



UNIVERSIDADE FEDERAL DO CEARÁ
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

Congestion Control Algorithms for 4G OFDMA-based Systems with Mixed Traffic Scenarios

Master of Science Thesis

Author

Evilásio Oliveira Lucena

Advisor

Prof. Dr. Walter da Cruz Freitas Júnior

FORTALEZA

2010

EVILÁSIO OLIVEIRA LUCENA

CONGESTION CONTROL ALGORITHMS FOR 4G OFDMA-BASED SYSTEMS WITH MIXED
TRAFFIC SCENARIOS

Dissertação submetida à Coordenação do Programa de Pós-graduação em Engenharia de Teleinformática, da Universidade Federal do Ceará, como parte dos requisitos para obtenção do grau de Mestre em Engenharia de Teleinformática.

Área de concentração: Sinais e Sistemas

Orientador: Prof. Dr. Walter da Cruz Freitas Júnior

FORTALEZA

2010

Ficha elaborada pela bibliotecária Umbelina Caldas Neta - CRB558-CE

L986c Lucena, Evilásio Oliveira
Congestion control algorithms for 4G OFDMA – based systems with mixed traffic scenarios = Algoritmos de controle de carga para sistemas 4G OFDMA com tráfego misto / Evilásio Oliveira Lucena, 2010.
72f. ; il.; enc.

Orientador: Prof. Dr. Walter da Cruz Freitas Júnior
Área de concentração: Sinais e sistemas
Dissertação (Mestrado) - Universidade Federal do Ceará, Departamento de Engenharia de Teleinformática, Fortaleza, 2010.

1. Teleinformática. 2. Sinais e Sistemas. 3. Sistemas de comunicação sem fio. 4. Processamento de sinais. I. Freitas Júnior, Walter da Cruz (orient.). II. Universidade Federal do Ceará – Programa de Pós- Graduação em Engenharia de Teleinformática. III. Título.

CDD 621.38

EVILÁSIO OLIVEIRA LUCENA

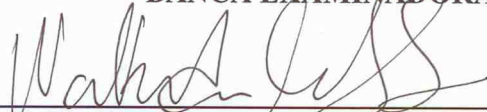
CONGESTION CONTROL ALGORITHMS FOR 4G OFDMA-BASED SYSTEMS WITH
MIXED TRAFFIC SCENARIOS

Dissertação submetida à Coordenação do Programa de Pós-Graduação em Engenharia de Teleinformática, da Universidade Federal do Ceará, como requisito parcial para a obtenção do grau de Mestre em Engenharia de Teleinformática.

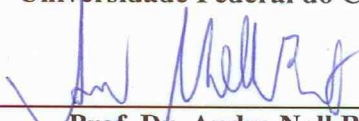
Área de concentração _____.

Aprovada em 23/04/2010.

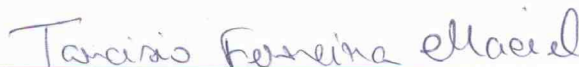
BANCA EXAMINADORA



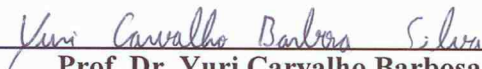
Prof. Dr. Walter da Cruz Freitas Júnior (Orientador)
Universidade Federal do Ceará -UFC



Prof. Dr. Andre Noll Barreto
Universidade de Brasília -UnB



Prof. Dr. Tarcísio Ferreira Maciel
Universidade Federal do Ceará -UFC



Prof. Dr. Yuri Carvalho Barbosa
Universidade Federal do Ceará -UFC

*To the loving memory of my father and to my
mother, who has educated me.*

Acknowledgements

First and foremost, I would like to thank God for giving me the life and for always guiding me through hard, but safe ways. I would also like to thank my family, specially my mother Rosália, as well as my siblings for their love and support through all my life. I also thank my aunts "Nir" and "Nise", and my uncle Heladio, for their help always I need.

I would like to thank Professor Rodrigo Cavalcanti for believing me and for giving me the opportunity to work at GTEL. I would also like to thank all UFC.22 staff, but specially Alex Silva, Rafael Lima and Professors Walter Freitas and Tarcisio Maciel for their guidance in my research.

I am also grateful to the following colleagues from GTEL, with whom I spent hours of studies, and whose suggestions and encouragement were invaluable throughout my Master's course: Igor Guerreiro, Ícaro Silva, Lígia Sousa and Darlan Cavalcante. I would also like to thank Patrícia Lana for giving me lots of smart tips on how to write correctly and politely in English.

Further thanks to the National Council for Scientific and Technological Development (CNPq) for giving me an initial financial support through Master's scholarship during some months, and to the Research and Development Centre, ERICSSON Telecomunicações S/A, for also giving me financial support under EDB/UFC.22 Technical Cooperation contract.

Last but not least, my special thanks to some friends, who always help me with their friendship: Levy, Júlio Cesar, Pedro André, Décio, "Farelo", Camilo, Franco, Davyd, Leonardo and Anna Karla (or just "Ly"). I would also like to thank to all my friends from Shalom, specially Raylanno, Randara, "Ninho", Jader, Pedro, and specially its main founder Moisés Azevedo, and its cofounder Emmir Nogueira, for their example of faith and humbleness.

Resumo

Um conjunto de algoritmos adaptativos para controle de carga para redes com requerimento e suporte à 4a Geração (ou, simplesmente, 4G) empregando a técnica de Múltiplo Acesso por Divisão de Frequências Ortogonais (conhecido por OFDMA, do inglês *Orthogonal Frequency Division Multiple Access*) para o enlace direto são estudados nesta dissertação de mestrado. Este conjunto de algoritmos adaptativos para controle de carga é a generalização do conjunto proposto por Rodrigues, E.B e Cavalcanti, F.R.P. Esta generalização foi motivada pela adoção da técnica Orthogonal Frequency Division Multiple Access (OFDMA) por sistemas 4G, tais como Worldwide Interoperability for Microwave Access (WiMAX) and Long Term Evolution (LTE)-Advanced. Apesar do desempenho dos usuários do serviço de voz sobre IP (conhecido por VoIP, do inglês *Voice over IP*) ter melhorado com a adoção do conjunto de algoritmos adaptativos, a taxa de erro de quadros (conhecida como FER, do inglês *Frame Error Rate*) pode ser melhor controlada se seu aumento for previsto.

Um dos fatores que podem provocar o aumento da FER dos usuários VoIP é a perda de pacote. Um pacote é considerado perdido quando não é transmitido antes do tempo limite imposto pelas restrições de atraso. Em consequência, além da generalização do conjunto de algoritmos adaptativos para controle de carga para redes com requerimento e suporte a sistemas 4G empregando OFDMA no enlace direto, a principal contribuição apresentada nesta dissertação de mestrado é uma ferramenta que prevê a sobrecarga por meio do atraso adicional a esse conjunto de algoritmos adaptativos.

Por fim, é mostrado que esse conjunto de algoritmos com a ferramenta adicional de previsão de sobrecarga baseada no atraso não apenas garante a provisão da qualidade do serviço (conhecida por QoS, do inglês *Quality of Service*) para o VoIP, impondo somente uma pequena degradação aos usuários de tempo não-real (como usuários que acessam à internet), como também reduz picos de FER.

Palavras-chave: Predição de sobrecarga, Serviços de Tempo Real, Atraso.

Abstract

In this master thesis, we discuss an adaptive Congestion Control (CC) framework for networks employing OFDMA in the downlink. The adaptive CC is a generalization of a CC framework proposed before by Rodrigues, E.B and Cavalcanti, F.R.P. This generalization was motivated by the adoption of OFDMA by Fourth Generation (4G) systems, such as WiMAX and LTE-Advanced. Although the performance of the Voice over IP (VoIP) service can be improved with the adoption of adaptive CC, the VoIP Frame Erasure Rate (FER) can be better controlled if we could predict its build-up.

One of the factors that can increase the FER of VoIP flows is the packet loss for example, when a packet is not transmitted before the deadline imposed by delay constraints. As a result, besides the framework generalization for OFDMA 4G systems, the main contribution of this master thesis is an additional feature to the adaptive CC framework for overload prediction based on delay. In conclusion, we show that the generalized framework with the additional feature does not only guarantee the Quality of Service (QoS) fulfillment of VoIP, while imposing only a small performance degradation to the World Wide Web (WWW) service, but also reduces FER peaks.

Key-words: Congestion Control, Overload Prediction, Real Time (RT) Services, Delay, OFDMA.

List of Acronyms

2G	Second Generation
3GPP	3rd. Generation Partnership Project
3G	Third Generation
4G	Fourth Generation
AC	Admission Control
AD	Attack-Decay
ADSL	Asymmetric Digital Subscriber Line
AMPS	Advanced Mobile Phone Service
AMR	Adaptive Multirate
ARMA	Autoregressive Moving Average
B3G	Beyond 3G
BS	Base Station
CAPEX	Capital Expenditure
CEPT	European Conference of Postal and Telecommunications Administrations
CC	Congestion Control
CDF	Cumulative Distribution Function
CDMA	Code-Division Multiple Access
EDB	Estimated Delay Based
EFLC	Error Feedback Based Load Control
eNB	Enhanced Node B
ETSI	European Telecommunication Standards Institute
EU	Enhanced Uplink
FDD	Frequency-Division Duplexing
FDM	Frequency Division Multiplexing

FDMA	Frequency Division Multiple Access
FER	Frame Erasure Rate
GSM	Global System for Mobile Communications
GPRS	General Packet Radio Services
HSDPA	High-Speed Downlink Packet Access
HSPA	High Speed Packet Access
IP	Internet Protocol
ISI	Inter Symbol Interference
JLC	Jump Based Load Control
LC	Load Control
LTE	Long Term Evolution
MDB	Measured Delay Based
MPF	Multicarrier Proportional Fair
NMT	Nordic Mobile Telephony
NRT	Non-Real Time
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OLPC	Outer-loop Power Control
OPEX	Operational Expenditure
PDC	Personal Digital Cellular
PF	Proportional Fair
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadri-Phase Shift Keying
RNC	Radio Network Controller
RRM	Radio Resource Management
RRU	Radio Resource Unit
RT	Real Time
SAC	Session Admission Control
SAE	System Architecture Evolution
SES	Simple Exponential Smoothing
SMS	Short Message Service

SNR	Signal-to-Noise Ratio
TD	Time Division
TDD	Time-Division Duplexing
TDMA	Time Division Multiple Access
TTI	Transmission Time Interval
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
UTRA	Universal Terrestrial Radio Access
UTRAN	Universal Terrestrial Radio Access Network
VoIP	Voice over IP
WCDMA	Wideband CDMA
WiMAX	Worldwide Interoperability for Microwave Access
WMPF	Weighted Multicarrier Proportional Fair
WPF	Weighted Proportional Fair
WWW	World Wide Web

Contents

1 Introduction	16
1.1 Context of the Problem	16
1.2 Motivation and Objectives	17
1.3 Scientific Production and Contributions	17
1.4 Organisation of this Master Thesis	18
2 Fundamentals: Wireless Systems, Services and RRM Strategies	19
2.1 Brief History of Evolution of Mobile Generations	19
2.2 Basics of Radio Resource Management	21
2.3 Orthogonal Frequency Division Multiple Access (OFDMA)	27
2.4 State-of-the-Art	29
3 Adaptive Congestion Control Framework	31
3.1 Admission control	31
3.1.1 Measured Delay Based (MDB)	32
3.2 Scheduling	33
3.2.1 Weighted Multicarrier Proportional Fair (WMPF)	33
3.3 Load control	34
3.3.1 Error Feedback Based Load Control (EFLC)	34
3.4 Overload Prediction Based on Delay	35
3.4.1 Motivation	36
3.4.2 Formulation	37
4 Simulation Results	38
4.1 Simulation tool	38
4.2 Simulation parameters	38
4.2.1 Discussion	39
4.2.2 Saving Results	39
4.3 Performance metrics and simulation scenarios	41
4.4 Results	43
4.4.1 Dynamic of Parameters in Time and their behavior	43
4.4.2 Results	45
4.4.3 Study Case of Service Balancing	63
5 Conclusions and Perspectives	66

Appendix A Traffic Modeling

67

Bibliography

69

List of Figures

2.1	Evolution of 4G Technology.	21
2.2	Congestion control in the presence of difference service types.	24
2.3	Operation of short term RRM strategies with NRT load.	25
2.4	Frequency-time representation of an OFDM signal.	28
2.5	Frequency-time resource grid in OFDMA.	29
3.1	Session admission control scheme applied.	32
3.2	Time variation of FER_{VoIP}^{filt} , α and β	35
3.3	FER.	36
3.4	Delay Prediction.	37
4.1	Flows convergence.	41
4.2	Time variation of α , β and FER_{VoIP}^{filt} at the load of 0.875 Users/s.	43
4.3	Time variation of α , β and FER_{VoIP}^{filt} at the load of 1 Users/s.	44
4.4	Time variation of α , β and FER_{VoIP}^{filt} at the load of 1.125 Users/s.	44
4.5	Time variation of α , β and FER_{VoIP}^{filt} at the load of 1.25 Users/s.	45
4.6	Blocking Rate of flows of both services with mix of 25% of VoIP flows and 75% of WWW ones.	46
4.7	Blocking Rate of flows of both services with mix of 50% of VoIP flows and 50% of WWW ones.	47
4.8	Blocking Rate of flows of both services with mix of 75% of VoIP flows and 25% of WWW ones.	47
4.9	Throughput of WWW flows for mix of 25% of VoIP flows and 75% of WWW ones with <i>Non adaptive CC</i> framework.	48
4.10	Throughput of WWW flows for mix of 50% of VoIP flows and 50% of WWW ones with <i>Non adaptive CC</i> framework.	48
4.11	Throughput of WWW flows for mix of 75% of VoIP flows and 25% of WWW ones with <i>Non adaptive CC</i> framework.	49
4.12	Throughput of WWW flows for mix of 25% of VoIP flows and 75% of WWW ones with <i>Adaptive CC</i> framework.	49
4.13	Throughput of WWW flows for mix of 50% of VoIP flows and 50% of WWW ones with <i>Adaptive CC</i> framework.	50
4.14	Throughput of WWW flows for mix of 75% of VoIP flows and 25% of WWW ones with <i>Non adaptive CC</i> framework.	50
4.15	Throughput of WWW flows for mix of 25% of VoIP flows and 75% of WWW ones with <i>Delay-based Prediction</i> framework.	51

4.16	Throughput of WWW flows for mix of 50% of VoIP flows and 50% of WWW ones with <i>Delay-based Prediction</i> framework.	51
4.17	Throughput of WWW flows for mix of 75% of VoIP flows and 25% of WWW ones with <i>Delay-based Prediction</i> framework.	52
4.18	Comparison of WWW flows' Throughput for mix of 25% of VoIP flows and 75% of WWW ones between different frameworks.	52
4.19	Comparison of WWW flows' Throughput for mix of 50% of VoIP flows and 50% of WWW ones between different frameworks.	53
4.20	Comparison of WWW flows' Throughput for mix of 75% of VoIP flows and 25% of WWW ones between different frameworks.	53
4.21	Time variation of FER_{VoIP}^{filt} and D_{VoIP}^{filt} for mix 50% of VoIP and 50% of WWW flows.	54
4.22	Time variation of FER_{VoIP}^{filt} and D_{VoIP}^{filt} for mix 50% of VoIP and 50% of WWW flows at load of 1.125 Users/s.	55
4.23	Time variation of FER_{VoIP}^{filt} and D_{VoIP}^{filt} for mix 50% of VoIP and 50% of WWW flows at load of 1.25 Users/s.	55
4.24	Comparison of CDF of Mean delay with mix 25% of VoIP and 75% of WWW flows for three different frameworks: <i>Non adaptive CC</i> , <i>Adaptive CC</i> and <i>Delay-based Prediction</i>	56
4.25	Comparison of CDF of Mean delay with mix 50% of VoIP and 50% of WWW flows for three different frameworks: <i>Non adaptive CC</i> , <i>Adaptive CC</i> and <i>Delay-based Prediction</i>	57
4.26	Comparison of CDF of Mean delay with mix 75% of VoIP and 25% of WWW flows for three different frameworks: <i>Non adaptive CC</i> , <i>Adaptive CC</i> and <i>Delay-based Prediction</i>	57
4.27	CDF of FER for mix 25% of VoIP and 75% of WWW flows.	58
4.28	CDF of FER for mix 50% of VoIP and 50% of WWW flows.	59
4.29	CDF of FER for mix 75% of VoIP and 25% of WWW flows.	59
4.30	Satisfaction for mix 25% of VoIP and 75% of WWW flows.	60
4.31	Satisfaction for mix 50% of VoIP and 50% of WWW flows.	61
4.32	Satisfaction for mix 75% of VoIP and 25% of WWW flows.	61
4.33	Joint Capacity with the three different frameworks.	62
4.34	Time variation of α , β and FER_{VoIP}^{filt} at the load of 0.875 Users/s.	63
4.35	Time variation of α , β and FER_{VoIP}^{filt} at the load of 1 Users/s.	63
4.36	Time variation of α , β and FER_{VoIP}^{filt} at the load of 1.125 Users/s.	64
4.37	Time variation of α , β and FER_{VoIP}^{filt} at the load of 1.25 Users/s.	64
4.38	Satisfaction for mix 50% of VoIP and 50% of WWW flows.	65
A.1	Two-State Voice Traffic Model.	67

List of Tables

4.1	Main parameters of the simulation tool.	40
A.1	Parameters of the Two-State Voice Traffic Model.	68
A.2	WWW Session Model 2	68

Chapter

1

Introduction

1.1 Context of the Problem

In the world of telecommunications, people today are more connected and more mobile than ever. We have more devices and more ways to stay in touch with one another. The Internet and wireline worlds are experiencing a rapid convergence of Internet Protocol (IP) video, audio, and data into completely new applications. Users want that same on-demand access and Internet, multimedia experience, and content anywhere from any device [1].

A quality of service framework is a fundamental component of a Fourth Generation (4G) broadband wireless network for satisfactory service delivery of evolving Internet applications to end users, and managing the network resources. Today's popular mobile Internet applications, such as voice, gaming, streaming, and social networking services, have diverse traffic characteristics and, consequently, different Quality of Service (QoS) requirements. A rather flexible QoS framework is highly desirable to be future-proof to deliver the incumbent as well as emerging mobile Internet applications [2]. One strategy to guarantee the desired QoS for specific services is the utilization of prioritization among services in some functionalities such as scheduling [3].

However, there are some situations in which scheduling alone cannot guarantee QoS for any kind of services, e.g., when the system is overloaded or in congestion conditions. Congestion situations (overload and/or outage) can be caused by different factors such as random behavior of external interference, different mobility profiles and geographical location of mobile terminals. These factors can cause variations in cell load and in the QoS experienced by the users. In these situations, the prioritization should be applied in a broader sense by means of Congestion Control (CC) algorithms.

In [4] a QoS-driven adaptive CC framework is presented. That framework joins the functionalities of scheduling, Admission Control (AC) and Load Control (LC), and it has as a main goal to guarantee the QoS of a certain kind of flows in the High-Speed Downlink Packet Access (HSDPA) system in multiservice scenarios.

In this work we propose a generalization of the CC framework presented in [4] for networks employing Orthogonal Frequency Division Multiple Access (OFDMA) in the downlink. This generalization is motivated by the adoption of OFDMA by 4G systems, such as Worldwide Interoperability for Microwave Access (WiMAX) and Long Term Evolution (LTE)-Advanced systems. As we will show later in this document, although the performance of the Voice over IP (VoIP) service can be improved with the adoption of *Adaptive CC*, the VoIP Frame Erasure Rate (FER) can be better controlled if we could predict its build-up. One of the factors that

can increase the FER of VoIP flows is the packet discarding. In general, packets are discarded when they have excessive delays.

1.2 Motivation and Objectives

In this work we propose two main contributions:

- ▶ Generalization of the CC framework presented in [4] for networks employing OFDMA in the downlink. The generalized framework is called hereafter *Adaptive CC*. This generalization is motivated by the adoption of OFDMA by 4G systems, such as WiMAX and LTE-Advanced systems. The objective here is to benefit from frequency and multiuser diversities keeping QoS guaranteed for all services.
- ▶ A new feature based on delay to be added to the generalized framework to predict an overload situation. In order to prevent high FER peaks, the generalized framework with the feature of overload prediction based on delay is called hereafter *Delay-based Prediction*.

1.3 Scientific Production and Contributions

Troughout the Master's course we have contributed with the following publications. A list with two conference papers follow below:

- ▶ E. O. Lucena, F. R. M. Lima, W. C. Freitas Jr. and F. R. P. Cavalcanti, "Overload Prediction Based on Delay in Wireless OFDMA Systems", Submitted to the IEEE Global Communications Conference (IEEE Globecom 2010), Miami, Florida, USA, December 2010.
- ▶ E. O. Lucena, F. R. M. Lima, W. C. Freitas Jr. and F. R. P. Cavalcanti, "Congestion Control Framework for Real-Time Services in OFDMA-based Systems", 27th Brazilian Symposium on Communications (SBrT'09), Blumenau, Santa Catarina, Brazil, September-October 2009.

This master thesis has been conceived in the context of UFC.22 research project, that is a cooperation between GTEL and Ericsson Research. Thus, three technical reports have been produced during the period of the master's course and one is in the process of writing. The list follows below:

- ▶ E. O. Lucena, F. R. M. Lima, W. C. Freitas Jr. and F. R. P. Cavalcanti, "Congestion Control Studies in OFDMA-based Systems", Final Technical Report (FR) of UFC.22 Project, To be delivered in July-August 2010.
- ▶ E. O. Lucena, F. R. M. Lima, W. C. Freitas Jr. and F. R. P. Cavalcanti, "Congestion Control Studies in OFDMA-based Systems", E. O. Lucena, F. R. M. Lima, Third Intermediate Technical Report (TR03) of UFC.22 Project, February 2010.
- ▶ E. O. Lucena, F. R. M. Lima, W. C. Freitas Jr. and F. R. P. Cavalcanti, "Congestion Control Studies in OFDMA-based Systems", Second Intermediate Technical Report (TR02) of UFC.22 Project, August 2009.
- ▶ E. O. Lucena, A. P. Silva, W. C. Freitas Jr. and F. R. P. Cavalcanti, "Congestion Control Studies over Dynamic OFDMA Simulator", First Intermediate Technical Report (TR01) of UFC.22 Project, February 2009.

1.4 Organisation of this Master Thesis

The remainder of this document is organized as follows:

Chapter 2 – In this chapter we provide a background to the good understanding of this work.

We start with a brief history of the evolution of mobile generations. After that, we show some concepts to talk about Radio Resource Management (RRM). The benefits provided by the adoption of OFDMA as well the most relevant works related to this Master's Thesis are also discussed in this chapter.

Chapter 3 – In this chapter the main contributions of this master's thesis are concentrated.

Besides the *Adaptive CC* framework we show the overload prediction based on delay feature.

Chapter 4 – In this chapter we present the performance evaluation and its results in a case

study with a mix of services: VoIP as the Real Time (RT) and World Wide Web (WWW) as the Non-Real Time (NRT).

Chapter 5 – Finally the main conclusions and perspectives are summarized in this chapter.

Chapter 2

Fundamentals: Wireless Systems, Services and RRM Strategies

2.1 Brief History of Evolution of Mobile Generations

In 1947 AT&T [5] introduced the concept of cellular telephony (without reusing radio frequencies yet), which became fundamental to all subsequent mobile-communication systems. In spite of the limitations of the service, there were systems deployed in many countries during the 1950s and 1960s, but the users counted only in thousands at the most. In 1969 the Bell System made commercial cellular radio operational by employing frequency reuse for the first time, but the big uptake of subscribers and usage came when mobile communication became an international concern and the industry was invited into the development process [6]. The first international mobile communication system was the analog Nordic Mobile Telephony (NMT) system which was introduced in the Nordic countries in 1981, at the same time as analog Advanced Mobile Phone Service (AMPS) was introduced in North America. They had in common that their equipments were still bulky and voice quality often inconsistent, with 'cross-talk' between users being a common problem.

With an international system such as NMT came the concept of 'roaming', giving a service also for users traveling outside the area of their 'home' operator. This also gave a larger market for the mobile phones, attracting more companies into the mobile communications business.

With the advent of digital communication during the 1980s, the opportunity to develop a Second Generation (2G) of mobile-communication standards and systems, based on digital technology, surfaced. With digital technology came an opportunity to increase the capacity of the systems, to give a more consistent Quality of Service (QoS), and to develop much more attractive truly mobile devices.

In Europe, the telecommunication administrations in European Conference of Postal and Telecommunications Administrations (CEPT)¹ initiated the Global System for Mobile Communications (GSM) project to develop a pan-European mobile-telephony system. The GSM activities were in 1989 continued within the newly formed European Telecommunication Standards Institute (ETSI). After evaluations of Time Division Multiple Access (TDMA), Code-Division Multiple Access (CDMA), and Frequency Division Multiple Access (FDMA)-based proposals in the mid 1980s, the final GSM standard was built on TDMA and FDMA.

All these standards were 'narrowband' in the sense that they targeted

¹The CEPT consists of the telecom administrations from 48 countries.

'low-bandwidth' services such as voice. With the 2G digital mobile communications came also the opportunity to provide data services over the mobile-communication networks. The primary data services introduced in 2G were text messaging as Short Message Service (SMS) and circuit-switched data services enabling e-mail and other data applications. The main 2G mobile telephony standards are GSM, CDMA and TDMA [7]. The peak data rates in GSM standard were initially 9.6 kbps. Higher data rates were introduced later in evolved 2G systems by assigning multiple time slots to a user and by adaptive modulation and coding schemes.

Packet data over cellular systems became a reality during the second half of the 1990s, with General Packet Radio Services (GPRS) introduced by GSM and to other cellular technologies such as the Japanese Personal Digital Cellular (PDC) standard. These technologies are often referred to as 2.5G.

With the advent of Third Generation (3G) and the higher-bandwidth radio interface of Universal Terrestrial Radio Access (UTRA) came possibilities for a range of new services that were only hinted at with 2G and 2.5G. The 3G radio access development is today handled in 3rd. Generation Partnership Project (3GPP). However, the initial steps for 3G were taken in the early 1990s, long time before 3GPP was formed.

The 3GPP is the standards-developing body that specifies the 3G UTRA and GSM systems. The 3GPP documents are divided into releases, where each release has a set of features added compared to the previous ones.

The outcome of the ETSI process in early 1998 was the selection of Wideband CDMA (WCDMA) as the technology for Universal Mobile Telecommunications System (UMTS) in the paired spectrum (Frequency-Division Duplexing (FDD)) and Time Division (TD)-CDMA for the unpaired spectrum (Time-Division Duplexing (TDD)).

The first major addition of radio access features to WCDMA is Release 5 with High-Speed Downlink Packet Access (HSDPA) and Release 6 with Enhanced Uplink (EU). These two are together referred to as High Speed Packet Access (HSPA). With HSPA, UTRA goes beyond the definition of 3G mobile system and also encompasses broadband mobile data.

The mobile networks are continuously evolving in order to support more users, to achieve higher data rates, and to provide new (multimedia) services.

This constant evolution makes possible a scenario where mobile networks are able to compete with fixed (wired) networks for the broadband market. With the inclusion of an Evolved Universal Terrestrial Radio Access Network (UTRAN) (Long Term Evolution (LTE)) and the related System Architecture Evolution (SAE) in Release 8, further steps are taken in terms of broadband capabilities. Within 3GPP, LTE-advanced is seen as the next major step in the evolution of LTE, which is very similar to HSPA being the first major step in the evolution of the WCDMA radio access. It is generally anticipated that LTE-Advanced will coincide with LTE Release 10 with the intermediate Release 9 mainly implying minor updates to the current LTE specifications. With the initiation of the LTE-Advanced ramping up, this smooth transition to Fourth Generation (4G) radio access is now ongoing. 3GPP continues to study further advancements for the Evolved UTRA networks. See [8] for further details about the evolution of mobile generations and also [9] for further details about LTE-Advanced.

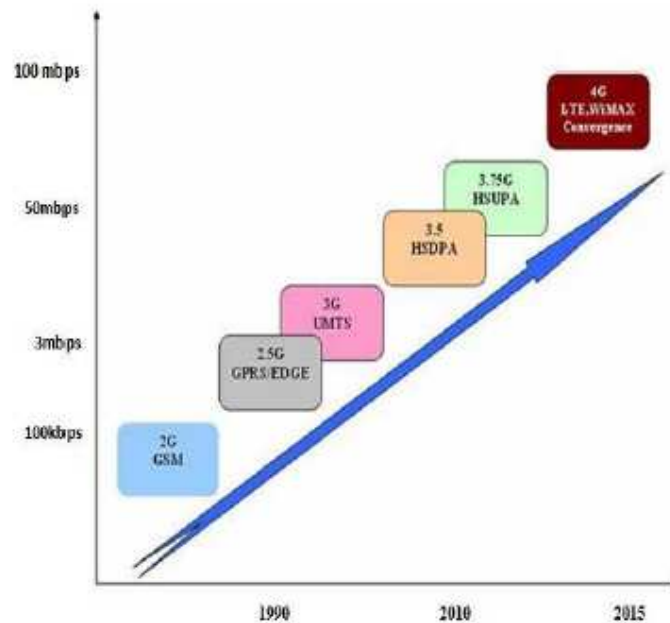


Figure 2.1: Evolution of 4G Technology.

2.2 Basics of Radio Resource Management

RT and NRT Services

Services and applications can be classified in terms of several aspects such as directionality (unidirectional or bi-directional), symmetry of the communications (symmetric or asymmetric), interactivity, number of parties and delivery requirements [10]. Specifically, when delivery requirements are concerned, the services are categorized as Non-Real Time (NRT) or Real Time (RT). RT services require a short time response between the communicating parts and, in general, impose strict requirements regarding packet delay and jitter. Voice over IP (VoIP) and online games are examples of services of this class. On the other hand, NRT services do not have tight requirements concerning packet delay although high packet delays are unacceptable. The major constraint of NRT services is the information integrity, i.e., information loss is not tolerable.

With the deployment of 3G mobile communication systems in process, the interest of many research bodies shifts towards future systems beyond 3G. Depending on the time such new systems are planned to be introduced and on the characteristic of improving or replacing existing systems they are called Beyond 3G (B3G) [11]. There is no formal definition for what 4G is; however, there are certain objectives that are projected for 4G. These objectives include: that 4G will be a fully IP-based integrated system. 4G will be capable of providing between 100 Mbit/s and 1 Gbit/s speeds both indoors and outdoors, with premium quality and high security [12]. The term 4G is used broadly to include several types of broadband wireless

access communication systems, not only cellular telephone systems. While neither standards bodies nor carriers have concretely defined or agreed upon what exactly 4G will be, fourth generation networks are likely to use a combination of WiMAX and Wi-Fi technologies [13].

With 4G a range of new services will be available. If a user want to be able to access the network from lots of different platforms: cell phones, laptops, PDAs he is free to do so in 4G which delivers connectivity intelligent and flexible enough to support some RT services such as streaming video, VoIP telephony, still or moving images, e-mail and other classified as NRT such as Web browsing, e-commerce, and location-based services through a wide variety of devices. That means Freedom for consumers [11].

Quality of service (QoS) is another feature that can be adapted for different users depending on their specific application, such as voice, streaming video, or internet access for example [13].

Radio resource management

A Radio Resource Unit (RRU) is defined here by the set of basic physical transmission parameters necessary to support a signal waveform transporting end user information corresponding to a reference service. For example, in FDMA, an RRU is equivalent to a certain bandwidth within a given carrier frequency. In TDMA, an RRU is equivalent to a pair consisting of a carrier frequency and a time slot. In CDMA, an RRU is defined by a carrier frequency, a code sequence and a power level. It is worth noting that, in a multiservice scenario, each service may require different amounts of RRUs due to the maximum bit data rates achievable with adaptive modulation coding schemes. Services with higher bit rates will, consequently, require more RRUs.

Radio Resource and QoS management functionalities are very important in the framework of 3G and beyond 3G systems because the system relies on them to guarantee a certain target QoS, maintain the planned coverage and offer a high capacity, objectives which tend to be contradictory (e.g. capacity may be increased at the expense of a coverage reduction; capacity may be increased at the expense of a QoS reduction, etc.). Radio Resource Management (RRM) functions can be implemented in many different algorithms, and this impacts on the overall system efficiency and on the operator infrastructure cost. Additionally, RRM strategies are not subject to standardisation, so there can be a differentiation issue among manufacturers and operators.

Since the different RRM functions will track different radio interface elements and effects, RRM functions can be classified according to the time scales they use to be activated and executed. Since short or long term time scales variations are relative concepts, the approach preferred here is to associate typical time scale activation periods with the different RRM functions. Time scales between consecutive activations of the algorithm are the time between when an action is carried out by a specific RRM algorithm and the next time that the same algorithm needs to operate. We present following some RRM functions starting with the less dynamic ones:

Admission control

Admission control decides the admission or rejection of requests for set-up and reconfiguration of radio bearers. The request should be admitted provided that the QoS requirements can be met and that the QoS requirements of the already accepted connections are not affected by the new request acceptance.

Since the maximum cell capacity is intrinsically connected to the amount of interference or, equivalently, the cell load level, the use of admission control algorithms is based on

measurements and/or estimates of the current network load situation as well as on the estimation of the load increase that the acceptance of the request would cause.

It is worth noting that admission control decisions are taken at the specific moment a new request is performed, so that the decision may be based on the radio network situation at that time as well as on the recent past history. Nevertheless, the admission decision can in no way anticipate exactly the future network load, so that additional radio resource management functions are necessary to cope with the dynamic network evolution and to keep the QoS requirements under control.

The randomness associated with a cellular radio environment (e.g. propagation, mobility, traffic, etc.) allows admission control to play the role of thick tuning in the management of the radio resources. If decisions are too soft, and too many users are being accepted, an overload situation may follow and further RRM mechanisms will need to be activated. If decisions are too strict, and too few users are being accepted, the operator will be losing revenue and a tuning of the admission control algorithm will be necessary.

In contrast to 2G, where the network is accessed mostly by real time voice users with equal quality requirements, in 3G WCDMA multimedia services (e.g. video-telephony, streaming video, web browsing, etc.) with diverse QoS requirements (e.g. business segment, consumer segment, etc.) are expected. Therefore, admission control algorithms must take into consideration that the amount of radio resources needed for each connection request will vary. Similarly, the QoS requirements in terms of RT or NRT transmission should also be considered in an efficient admission control algorithm. Clearly, admission conditions for NRT traffic can be more relaxed on the assumption that the additional RRM mechanisms complementing admission control will be able to limit non real time transmissions when the air interface load is excessive.

Congestion control

Congestion control, also denoted as load control, faces situations in which the QoS guarantees are at risk due to the evolution of system dynamics (mobility aspects, increase in interference, traffic variability, etc.). For example, if several users in a cell suddenly move far from the Enhanced Node B (eNB), there may not be enough power to satisfy all the links' qualities simultaneously and some actions are required to cope with this situation. Note that, although a strict admission control could be carried out, as long as the radio network behaviour has strong random components, there is always some probability that these overload situations occur and, consequently, congestion control mechanisms must be included in the set of radio resource management techniques.

Congestion situations in the radio interface are caused by excessive interference. Thus, congestion control algorithms need continuously to monitor the network status in order to correct overload situations when they are present. The monitoring will be based on network measurements, such as downlink transmitted power, uplink cell load factor, etc., which need to be suitably averaged to avoid both false congestion detections (i.e. triggering congestion resolution mechanisms when the air interface is not really overloaded) and congestion non-detection (i.e. do not trigger congestion resolution mechanisms when the air interface is really overloaded). Additionally, the congestion control algorithm needs to exhibit a fast reactivity under overload conditions in order to prevent degradation of the quality of the connections.

Similarly to the admission control algorithm, congestion control is closely related to other RRM functions. In particular, congestion resolution actions will be supported, for example,

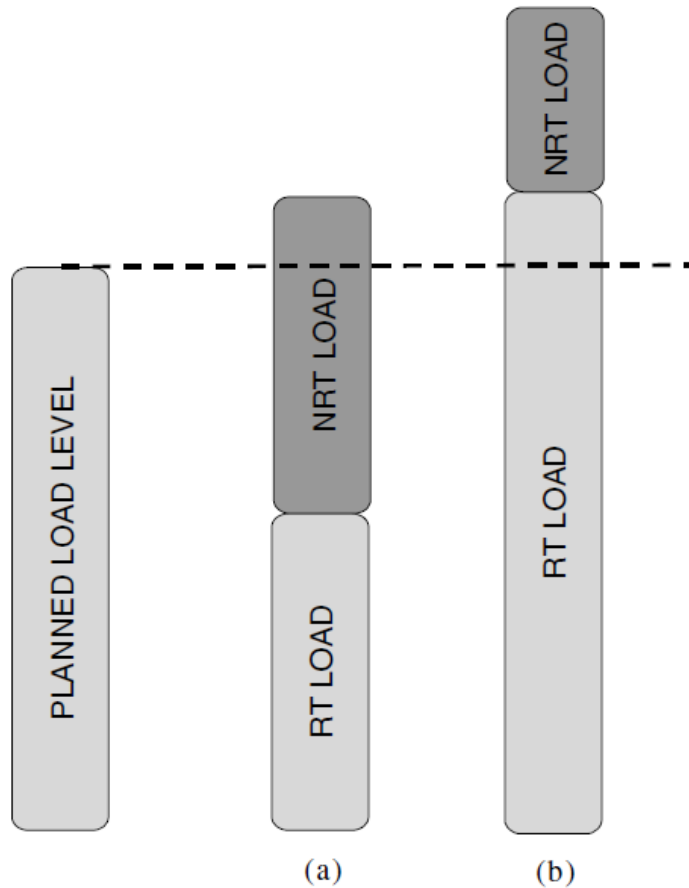


Figure 2.2: Congestion control in the presence of different service types.

by admission control (e.g. by refusing new connections while the network is congested) or handover (e.g. by transferring connections from a congested cell to neighbouring cells). More precisely, the actions to be carried out may depend on the situation, the origin of the congestion and the services mix present at the congestion time. As reflected in Figure 2.2, a maximum load level is planned in the network. In general, the current load will be the result of both RT plus NRT services transmissions. If the RT traffic load is below the planned value, it is possible to solve the congestion situation by acting on NRT users, as shown by Figure 2.2(a). Congestion may be alleviated by reducing the transmission rates of a number of NRT users. An extreme case would be the inhibiting of all NRT transmissions in the cell. If this is not enough, the help of neighbouring cells could be enlisted, by reducing NRT intercell interference contribution. If this is still not enough, as shown in Figure 2.2b, it would be necessary to reduce the RT load. In the case of conversational services, this should be accomplished by dropping some calls.

Packet scheduling

The packet scheduling algorithm is devoted to deciding the suitable radio transmission parameters for each connection in a reduced time scale and in a very dynamic way. It operates on a frame by frame (or TTI) basis to take advantage of the short term variations in the interference level. Taking into account its operation in short periods of time this strategy is referred to as short term RRM strategy. Its operation is illustrated in Figure 2.3. Specifically, Figure 2.3(a) reflects a situation in which the current load, including both RT and NRT users, is below the planned load level, thus a certain spare capacity exists in the cell. Therefore, the

purpose of short term RRM strategies will be to bring the cell to the situation reflected in Figure 2.3(b), in which the spare capacity has been filled with a certain amount of NRT load. This can be achieved by allowing the transmission of other users not included in Figure 2.3(a) and/or the increase in the bit rate of other users already included in Figure 2.3(a).

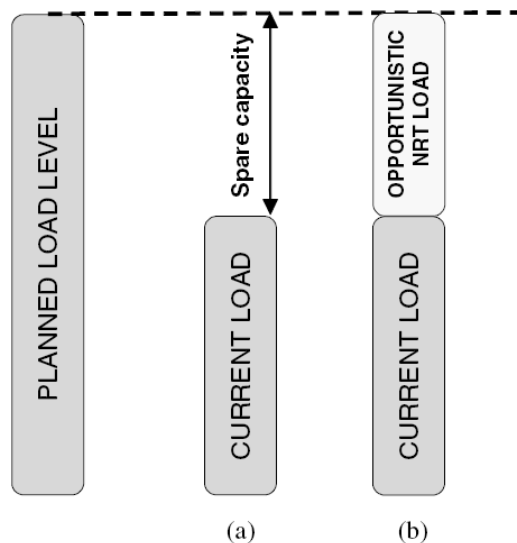


Figure 2.3: Operation of short term RRM strategies with NRT load.

The packet scheduling strategy may follow a time-scheduling approach (i.e. multiplex a low number of users simultaneously with relatively high bit rates), a code-scheduling approach (i.e. multiplex a high number of users simultaneously with relatively low bit rates) and combinations of both. Furthermore, prioritisation mechanisms can be considered in the scheduling algorithm. Priorities can be established, for example, at service class level (e.g. interactive traffic is assigned higher priority than background traffic) or at user type level (e.g. business users are assigned higher priority than consumer users).

Interactions among RRM functions

Congestion should not occur very often if the RRM parameters are correctly adjusted. Nevertheless, when congestion does occur, the reduction of the maximum bit rate is the first action that can be taken, thus modifying the transmission capabilities granted at the admission phase. Furthermore, for severe congestion situations, blocking new connection requests may help to prevent further overload in the network, so that interactions between congestion and admission control are also feasible. Similarly, the admission control algorithm may accept a high priority request relying on the fact that congestion mechanisms will be able to manage the possible resulting overload by taking actions on less prioritised users.

QoS provisioning for multimedia traffic in wireless environment in beyond 3G systems

The provision of QoS guarantees is a pressing need in wired and wireless networks as well as in distributed computing systems, particularly to support multimedia-enabled applications. Throughput, timeliness, reliability, and perceived quality are the foundations of what is known as QoS. The combination of QoS and wireless environment has been one of the hot topics in the telecommunications for a few years.

The research community is now directing its interest toward unified ways of looking at system design, optimization, and QoS issues to satisfy the requirements of next-generation

mobile and wireless Internet Protocol (IP) networks. The implementation of all IP mobile networks implies that IP QoS architectures and mechanisms need to be developed, because the existing best-effort based mechanisms are unable to cope with the application requirements. To provide a research basis for the definition of 4G systems, much work still has to be done.

Satisfaction Metrics

The concept of user satisfaction is very important when interpreting the system performance. There are many parameters to consider such as the service type, technical parameters and even economical issues (such as the price to use the wireless service) [14–16]. However, in this study we consider only technical aspects concerning the perceived quality by the end user.

In order to assure the desired QoS of telephony (RT service), 2G and 3G networks utilize circuit-switched connections, i.e., resources are dedicated to the flows² since the session initialization. However, in order to efficiently support multiple services on the same network the operators are upgrading their core network to the All-IP concept. Among the advantages of an All-IP architecture we can mention the efficient support to mass-market usage of any IP-based service and reduced Operational Expenditure (OPEX) and Capital Expenditure (CAPEX). Despite the advantages of this architecture, the QoS guarantees of RT services remains a challenging task. With All-IP, the resources are packet-switched, i.e., the resources are dynamically allocated to the flows according to the demands.

The basic routing philosophy on the Internet is “best-effort.” This attitude serves most users acceptably but it is not adequate for the time-sensitive, continuous stream transmission required for VoIP. QoS refers to the ability of a network to provide better, more predictable service to selected network traffic over various underlying technologies, including IP-routed networks. QoS features are implemented in network routers by:

- ▶ Supporting dedicated bandwidth;
- ▶ Improving loss characteristics;
- ▶ Avoiding and managing network congestion;
- ▶ Shaping network traffic;
- ▶ Setting traffic priorities across the network.

Voice applications have different characteristics and requirements from those of traditional data applications. Because they are innately real-time, voice applications tolerate minimal delay in delivery of their packets. Additionally, they are intolerant of packet loss, out-of-order packets, and jitter [17].

RRM algorithm evaluation by means of simulations

Normally, computer simulations constitute the preferred solution for the evaluation and validation of the different RRM strategies. Although it is also possible to evaluate them using analytical approaches, this usually requires a high number of simplifying assumptions that reduce the precision of the obtained results. Consequently, the use of analytical approaches is often limited to obtaining rough performances and general trends in RRM strategies. Simulations, however, allow obtaining more precise results provided that the

²A terminal can bear multiple service flows. However, without loss of generality, we consider that a terminal corresponds to a flow.

simulation models adequately capture the real system behaviour. In the RRM algorithm simulation methodology, there exists a clear trade-off between the computational complexity of the simulations and the level of detail considered in them. Therefore, higher numbers of simulated procedures and parameters means longer simulation times but also more precise results. Consequently, a suitable simulation model will be based on extracting the relevant procedures and parameters from the real systems that may have an impact on the evaluated RRM strategies.

Furthermore, to cope with the complexity trade-off, the simulation is usually split into two different types of simulations with different time resolutions. They are denoted as link and system-level simulations. The link-level simulation is responsible for obtaining the physical layer behaviour of the channel observed by a mobile user to communicate with its corresponding base station, either in the uplink or in the downlink. To this end, link-level simulators with a time resolution below the chip time are usually devised (e.g. they typically operate with four or eight samples of the received signals per chip). On the other hand, a system-level simulator evaluates the behaviour of the RRM algorithms in multi-cell, multi-user and multi-service scenarios resulting from the previous network planning procedure. To handle these complex scenarios in moderate simulation times, a system-level simulator makes use of the off-line results obtained by means of the link-level simulator to characterise each link of each user in each cell. Therefore, the system-level simulators typically operate on a slot-by-slot or a frame-by-frame basis rather than on a chip-by-chip basis.

System-level simulation tools must be able to combine information about the network configuration (e.g. cell sites, transmitted powers, etc.) with information about the position of the mobiles and the traffic that they are likely to generate, in order to build a realistic picture of the network in terms of its coverage and the offered QoS. Users are scattered around the network based on an expected traffic distribution. In the case of static simulators, users do not move and so the tool builds a snapshot of the network for a particular user distribution. Many snapshots with different distributions of users are run in order to obtain a composite view of the network performance. However, in the case of dynamic simulations, the users move around and generate traffic, so they behave as much as possible like real users. Consequently, dynamic simulators allow better capturing of the real situations and provide more accurate results than static simulators.

Other RRM functionalities

Other RRM functions could be also applied in this solution, as code management, handover, power control as also interactions among these different RRM functions. For more details about other RRM functionalities see [18, 19].

2.3 Orthogonal Frequency Division Multiple Access (OFDMA)

OFDMA is a multiple access scheme based on Orthogonal Frequency Division Multiplexing (OFDM) [20]. OFDM is a transmission technology that has been utilized in wired and wireless communications. Asymmetric Digital Subscriber Line (ADSL) broadband access and power line communications are examples of applications of OFDM in wired systems. In wireless systems, the OFDM technology is utilized in IEEE 802.11 a/g and planned to be utilized in LTE and Mobile WiMAX.

In OFDM, the available frequency band for transmission is divided into several subcarriers that have narrower bandwidth than the channel coherence bandwidth, as in Frequency Division Multiplexing (FDM) systems. However, the subcarriers in OFDM are designed to

be orthogonal among each other, which leads to higher spectral efficiency than FDM as it is illustrated in Fig. 2.4. The narrowband subcarriers also imply simplified equalization process because of the flat fading channel experienced in each subcarrier. Besides that, as the data rate transmitted in each subcarrier is low and consequently the modulated symbols are longer than the delay spreading, OFDM is robust against Inter Symbol Interference (ISI). In order to effectively mitigate the effects of ISI, a guard interval named cyclic prefix, that consists in a copy of part of the OFDM symbol, is inserted before the OFDM symbol transmission. More details about OFDM can be found in [21].

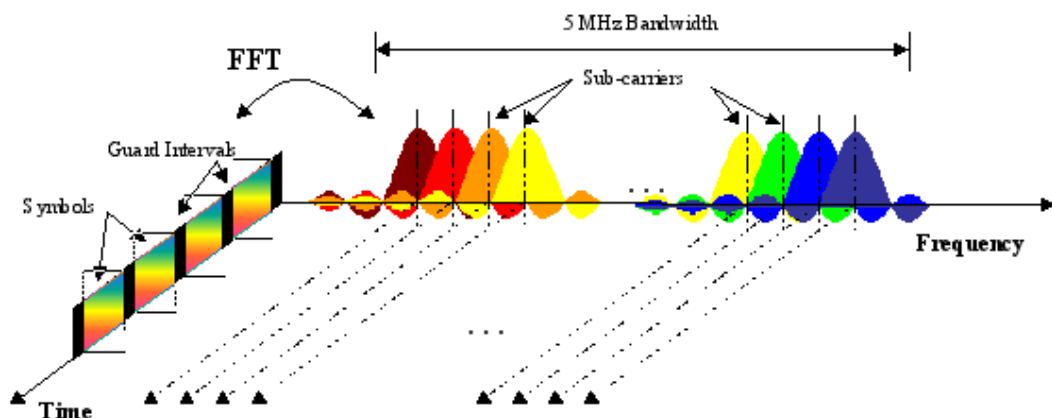


Figure 2.4: Frequency-time representation of an OFDM signal [22].

With OFDMA [23], the multiple access is achieved by the assignment of different subcarriers or block of them to individual User Equipments (UEs) at different time periods. More specifically, in OFDMA the system resources can be arranged in a time-frequency grid as shown in Fig. 2.5. In the frequency axis the granularity is defined by the subcarriers while in the time dimension it is defined by an OFDM symbol.

One of the advantages of an OFDMA-based system is the opportunity to benefit from frequency and multiuser diversities. Frequency diversity means that it is unlikely that all frequency resources in a link have the same channel quality. Multiuser diversity occurs due to the independence of UE channels caused by distinct UE positions in a cell, therefore, frequency resources in poor channel states for some UEs possibly will be in good channel conditions for other UEs. A mechanism for taking advantage of the frequency and multiuser diversities is the employment of scheduling algorithms. Scheduling algorithms are responsible for selecting which UEs will have access to the system resources and with which configuration. In this way, scheduling algorithms have a great impact on system performance.

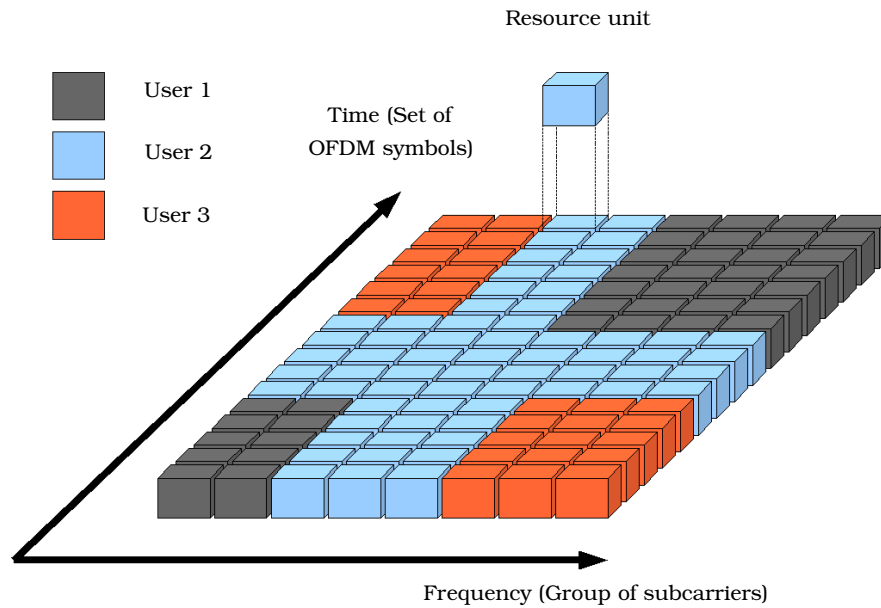


Figure 2.5: Frequency-time resource grid in OFDMA.

2.4 State-of-the-Art

The most relevant works related to this Master's thesis can be divided into three categories: those related to Admission Control (AC) algorithms, those related to scheduling, and those related to Load Control (LC) algorithms. The work [4] proposed a QoS-driven adaptive Congestion Control (CC) framework that joins the functionalities of scheduling, AC and LC. The objective of that framework is to guarantee the QoS of RT flows in the HSDPA system in multiservice scenarios.

The Session Admission Control (SAC) algorithm is an AC algorithm used in [4, 24] which is employed considering the delay as the resource to be shared among users in the system. In that work were studied two different schemes used to estimate the delay resource usage: Measured Delay Based (MDB) and Estimated Delay Based (EDB). EDB is a method to indirectly estimate the delay in a certain queue based on some other metrics. In [25] an AC scheme is proposed combining power and traffic measurements, like the number of transport blocks in the queue and the service rate of the base station. By using those metrics, an estimate of the delay can be calculated and compared to a limiting threshold in the admission process. As we have to monitor the system, a natural way for this is to perform measurements. Hence, in this work we use MDB. MDB uses an Attack-Decay (AD) filter that is a variation of the exponential filtering presented in [26].

In multi-user system sharing a time varying channel, a compromise between fairness and high system throughput is the well-known Proportional Fair (PF) [27] scheduling, which is extended to multi-carrier transmission systems by [28] and it is called Multicarrier Proportional Fair (MPF). In [3] the scheduling adopted is the Weighted Proportional Fair (WPF). The WPF algorithm works almost the same way as the classical PF scheduler. The only difference is a fixed multiplicative weight which is a QoS differentiation factor. The scheduling algorithm used in [4] is also the WPF, but it adapts the weights relative to each service to control the congestion in the RT service. Finally, the Weighted Multicarrier Proportional

Fair (WMPPF) scheduler is used in [29] as a natural generalization of WPF in an OFDMA-based system where the frequency diversity can be exploited by adding another dimension in that prioritization.

In [30] are proposed two LC algorithms: Error Feedback Based Load Control (EFLC) and Jump Based Load Control (JLC). The main objective of the LC algorithms is to make sure that the Frame Erasure Rate (FER) of the VoIP users connected to the system is around a target value. The proposal of the JLC algorithm was inspired from the algorithm presented in [31]. The priority margins in the JLC are updated according to the well-known Outer-loop Power Control (OLPC) jump algorithm proposed in [32]. The proposal of the EFLC algorithm was based on the work developed in [33]. The main differences between the JLC and EFLC algorithms are the step size for the adaptation of the α and β parameters, and the way the algorithms decide whether the VoIP QoS requirement was met or not. The EFLC uses a dynamic step size whose calculation is directly derived from the VoIP QoS metric - FER - that is being targeted. Thus, it is expected that EFLC is able to perform a more fine-tuned control of the VoIP FER towards the desired value.

Based on those reference works, this Master's thesis aims at providing an adaptive CC for networks employing OFDMA in the downlink with mixed traffic. This work generalizes the framework proposed in [4] to OFDMA-based system where the frequency and multiuser diversities can be exploited. Besides this, in this work a new feature based on delay is added to the generalized framework to predict an overload situation. In the next chapter, the generalized adaptive CC framework as well as the overload prediction feature will be described in details.

Chapter 3

Adaptive Congestion Control Framework

Adaptive Congestion Control (CC) comprises in a coordinated manner the operation of three algorithms as follows:

- ▶ **Admission Control:** responsible for granting or denying the access of a new flow to the system. The Session Admission Control (SAC) algorithm is presented in more details in the section 3.1.
- ▶ **Scheduling:** responsible for defining which flows will be scheduled, determining their required data rate at the current Transmission Time Interval (TTI) and which resources will be assigned to the selected flows. See section 3.2 for more details of the Weighted Multicarrier Proportional Fair (WMPF) scheduler.
- ▶ **Load Control:** responsible for the adaptation and filtering of the priority margins α and β and so to establish different priorities between the services. For more details about their relation in both Admission Control (AC) and scheduling algorithms see section 3.3.

For more details about the original proposal of the *Adaptive CC* without its generalization for networks employing Orthogonal Frequency Division Multiple Access (OFDMA) in the downlink see [4, 34].

3.1 Admission control

AC algorithms are responsible for granting or denying the access of a new flow to the system. The criterion used for decision can be the availability of physical resources or the service quality of the connected flows. AC algorithms are important tools to control the congestion in a system. By rejecting new flows, the Quality of Service (QoS) of the already connected flows can be maintained. In this work, a SAC [24] scheme is employed to guarantee the quality of a single priority service in a mix with other services. The service with high priority is the Real Time (RT) one. The SAC scheme is presented in Figure 3.1.

The SAC algorithm considers delay as the resource to be shared among flows in the system. In order to do that, the packet delays of the RT traffic are regularly measured and filtered. There are two admission thresholds depending on the service type: D_{RT}^{th} for the RT service and D_{Other}^{th} for other low-priority services. Therefore, when a new RT flow tries to access the system, the SAC algorithm will check if the filtered delay of the RT service is greater or lower

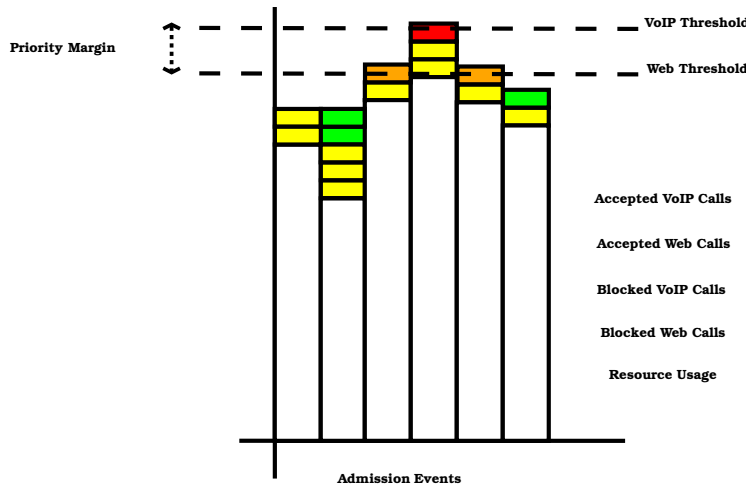


Figure 3.1: Session admission control scheme applied.

than the RT admission threshold. If greater, the new flow is rejected; if lower the flow is admitted. The procedure is the same if the new flow is of an Non-Real Time (NRT) service.

According to the admission thresholds, a service can be more prioritized than the others. Although in [24] these admission thresholds are fixed, the main idea in the *Adaptive CC* framework is to adapt them according to the congestion status of an RT service.

In order to prioritize a service, we should introduce a new variable: the SAC priority margin, α . We define this priority margin (in decibel) as

$$\alpha[k] = 10 \log_{10} \left(\frac{D_{Other}^{th}[k]}{D_{RT}^{th}} \right), \quad (3.1)$$

where $\alpha[k]$ is the SAC priority margin at TTI k .

As we can see, this variable defines a ratio between the admission threshold of two services and it can define a difference in their priorities once D_{RT}^{th} can assume a fixed value and α can assume different values at different TTIs. The Load Control (LC) algorithm is responsible for the adaptation of α and so to establish different priorities between the services as it will be discussed in section 3.3.

3.1.1 Measured Delay Based (MDB)

It is the scheme used to estimate the delay resource usage studied in this work. A natural way to monitor a specific system behavior is to perform measurements. Therefore, the delay must be measured in the sense that only one metric should be able to represent reliably the system performance, concerning delay experienced by all packets from all users in the cell. The measurement reports from the Enhanced Node B (eNB) are already standardized and must use an exponential filtering approach based on Autoregressive Moving Average (ARMA) filters for measurements from the physical layer [35]. The problem with this approach is that the filtering process tracks the average and the average delay for VoIP traffic does not increase significantly with increasing load, while the user satisfaction is highly affected by the load. In [26], a variation of the exponential filtering is considered to be implemented to filter the received reports. It is called Attack-Decay (AD) filter. The update of the filtered measurement is done in accordance with the following expressions:

$$D_n = M_n - F_{n-1}, \quad (3.2)$$

$$\text{if } D_n > 0 \text{ then } F_n = F_{n-1} + D_n \cdot f_{up}, \quad (3.3)$$

$$\text{else if } D_n < 0 \text{ then } F_n = F_{n-1} + D_n \cdot f_{down}, \quad (3.4)$$

where F_{n-1} is the old filtered measurement result, F_n is the updated filtered measurement and M_n is the new sample that is feeding the filter. This filter has the advantage of being capable to track any percentile just by setting the f_{up} and f_{down} , attack and decay factors, respectively.

That is the motivation of using the AD filter to evaluate the delay measurements of Voice over IP (VoIP) packets. Each transmitted or discarded packet provides a delay sample that feeds the AD filter. The updated filtered value in each cell is reported to the Radio Network Controller (RNC) every report interval, in order to be used for admission purposes.

3.2 Scheduling

In multi-user system sharing a time varying channel, a compromise between fairness and high system throughput is the well-known Proportional Fair (PF) [27] scheduling, which is extended to multi-carrier transmission systems by [28] and it is called Multicarrier Proportional Fair (MPF). In [3] the scheduling adopted is the Weighted Proportional Fair (WPF). The WPF algorithm works almost the same way as the classical PF scheduler. The only difference is a fixed multiplicative weight which is a QoS differentiation factor.

The scheduling algorithm used in [4] is also the WPF. With WPF, the (single) scheduled flow is the one with highest priority. The priority of flow j at TTI k is given by

$$p_j[k] = w_j[k] \cdot \left(\frac{r_j[k]}{t_j[k]} \right), \quad (3.5)$$

where $w_j[k]$ represents a service-dependent weight, $r_j[k]$ and $t_j[k]$ are the supported and filtered data rates of flow j at TTI k (according to the channel state), respectively, and that provides a history of the allocated data rates in the past. If the flow is from an RT session, $w_j[k]$ is set to W_{RT} . On the other hand, if the flow is from another service the priority is equal to W_{Other} . Therefore, by setting different values to these weights (W_{RT} and W_{Other}) some prioritization among the services can be accomplished. As in the SAC scheme, these weights are fixed in [3]. The *Adaptive CC* framework adapts these weights to control the congestion in the RT service.

3.2.1 Weighted Multicarrier Proportional Fair (WMPF)

In an OFDMA-based system, the frequency diversity can be exploited by adding another dimension in that prioritization. Therefore, we have adopted the WMPF scheduler [29] that is a natural generalization of WPF. The prioritization in WMPF is given by

$$p_{j,n}[k] = w_j[k] \cdot \left(\frac{r_{j,n}[k]}{t_j[k]} \right), \quad (3.6)$$

where $p_{j,n}[k]$ is the priority of flow j in subcarrier n at TTI k and $r_{j,n}[k]$ is the supported data rate of flow j in subcarrier n at TTI k (according to the channel state of subcarrier n). The filtered data rate $t_j[k]$ is given by

$$t_j[k] = \mu \cdot \varphi_j[k] + (1 - \mu) \cdot t_j[k - 1], \quad (3.7)$$

where μ is an exponential filter constant, $\varphi_j[k]$ is the allocated rate of flow j at TTI k if j is scheduled. If j is not scheduled at TTI k $\varphi_j[k]$ assumes a value 0 in this TTI. If μ assumes a high value, so the allocated rate φ calculated at TTI k becomes more important than the throughput past values. The filtered rate is a way of giving resources for those who has not transmitted a lot before. Hence, it would be more interesting a lower value to μ .

From the priorities $p_{j,n}[k]$ we can build the priority matrix \mathbf{P} . The flow selection consists in assigning the pair flow-subcarrier corresponding to the largest entry in the priority matrix \mathbf{P} . In this way, multiple flows can be scheduled simultaneously with potentially higher data rates. The pseudo-code of WMPF is presented in Algorithm 3.1.

Algorithm 3.1 WMPF algorithm

```

1:  $\mathcal{A}$  Set of assigned flow-subcarrier pairs
2:  $\mathcal{N} = \{1, \dots, N\}$  Set of available subcarriers
3:  $\mathcal{J} = \{1, \dots, J\}$  Set of active flows
4: while ( $\mathcal{N}$  is not empty) do
5:    $(j^*, n^*) \leftarrow \arg \max_{j \in \mathcal{J}, n \in \mathcal{N}} \{\mathbf{P}\}$ 
6:    $\mathcal{N} \leftarrow \mathcal{N} - \{n^*\}$ 
7:    $\mathcal{A} \leftarrow \mathcal{A} \cup \{j^*, n^*\}$ 
8: end while

```

After defining the priority $p_{j,n}[k]$, we should introduce a new variable: the WMPF priority margin, $\beta[k]$. We define this priority margin (in decibel) as

$$\beta[k] = 10 \log_{10} \left(\frac{W_{Other}[k]}{W_{RT}} \right), \quad (3.8)$$

where W_{RT} and W_{Other} represent a service-dependent weight of the RT flow and other service, respectively and $\beta[k]$ is the value of β at TTI k .

As we can see, this variable defines a ratio between the weight of two services and it can define a difference in their priorities once W_{RT} can assume a fixed value and β can assume different values in different TTIs. The LC algorithm is responsible for the adaptation of β and so to establish different priorities between the services as it will be discussed in section 3.3.

3.3 Load control

Considering that D_{RT}^{th} and W_{RT} are fixed reference values, the dynamic adaptation of the priority margins α and β can control the prioritization of the RT service over the other services. The main idea of the LC algorithm is to adapt these priority margins according to the QoS of the ongoing sessions of the high priority service. If the QoS of the sessions of the RT service is not being fulfilled, the LC algorithm decreases the priority margins. As a consequence, the sessions of this RT service will be scheduled more often and the system will decrease the number of admitted sessions of other services in order to protect the ongoing sessions of the RT service.

3.3.1 Error Feedback Based Load Control (EFLC)

The EFLC algorithm was proposed by [30] and was based on the work developed in [33]. Since all downlink RT traffic will be scheduled in the same place, it can calculate the a RT service measure averaged over all the connected RT users. The adaptation of the priority margin at each TTI k is calculated by means of a RT service metric. Without loss of generality, we consider the Frame Erasure Rate (FER) as the main performance metric of RT services. However, another metric could be used instead. The adaptation of the priority margin $\alpha[k]$ can

be given as follows:

$$\alpha[k] = \min \{ \max \{ \alpha_{\min}, \alpha[k-1] - \sigma_{\alpha} \cdot e[k] \}, \alpha_{\max} \}, \quad (3.9)$$

where $e[k]$ is given by

$$e[k] = FER_{RT}^{filt}[k] - FER_{RT}^{target}. \quad (3.10)$$

The FER considers a ratio of number of lost frames (or packets) and the total number of generated packets. FER_{RT}^{filt} is the filtered FER in the last control interval and FER_{RT}^{target} is the target FER to experience a good QoS. The filtered FER $FER_{RT}^{filt}[k]$ is obtained by applying a Simple Exponential Smoothing (SES) filter to the time series comprised by the average FER in each TTI [36]. It is important to observe that $\beta[k]$ is adapted in the same way as $\alpha[k]$ where β_{\min} , β_{\max} and σ_{β} are replaced by α_{\min} , α_{\max} and σ_{α} , respectively. α_{\min} , α_{\max} , β_{\min} and β_{\max} are the minimum and maximum values in dB of the $\alpha[k]$ and $\beta[k]$ parameters, respectively. The fixed parameters σ_{α} and σ_{β} control the adaptation speed of the priority margins $\alpha[k]$ and $\beta[k]$, respectively.

In order to see the time variation of FER_{VoIP}^{filt} and of the priority margins α , β , we present the Figure 3.2. We can see that when the value of FER_{VoIP}^{filt} is higher than VoIP FER threshold (represented in this figure as 0.01), we can notice the adaptation of α and β .

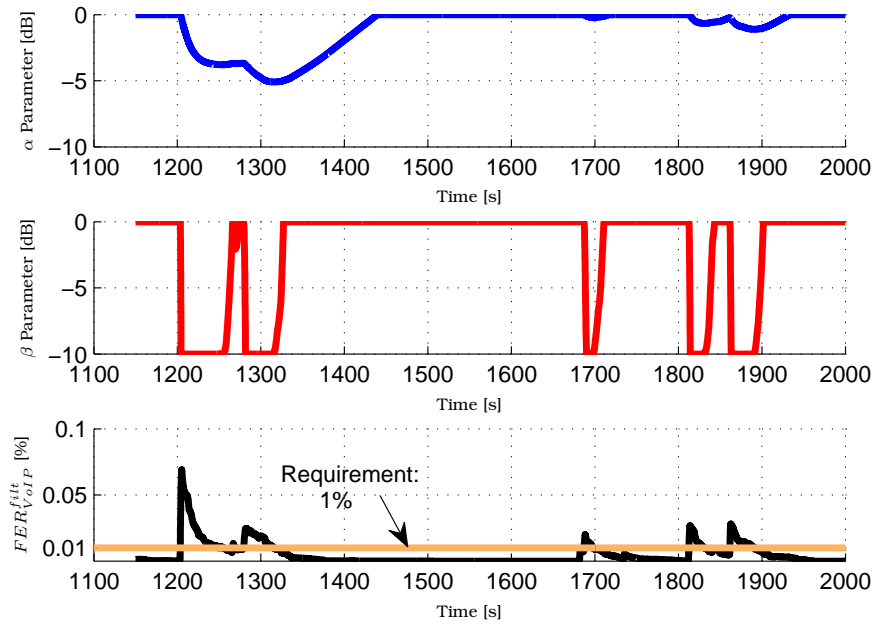


Figure 3.2: The time variation of FER_{VoIP}^{filt} and Adaptation of SAC and WMPF priorities margins.

From Equations (3.9) and (3.10), we can see that when the QoS of VoIP sessions in the system is worse than expected ($FER_{VoIP}^{filt}[k] > FER_{VoIP}^{target}$), the LC algorithm will decrease the priority margins. By decreasing the priority margins the VoIP sessions are prioritized in both AC and scheduling algorithms.

3.4 Overload Prediction Based on Delay

In this section we present the *Delay-based Prediction* framework, which predicts an overload situation based on delay measurements. This section is organized as follows:

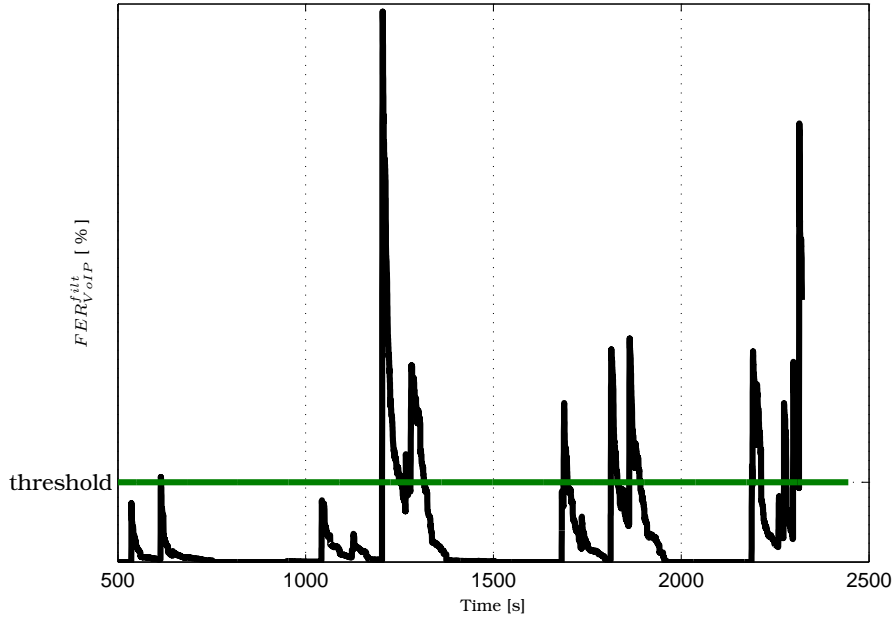


Figure 3.3: Time variation of FER.

- **Motivation:** In 3.4.1 we show the reason of the overload prediction based on delay study.
- **Formulation:** In section 3.4.2 we find more details about the formulation of the delay prediction.

3.4.1 Motivation

The *Adaptive CC* presented in the previous section is capable of protecting the QoS of the RT services by monitoring a service specific metric (in our case, the FER). The prioritization among services in scheduling, AC and LC are changed in order to *react* when an overload situation is detected. However, as the reaction of the *Adaptive CC* takes place when the overload already exist, the system will get back to normal load conditions only after a certain period. Within this period, RT sessions can be compromised due to poor QoS experience. Therefore, we believe that the QoS of RT sessions could be even protected if a *predictive* capacity would be added to the *Adaptive CC*.

The packet delay plays an important role in the perceived QoS of RT services. Usually, the packets of RT services have stringent delay requirements. In case the packet delay deadlines are violated, these packets are usually discarded by upper protocol layers and the overall QoS experienced by the end user is degraded. Therefore, if in average the packet delays of the active RT flows are increasing, this is a strong indication that packet discard will happen. With this in mind, our second contribution in this work is the addition of a new feature to the *adaptive CC* that is the capability of predicting and avoiding overload situations by using measurements of the packet delays of RT flows.

Another interesting subject is the early warning of increasing delay in VoIP packets on the system for the *Adaptive CC* before FER really builds up. We can see in Figure 3.3 the time variation of FER. An important insight is the role of the delay in the behavior of the FER. When the delay of all VoIP packets in the system are getting close a threshold, it is possible to predict that a packet can be lost before it happens and so the system can react before FER

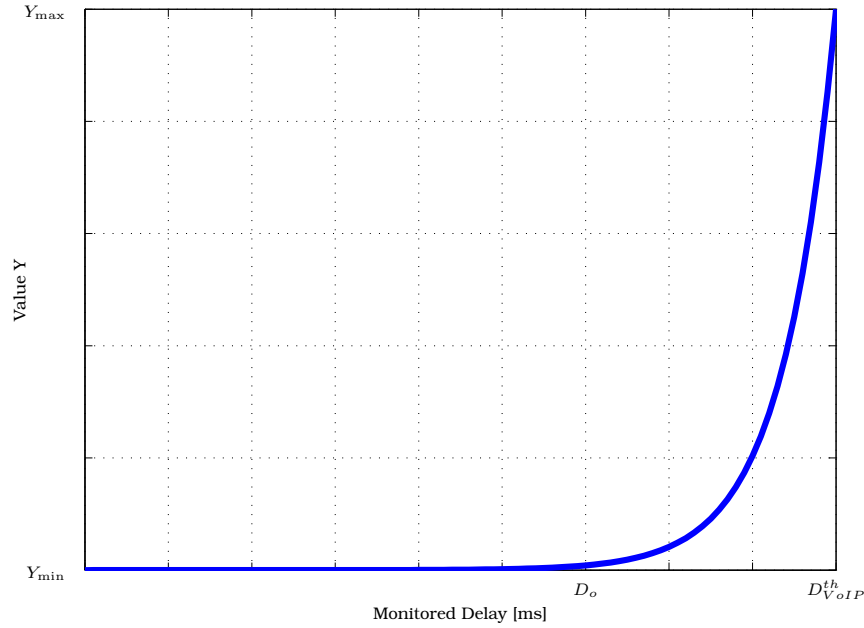


Figure 3.4: Behavior of the Delay Prediction Variable Y .

really builds up.

3.4.2 Formulation

The prediction can be represented by a variable that could be added to the priority margins α and/or β so that scheduling weights and/or admission thresholds can start to be changed as the delay increases and gets close a threshold before the loss packet occurs and the FER really increases.

Figure 3.4 illustrates the behavior of the delay prediction variable Y that is calculated as:

$$Y = Y_{\min} - M + M \cdot \exp\left(\frac{d}{D_{VoIP}^{th}} \cdot \ln\left(\frac{Y_{\max} - Y_{\min} + M}{M}\right)\right) \quad (3.11)$$

where M is a constant responsible for the slope of the exponential curve, d is the delay measured, D_{VoIP}^{th} is the VoIP SAC delay threshold, Y is an adjustment factor to the filtered FER, Y_{\min} and Y_{\max} are fixed parameters that indicate the minimum and the maximum values of Y , respectively.

The variable Y works adding a value to the FER_{VoIP}^{filt} with the behavior presented in Figure 3.4. Hence, the Equation (3.10) can be modified using the Equation (3.11) and yield

$$e[k] = \left(FER_{VoIP}^{filt}[k] + Y\right) - FER_{VoIP}^{target}. \quad (3.12)$$

It is worth noticing that Y increases only when the monitored delay is higher than a threshold D_0 , and so $e[k]$ also increases, simulating a higher FER. Thus, we force a reaction of the system before the increase of the FER and the larger $e[k]$ is, the faster the *Delay-based Prediction* framework works.

Chapter 4

Simulation Results

In this section, we present a performance evaluation of the *Delay-based Prediction* framework. One of the objectives in this study is to evaluate how the Frame Erasure Rate (FER) could be affected with the insertion of the delay prediction variable Y of Equation (3.11) into Equation (3.10) as shown in Equation (3.12). The performance is compared with a framework without overload prediction called *Adaptive Congestion Control (CC)* and to a reference framework called *Non Adaptive CC*, where no CC solution is applied, i.e., in which the Weighted Multicarrier Proportional Fair (WMPF) and Session Admission Control (SAC) priority margins are not adapted, but are fixed at 0dB. Details about the simulation tool are presented in section 4.1 while the main parameters used to obtain the results are presented in section 4.2. In section 4.3 we define the performance metrics used in this study. Finally, in section 4.4 we show and analyze the simulation results.

4.1 Simulation tool

A dynamic simulation tool was conceived in order to evaluate and validate the generalization of the CC framework as well as the new feature based on delay. This simulation tool was developed in Matlab® [37] and incorporates several characteristics such as detailed models for traffic generation of Real Time (RT) and Non-Real Time (NRT) services, higher layer protocols, radio propagation characteristics and mobility profiles. It is important to emphasize that some radio parameters (e.g. Orthogonal Frequency Division Multiple Access (OFDMA) as the downlink scheme, the channel bandwidth, multiple services and the available modulations) are chosen in order to simulate Beyond 3G (B3G) scenarios such as 3GPP Long Term Evolution (LTE). More informations about LTE and LTE-Advanced can be found in [38] and [39].

4.2 Simulation parameters

The simulation tool models the main aspects of a single-cell OFDMA-based system. The simulation modelling includes aspects such as propagation phenomena (e.g., path loss, shadowing and fast fading), service applications and link adaptation. The channel is free of errors, i.e., a packet can be lost only when its delay is higher than a discard timer. This may happen due to queuing time related to the scheduling policy or due to network congestion. For the sake of simplicity there is no retransmission modelled in this work. If retransmissions were considered, this would imply in a greater number of packets to be transmitted and it would increase the network congestion and also need more complex scheduling algorithms.

The results can be obtained with different congestion levels. The congestion levels are

related to the interval time between the arrival of different flows in the system. Concerning traffic modelling we consider a mixed traffic scenario with Voice over IP (VoIP) as the RT service and World Wide Web (WWW) as the NRT service. The flows arrive in the system according to a Poisson distribution. It is always considered in this work that when a RT or NRT flow arrives in the system, it has at least a packet ready to transmit. Following the new flow arrival some propagation parameters are calculated based on its position inside the cell such as its path loss, its shadowing and its path gain. The propagation conditions changes once the position of the flow also changes. We assume that the data symbols are independently modulated and transmitted over a high number of closely spaced orthogonal subcarriers. The modulation schemes Quadri-Phase Shift Keying (QPSK), 16 Quadrature Amplitude Modulation (QAM), and 64 QAM are available and are chosen depending on the flow Signal-to-Noise Ratio (SNR) that varies with the propagation conditions. A session is finished when a flow has no more packets to transmit. The simulation ends when a certain number of sessions is achieved.

The main parameters used for these simulations are presented in Table 4.1.

4.2.1 Discussion

In this section is presented a discussion about some important parameters in this work.

4.2.1.1 Flows' arrival Rate

In this work, dynamic simulations are performed to obtain the results. Each simulation has a load which informs the flows' arrival rate, in other words, the flows birth's rate. For all simulations performed in this work 0.125 Users/s is the lowest flows' arrival rate simulated (It means that every 8 s arrives a new flow to the system) and 1.25 Users/s is the highest flows' arrival rate simulated in this work (It means that every 0.8 s arrives a new flow to the system).

4.2.1.2 Stop Condition

The criterion chosen to stop a simulation in this work is setting a maximum number of sessions for each service in a multiple service simulation. Clearly, in the case of single service simulations only the studied service is considered. For all simulations performed in this work 1000 sessions of both services should finish. If the maximum number of sessions set is greater, the time simulation will be also greater.

4.2.2 Saving Results

Warm-up period is a start-up period of a simulation run where an initially empty system is processing entites at a rate that is different from the one observed when the system has reached a steady state. Data collected during the warm-up period is discarded if steady state performance measures are desired. Data start to be saved after a warm-up period. The warm-up period was defined after some tests, after finding the flows convergence in time. Figure 4.1 represents the time variation of the number of flows during a complete simulation. It is possible to observe that the number of VoIP flows is not so different of WWW ones as expected for mix 50% of VoIP and 50% of WWW flows.

It is important to mention that the number of flows varies depending on the load, the service mixes and all the parameters related to the load control framework that can change among different simulations.

Table 4.1: Main parameters of the simulation tool.

Parameter	Value	Unit
General parameters		
Transmission Time Interval (TTI)	1	ms
Warm-up period	500	s
Period of saved results updating	100	s
Flows' arrival Rate	[0.125 0.25 0.375 0.5 0.625 0.75 0.875 1 1.125 1.25]	Users/s
Stop Condition	1000 complete sessions of each service	-
RT service	VoIP	-
NRT service	WWW	-
Mix of services	0 % VoIP 100 % WWW 25 % VoIP 75 % WWW 50 % VoIP 50 % WWW 75 % VoIP 25 % WWW 100 % VoIP 0 % WWW	- - -
System parameters		
Carrier frequency	2	GHz
Number of Subcarriers	200	-
Subcarrier spacing	15	kHz
Cell radius	500	m
Minimum distance from Base Station (BS)	100	m
Maximum BS power	5	W
White noise power density	-174	dBm/Hz
Propagation parameters		
Fast Fading speed	3	km/h
Path loss [40]	$128 + 37.6 \cdot \log_{10}(d)$	dB
Standard deviation of lognormal shadow fading	8	dB
Small-scale fading	Multiple path Rayleigh	-
Average tap powers	$10 \cdot \exp([-5.7 - 7.6 - 10.1 - 10.2 - 10.2 \dots$... -11.5 - 13.4 - 16.3 - 16.9 - 17.1 - 17.4... ...-19.0 - 19.0 - 19.8 - 21.5 - 21.6 - 22.1 ...-22.6 - 23.5 - 24.3]/10)	W ...
Average tap delays	[0 0.217 0.512 0.514 0.517 0.674 0.882 1.230 1.287 1.311 1.349 1.533 1.535 1.622 1.818 1.836 1.884 1.943 2.048 2.140] $\cdot 10^{-6}$	s
Link-to-System parameters		
Possible Rates	[1 2 3 4 5 6]	kbps
Reference Values for mapping the SNR	[0 185.5 619 1801 5526 22860]	dB
Traffic parameters		
VoIP traffic model	according to [40]	-
Average VoIP session	60	s
Average VoIP packet call time	3	s
VoIP packet generation time	20	ms
VoIP packet size	256	bits
VoIP FER threshold	1	%
VoIP satisfaction requirement	95	%
WWW traffic model	according to [41]	-
WWW packet call mean size	4100	bytes
WWW packet call standard deviation	30000	bytes
WWW packet call maximum size	100000	bytes
WWW maximum reading time	120	s
WWW throughput threshold	128	kbps
WWW satisfaction requirement	90	%
Congestion Control parameters		
Attack factor f_{up}	0.5	-
Decay factor f_{down}	0.05	-
Exponential filter constant μ	0.2	-
VoIP SAC delay threshold (D_{VoIP}^{th})	100	ms
VoIP WMPF priority weight (W_{VoIP}^{prio})	1	-
Time basis for adaptation of α	100	ms
Time basis for adaptation of β	1	ms
Maximum value of α and β ($\alpha_{max}, \beta_{max}$)	0	dB
Minimum value of α and β ($\alpha_{min}, \beta_{min}$)	-10	dB
SAC step size (σ_{α})	0.5	dB
WMPF step size (σ_{β})	0.5	dB
Overload Prediction Based on Delay parameters		
Maximum value of adjustment factor to FER_{VoIP}^{filt} (Y_{max})	0.05	-
Minimum value of adjustment factor to FER_{VoIP}^{filt} (Y_{min})	0	-
Constant responsible for the slope of the exponential curve M	$3 \cdot 10^{-8}$	-
Time basis for adaptation of Y	100	ms

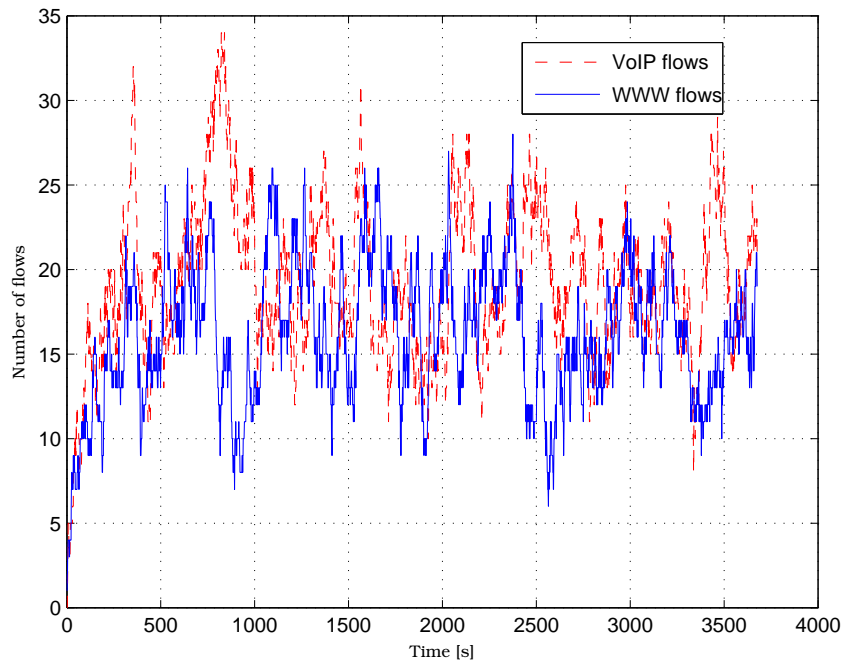


Figure 4.1: Flows Convergence for mix 50% of VoIP and 50% of WWW flows.

4.3 Performance metrics and simulation scenarios

The following metrics are used in the performance evaluation for all present results:

VoIP delay

The delay of the VoIP service considered in this work. The delay of each VoIP packet is logged for further results evaluation.

Frame Erasure Rate

The FER metric is concerned with the VoIP service only. VoIP packet loss can occur when it arrives at the scheduler too late, i.e., its delay is higher than a discard timer. It is relevant to say that the FER metric calculation takes into account the ratio of the number of lost packets (voice frames) and the total number of generated packets (voice frames).

Blocking

The blocking rate is the ratio between the number of calls, whose access to the system was denied by the AC policy and the total number of call requests. It is possible to discriminate blocking metrics for each service class, in case there is any priority handling in the AC algorithm.

WWW throughput

During an interactive service class (WWW) session, a user may download several WWW pages (packet calls). The size in bits of all the WWW pages downloaded during the lifetime of the user are summed and divided by the activity time of the session. This is the WWW session throughput which is used as a performance metric.

User satisfaction

The satisfaction is reached if the service provided fulfills the service requirements. The definition of satisfaction depends on the service class the user belongs to. A VoIP user is assumed as satisfied if it is not blocked by the Admission Control (AC) functionality and has a FER lower or equal to 1% [42], reflecting a good perceived speech quality provided by the Adaptive Multirate (AMR) codec with 2% FER (1% guaranteed for each link direction). A WWW data user is regarded as satisfied when it is not blocked and its session throughput is higher than or equal to 128 kbps [24]. .

Offered load

We define the offered load as the mean flow arrival rate (in number of flows per second) in the system. This is an input parameter to the Poisson processes used to model flow arrivals. The system capacity (offered load) will be represented by the estimated total number of users of all service classes (VoIP and WWW) per minute in each cell (sector). This estimative considers the Poisson arrival rate and the mean session duration (holding time) of each service class. The mean session duration of each service can be derived from the traffic models.

System capacity regions

The system capacity regions are defined as the set of expected number of users (offered load) for which acceptable system-level quality is sustained for all service classes [43]. The capacity region is constructed varying the traffic mix among the considered service classes, including single service evaluations. The system-level Quality of Service (QoS) limits considered in this study for the VoIP service is 95%, and for the WWW service is 90%. These values were chosen based on typical values used in system-level performance evaluations, and also based on the user satisfaction results, since the studied packet scheduling algorithms presented remarkable differences in system capacity for these values of QoS limits.

4.4 Results

4.4.1 Dynamic of Parameters in Time and their behavior

In this first part we show the time variation of α , β and FER_{VoIP}^{filt} . As explained before in Equation 3.10 when FER_{RT}^{filt} is greater than FER_{RT}^{target} then $e[k]$ assumes a value greater than 0 dB, and, consequently, $\alpha[k]$ and $\beta[k]$ assume negative values giving priority in blocking a new WWW flow that desires access the system or giving priority in the scheduling to VoIP flows, respectively. The following results presented in Figures 4.2, 4.3, 4.4 and 4.5 show α and β assuming negative values and coming back to values next to 0 dB, e.g., giving less priority to the VoIP flows.

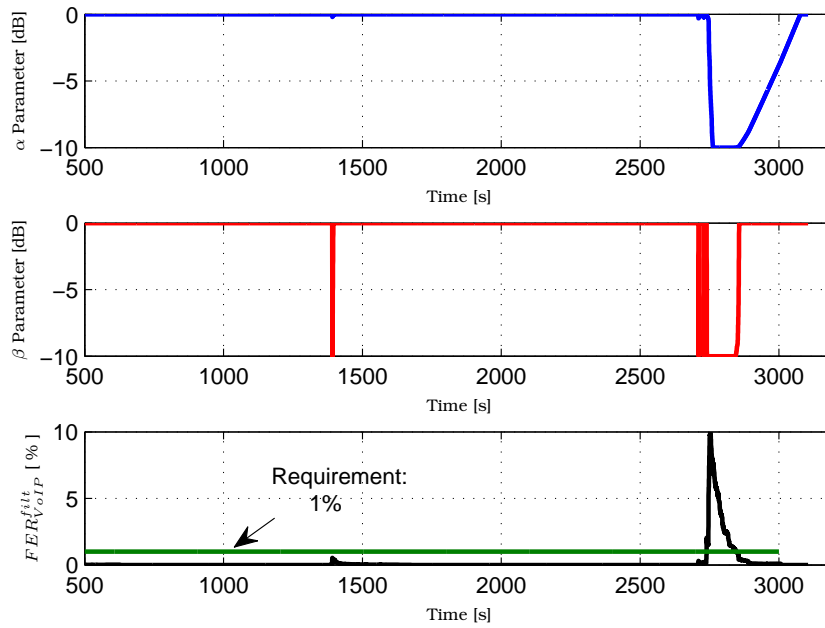


Figure 4.2: Time variation of α , β and FER_{VoIP}^{filt} at the load of 0.875 Users/s.

In Figures 4.2 and 4.3 we can observe that the FER_{VoIP}^{filt} is well controlled, and it keeps below the FER_{RT}^{target} (here represented by the green line with value of 1%). For this reason, α and β are kept near to 0 dB, assuming negative values in rare moments.

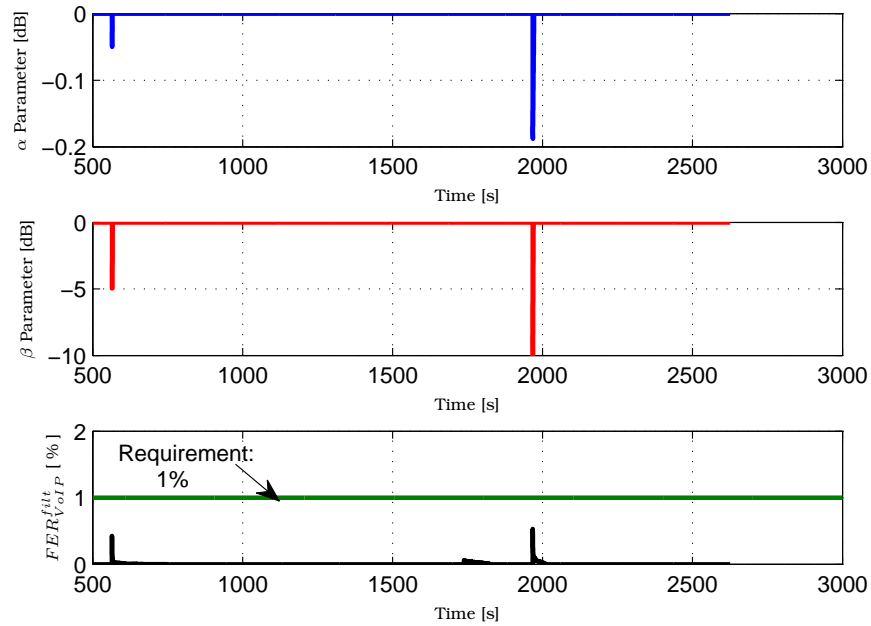


Figure 4.3: Time variation of α , β and FER_{VoIP}^{filt} at the load of 1 Users/s.

Regarding Figures 4.4 and 4.5, as they are obtained with very high loads, the FER_{VoIP}^{filt} is usually greater than FER_{VoIP}^{filt} forcing the system to react. Hence, α and β assume a lot of times negative values. It is important to note that β answers and adapts more rapidly reaching the extreme values (0 dB and -10 dB) more often than α does.

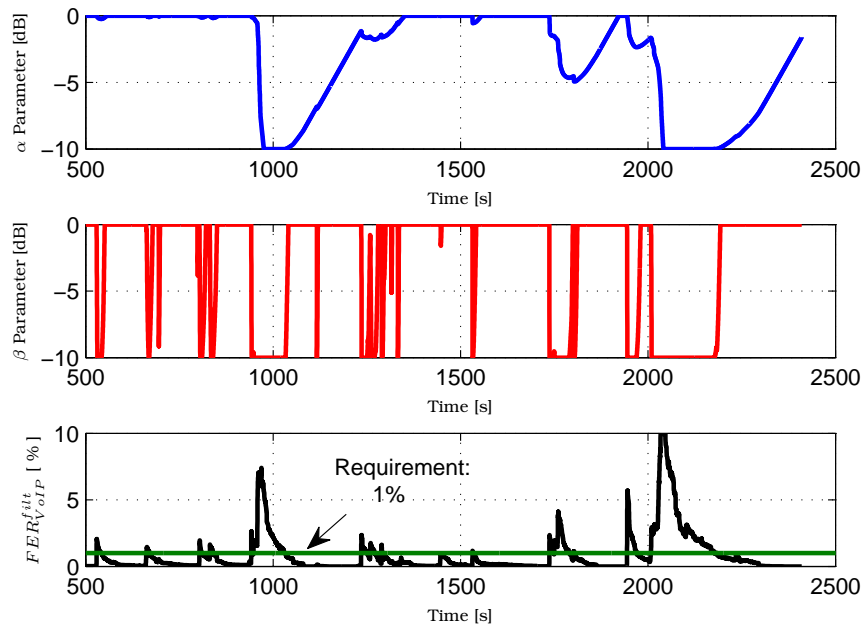


Figure 4.4: Time variation of α , β and FER_{VoIP}^{filt} at the load of 1.125 Users/s.

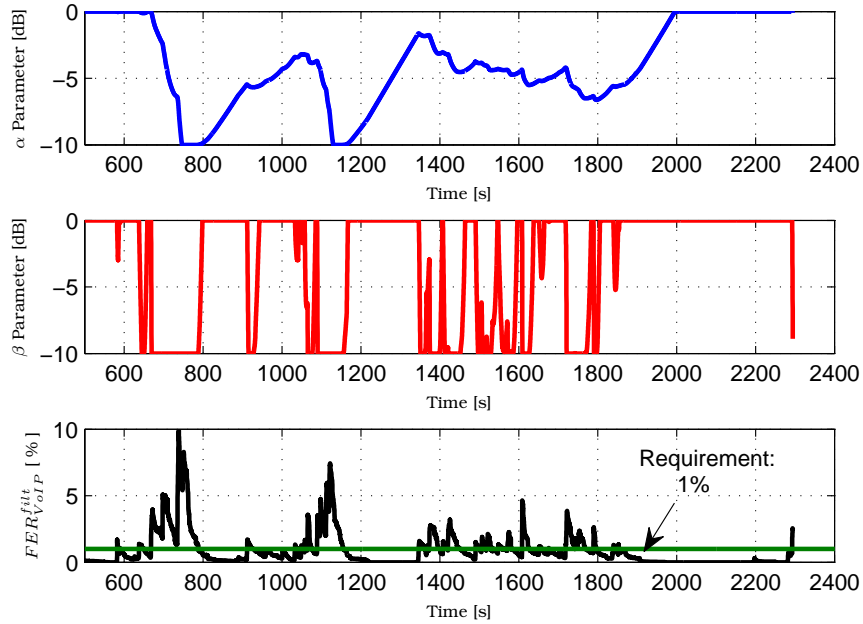


Figure 4.5: Time variation of α , β and FER_{VoIP}^{filt} at the load of 1.25 Users/s.

4.4.2 Results

In this section we present some results for different mixes and loads when the three different frameworks are performed. First of all, in Figures 4.6, 4.7 and 4.8, we can see the blocking rate and discuss how much the AC functionality by rejecting new flows, can keep the QoS of the already connected flows.

Blocking Rate

In Figure 4.6 we have more WWW flows than VoIP ones. The performance of both services in *Non adaptive CC* is similar. This happens because there is no priority service in this case. We can also observe that when the *Adaptive CC* framework is applied the blocking rate is lower than that when *Non adaptive CC* is performed. The reason for that is that as the priorities margins are updated the framework reacts protecting more the ongoing sessions of VoIP flows. As a consequence, less flows are blocked in order to reach the QoS of VoIP. In the end, when *Delay-based Prediction* is performed we find the lowest values of blocking rates between this service mix for both services. It is important to emphasize that, since in this case the percentage of VoIP flows is less than that of WWW ones, keeping the VoIP flows satisfied is an easy task, so that the blocking rate assume low values.

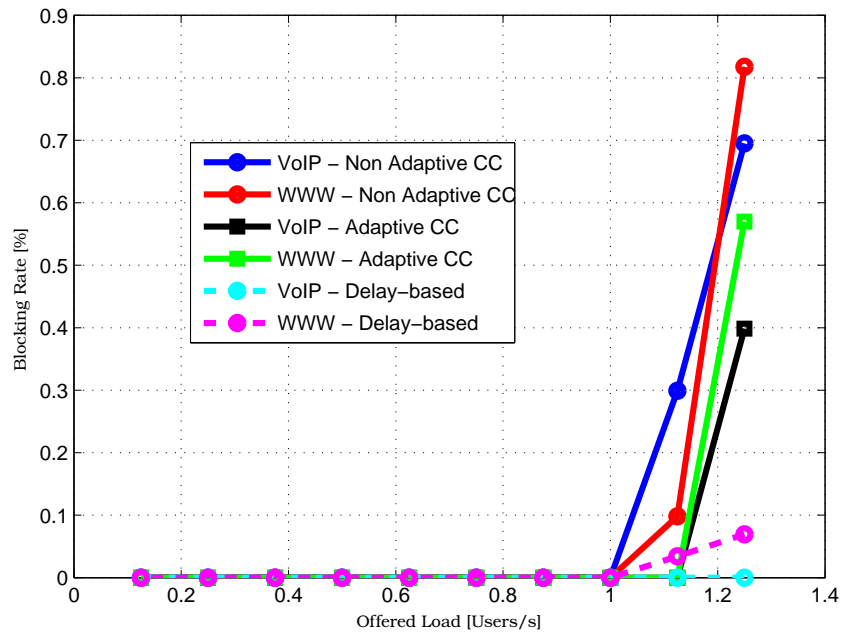


Figure 4.6: Blocking Rate of flows of both services for three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* for mix 25% of VoIP and 75% of WWW flows.

In Figures 4.7 and 4.8 the values of VoIP flows blocking rate in *Delay-based Prediction* continue to be the lowest between all the others, but those related to WWW flows when *Delay-based Prediction* is performed have the largest values. The reason for this is that we find more VoIP flows here than before, so keeping the ongoing VoIP sessions near to the target values becomes a hard job. As a consequence, the AC acts blocking more the less priority service, in this case, WWW.

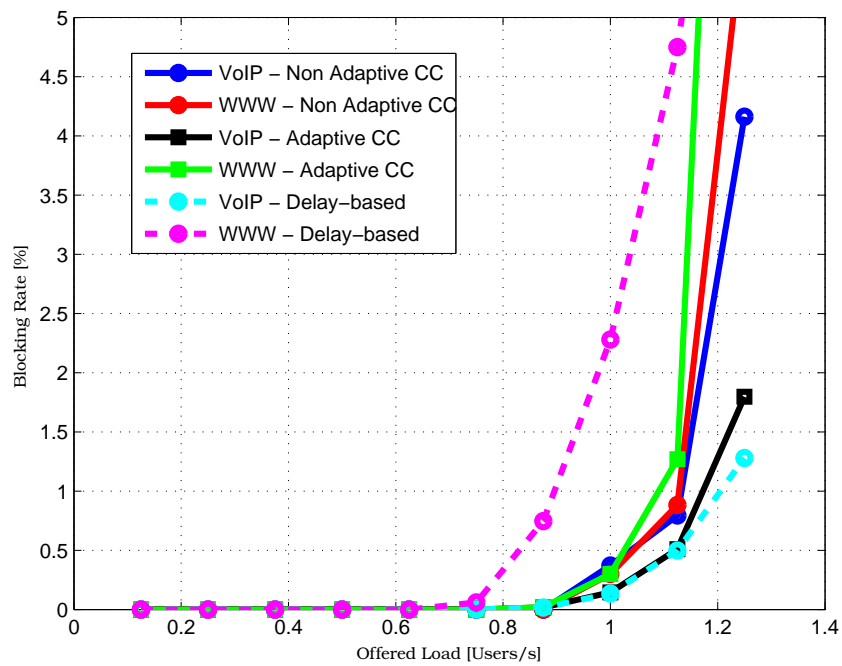


Figure 4.7: Blocking Rate of flows of both services for three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* for mix 50% of VoIP and 50% of WWW flows.

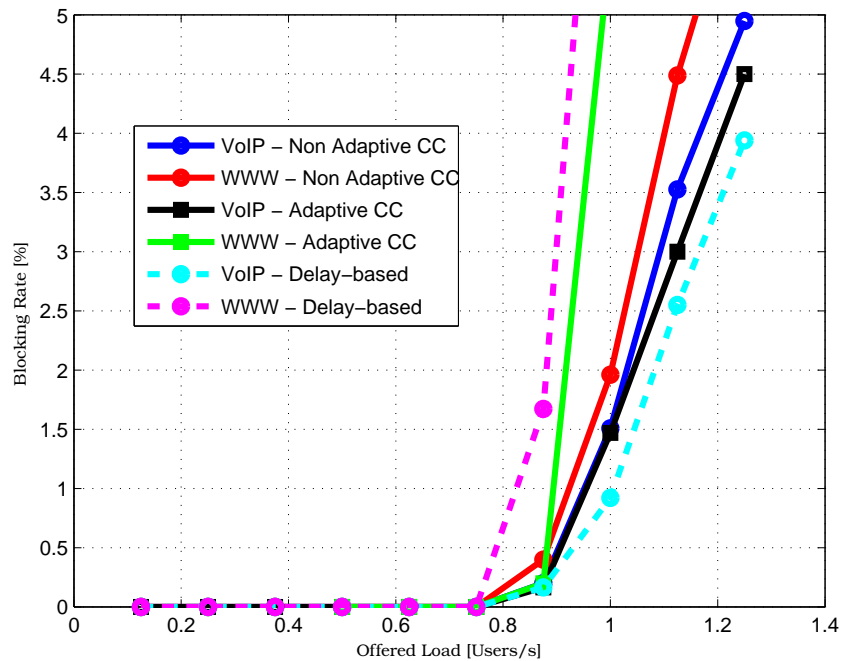


Figure 4.8: Blocking Rate of flows of both services for three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* for mix 75% of VoIP and 25% of WWW flows.

Throughput

Later we show a group of throughput figures, for each service mix and for each framework. In these figures, our objective is to show that the WWW service has similar performance in all the three frameworks. In Figures 4.9, 4.10 and 4.11, we can see the throughput for different service mixes when *Non adaptive CC* framework is adopted. As expected, the higher the offered load, the lower the average throughput is. It is also important to emphasize that for the three service mixes the target level of satisfaction to the WWW service (90%) is found around the 0.5 Users/s load.

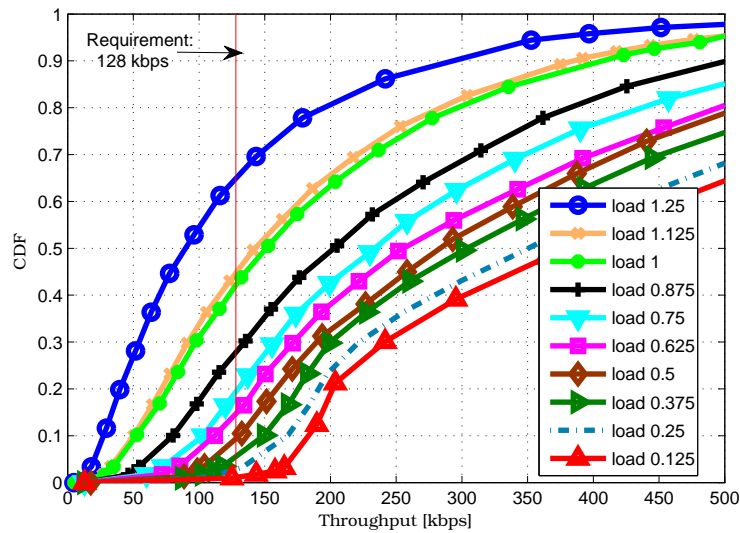


Figure 4.9: Throughput of WWWW flows for mix of 25% of VoIP flows and 75% of WWW ones with *Non adaptive CC* framework.

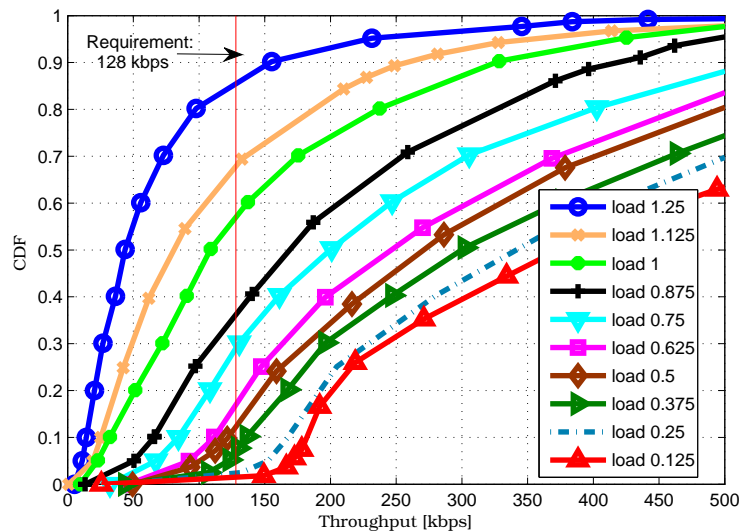


Figure 4.10: Throughput of WWWW flows for mix of 50% of VoIP flows and 50% of WWW ones with *Non adaptive CC* framework.

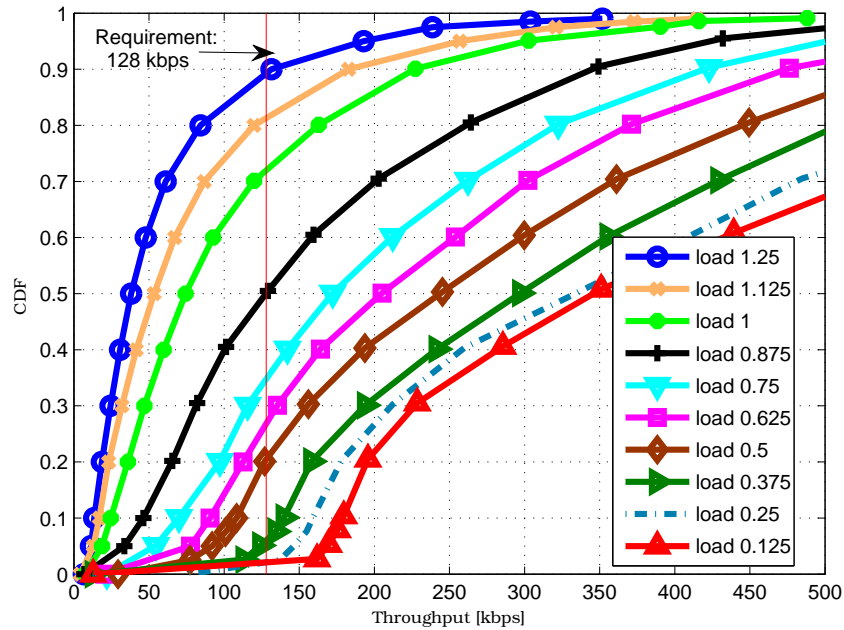


Figure 4.11: Throughput of WWWW flows for mix of 75% of VoIP flows and 25% of WWWW ones with *Non adaptive CC* framework.

In Figures 4.12, 4.13 and 4.14, we can see the throughput for different service mixes when *Adaptive CC* framework is applied. In spite of these results being concerned with another framework, the same comments discussed before can be possible also here. Also, the higher the offered load, the lower the average throughput is. The target satisfaction level of the WWWW service (90%) for the three service mixes is also around the 0.5 Users/s load.

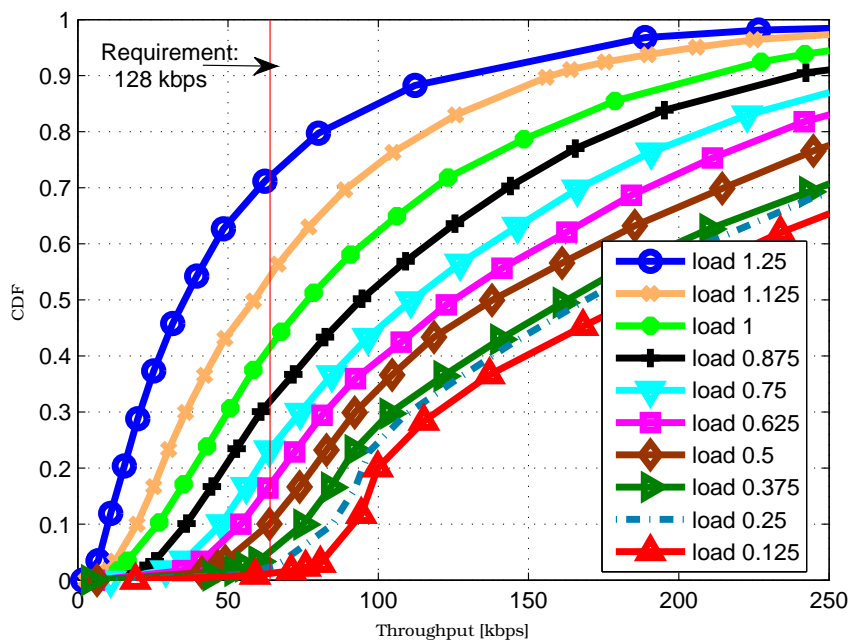


Figure 4.12: Throughput of WWWW flows for mix of 25% of VoIP flows and 75% of WWWW ones with *Adaptive CC* framework.

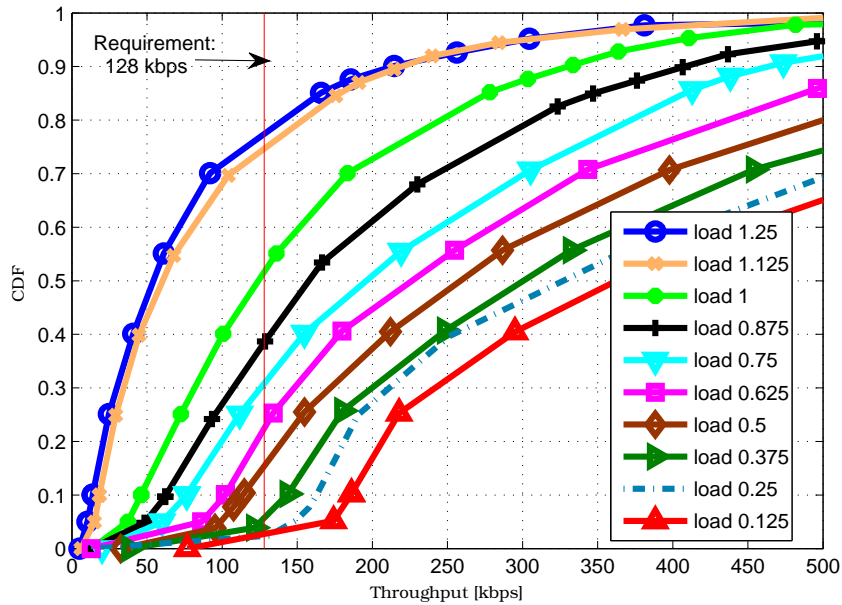


Figure 4.13: Throughput of WWWW flows for mix of 50% of VoIP flows and 50% of WWWW ones with *Adaptive CC* framework.

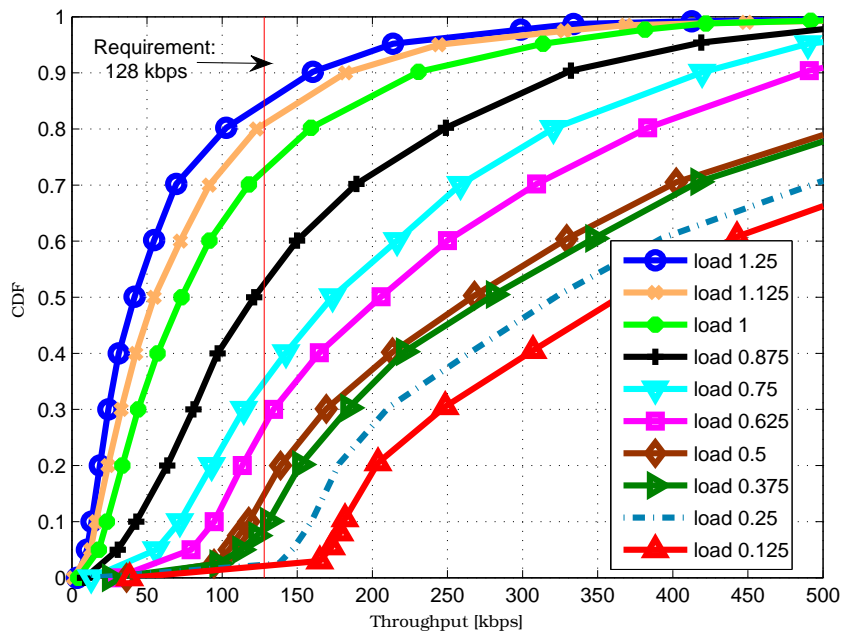


Figure 4.14: Throughput of WWWW flows for mix of 75% of VoIP flows and 25% of WWWW ones with *Adaptive CC* framework.

In Figures 4.15, 4.16 and 4.17, we can see the throughput for different service mixes when *Delay-based Prediction* framework is performed. Here, the target satisfaction level of the WWWW service (90%) for the three service mixes is also around the 0.5 Users/s load. In conclusion, we can note that the performance of WWWW flows are similar when different service mixes are compared in all the three frameworks.

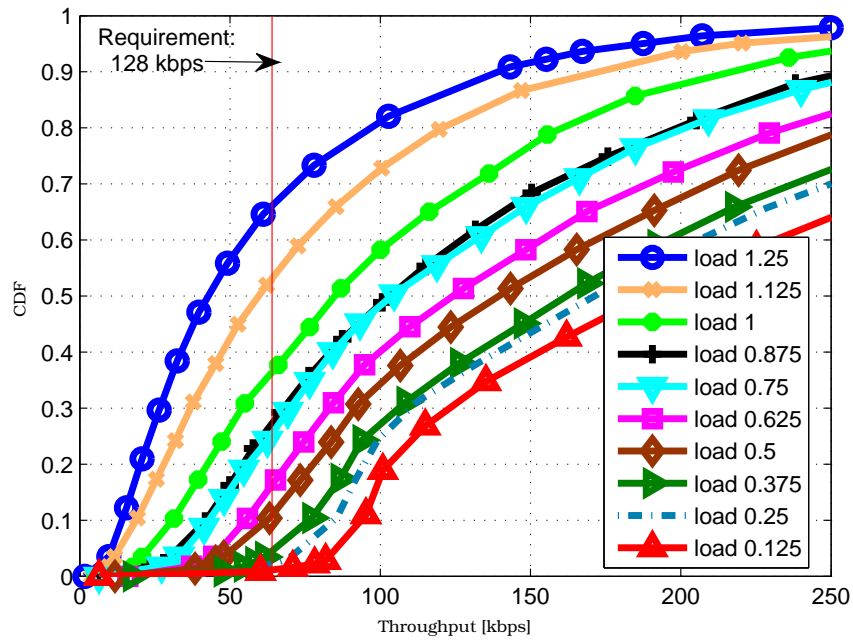


Figure 4.15: Throughput of WWW flows for mix of 25% of VoIP flows and 75% of WWWW ones with *Delay-based Prediction* framework.

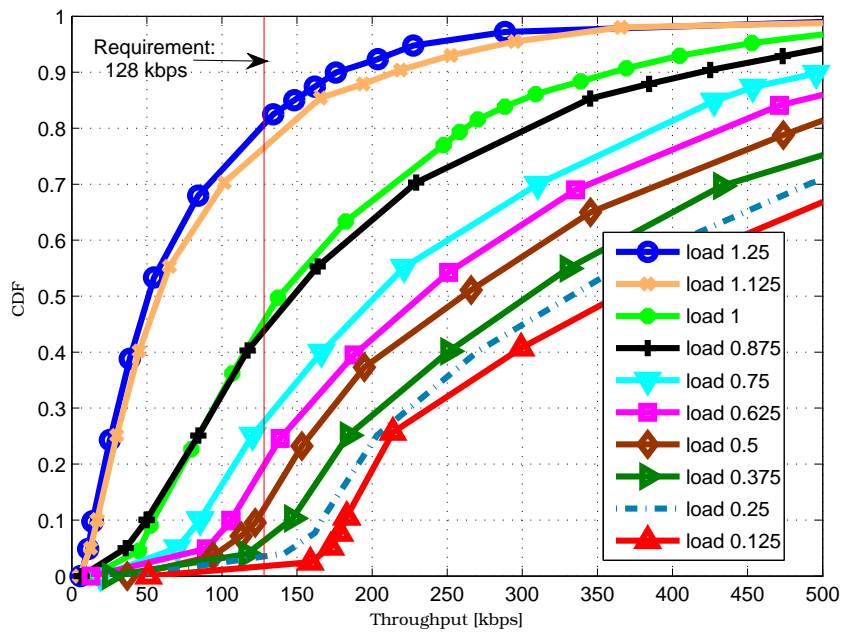


Figure 4.16: Throughput of WWWW flows for mix of 50% of VoIP flows and 50% of WWWW ones with *Delay-based Prediction* framework.

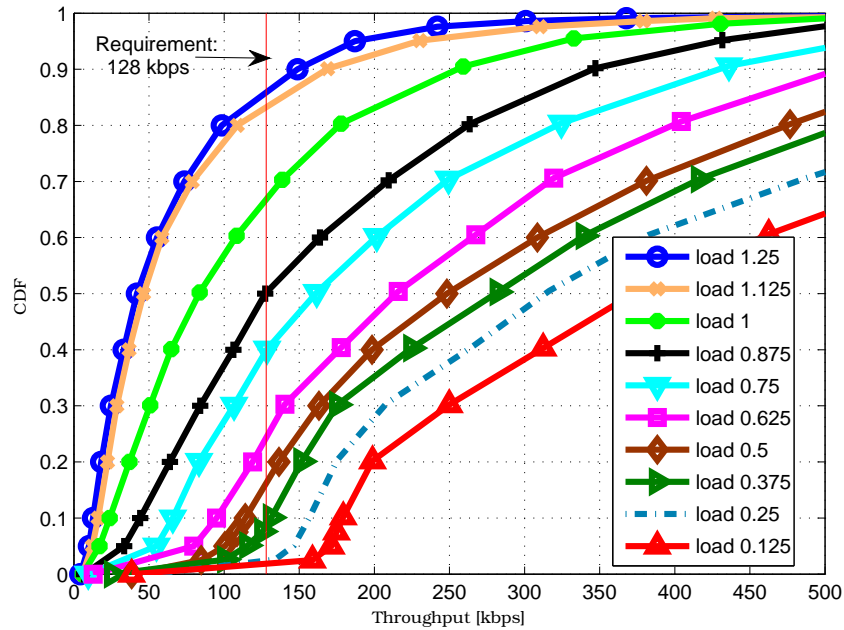


Figure 4.17: Throughput of WWWW flows for mix of 75% of VoIP flows and 25% of WWWW ones with *Delay-based Prediction* framework.

In order to know the performance of WWWW service concerning the throughput, in the Figures 4.18, 4.19 and 4.20 we show a comparison for different service mixes and different loads. In Figure 4.18 we can see at 1.125 Users/s load that the performance when the three frameworks are compared are similar. The same behavior is found at 0.75 Users/s load.

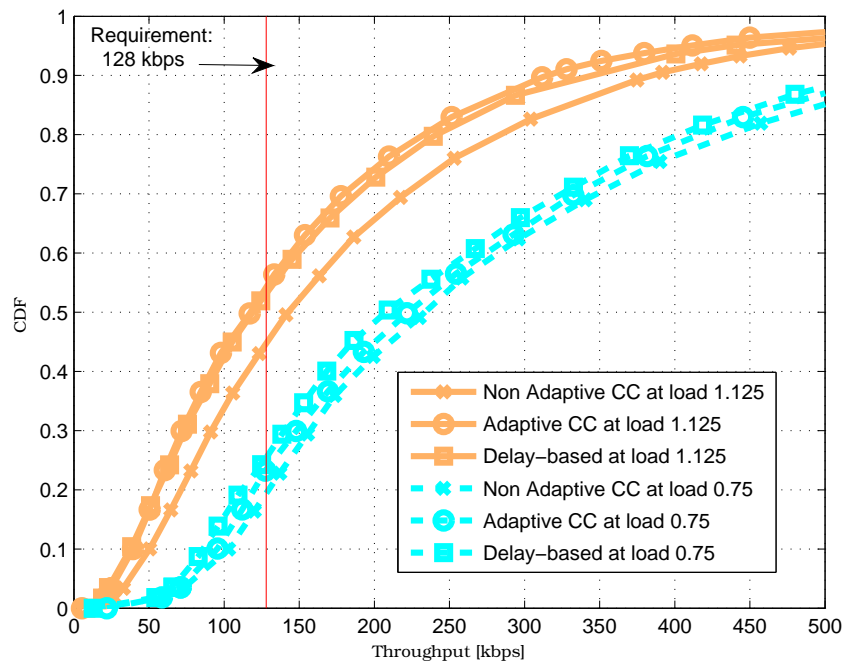


Figure 4.18: Comparison of WWWW flows' Throughput for mix of 25% of VoIP flows and 75% of WWWW ones between *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* frameworks for different loads.

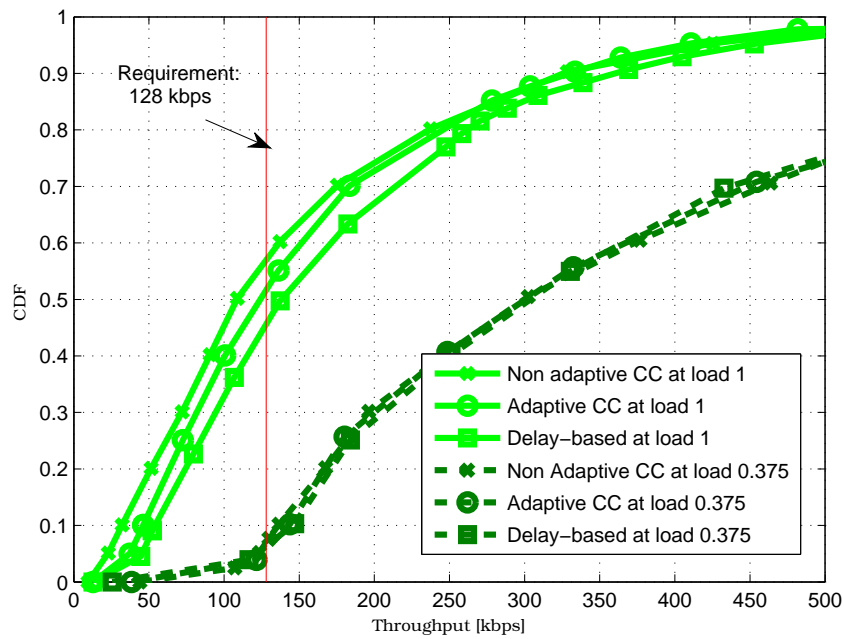


Figure 4.19: Comparison of WWWW flows' Throughput for mix of 50% of VoIP flows and 50% of WWWW ones between *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* frameworks for different loads.

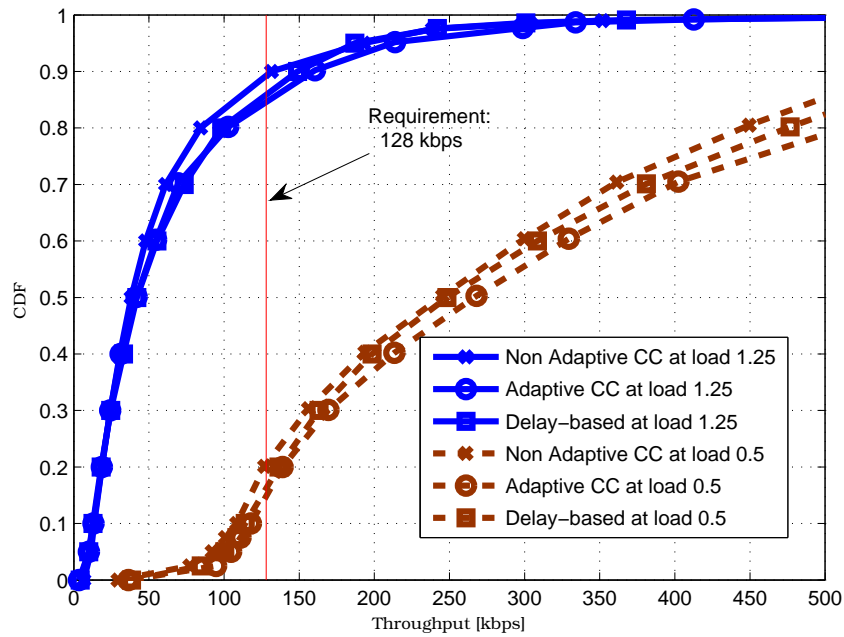


Figure 4.20: Comparison of WWWW flows' Throughput for mix of 75% of VoIP flows and 25% of WWWW ones between *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* frameworks for different loads.

Filtered Delay and Filtered FER

Subsequently we show the time variation of the filtered delay D_{VoIP}^{filt} and how it is linked with the FER_{VoIP}^{filt} . The goal is to show how important the delay can be in a CC solution. In Figures 4.21, 4.22 and 4.23, we show how is the time variation of the filtered delay D_{VoIP}^{filt} and how it is linked with the FER_{VoIP}^{filt} . First of all, we can see in these figures that there is a strong correlation between the packet delays and the QoS of VoIP flows represented by the FER. In fact, peaks of the FER_{VoIP}^{filt} are usually preceded by high values of D_{VoIP}^{filt} for the three presented frameworks.

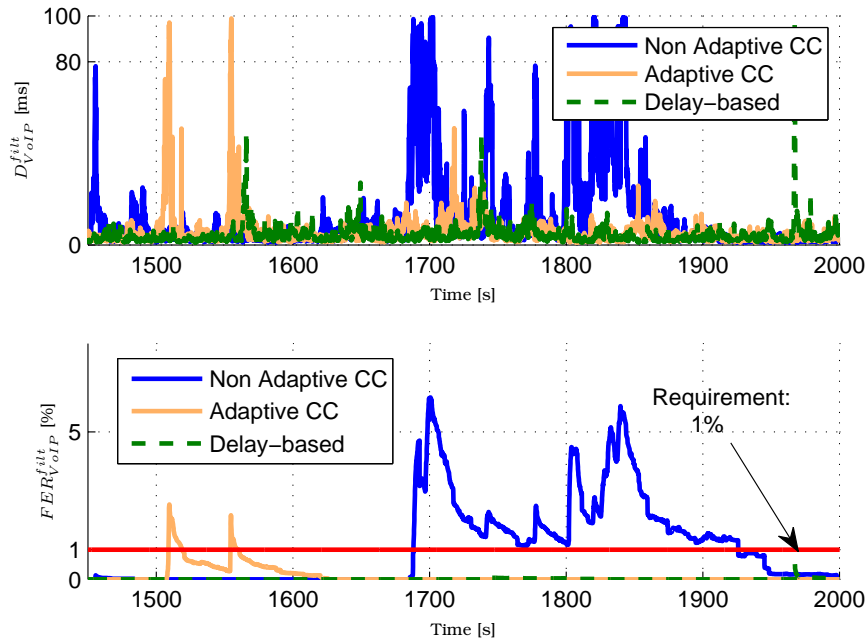


Figure 4.21: Time variation of FER_{VoIP}^{filt} and D_{VoIP}^{filt} in three different frameworks *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* at load of 1 Users/s for mix 50% of VoIP and 50% of WWW flows.

When the performance of the frameworks are concerned, the results presented in Figures 4.21, 4.22 and 4.23 also provide important insights. We can see that the three presented frameworks succeed in controlling an overload situation characterized by $FER_{VoIP}^{filt} > 1\%$ in this case. However, the feature proposed in *Delay-based Prediction* framework presents an important difference, when compared with the two other frameworks: the overload predictive capacity. The FER_{VoIP}^{filt} for the *Non Adaptive CC* and *Adaptive CC* frameworks is controlled only when the FER_{VoIP}^{filt} is higher than the FER_{VoIP}^{target} . As we can see in this figure, although the FER_{VoIP}^{filt} presents an increase when the D_{VoIP}^{filt} is high with the *Delay-based Prediction* framework, the FER_{VoIP}^{filt} does not exceed the FER_{VoIP}^{target} . This is achieved by the use of packet delay measurements in order to preview overload in the VoIP service.

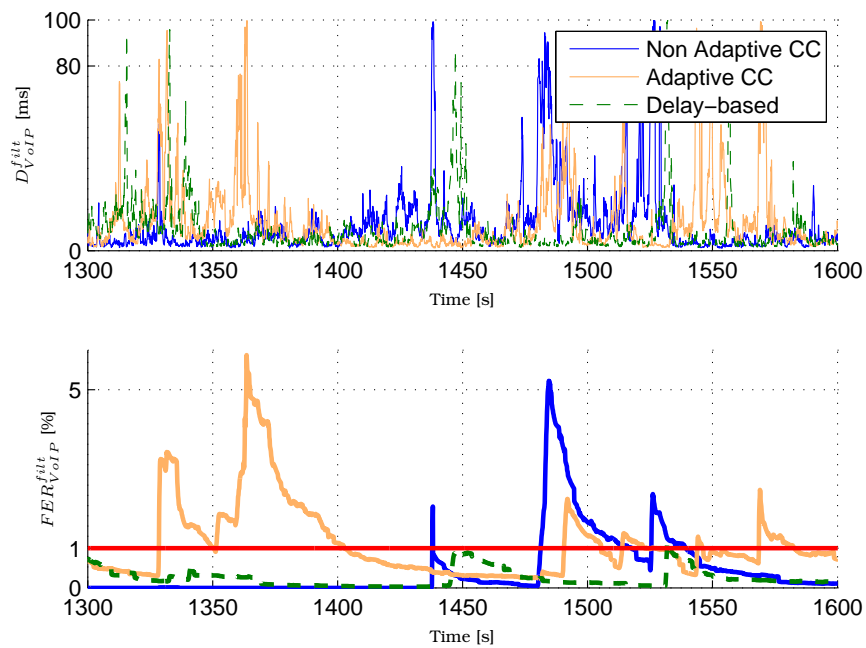


Figure 4.22: Time variation of FER_{VoIP}^{filt} and D_{VoIP}^{filt} in three different frameworks *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* at load of 1.125 Users/s for mix 50% of VoIP and 50% of WWW flows.

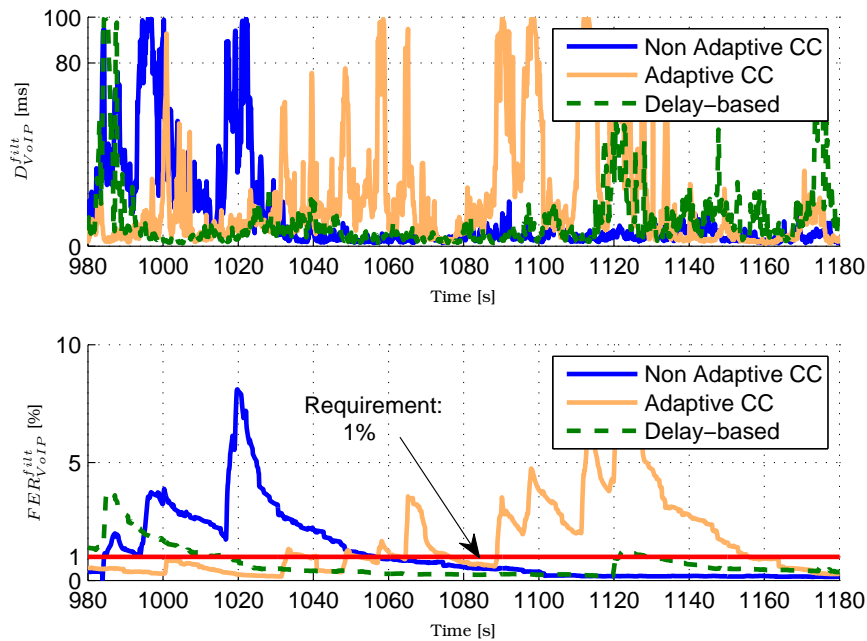


Figure 4.23: Time variation of FER_{VoIP}^{filt} and D_{VoIP}^{filt} in three different frameworks *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* at load of 1.25 Users/s for mix 50% of VoIP and 50% of WWW flows.

Mean Delay

In order to complement the information provided by the Figures of time variation of FER_{VoIP}^{filt} and D_{VoIP}^{filt} , we show in Figures 4.24, 4.25 and 4.26 the performance gains of the *Delay-based Prediction* compared with the two other frameworks in reducing the mean FER_{VoIP}^{filt} for different service mixes. This result shows that our proposed *Delay-based Prediction* is able to control the FER in overload conditions improving QoS experienced for the end user.

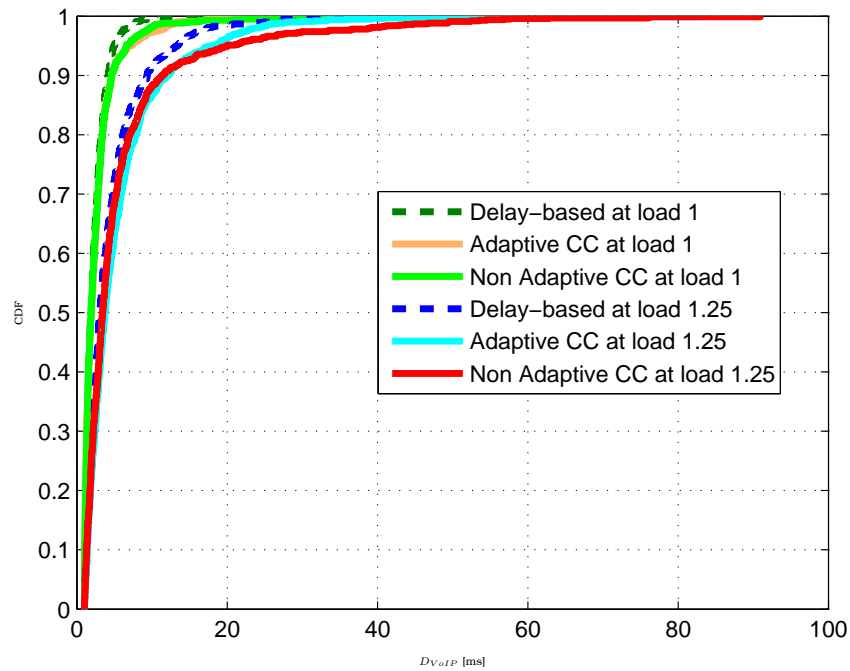


Figure 4.24: Comparison of CDF of Mean delay with mix 25% of VoIP and 75% of WWW flows for three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction*.

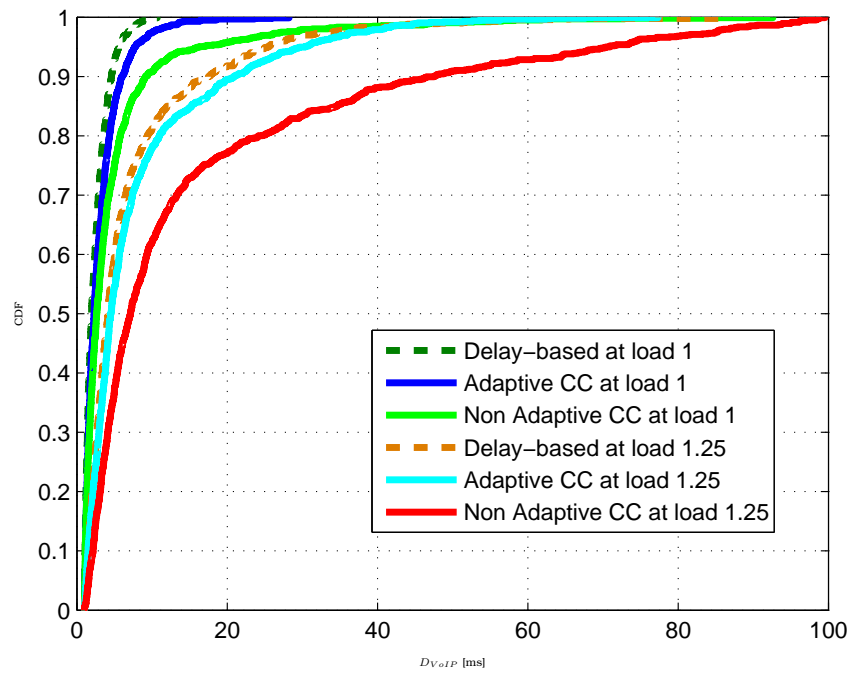


Figure 4.25: Comparison of CDF of Mean delay with mix 50% of VoIP and 50% of WWW flows for three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction*.

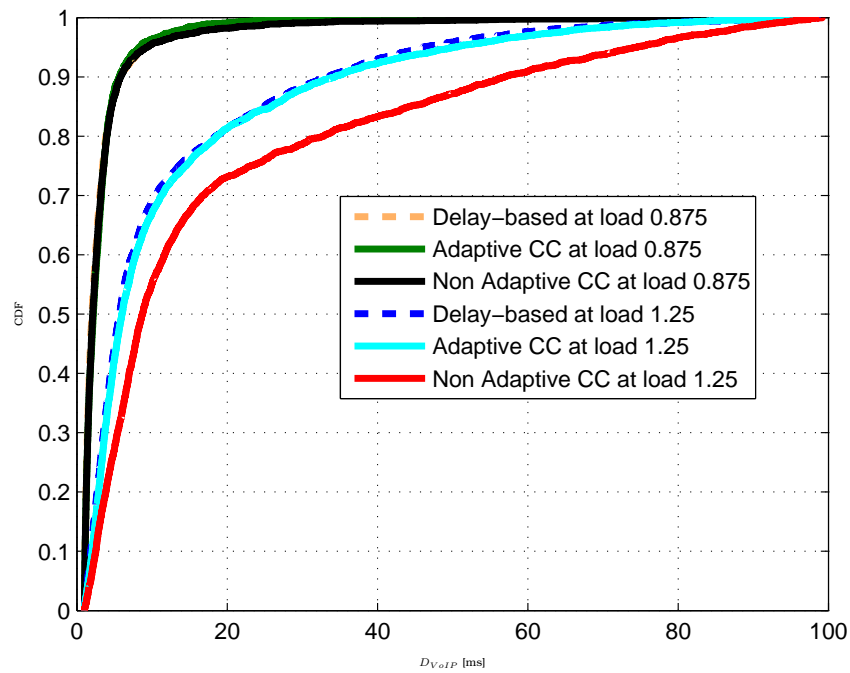


Figure 4.26: Comparison of CDF of Mean delay with mix 75% of VoIP and 25% of WWW flows for three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction*.

FER

In Figures 4.27, 4.28 and 4.29 we present the Cumulative Distribution Function (CDF) of FER of the VoIP flows that had been connected to the system for different mixes in three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* for different loads. We intend to show that more connected VoIP flows are satisfied when we apply the CC solution and even more are satisfied when we add the overload prediction feature. In Figure 4.27 we can conclude that *Delay-based Prediction* improves the number of flows with the VoIP FER less than the requirement in 2.44% if compared with *Adaptive CC* and 8.28% if compared with *Non adaptive CC* for mix of 25% of VoIP and 75% of WWW flows at load 1.375 Users/s. In Figure 4.28 the number of flows with the VoIP FER less than the requirement increased with *Delay-based Prediction* in 1.89% if compared with *Adaptive CC* and 9.75% if compared with *Non adaptive CC* for mix of 50% of VoIP and 50% of WWW flows at load 1.25 Users/s. In Figure 4.29 we present as well a simulation result for mix of 75% of VoIP and 25% of WWW flows at load 0.875 Users/s. The reason for these gains is that the framework measures frequently the delay and so it is possible to know when the system is near to be overloaded. Hence, the *Delay-based Prediction* framework can react before the FER increases.

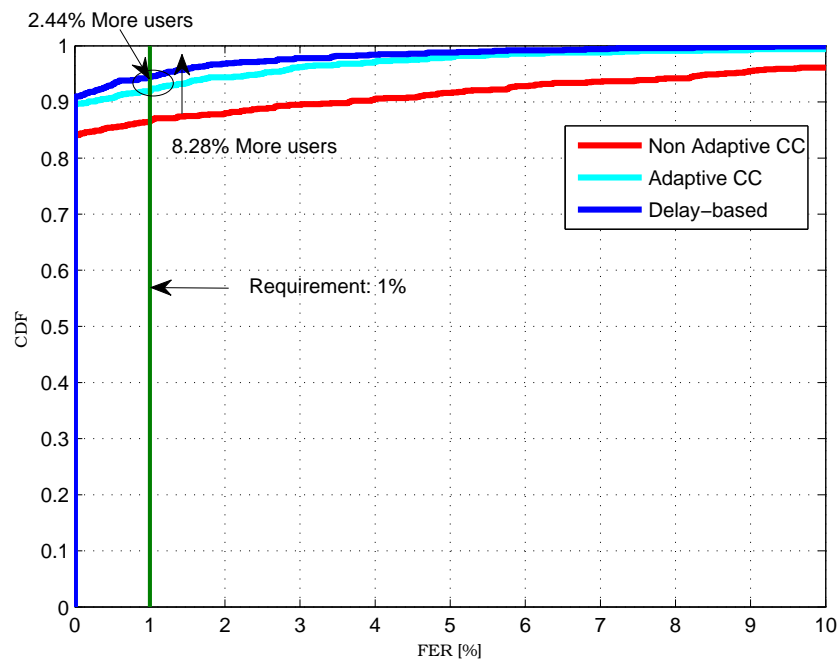


Figure 4.27: CDF FER for VoIP users in three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* at load of 1.375 Users/s for mix 25% of VoIP and 75% of WWW flows.

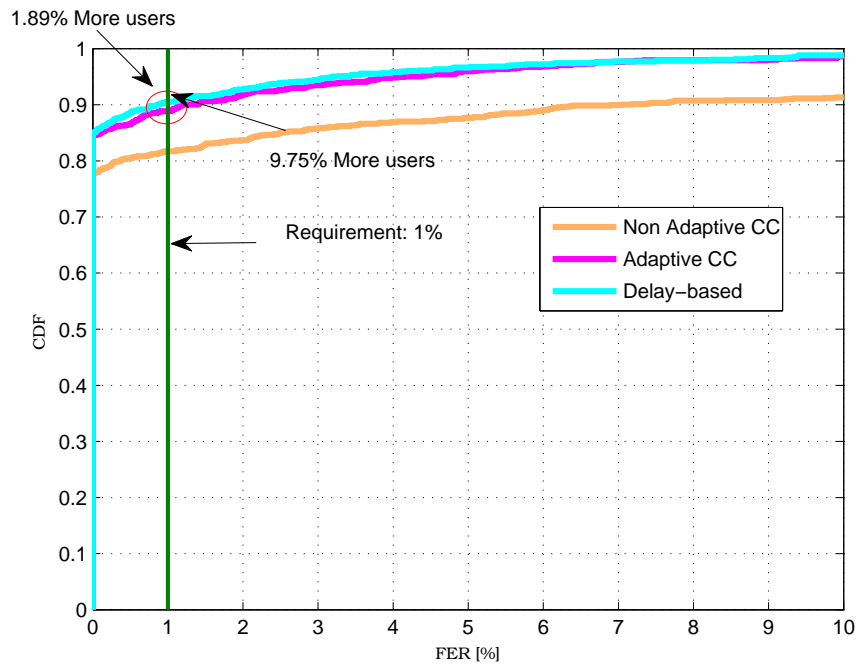


Figure 4.28: CDF FER for VoIP users in three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* at load of 1.25 Users/s for mix 50% of VoIP and 50% of WWW flows.

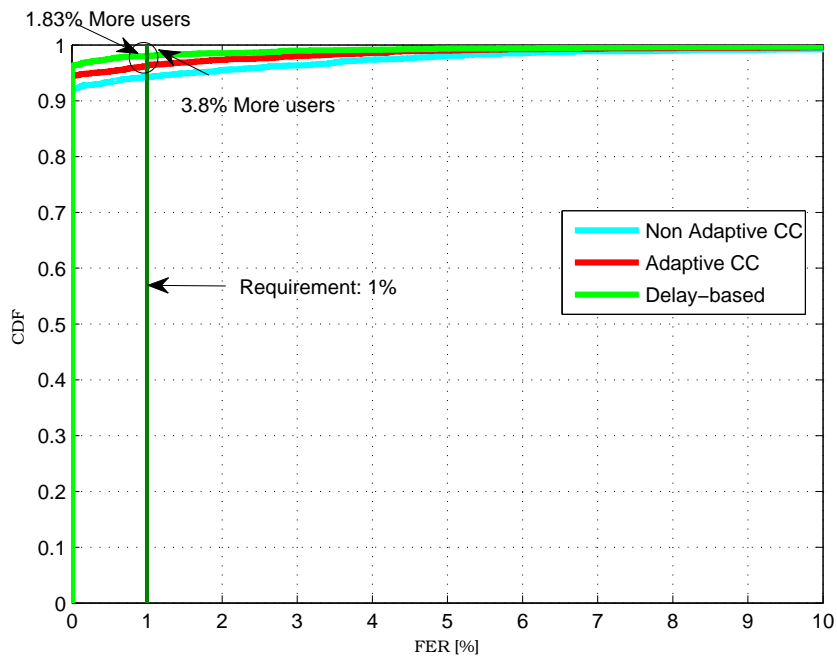


Figure 4.29: CDF FER for VoIP users in three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* at load of 0.875 Users/s for mix 75% of VoIP and 25% of WWW flows.

Satisfaction

Another important result presented in this study is the satisfaction ratio. In Figures 4.30, 4.31 and 4.32 we present the satisfaction ratio for different service mixes also in three scenarios: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction*. We can notice that the higher the percentage of VoIP flows, the greater the performance gain is for VoIP if we compare the *Delay-based Prediction* framework and the *Adaptive CC*. The reason for this is that the WWW demands a lot of resources, so the lower the number of WWW flows, the greater the performance gain in VoIP service will be. The performance gain obtained in *Delay-based Prediction* framework occurs because the system monitors the delay and so it is possible to predict that a VoIP packet can be lost before it happens and so the system can react before FER builds up.

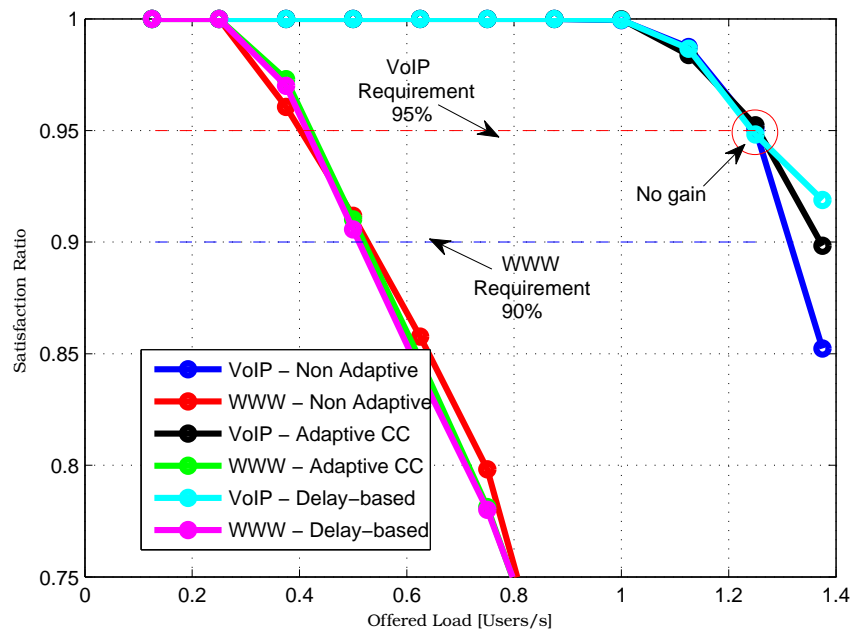


Figure 4.30: Satisfaction for three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* for mix 25% of VoIP and 75% of WWW flows.

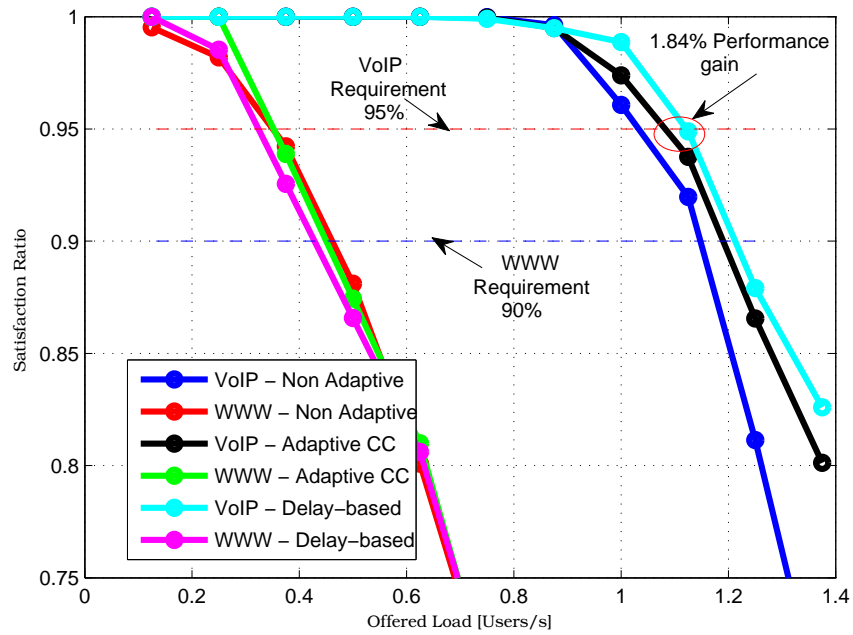


Figure 4.31: Satisfaction for three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* for mix 50% of VoIP and 50% of WWW flows.

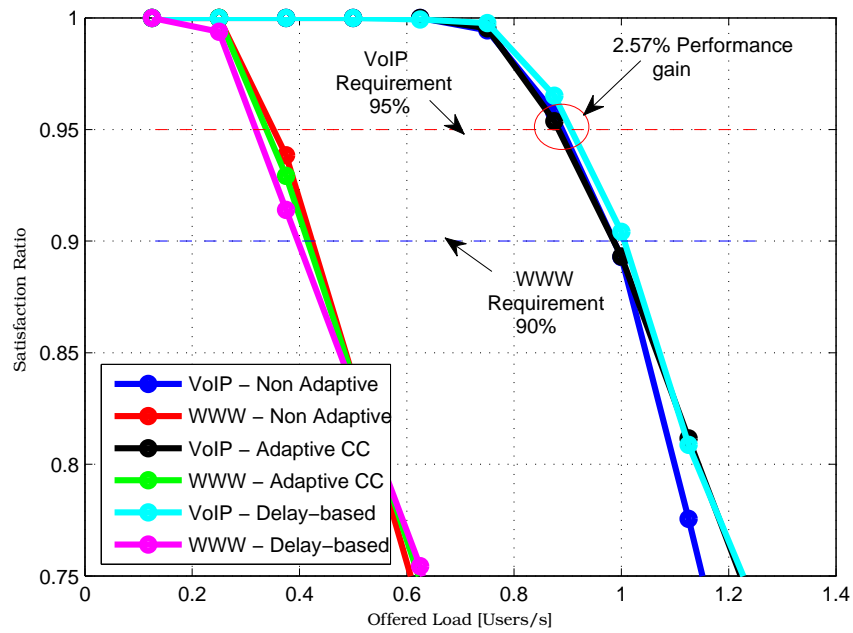


Figure 4.32: Satisfaction for three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* for mix 75% of VoIP and 25% of WWW flows.

Joint Capacity

Finally, in Figure 4.33 we present a joint capacity result. This capacity region was constructed varying the traffic mix among VoIP and WWW services, including single service evaluations. As we have just seen in the satisfaction curves, the WWW is the more restrictive service that impose limits in the joint capacity. Thus, it is not expected a gain in this joint capacity. By regarding this figure, we can conclude that *Delay-based Prediction* imposes only a small performance degradation in order to guarantee the QoS fulfillment of VoIP.

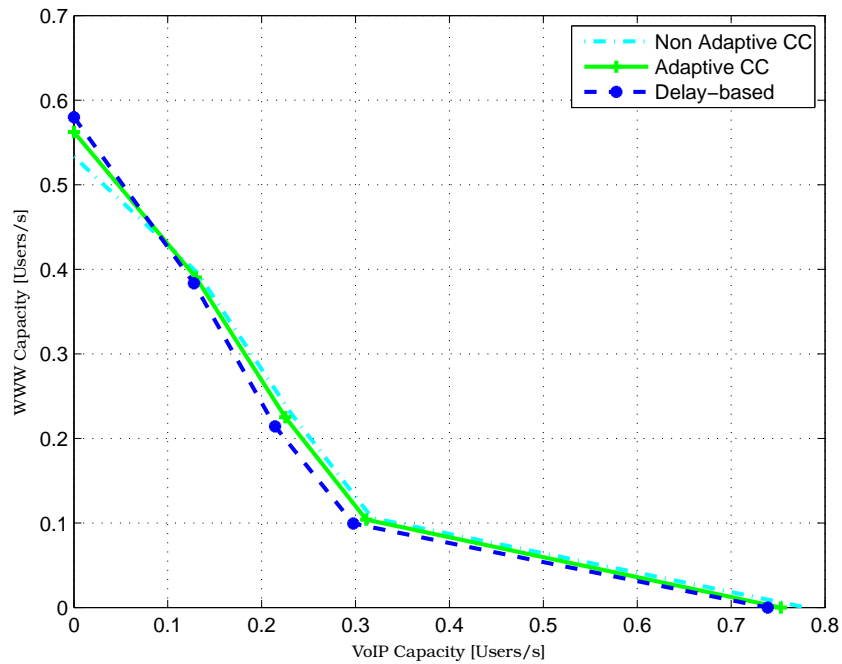


Figure 4.33: Joint Capacity with three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction*.

4.4.3 Study Case of Service Balancing

In this last part we show the time variation of α , β and FER_{VoIP}^{filt} . The following results presented in Figures 4.34, 4.35, 4.36, 4.37 and 4.38 show a study case when α is varying between 10 and -10 and β between 10 and -20 . In this study case, we intend to know what happens when α and β also assume positive values, e.g. when FER_{RT}^{filt} is well controlled and so the system can give more priority to the WWW flows.

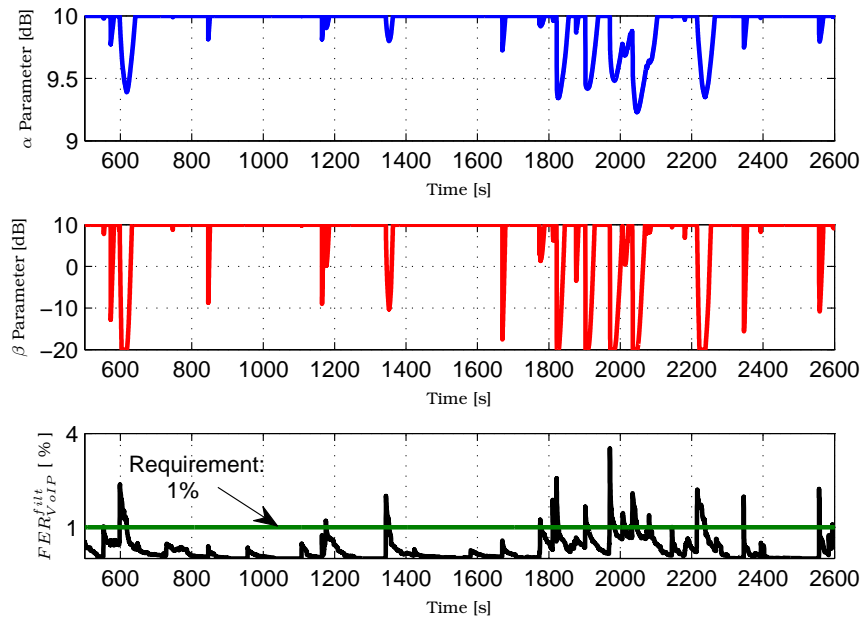


Figure 4.34: Time variation of α , β and FER_{VoIP}^{filt} at the load of 0.875 Users/s.

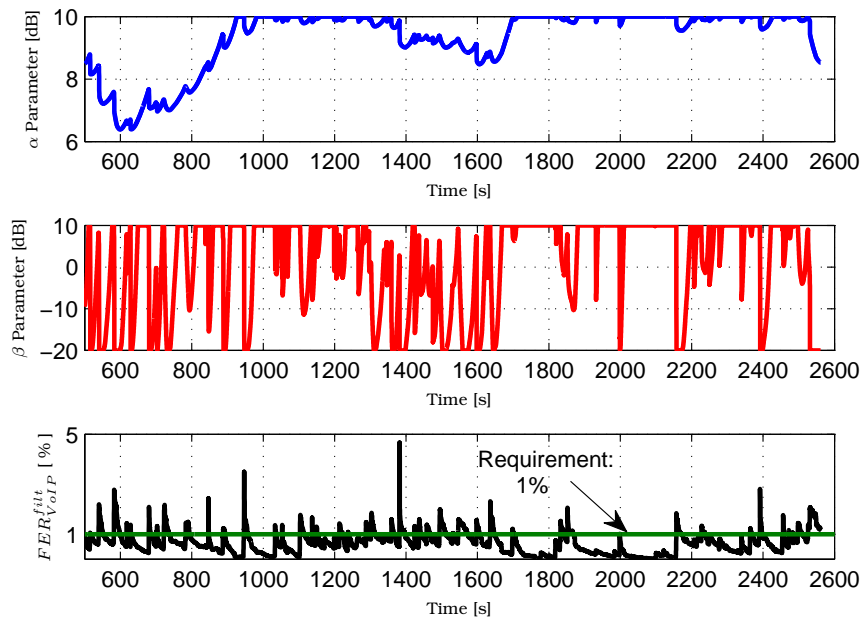


Figure 4.35: Time variation of α , β and FER_{VoIP}^{filt} at the load of 1 Users/s.

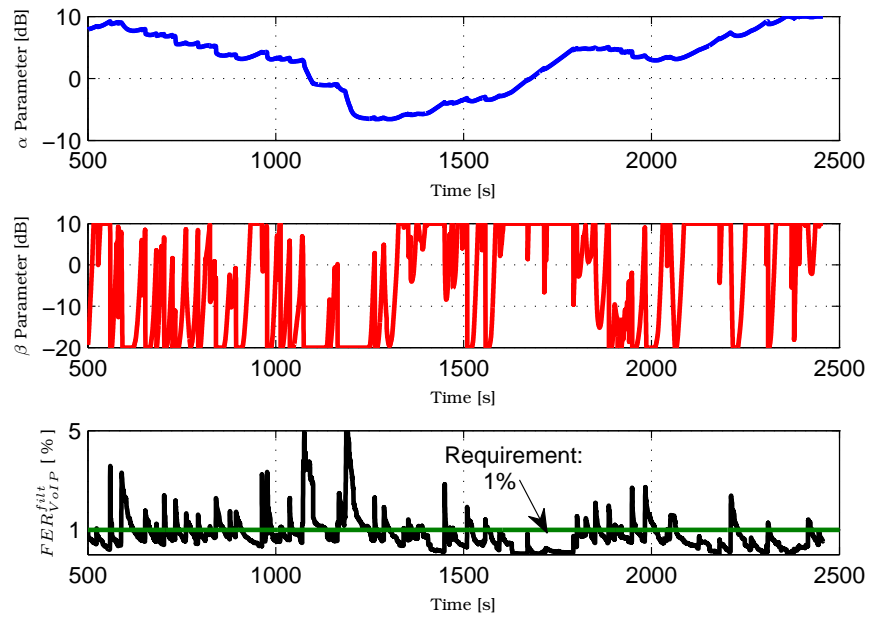


Figure 4.36: Time variation of α , β and FER_{VoIP}^{filt} at the load of 1.125 Users/s.

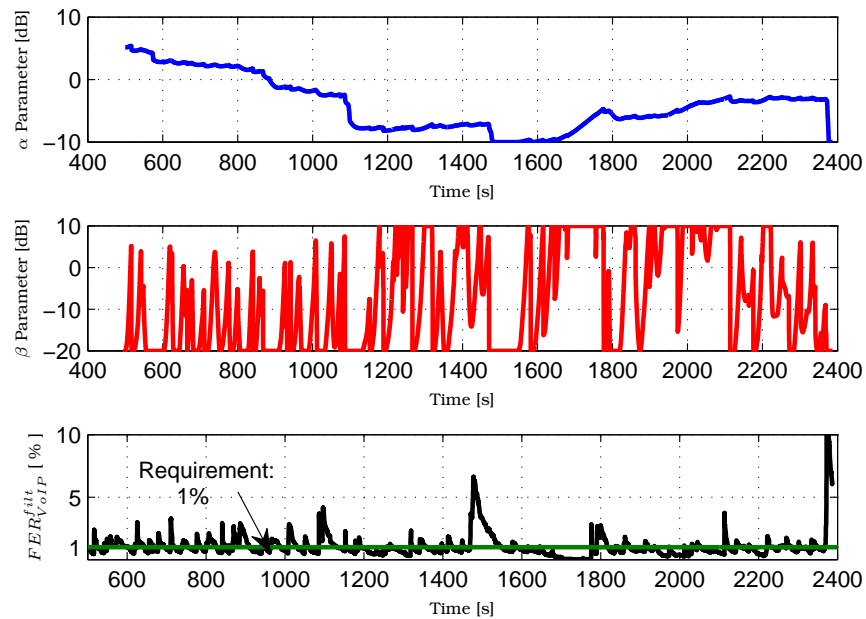


Figure 4.37: Time variation of α , β and FER_{VoIP}^{filt} at the load of 1.25 Users/s.

In this study case, as α and β also assume positive values, WWW flows can be also prioritized when FER_{VoIP}^{filt} is below the FER_{VoIP}^{target} . In Figure 4.38 we can observe that as the WWW also can be sometimes prioritized during the simulation, the system resources of both services seems to be better shared. Hence, with this result we can also increase the system capacity.

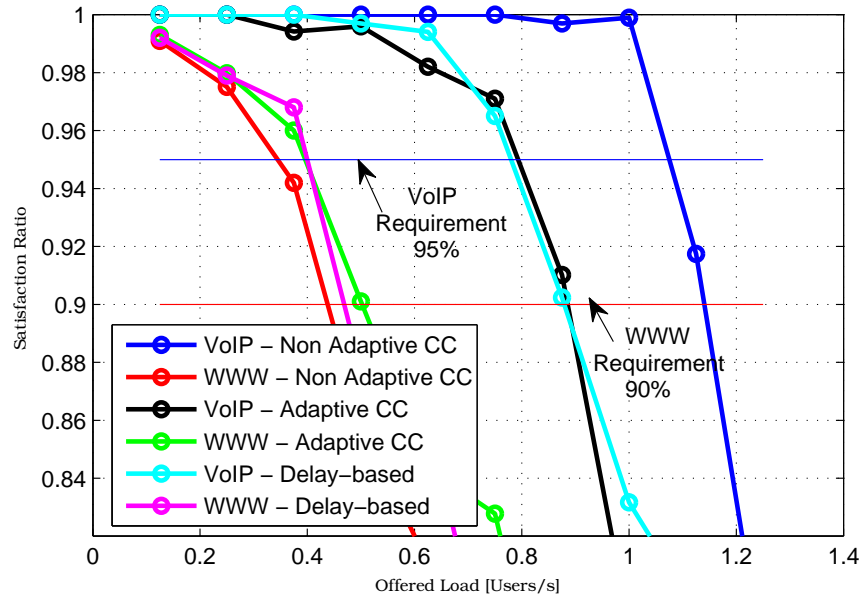


Figure 4.38: Satisfaction for three different frameworks: *Non adaptive CC*, *Adaptive CC* and *Delay-based Prediction* for mix 50% of VoIP and 50% of WWW flows.

Chapter 5

Conclusions and Perspectives

In this work the simulation tool models the main aspects of a single-cell Orthogonal Frequency Division Multiple Access (OFDMA)-based system. The simulation modelling includes aspects such as propagation phenomena (e.g., path loss, shadowing and fast fading), service applications, link adaptation, a channel free of errors and detailed models for traffic generation of Real Time (RT) and Non-Real Time (NRT) services, higher layer protocols, and mobility profiles. We presented two contributions to protect the Quality of Service (QoS) of RT services in a mixed traffic scenario. The first contribution is the generalization of the Congestion Control (CC) framework proposed in [4] for networks employing OFDMA in the downlink. In the second contribution we proposed a new feature to be added to the generalized framework. The *Delay-based Prediction* framework besides guaranteeing the QoS fulfillment of a RT service also prevent high peaks of Frame Erasure Rates (FERs) by using the packet-delay prediction capability. By analyzing simulation results, when the blocking rate begins to increase its value, the system is already overloaded, and the satisfaction of the World Wide Web (WWW) flows is already below the threshold of 90%. We can conclude that the overload prediction based on delay can efficiently avoid the increase of FER before it really builds up and prevent QoS degradation of RT flows imposing only a small performance degradation to the WWW service.

Finally, in the study case of service balancing, when α and β also assume positive values, the WWW flows can be prioritized when FER_{RT}^{filt} is well controlled and less than FER_{RT}^{target} . In conclusion, both services could be prioritized so that the system resources could be better shared and the system capacity increased. As a perspective of this work, we intend to study a service balancing QoS-based framework.

Traffic Modeling

The application layer is the highest layer in the Open Systems Interconnection (OSI) reference model. One important aspect of the application layer is the statistic nature of each traffic type, which is emulated by a suitable traffic model. These traffic models are very important for the performance analysis of the wireless systems.

In the traffic model for *voice*, voice activity is modeled through a two-state or three-state Markov chain, as described in [44]. The two-state traffic model assumes the use of a slow voice activity detector. This activity pattern is illustrated in figure A.1. An average duration of 60s is assumed for the voice calls. The table A.1 summarizes the parameters of the model of two-state voice traffic. For the *interactive* WWW service, the model is described in table A.2.

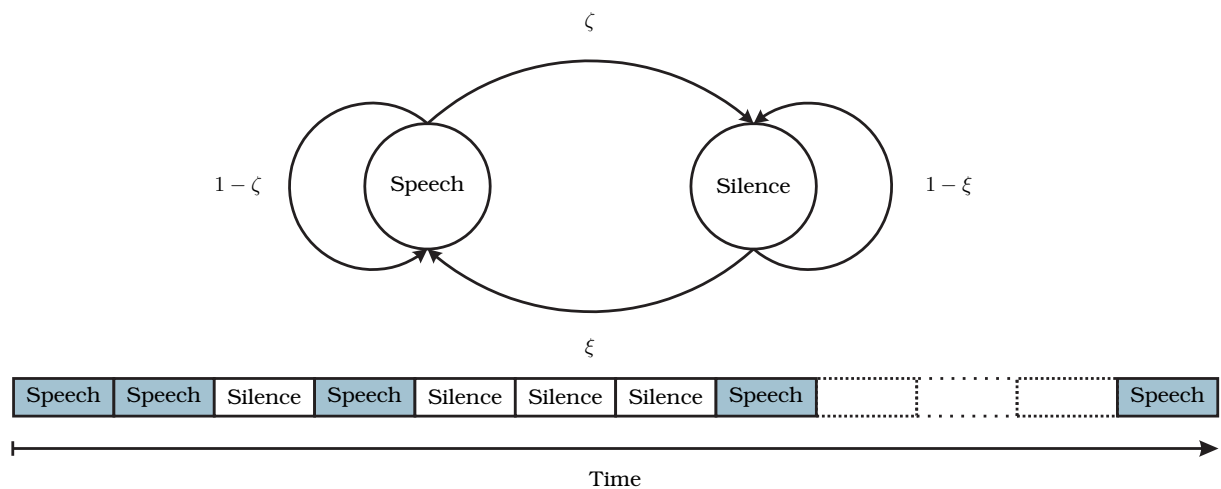


Figure A.1: Two-State Voice Traffic Model.

Table A.1: Parameters of the Two-State Voice Traffic Model.

Parameter	Value	Unit
Number of States	2 (speech (on) and silent (off))	-
Average duration of speech period (t_{on})	3	s
Average duration of silent period (t_{off})	3	s
Average call duration	60	s

Table A.2: WWW Session Model 2 [41].

Description	Distribution	Parameters
Number of packet calls per session	Geometric	10 (mean)
Number of packets per packet call	deterministic	1
Reading time per packet calls	Truncated Pareto	$\alpha = 3.2$ $k = 3.45$ $m = 120s$
Packet call size (byte)	Lognormal	$\mu = 4, 100$ $\sigma = 30, 000$
Maximum packet size	deterministic	100,000bytes

Bibliography

- [1] D. McQueen, "The Momentum Behind LTE Adoption [3GPP LTE]," *Communications Magazine, IEEE*, vol. 47, no. 2, pp. 44–45, february 2009.
- [2] M. Alasti, B. Neekzad, J. Hui, and R. Vannithamby, "Quality of Service in WiMAX and LTE Networks [Topics in Wireless Communications]," *Communications Magazine, IEEE*, vol. 48, no. 5, pp. 104–111, may 2010.
- [3] A. R. Braga, E. B. Rodrigues, and F. R. P. Cavalcanti, "Packet Scheduling for VoIP over HSDPA in Mixed Traffic Scenarios," in *Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on*, Helsinki, September 2006, pp. 1–5.
- [4] E. B. Rodrigues and F. R. P. Cavalcanti, "QoS-Driven Adaptive Congestion Control for Voice over IP in Multiservice Wireless Cellular Networks," *IEEE Communications Magazine*, vol. 46, no. 1, pp. 100–107, January 2008.
- [5] AT&T, "Milestones in AT&T History," 2005. [Online]. Available: <http://www.corp.att.com/history/milestones.html>
- [6] T. Farely and K. Schmidt, "Digital Wireless Basics: Frequency Reuse," 2006. [Online]. Available: http://www.privateline.com/mt_digitalbasics/
- [7] Kioskea, "Mobile telephony," 2008. [Online]. Available: <http://en.kioskea.net/contents/telephonie-mobile/reseaux-mobiles.php3>
- [8] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*, 2nd ed. Academic Press, 2008.
- [9] 3GPP, "Requirements for Further Advancements for E-UTRA (LTE-Advanced)," 3rd Generation Partnership Project, Tech. Rep. TR 36.913 V9.0.0 - Release 9, December 2009.
- [10] F. J. Velez and L. M. Correia, "Classification and Characterisation of Mobile Broadband Services," in *Vehicular Technology Conference, 2000. IEEE VTS-Fall VTC 2000. 52nd*, vol. 3, September 2000, pp. 1417–1423.
- [11] A. Khan, M. Qadeer, J. Ansari, and S. Waheed, "4G as a Next Generation Wireless Network," in *Future Computer and Communication, 2009. ICFCC 2009. International Conference on*, 3-5 2009, pp. 334–338.
- [12] R. Young Kyun, Kim; Prasad, *4G Roadmap and Emerging Communication Technologies*. Artech House, 2006.

- [13] J. Govil, "An Empirical Feasibility Study of 4G's Key Technologies," in *Electro/Information Technology, 2008. EIT 2008. IEEE International Conference on*, 18-20 2008, pp. 267–270.
- [14] N. Enderle and X. Lagrange, "User Satisfaction Models and Scheduling Algorithms for Packet-Switched Services in UMTS," in *Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual*, vol. 3, 2003, pp. 1704–1709.
- [15] L. Badia, M. Boaretto, and M. Zorzi, "A Users' Satisfaction driven Scheduling Strategy for Wireless Multimedia QoS," in *Proceedings QoSIS 2003, Stockholm, Sweden*, October 2003.
- [16] D. Chalmers and M. Sloman, "A Survey of Quality of Service in Mobile Computing Environments," *IEEE Communications Surveys*, April 1999.
- [17] P. Mehta and S. Udani, "Voice over IP: Sounding Good on the Internet," *Potentials, IEEE*, vol. 20, no. 4, pp. 36–40, October 2001.
- [18] J. Pérez-Romero, O. Sallent, R. Agustí, and M. A. Díaz-Guerra, *Radio Resource Management Strategies in UMTS*, 1st ed. John Wiley & Sons Ltd, October 2005.
- [19] Y. Koucheryavy, G. Giambene, D. Staehle, F. Barcelo-Arroyo, B. T. V. Siris, and Editors, *Traffic and QoS Management in Wireless Multimedia Networks - COST 290 Final Report*. Springer, 2009.
- [20] E. L. Pinto and C. P. Albuquerque, "A Técnica de Transmissão OFDM," *Revista Científica Periódica - Telecomunicações*, vol. 5, June 2002.
- [21] Y. G. Li and G. L. Stüber, *Orthogonal Frequency Division Multiplexing for Wireless Communications*, 1st ed. Springer, December 2005.
- [22] 3GPP, "Feasibility Study for Evolved Universal Terrestrial Radio Access (UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN)," 3rd Generation Partnership Project, Tech. Rep. TR 25.912 V7.2.0 - Release 7, September 2006.
- [23] M. Sternad, T. Svensson, T. Ottosson, A. Ahlen, A. Svensson, and A. Brunstrom, "Towards Systems Beyond 3G Based on Adaptive OFDMA Transmission," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2432–2455, December 2007.
- [24] A. R. Braga, S. Wänstedt, and M. Ericson, "Admission Control for VoIP over HSDPA in a Mixed Traffic Scenario," in *Telecommunications Symposium, 2006 International*, Fortaleza, Ceara, September 2006, pp. 71–76.
- [25] M. Kazmi, P. Godlewski, and C. Cordier, "Admission Control Strategy and Scheduling Algorithms for Downlink Packet Transmission in WCDMA," vol. 2, 2000, pp. 674–680 vol.2.
- [26] E. B. Rodrigues and J. Olsson, "Admission Control for Streaming Services over HSDPA," in *AICT-SAPIR-ELETE '05: Proceedings of the Advanced Industrial Conference on Telecommunications/Service Assurance with Partial and Intermittent Resources Conference/E-Learning on Telecommunications Workshop*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 255–260.
- [27] A. Furuskär, "Packet Scheduling and Quality of Service in HSDPA," Department of Communication Technology - Institute of Electronic Systems - Aalborg University, Tech. Rep. Ph.D. dissertation, October 2003.

- [28] H. Kim and Y. Han, "A Proportional Fair Scheduling for Multicarrier Transmission Systems," *Communications Letters, IEEE*, vol. 9, no. 3, pp. 210 – 212, March 2005.
- [29] Y. C. L. Yanhui, W. Chunming and T. Guangxin, "Downlink Scheduling and Radio Resource Allocation in Adaptive OFDMA Wireless Communication Systems for User-Individual QoS," in *Transactions on Engineering, Computing and Technology - World Enformatika Society*, no. 2, March 2006, pp. 426–440.
- [30] E. B. Rodrigues, F. R. M. Lima, and F. R. P. Cavalcanti, "Load Control for VoIP over HSDPA in Mixed Traffic Scenarios," in *Personal*, Athens, Greece, September 2007.
- [31] J. Gu and X. Che, "Adaptive Uplink Load Control in CDMA Systems," vol. 2, September 2005, pp. 1193 – 1196.
- [32] A. Sampath, P. S. Kumar, and J. M. Holtzman, "On Setting Reverse Link Target SIR in a CDMA system," vol. 2, May 1997, pp. 929 –933 vol.2.
- [33] S. K. Kasera, R. Ramjee, S. R. Thuel, and X. Wang, "Congestion Control Policies for IP-based CDMA Radio Access Networks," *Mobile Computing, IEEE Transactions on*, vol. 4, no. 4, pp. 349 – 362, july-aug. 2005.
- [34] F. R. P. Cavalcanti, S. Andersson, and Editors, *Optimizing Wireless Communication Systems*. Springer, 2009.
- [35] 3GPP, "UTRAN Iub Interface NBAP Signalling," 3rd Generation Partnership Project, Tech. Rep. TS 25.433 V6.4.0 - Release 6, January 2005.
- [36] R. E. Goot, U. Mahalab, and R. Cohen, "Nonlinear Exponential Smoothing (NLES) Algorithm for Noise Filtering and Edge Preservation," in *HAIT Journal of Science and Engineering*, vol. 2, May 2005.
- [37] "Matlab - the language of technical computing," 2009. [Online]. Available: <http://www.mathworks.com/products/matlab/>
- [38] 3GPP, "3GPP Specification Series." [Online]. Available: <http://www.3gpp.org/ftp/Specs/html-info/36-series.htm>
- [39] C. Yue-yun and D. Xiao-hui, "A Novel Sub-carrier Allocation Algorithm in 3G LTE System," in *Wireless, Mobile and Sensor Networks, 2007. (CCWMSN07). IET Conference on*, 12-14 2007, pp. 474 –477.
- [40] 3GPP, "Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS," UMTS/ETSI, Tech. Rep. TR 101.112 v3.2.0, April 1998.
- [41] C. Johansson, L. D. Verdier, and F. Khan, "Performance of Different Scheduling Strategies in a Packet Radio System," in *Universal Personal Communications, 1998. ICUPC '98. IEEE 1998 International Conference on*, vol. 1, Florence, October 1998, pp. 267–271.
- [42] 3GPP, "Performance Characterization of the Adaptive Multi-Rate Speech Codec," 3rd Generation Partnership Project, Tech. Rep. TS 25.975 V6.0.0 - Release 6, December 2004.
- [43] A. Furuskär, "Radio Resource Sharing and Bearer Service Allocation for Multi-bearer Service, Multi-access Wireless Networks - Methods to Improve Capacity," Royal Institute of Technology - KTH, Tech. Rep. Ph.D. dissertation, 2003.

-
- [44] D. J. Goodman, "Efficiency of Packet Reservation Multiple Access," *IEEE Transactions on Vehicular Technology*, pp. 170–176, February 1991.