127–2159, 2006. ix, 6, 7

[2] P. Mitran, L.B. Le, and C. Rosenberg. Queue-aware resource allocation for downlink ofdma cognitive radio networks. *Wireless Communications, IEEE Transactions on*, 9(10):3100–3111, 2010. ix, 2, 17, 18, 23, 25, 34, 36, 37, 38, 41, 89, 99, 100

[3] F.C.C.S.P.T. Force. Report of the spectrum efficiency working group. *Washington DC, November of*, 2002. 1

[4] J. Mitola III and G.Q. Maguire Jr. Cognitive radio: making software radios more personal. *Personal Communications, IEEE*, 6(4):13–18, 1999. 1, 6

[5] T.A. Weiss and F.K. Jondral. Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency. *Communications Magazine, IEEE*, 42(3):S8–14, 2004. 1, 17

[6] X. Kang, H. Garg, Y.C. Liang, and R. Zhang. Optimal power allocation for ofdm-based cognitive radio with new primary transmission protection criteria. *Wireless Communications, IEEE Transactions on*, 9(6):2066–2075, 2010. 2, 9, 10, 89

[7] Y. Zhang and C. Leung. A distributed algorithm for resource allocation in ofdm cognitive radio systems. *Vehicular Technology, IEEE Transactions on*, 60(2):546–554, 2011. 2

[8] T. Qin and C. Cyril Leung. Fair adaptive resource allocation for multiuser ofdm cognitive radio systems. In *Communications and Networking in China, 2007. CHINACOM'07. Second International Conference on*, pages 115–119. IEEE, 2007. 2, 9, 10, 89

[9] C.H. Chen and C.L. Wang. Joint subcarrier and power allocation in multiuser ofdm-based cognitive radio systems. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–5. IEEE, 2010. 2, 89

[10] Jie Zhang, Tiejun Lv, Hui Gao, and Yueming Lu. Joint uplink power and subchannel allocation in cognitive radio network. In *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*, pages 306–311. IEEE, 2012. 2, 89

[11] Mathieu Lessinnes, J-M Dricot, Philippe De Doncker, Luc Vandendorpe, and Francois Horlin. Dynamic resource allocation for mimo cognitive networks with low control traffic and low computational complexity. *Vehicular Technology, IEEE Transactions on*, 62(4):1732–1740, 2013. 2, 89

[12] A.L. Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005. 3, 10, 12, 59, 65

[13] M. Fiedler, T. Hossfeld, and P. Tran-Gia. A generic quantitative relationship between quality of experience and quality of service. *Network, IEEE*, 24(2):36–41, 2010. 4, 15, 16, 92, 93, 120

[14] M. Siller and J. Woods. Improving quality of experience for multimedia services by qos arbitration on a qoe framework. In *International Conference on Packet Video*, pages 28–29. Citeseer, 2003. 4

[15] D. Soldani, D. Chiavelli, J. Laiho, M. Li, N. Muhammad, G. Giambiasi, and C. Rodriquez. *QoE and QoS Monitoring*. Wiley Online Library, 2006. 4

[16] M. Siller and J. Woods. Using an agent based platform to map quality of service to experience in conventional and active networks. *IEE Proceedings-Communications*, 153:828, 2006. 4

[17] Global mobile data traffic forecast update, 2011-2016. 6, 69

[18] Andrea Goldsmith, Syed Ali Jafar, Ivana Maric, and Sudhir Srinivasa. Breaking spectrum gridlock with cognitive radios: An information theoretic perspective. *Proceedings of the IEEE*, 97(5):894–914, 2009. 8

[19] Chao Wang, Ming Xiao, and Lars Rasmussen. Performance analysis of coded secondary relaying in overlay cognitive radio networks. In *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*, pages 294–299. IEEE, 2012. 8

[20] Goochul Chung, Sriram Sridharan, Sriram Vishwanath, and Chan Soo Hwang. On the capacity of overlay cognitive radios with partial cognition. *Information Theory, IEEE Transactions on*, 58(5):2935–2949, 2012. 8

[21] Deepak R Joshi, Dimitrie C Popescu, and Octavia A Dobre. Joint spectral shaping and power control in spectrum overlay cognitive radio systems. *Communications, IEEE Transactions on*, 60(9):2396–2401, 2012. 8

[22] Han Hu and Qi Zhu. Dynamic spectrum access in underlay cognitive radio system with sinr constraints. In *Wireless Communications, Networking and Mobile Computing, 2009. WiCom'09. 5th International Conference on*, pages 1–4. IEEE, 2009. 8, 18

[23] Hela Hakim, Hatem Boujemaa, and Wessam Ajib. Performance comparison between adaptive and fixed transmit power in underlay cognitive radio networks. 2013. 8, 18

[24] Myung Keun Yoon, Ki Hwan Lee, and Ju Bin Song. Performance analysis of distributed cooperative spectrum sensing for underlay cognitive radio. In *Advanced Communication Technology, 2009. ICACT 2009. 11th International Conference on*, volume 1, pages 338–343. IEEE, 2009. 8, 18

[25] Shu-Hsien Wang, Chih-Yu Hsu, and Y-WP Hong. Distributed exploitation of spectrum and channel state information for channel reservation and selection in interweave cognitive radio networks. *Wireless Communications, IEEE Transactions on*, 12(7):3458–3472, 2013. 8

[26] A Swindlehurst et al. Prescient precoding in heterogeneous dsa networks with both underlay and interweave mimo cognitive radios. 2013. 8

[27] D. Cabric, S.M. Mishra, and R.W. Brodersen. Implementation issues in spectrum sensing for cognitive radios. In *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Eighth Asilomar Conference on*, volume 1, pages 772–776. IEEE, 2004. 8

[28] T. Yucek and H. Arslan. A survey of spectrum sensing algorithms for cognitive radio applications. *Communications Surveys & Tutorials, IEEE*, 11:116–130, 2009. 8

[29] G. Ganesan and Y. Li. Cooperative spectrum sensing in cognitive radio, part i: Two user networks. *Wireless Communications, IEEE Transactions on*, 6:2204–2213, 2007. 8

[30] G. Taricco. Optimization of linear cooperative spectrum sensing for cognitive radio networks. *Selected Topics in Signal Processing, IEEE Journal of*, 5(1):77–86, 2011. 8

[31] G. Ganesan and Y. Li. Cooperative spectrum sensing in cognitive radio networks. In *New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on*, pages 137–143. IEEE, 2005. 8

[32] M.M. Rashid, J. Hossain, E. Hossain, and V.K. Bhargava. Opportunistic spectrum access in cognitive radio networks: a queueing analytic model and admission controller design. In *Global Telecommunications Conference, 2007. GLOBECOM'07. IEEE*, pages 4647–4652. IEEE, 2007. 10, 89

[33] C. Pei-Pei, Z. Qin-Yu, and W. Ye. Qos based resource allocation for ofdm-based cognitive radio systems with mixed services. In *Wireless Communications Networking and Mobile Computing (WiCOM), 2010 6th International Conference on*, pages 1–5. IEEE, 2010. 11, 89

[34] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz. Primary users in cellular networks: A large-scale measurement study. In *New Frontiers in Dynamic*

Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on, pages 1–11. IEEE, 2008. 11, 58, 97

[35] L. Csurgai-Horvath and J. Bito. Primary and secondary user activity models for cognitive wireless network. In *Telecommunications (ConTEL), Proceedings of the 2011 11th International Conference on*, pages 301–306. IEEE, 2011. 12, 58

[36] Bruce D Fritchman. A binary channel characterization using partitioned markov chains. *Information Theory, IEEE Transactions on*, 13(2):221–227, 1967. 12

[37] J. Riihijarvi, J. Nasreddine, and P. Mahonen. Impact of primary user activity patterns on spatial spectrum reuse opportunities. In *Wireless Conference (EW), 2010 European*, pages 962–968. IEEE, 2010. 13, 58

[38] D.T. Ngo, C. Tellambura, and H.H. Nguyen. Resource allocation for ofdma-based cognitive radio multicast networks with primary user activity consideration. *Vehicular Technology, IEEE Transactions on*, 59(4):1668–1679, 2010. 13, 58

[39] International Telecommunication Union. Telecommunication Standardization Sector. Definition of quality of experience (qoe). (COM 12 LS 62 E), 2007. 13, 91, 92

[40] International Telecommunication Union. Telecommunication Standardization Sector. Quality of service and dependability vocabulary. (E.800), 1998. 14

[41] International Telecommunication Union. Telecommunication Standardization Sector. Communications quality of service: A framework and definitions. (G.1000), 2001. 14

[42] International Telecommunication Union. Telecommunication Standardization Sector. The e-model, a computational model for use in transmission planning. (G.107), 1998. 15, 92, 118

[43] R.G. Cole and J.H. Rosenbluth. Voice over ip performance monitoring. *ACM SIGCOMM Computer Communication Review*, 31(2):9–24, 2001. 15, 16, 92, 93, 118

[44] A.W. Rix, J.G. Beerends, M.P. Hollier, and A.P. Hekstra. Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 2, pages 749–752. Ieee, 2001. 15

[45] Z. Qiao, L. Sun, and E. Ifeachor. Case study of pesq performance in live wireless mobile voip environment. In *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pages 1–6. IEEE, 2008. 16

[46] S. Singh, H.P. Singh, and J. Singh. Spectral analysis of speech quality in voip for g. 729a and amr-wb speech coders. In *Computational Intelligence, Communication Systems and Networks (CICSyN), 2010 Second International Conference on*, pages 182–187. IEEE, 2010. 16

[47] I. Tsompanidis, G. Fortetsanakis, T. Hirvonen, and M. Papadopouli. Analyzing the impact of various wireless network conditions on the perceived quality of voip. In *Local and Metropolitan Area Networks (LANMAN), 2010 17th IEEE Workshop on*, pages 1–6. IEEE, 2010. 16

[48] S. Li, T.H. Luan, and X. Shen. Channel allocation for smooth video delivery over cognitive radio networks. In *GLOBECOM 2010, 2010 IEEE Global Telecommunications Conference*, pages 1–5. IEEE, 2010. 16

[49] C.N. Ververidis, J. Riihijarvi, and P. Mahonen. Evaluation of quality of experience for video streaming over dynamic spectrum access systems. In *World of Wireless Mobile and Multimedia Networks (WoWMoM), 2010 IEEE International Symposium on a*, pages 1–8. IEEE, 2010. 16

[50] Ke-Lin Du and Madisetti NS Swamy. *Wireless communication systems: from RF subsystems to 4G enabling technologies*. Cambridge University Press, 2010. 21

[51] Ali Ghrayeb and Tolga M Duman. Performance analysis of mimo systems with antenna selection over quasi-static fading channels. *IEEE Transactions on vehicular technology*, 52(2):281–288, 2003. 21

[52] Youjian Liu, Kin N Lau, Oscar Y Takeshita, and Michael P Fitz. Optimal rate allocation for superposition coding in quasi-static fading channels. In *Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on*, page 111. IEEE, 2002. 21

[53] Qingwen Liu, Shengli Zhou, and Georgios B Giannakis. Cross-layer combining of adaptive modulation and coding with truncated arq over wireless links. *Wireless Communications, IEEE Transactions on*, 3(5):1746–1755, 2004. 22

[54] Ming Jia, Jianglei Ma, Wen Tong, Dong-Sheng Yu, and Peiying Zhu. Adaptive modulation and coding, September 5 2006. US Patent 7,103,325. 22

[55] Hermann Bischl, Hartmut Brandt, Tomaso de Cola, Riccardo De Gaudenzi, Ernst Eberlein, Nicolas Girault, Eric Alberty, Stefan Lipp, Rita Rinaldo, Bjarne Rislow, et al. Adaptive coding and modulation for satellite broadband networks: From theory to practice. *International Journal of Satellite Communications and Networking*, 28(2):59–111, 2010. 22

[56] Long Bao Le, Patrick Mitran, and Catherine Rosenberg. Queue-aware subchannel and power allocation for downlink ofdm-based cognitive radio networks. In *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, pages 1–6. IEEE, 2009. 23

[57] Rajamani Ganesh and K Joseph. Effect of non-uniform traffic distributions on performance of a cellular cdma system. In *Universal Personal Communications Record, 1997. Conference Record., 1997 IEEE 6th International Conference on*, volume 2, pages 598–602. IEEE, 1997. 25

[58] Xue Tang and Hongwen Yang. Effect of user distribution on the capacity of cellular networks. In *2012 National Conference on Information Technology and Computer Science*. Atlantis Press, 2012. 25

[59] Mohamed Elalem and Lian Zhao. Realistic user distribution and its impact on capacity and coverage for a wcdma mobile network. In *Sarnoff Symposium, 2009. SARNOFF'09. IEEE*, pages 1–5. IEEE, 2009. 25

[60] Jean Paul Linnartz. Distribution of amplitude and power for rician fading, 1999. 27

[61] Pierre LEcuyer. Non-uniform random variate generations. In *International Encyclopedia of Statistical Science*, pages 991–995. Springer, 2011. 27

[62] Predrag R Jelenković and Aurel A Lazar. Asymptotic results for multiplexing subexponential on-off processes. *Advances in Applied Probability*, 31(2):394–421, 1999. 59

[63] Nicholas G Duffield and Neil O'connell. Large deviations and overflow probabilities for the general single-server queue, with applications. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 118, pages 363–374. Cambridge Univ Press, 1995. 59

[64] Cisco. Cisco visual networking index: Global mobile data traffic forecast update, 20132018, 2014. 69

[65] FOCUS GROUP ON IPTV International Telecommunication Union. Telecommunication Standardization Sector. Report of the 2nd focus group on ip television (iptv) meeting. (FG IPTV-R-0014), 2006. 69

[66] Soohong Park and Seong-Ho Jeong. Mobile iptv: Approaches, challenges, standards, and qos support. *Internet Computing, IEEE*, 13(3):23–31, 2009. 69

[67] Ismail Djama and Toufik Ahmed. A cross-layer interworking of dvb-t and wlan for mobile iptv service delivery. *Broadcasting, IEEE Transactions on*, 53(1):382–390, 2007. 69

[68] James She, Fen Hou, Pin-Han Ho, and Liang-Liang Xie. Iptv over wimax: Key success factors, challenges, and solutions [advances in mobile multimedia]. *Communications Magazine, IEEE*, 45(8):87–93, 2007. 69

[69] M.W. Garrett and W. Willinger. Analysis, modeling and generation of self-similar vbr video traffic. In *ACM SIGCOMM Computer Communication Review*, volume 24, pages 269–280. ACM, 1994. 72

[70] S. Ma and C. Ji. Modeling heterogeneous network traffic in wavelet domain. *IEEE/ACM Transactions on Networking (TON)*, 9(5):634–649, 2001. 72, 76

[71] O. Lazaro, D. Girma, and J. Dunlop. Real-time generation of synthetic mpeg-4 video traffic using wavelets. In *Vehicular Technology Conference, 2001. VTC 2001 Fall. IEEE VTS 54th*, volume 1, pages 418–422. IEEE, 2001. 72

[72] C. Huang, M. Devetsikiotis, I. Lambadaris, and A.R. Kaye. Modeling and simulation of self-similar variable bit rate compressed video: a unified approach. In *ACM SIGCOMM Computer Communication Review*, volume 25, pages 114–125. ACM, 1995. 72, 75

[73] Hong Fei and Wu Zhimei. Multifractal analysis and model of the mpeg-4 video traffic. In *Performance, Computing, and Communications Conference, 2003. Conference Proceedings of the 2003 IEEE International*, pages 463–467. IEEE, 2003. 76

[74] Video trace library, http://web.cecs.pdx.edu/ wuchi/video/mjpeg/index.html. 77, 81

[75] Mahmood M. Zonoozi and Prem Dassanayake. User mobility modeling and characterization of mobility patterns. *Selected Areas in Communications, IEEE Journal on*, 15(7):1239–1252, 1997. 103

[76] International Telecommunication Union. Telecommunication Standardization Sector. *Methods for Subjective Determination of Transmission Quality*. International Telecommunication Union, 1996. 104

[77] International Telecommunication Union. Telecommunication Standardization Sector. Estimating end-to-end performance in ip networks for data applications. (G.1030), 2005. 120

[78] Ling Xu. *PhD Thesis: Planning Simulation Run Length in Packet Queues in Communications Networks*. School of EECS, Queen Mary University of London, 2013. 122