Dynamic Adaptation Algorithm for Multimedia Delivery in Wireless Networks

Teză destinată obţinerii
titlului ştiinţific de doctor inginer
la
Universitatea "Politehnica" din Timişoara
în domeniul INGINERIE ELECTRONICĂ ŞI TELECOMUNICAŢII
de către

Ing. Vasile Horia Muntean

Conducător ştiinţific: prof.univ.dr.ing Marius Oteşteanu

Referenţi ştiinţifici: prof.univ.dr.ing. Cristina Hava-Muntean
prof.univ.dr.ing. Virgil Dobrotă
prof.univ.dr.ing. Florin Alexa

Ziua susţinerii tezei: 21 Decembrie 2012

Seriile Teze de doctorat ale UPT sunt:

1. Automatică
2. Chimie
3. Energetică
4. Ingineria Chimică
5. Inginerie Civilă
6. Inginerie Electrică
7. Inginerie Electronică şi Telecomunicaţii
8. Inginerie Industrială
9. Inginerie Mecanică
10. Ştiinţa Calculatoarelor
11. Ştiinţa şi Ingineria Materialelor

Universitatea „Politehnica” din Timişoara a iniţiat seriile de mai sus în scopul diseminării expertizei, cunoştinţelor şi rezultatelor cercetărilor întreprinse în cadrul şcolii doctorale a universităţii. Seriile conţin, potrivit H.B.Ex.S Nr. 14 / 14.07.2006, tezele de doctorat susţinute în universitate începând cu 1 octombrie 2006.

România, 300159 Timişoara, Bd. Republicii 9,
tel. 0256 403823, fax. 0256 403221
e-mail: editura@edipol.upt.ro

Foreword

First, I want to express my gratitude to my thesis supervisor, Prof. Dr. Marius Oteşteanu, whose guidance, understanding and continuous technical and moral support proved to be crucial to the realization of this thesis.

I would also wish to address special thanks to Dr. Gabriel-Miro Muntean, Director of the Network Innovations Centre, Dublin City University, for the precious insights and technical advices offered during my research. Furthermore, I would like to thank my colleagues from Alcatel-Lucent for all fruitful discussions and good advices offered during the last years.

Finally, I would like to thank my parents for their unlimited, ultra-supportive encouragement, sacrifices and unconditional love throughout my entire life. Last but not least, many thanks to my sister Ada and to my special person Roxana, who both offered me their unconditional support all this time.

Timişoara, December 2012 Muntean Vasile Horia

Muntean, Vasile Horia

Dynamic Adaptation Algorithm for Multimedia Delivery in Wireless Networks

Teze de doctorat ale UPT, Seria 7, Nr. 58, Editura Politehnica, 2012, 144 pagini, 75 figuri, 32 tabele.

Cuvinte cheie:
wireless communications, multimedia streaming, 3GPP LTE networks, IEEE 802.11 WLAN, adaptation algorithms, QoS mapping schemes

Abstract:

In the evolution of mobile and wireless communications, four generations can be identified. In 1980s, the first generation (1G) provided support for analog communication at very low bitrates. In 1990s, the second generation (2G) marked the transition to digital communications and an increase in bitrates, whereas in 2000s the third generation (3G) saw the introduction of broadband to mobile communications, leading the way towards the beyond 3G generation (B3G) and forth generation (4G). 4G represents the grand vision of mobile communications, providing mobile users access to advanced applications comparable to those offered by wired networks and, most importantly, maintaining the mobility support which is impossible in wired networks.

The evolution of applications was also extremely rapid, but most importantly was inter-twined with that of the computing devices on which they run. In early years the addition of the graphical user interface to the basic word processing applications was considered an important leap in the development of applications. The latest addition to this concentrated effort was the support for mobility, which offered a rich new avenue for applications development, but put significant pressure on finding solutions, mostly in terms of quality-oriented data communications.

In this context, the thesis covers important aspects of the emerging technologies in the heterogeneous next generation network environment with focus on wireless communications and quality of multimedia. It describes in details several important communications standards, identifies research issues and challenges and presents a survey of the proposed solutions in the literature. The thesis introduces the newly proposed Dynamic Quality-Oriented Adaptation Scheme (DQOAS), a user-oriented adaptation mechanism that improves user quality of experience (QoE) while also enabling a higher number of simultaneously connected clients to communicate in an IEEE 802.11 or a 3GPP LTE network. DQOAS enhances an existing solution, the Quality-Oriented Adaptation Scheme (QOAS) by adding different user QoE expectation levels in the adaptation process. By making use of these levels, DQOAS will perform a differentiated treatment of the users in the adaptive process, based not only on network conditions, but also on user requirements in terms of their QoE. When DQOAS is deployed in an LTE environment, its intrinsic prioritization offers a mapping alternative for LTE QoS parameters in case of a traffic mix generated by one application, in order to improve end-user QoE.

Contents

INDEX

3GPP - 3rd Generation Partnership Project
AC - Access Category
ACK - Acknowledgment frame
AIAD - Adaptive Increase and Adaptive Decrease
AIFS - Arbitration Interframe Space
AIFSN - Arbitration Interframe Space Number
AIMD - Adaptive Increase and Multiplicative Decrease
AIPD - Adaptive Increase and loss Proportional Decrease
AM - Adaptation Model
AMC - Adaptive Modulation and Coding
AN - Access Network
AP - Access Point
AQM - Active Queue Management
ARQ - Automatic Request
AVP - Audio Visual Profile
AVPF - Audio Visual Profile with Feedback
BLER - Block Error Ratio
BSE - Bandwidth Share Estimate
BSS - Basic Service Set
CAPs - Controlled Access Phases
CFP - Contention Free Period
CN - Core Network
CP - Contention Period
CP - Cyclic Prefix
CQI - Channel Quality Information
CRC – Cyclic Redundancy Check
CSMA/CA - Carrier Sense Multiple Access with Collision Avoidance
CW- Contention Window
DCF - Distributed Coordination Function
DFS - Distributed Fair Scheduling
DiffServ - Differentiated Services
DIFS - Distributed Inter-Frame Space
DL - Downlink
DM - Domain Model
DQOAS - Dynamic Quality-Oriented Adaptation Scheme
DS - Distribution System
DSSS - Direct Sequence Spread Spectrum
DWFQ - Distributed Weighted Fair Queue
ECN - Explicit Congestion Notification
EDCA - Enhanced Distributed Channel Access
EM - Experience Model
EPC - Evolved Packet Core
EPS - Evolved Packet System
ESS - Extended Service Set
E-UTRAN - Evolved UMTS Terrestrial Radio Access
FDD - Frequency-Division Duplex
FDM - Frequency Division Multiplexing
FHSS - Frequency Hopping Spread Spectrum
FIFO - First-in, First-out
GOP - Group of Pictures
GSM - Global System for Mobile Communications
HARQ - Hybrid ARQ
HC - Hybrid Coordinator
HCCA - HCF Controlled Channel Access
HCF - Hybrid Coordination Function
HSDPA - High-Speed Downlink Packet Access
HSPA - High-Speed Packet Access

IBSS - Independent Basic Service Set
IETF - Internet Engineering Task Force
IntServ - Integrated Services
IP - Internet protocol
IPv4 - IP version 4
IPv6 - IP version 6
LAN - Local Area Network
LDA - Loss-Delay Adaptation Algorithm
LLC - Logical Link Control
LTE - Long Term Evolution
MAC - Media Access Control
MCS - Modulation and Coding Scheme
MIMO - Multiple Input Multiple Output
MME - Mobility Management Entity
MPDU - MAC Protocol Data Unit
MSDU - MAC Service Data Unit
MT - Maximum Throughput scheduler
MTU - Maximum Transfer Unit
MULTFRC - Multi TFRC
NAS - Non Access Stratum
NAV - Network Allocation Vector
NGN - Next Generation Networks
NS - Network Simulator
ODFM - Orthogonal Frequency Division Multiplexing
OFDMA - Orthogonal Frequency Division Multiple Access
OLSM - Open Loop Spatial Multiplexing
OPEX - Operating expenditure
PAMAH - Performance-Aware Multimedia-based Adaptive Hypermedia
PAPR - Peak-to-Average Power Ratio
PCF - Point Coordination Function
PDCP - Packet Data Convergence Protocol
PDN - Packet Data Network
PDV - Packet Delay Variation
PF - Proportional Fair scheduler
PHY - Physical Layer
PIFS - Priority Inter-Frame Spacing
PM - Performance Model
PRB - Physical Resource Block
QAM- Quadrature Amplitude Modulation
QOAS - Quality-Oriented Adaptation Scheme
QoDGS - Quality of Delivery Grading Scheme
QoE - Quality of Experience
QoS - Quality of Service
QPSK - Quadrature Phase-Shift Keying
RAP - Rate Adaptation Protocol
RE - Resource Element
RLC - Radio Link Control
RR - Round Robin scheduler
RRC - Radio Resource Connection
RSVP - Resource Reservation Protocol
RTCP - Real-Time Transport Control Protocol
RTP - Real-Time Transfer Protocol
RTS/CTS - Request-To-Send/Clear-To-Send
RTT - Round Time Trip
SAS - Server Arbitration Scheme
SC-FDMA - Single Carrier-Frequency Division Multiple Access
SDMA - Spatial Division Multiple Access
SDU - Service Data Unit
S-GW - Serving Gateway
SIFS - Short Inter-Frame Space
SINR - Signal to Interference and Noise Ratio
SISO - Single Input Single Output
SMCC - Streaming Media Congestion Control
SPI - Service Priority Information

List of Tables

List of Figures

1. INTRODUCTION

This chapter introduces the thesis to the reader, presenting the context, the problem statement and the goals of this work. The chapter ends with a brief description of the thesis structure.

1.1 Overview

Multimedia content is being delivered via all types of networks to viewers found in a variety of locations and using different types of devices. Increasing the performance of multimedia stream delivery requires overcoming many technological challenges, all of them having a direct effect on the user perceived quality of experience (QoE). This quality of experience influences in turn the quality of any application. It is important to realize that end-users are becoming increasingly quality-aware in their expectations. The problem of supporting high quality multimedia streaming is even more difficult when delivering multimedia streams over wireless networks, as wireless technologies offer lower bandwidth and the service is highly affected by environmental factors, traffic load and number of clients, as well as their location and mobility patterns.

Applications like live conferencing or e-learning have become important services offered over the Internet, especially in Corporate and Education environments, where many users are accessing the available multimedia content via wireless networks, using mobile devices. Most content is rich media-based and often puts significant pressure on the existing wireless networks in order to support high quality of delivery. In this context, offering a solution for improving user quality of experience when multimedia content is delivered over wireless networks is already a challenging task. Much research work has been done in the area of adaptive streaming-based schemes. Existing adaptive solutions offer a certain level of multimedia quality in variable network delivery conditions - TCP-Friendly Rate Control Protocol (TFRC) [1] and the enhanced Loss-Delay Adaptation Algorithm (LDA+) [2] but they are mainly designed to offer good results when streaming multimedia over wired networks. Using these algorithms over wireless networks returns medium/poor multimedia quality and do not include end-user perceived quality in the adaptation process.

Also, a variety of solutions have been proposed for streaming scalable multimedia content over IEEE 802.11 wireless networks [3] or IEEE 802.11 wireless ad-hoc networks [4]. Among these are adaptive algorithms that operate at the level of layers [4] or objects [5], fine-granular scalability schemes and perception-based approaches [6]. However none of these algorithms has considered in conjunction user preferences and interests as well as network delivery conditions.

1.2 Problems Statement

Consider a typical IEEE 802.11g Wireless Local Area Network (WLAN) with a number of devices attached. Access to the wireless medium being shared equally among these devices, they will compete for receiving a fair share of the available bandwidth. But what if, during an e-learning session for example, the person using the device is not satisfied by the obtained resources? Or what if one gets more physical resources than he requires for a minimum satisfaction level while others are below their acceptance level? The result of fairly sharing the bandwidth among connected devices in this case will lead to unsatisfied users while others may be satisfied but using more resources than their needs.

Figure 1 presents a case where 5 of the connected devices are involved in a streaming session, sharing a bandwidth of 3.2 Mbps. If the resources are divided equally, each user will receive the stream with a 0.64 Mbps transmission rate. Because every user has individual characteristics, one can assume that each has a minimum required bitrate level for the stream, in order to be satisfied with the quality. Taking into account the levels presented in the figure, only 3 users are satisfied, even if the shared bandwidth is theoretically enough to satisfy the needs of all five of them. It is in consequence necessary the presence of a delivery algorithm that considers the users' requirements when sharing the available radio resources among users in order to increase the satisfaction rate on the streaming process.

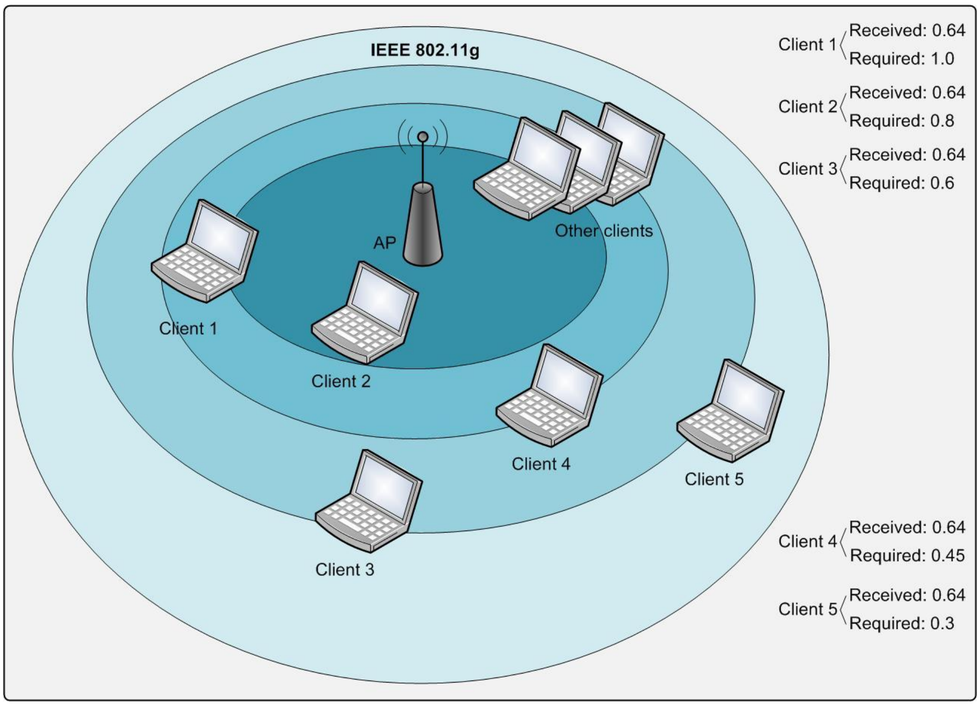


Figure 1 - Hypothetical streaming scenario (the received and required rates are expressed in Mbps)

A solution for this scenario would be to deliver the stream to each user at the minimum required throughput and after that, depending on the available radio resources, to dynamically increase it. If some of the users cannot be satisfied because of the limited bandwidth, the connection with those users will be terminated in order to free the resources for the other users. There will also be users terminating or requiring a connection, so the algorithm should dynamically adapt to any change in the delivery scenario, not only to the current network conditions and to actual user requirements.

For overcoming this problem, the author proposes the Dynamic Quality-Oriented Adaptation Scheme (DQOAS), a user-oriented adaptation mechanism designed to improve the quality of experience while also enabling a higher number of simultaneously connected clients to communicate in an IEEE 802.11 network. DQOAS algorithm enhances the Quality-Oriented Adaptation Scheme (QOAS) [8], [9] by adding different user QoE expectation levels in the adaptation process. Consequently, DQOAS will apply a differentiated treatment for the users, where the adaptive process will be based not only on network conditions but also on user requirements in terms of their QoE. Unlike QOAS which uses static potential bitrate adaptive levels, DQOAS dynamically adapts them to suit the delivery process.

This thesis also addresses a similar problem, but this time considering a Long Term Evolution (LTE) network. It is highly possible for the scenario presented above for IEEE 802.11 WLANs to take place in an LTE network, when the available radio resources are shared among users based just on the network parameters, affecting the perceived quality of the multimedia stream. Like before, DQOAS algorithm will be used in an LTE network to differentiate the users based on their particular characteristics, adapting the multimedia stream bitrate in accordance to their requirements and trying in the same time to improve the efficiency of the delivery.

A development of the problem presented above is when besides a multimedia stream, an application is generating other data flows for every user, each stream with a different traffic class and in consequence with a different priority. In this case, if the received quality for one of the streams is below users' expectations, he can consider the per-application QoE as not satisfactory, regardless of the efforts done by DQOAS or other adaptation mechanism to adapt the media-rich content.

To overcome this problem a new prioritization scheme is proposed and tested, in order to optimize the adaptation algorithm for LTE networks. This new QoS mapping scheme should update the priorities of streams coming from the same application to the highest priority value available (the stream with the highest priority will give the priority class for the entire application). This way, the packets coming from the same application will have the same queuing delay, possibly improving the QoE of the application as a whole. Used in conjunction with DQOAS, this new scheme may offer a mapping alternative for LTE QoS parameters in case on an application that generates a traffic mix (live conferencing, e-learning), in order to obtain an improved end-user QoE.

The research presented in this thesis continues the work performed in [7], enhancing the proposed DQOAS adaptation algorithm when used in LTE networks, by adding BLER and delay levels into the decision process. By adding these two network parameters, the algorithm is able to offer a better response to the variations introduced by the characteristics of the scheduler and the radio channel.
Another implemented improvement that can increase the algorithm's capacity to handle the continuously changing wireless conditions was to increase the original observation window for the interest factors considered in the level change decision performed by DQOAS. By increasing the observation period over a larger number of Transmission Time Intervals (TTIs), the undesired case when the algorithm triggers

a level change due to some short changes in the delivery conditions are avoided, thus reducing the unwanted fluctuations while maintaining a more stable throughput level in time. Another addition to DQOAS algorithm when used in LTE networks is the ability to evaluate and terminate an unsatisfactory multimedia session, based on the satisfaction coefficient variation. The advantages that may arise from this functionality are translated into a more efficient use of resources and a higher number of satisfied end-users.

1.3 Thesis Structure

The research presented in this thesis summarizes the efforts done to design and develop a new multimedia delivery algorithm as part of a bigger project that proposes a new e-learning adaptive system named Performance-Aware Multimedia-based Adaptive Hypermedia (PAMAH). PAMAHs' goal is to optimize users' Quality of Experience and their learning outcomes by automatically adapting content and navigational support based on both user interest and knowledge level and current network delivery conditions.

Chapter 2 gives an overview of the newest telecommunications technology on the market, LTE, detailing the physical and MAC layer processes and presenting the standardized methods for QoS provisioning.

In **chapter 3**, another popular telecommunications technology is described, with focus on the physical and MAC layers. Numerous solutions for QoS provisioning in IEEE 802.11 networks are presented and discussed, highlighting their advantages and disadvantages.

Chapter 4 gives a comprehensive presentation regarding the QoE concept for multimedia traffic and reviews different adaptation methods proposed by the research community, trying to identify solutions for the stated problems.

Chapter 5 introduces a novel solution developed initially for adaptive multimedia streaming over IEEE 802.11 Wireless LANs – Dynamic Quality-Oriented Adaptation Scheme (DQOAS), describing the design and the functioning stages of the proposed adaptation mechanism. The proposed algorithm concentrates on performing an optimal multimedia content management in accordance with end-user profile and preferences, as well as with actual network conditions.

The results of DQOAS delivery algorithm and its user-oriented approach in a IEEE 802.11 wireless LAN environment are presented in **chapter 6**, where different simulation scenarios are detailed and analyzed.

Based on the results obtained in IEEE 802.11 environment, it was of further interest to analyze the behavior of the algorithm over other wireless networks, like Next Generation Networks (NGNs). In this thesis Long Term Evolution (LTE) is the chosen technology as the next wide coverage wireless network. Initial simulation scenarios are ran and discussed and based on the results obtained, a new QoS parameters mapping scheme is used in order for DQOAS to perform optimally when used over LTE networks. The proposed mapping scheme together with the results obtained when DQOAS algorithm is used over LTE are presented in **chapter 7**.

Chapter 8 summarizes the research work and proposes further work in the area.

1.4 Publications

1. V. H. Muntean and G.-M. Muntean, "A novel adaptive multimedia delivery algorithm for increasing user quality of experience during wireless and mobile e-learning", IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, BMSB '09, 2009.
2. V. H. Muntean, M. Otesteanu and G.-M. Muntean, "QoS parameters mapping for the e-learning traffic mix in LTE networks", International Joint Conference on Computational Cybernetics and Technical Informatics (ICCC-CONTI), Timisoara, Romania, May 2010.
3. V. H. Muntean, M. Otesteanu and G.-M. Muntean, "DQOAS Performances for Traffic Mix Delivery over LTE Networks Using a New QoS Parameter Mapping Scheme", Scientific Bulletin of "Politehnica" University of Timisoara, Romania – Transactions on Automatic Control and Computer Science, ISSN 1224-600x, Ed. "Politehnica", Timisoara, Romania, vol.55,No. 3, September 2010, pp 161-170.
4. V. H. Muntean and M. Otesteanu, "WiMAX versus LTE - An overview of technical aspects for next generation networks technologies", 9th International Symposium on Electronics and Telecommunications (ISETC), Timisoara, Romania, November 2010.
5. V. H. Muntean and M. Otesteanu, "Techniques for improving the overall QoE for applications", Workshop "Cercetari doctorale in domeniul tehnic", POSDRU/88/1.5/S/50783, Craiova, Romania, February 2011.
6. V. H. Muntean and M. Otesteanu, "Performance evaluation of DQOAS algorithm in case of applications generating VoIP and video streaming when a new QoS prioritization scheme for LTE is used", IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Nuremberg, Germany, 8-10 June 2011.
7. V. H. Muntean, M. Otesteanu and G.-M. Muntean "QoS-oriented Multimedia Delivery over 4G Wireless Networks: Dynamic Quality-Oriented Adaptation Scheme - a user-oriented adaptation mechanism", LAP LAMBERT Academic Publishing, ISBN 978-3846545911, November 2011.
8. V. H. Muntean and M. Otesteanu, "QoS-oriented Multimedia Delivery Algorithm for Next Generation Wireless Networks", Workshop "Interdisciplinabilitatea si managementul cercetarii – Prezentarea rezultatelor obtinute de doctoranzi", POSDRU/88/1.5/S/50783, Timisoara, Romania, 24-25 November 2011.
9. V. H. Muntean and M. Otesteanu, "QoE-oriented multimedia delivery algorithm for e-learning in next generation wireless networks", Proceedings of the 8th International Scientific Conference "eLearning and Software for Education", Bucharest, Romania, April 26 - 27, 2012.

2. 3GPP LONG TERM EVOLUTION NETWORKS

First section of this chapter is a short introduction into wireless telecommunications, presenting the evolution of these systems, while section 2 describes in detail the architecture, the physical layer and MAC layer of the 3GPP LTE network technology, together with the specific QoS mechanisms.

2.1. Introduction

The evolution of wireless communications systems started long time ago, with the invention of the photophone (Figure 2) in 1880, when Alexander Graham Bell and his assistant, Charles Tainter, performed the world's first wireless telecommunication using modulated light beans. The principles of this invention, named by Bell himself as his "*greatest achievement*" [10], found their practical application one hundred years later, when fiber-optic communications came into use.

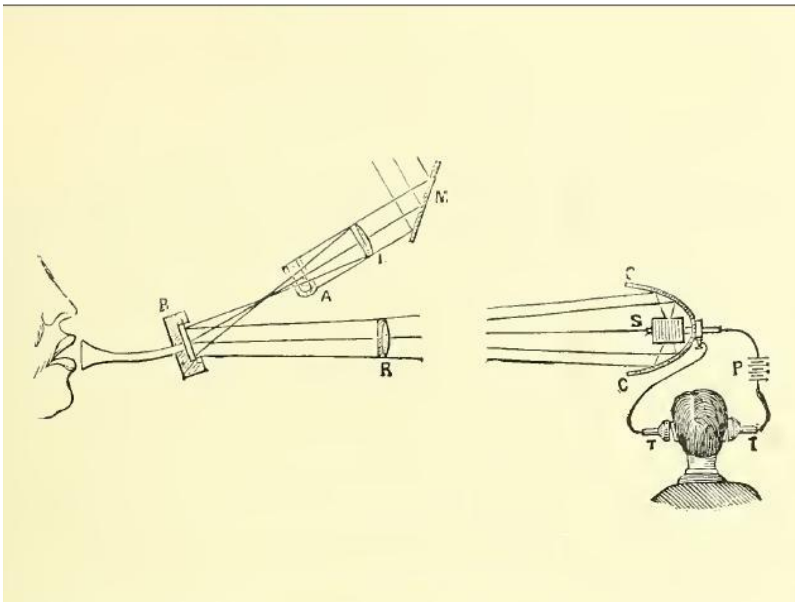


Figure 2 – Bell's Photophone (photo source: [10])

One major step forward in the field of wireless telecommunications happened in 1990, when the specifications for the first common cellular telephone system in Europe, **GSM** (Global System for Mobile Communications) [11], [12],

[13], were published by the European Telecommunications Standards Institute (ETSI). The GSM telecommunications system is still the world's largest mobile communication system and it represents the foundation on which two of the most important mobile data communications systems were developed: General Packet Radio Service (GPRS) and Enhanced Data Rates for GSM Evolution (EDGE). But in the rapid developing world of technologies in the wireless access domain, two of the standard's characteristics make GSM unsuited for data transmission: the small transfer rate (9.6 kbps) and the long time necessary to establish a connection (*n* x seconds). Taking into consideration that GSM uses a circuit switched transmission, the delay and jitter characteristics for a GSM data connection are good, but the bandwidth offered is too small for a realistic utilization, even when High-Speed Circuit-Switched Data (HSCSD) is used (9.6 kbps up to 57.6 kbps).

GPRS is a packet oriented data service for both GSM and TDMA (Time Division Multiple Access) users. The standard that was developed by ETSI, being now maintained by the 3rd Generation Partnership Project (3GPP), uses the standard modulation and frame structure used in GSM and TDMA and it is capable to dynamically reserve the radio resources only when there is available data to be sent. One important aspect that is introduced along with this standard is the possibility to differentiate and classify the traffic using the 3GPP QoS model (Rel 97/98). The proposed 3GPP QoS model contains the so called QoS profiles or QoS attributes that are defined end-to-end and are always associated with a bearer service [14]. The most important 3GPP QoS profiles in data streaming sessions are enumerated below [15]:

- Maximum bit rate (kbps)
- Guaranteed bit rate (kbps)
- Maximum SDU size (octets)
- SDU error ratio
- Transfer delay (ms)
- Traffic class (conversational, streaming, interactive and background)

In order to deliver increased data rates per radio channel compared to a GSM/GPRS connection, a new technology with six additional coding schemes was developed and standardized by 3GPP. Using this technology called **EDGE** [16], higher peak data rates were possible, while maintaining the same QoS profiles used in GPRS. The advantage of this technology is its backwards compatibility with GSM systems, making the GPRS → EDGE upgrade task easier for the mobile operators because no hardware changes were necessary in the network.

With the introduction of Universal Mobile Telecommunications System (**UMTS**), also known as 3G (third generation) mobile cellular system, the theoretical achievable peak data rates are increased to 350 kbps, while the same QoS profiles were maintained. Later on, the necessity for higher bandwidth led to the implementation of HSPA (High Speed Packet Access) and HSPA+ protocols that further improved and extended the existing telecommunications networks. New limits, as high as 168 Mbps, are set for the theoretical achievable bit rate during a downlink data transmission.

According to [17], the technical specifications for UMTS QoS attributes should meet a number of criteria, of which the most important are presented as follows: UMTS QoS mechanisms shall provide a mapping between application requirements and UMTS services, shall be able to interwork efficiently with existing

QoS schemes, shall support efficient resource utilization, shall support asymmetric bearers and shall provide control on a peer to peer basis between UE and 3G gateway node. In order to obtain a desired network QoS, a Bearer Service (BS) with defined characteristics and functionality has to be set between the two network elements involved in the data exchange. The BS includes aspects like control signaling, user plane transport and QoS management functionality in order to be able to provide the desired QoS. As shown in the BS layered architecture depicted in Figure 3, the BS on a specific layer is offering its services to the bearer on the next level while using the services provided by the layer below.

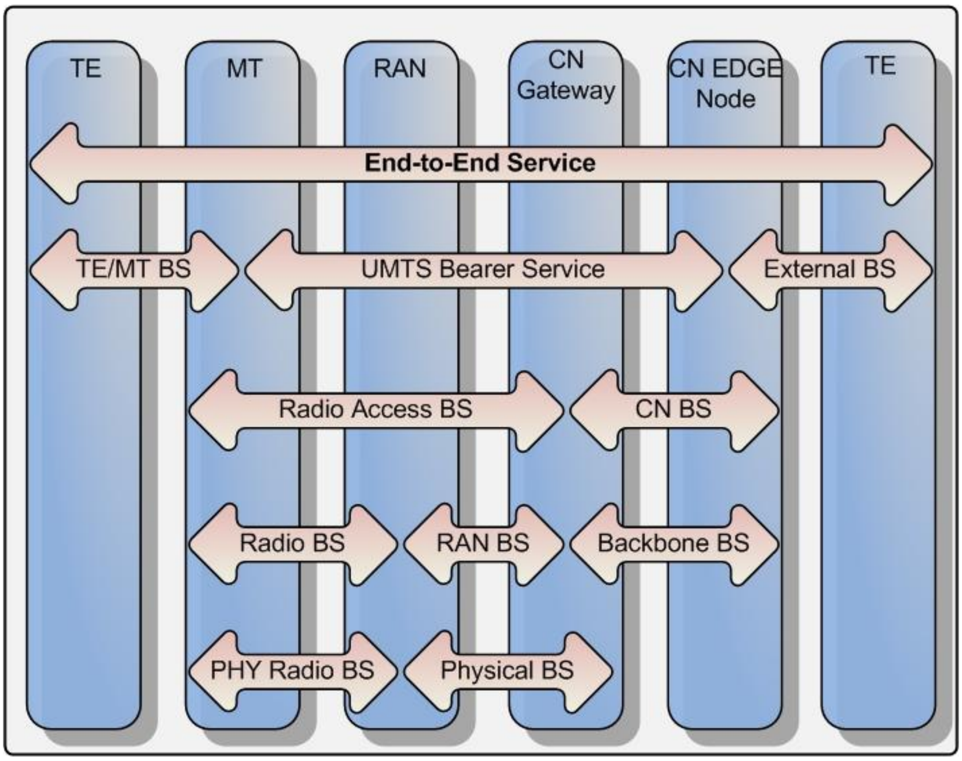


Figure 3 – UMTS QoS Architecture

When two UEs are involved in a data communication, the data flow between them has to pass across different bearer services of the UMTS network. The bearer service that provides in fact the QoS services offered by the operator is the UMTS BS. This bearer service is composed of two parts, the Radio Access BS, which provides confidential transport of user data and signaling, and the Core Network (CN) BS, which has the role to control and utilize the backbone network in order to provide the desired UMTS bearer service. The Radio BS is responsible with the radio interface transport and handles the part of the user flow that belongs to one sub-flow while the Radio Access Network (RAN) bearer service provides the transport between RAN and CN, together with the Physical BS. Radio bearer service and RAN bearer service are composing the Radio Access bearer.

Another family of the 3G standards is **CDMA2000** (Code Division Multiple Access 2000) [18], developed by 3GPP2 as an enhancement to a previous standard, IS-95, also known as CDMAOne. CDMA2000 family has four standards, representing the evolution of this technology: CDMA2000 1xRTT (Radio Transmission Technology), CDMA2000 1xEV-DO (Evolution – Data Optimized) Release 0 with Rev. A and Rev. B, UMB (Ultra Mobile Broadband) and CDMA2000 1xEV-DV (Evolution – Data/Voice).

CDMA2000 1xRTT doubles the voice capacity available in the older IS-95 networks, implementing also protocols for media access and QoS control which were not present before when the data was transmitted using a best effort RLP (Radio Link Protocol) delivery. QoS targets were not explicitly set and traffic classes like in UMTS were not used or defined but CDMA2000 is able to support the same applications and services [19], offering acceptable quality levels by using the radio link characteristics and the Mobile IP capabilities. For circuit switched applications, the SIP (Session Initiation Protocol) was used to transmit QoS requirements like service description: the initial request, bandwidth request, assured or non-assured mode, etc).

The technology uses a single pair of 1.25 MHz radio channels and is able to achieve a theoretical packet data speed of 153 kbps. But as the requirements for higher data rates grew, new technologies had to be developed to keep up with the evolution of the handheld devices. CDMA2000 1xEV-DO Rev. A was a response to those requests, being able to achieve peak data rates of 3.1 Mbps in downlink (DL) and 1.8 Mbps in uplink (UL), while Rev. B further increased the data speeds, while reducing the latency and the interference on the adjacent sectors [18]. The CDMA2000 1xEV-DV standard offered the same peak data rates speeds as CDMA2000 1xEV-DO but additionally it was able to simultaneously support on the same radio channel users from 1xRTT, 1xEv-DO and 1xEV-DV systems. Because of the lack of interest showed by the mobile operators on this technology, its development was suspended.

The next step for the CDMA2000 family was planned to be the introduction of the Ultra Mobile Broadband technology, as the next generation network (NGN), able to offer peak data rates up to 288 Mbps [20]. The development of this new technology was stopped in 2008 by the 3GPP2, deciding to favor the other NGN technology that was developed by 3GPP, LTE (Long Term Evolution), which had the possibility to perform seamless handovers to older networks like 1xRTT or 1xEV-DO. One major drawback of CDMA2000 family was that even if this standard used the same air interface as UMTS, W-CDMA (Wideband CDMA), the two standards are not compatible.

In parallel with the latest technologies developed by 3GPP and 3GPP2 to increase the data rates in wireless communications systems, a newly formed organization developed and proposed an alternative, able to provide broadband wireless access on a large scale coverage. This new technology was called **WiMAX** (Worldwide Interoperability for Microwave Access) and it was designed as an enhancement to the IEEE 802.11 WLAN (Wireless Local Area Network) by extending the wireless access to Wide Area Networks (WAN) and Metropolitan Area Networks (MAN). The initial version of WiMAX, IEEE 802.16-2004, was designed to provide broadband wireless connectivity to fixed and nomadic users only for the last mile. The coverage could go up to 50 km, allowing users to get broadband connectivity in NLOS (Non/Near Line Of Sight) conditions. The IEEE 802.16-2005 (Mobile WiMAX) standard comes with enhanced QoS and mobility up to 120 km/h, being specifically designed to fill the gap between wireless local area networks and high mobility cellular wide area networks. In order to obtain downlink peak data rates of up to 75 Mbps in mobility scenarios, the standard uses scalable OFDMA (Orthogonal Frequency Division Multiple Access) to dynamically modify FFT (Fast Fourier Transform) size, depending on the channel conditions [21].

Figure 4 illustrates the Network Reference Model (NRM) which in WiMAX includes the following logical entities: Subscriber Station (SS), Access Service Network (ASN) and Connectivity Service Network (CSN) together with the reference points for interconnecting the logical entities (R1-R5) [22].

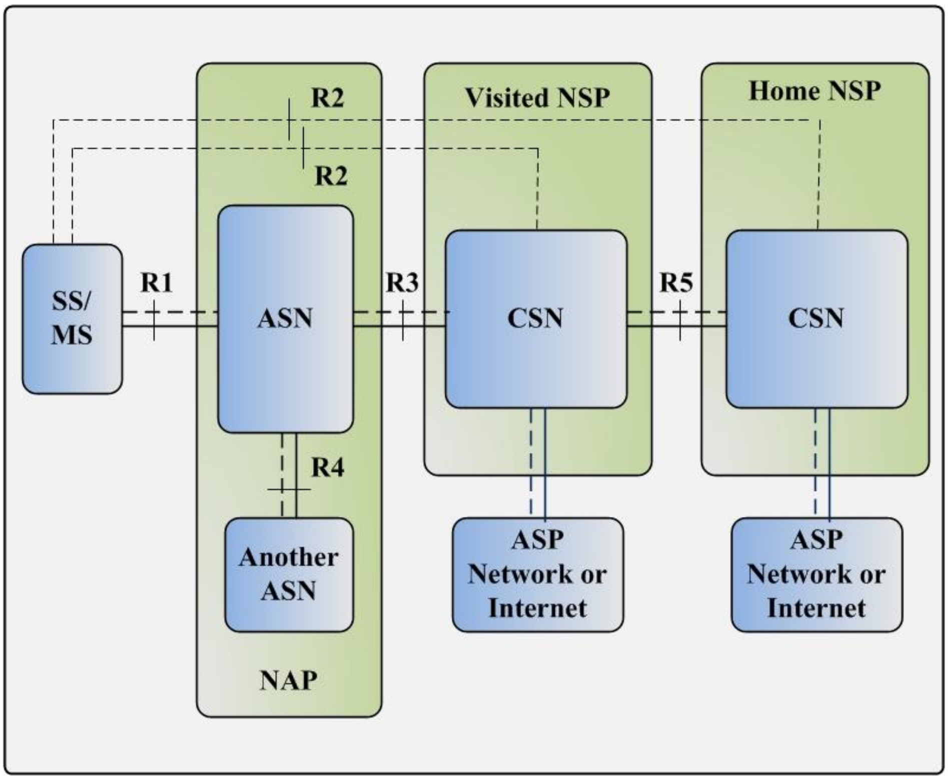


Figure 4 - Network Reference Model for WiMAX

The ASN holds one or more ASN Gateways (ASN GWs) and one or more Base Stations (BSs). The BS provides and manages resources over the air interface and is responsible for handover triggering while ASN-GW performs AAA (Authentication, Authorization and Accounting) client functionality, establishing and managing mobility tunnel with BSs and connections towards selected Connectivity Service Network (CSN) [21]. The CSN is defined as a set of network functions that provide IP connectivity to the WiMAX subscribers. The typical CSN comprises AAA proxy/servers, user databases, routers and Interworking gateway devices and is responsible of IP address Management, mobility, roaming and location management between ASN's and roaming between NSPs (Network Service Providers) by Inter-CSN tunneling.

WiMAX uses both license-exempt and licensed frequency bands, which are summarized in Table 1.

Table 1 - Reported frequency bands used for WiMAX

Regions	***Frequency Bands for WiMAX***	
	License Bands [GHz]	License-Exempt Bands [GHz]
USA	2.3 and 2.5	5.8
Europe	3.5 and 2.5	5.8
South East Asia	2.3, 2.5, 3.3 and 3.5	5.8
Middle East	3.5	5.8
Africa	3.5	5.8
South Central America	2.5 and 3.5	5.8

Peak data rates for WiMAX, 75 Mbps in DL and 25 Mbps in UL, are closely related to multiple antenna configurations (AAS-Adaptive Antenna Systems) and modulation schemes used (AMC-Adaptive Modulation & Coding).

In order to increase the capacity or to provide spatial diversity, Multiple antenna systems are being considered in all next generation cellular standards, including WiMAX: MIMO (Multiple Output Multiple Input), both Spatial Multiplexing (SM) and Space-Time/Frequency Block Coding (STBC/SFBC), and Beam-forming (BF). The key benefits of multiple antenna systems are the increased capacity, diversity, data rates and efficiency when compared to single antenna systems. In WiMAX, three types of multiple antenna techniques are utilized: SAS (Smart Antenna Systems), diversity techniques and MIMO which are further subdivided into open loop and closed loop power control systems.

WiMAX layered architecture described by IEEE 802.16e is composed of two layers: physical (PHY) and media access control (MAC). Every layer relies on the services provided by the layer below and can access these services by the means of a virtual interface, called Service Application Point (SAP). The 802.16 specifications subdivide the MAC layer into three sub-layers [23]: the convergence sub-layer (CS), the common part sub-layer (CPS) and the privacy sub-layer, as illustrated in Figure 5.

The MAC layer is responsible for assembling upper layer data into frames along with error detection and also attaches/detaches addresses to the fields upon transmission/ reception. The CS role is to take IP or ATM packets from upper layers through CS SAP and to classify then, since WiMAX supports two types of transmission modes: ATM and IEEE 802.3 (Ethernet). It also performs additional processing including frame compression, addressing frames according to IEEE 802.16, sending MSDUs (MAC Service Data Units) to CPS. CPS is the central part of MAC layer, defining the medium access method. It performs functions related to channel access, QoS requirements and connection establishment, and also organizing MSDUs from MAC SAP in MAC PDUs (Protocol Data Units), by doing segmentation and fragmentation. The privacy sub-layer has been designed to provide subscribers privacy across wireless network and strong protection for operators against theft of service. The physical layer takes MPDUs from PHY SAP and converts them into signals to be transmitted across the air interface [24].

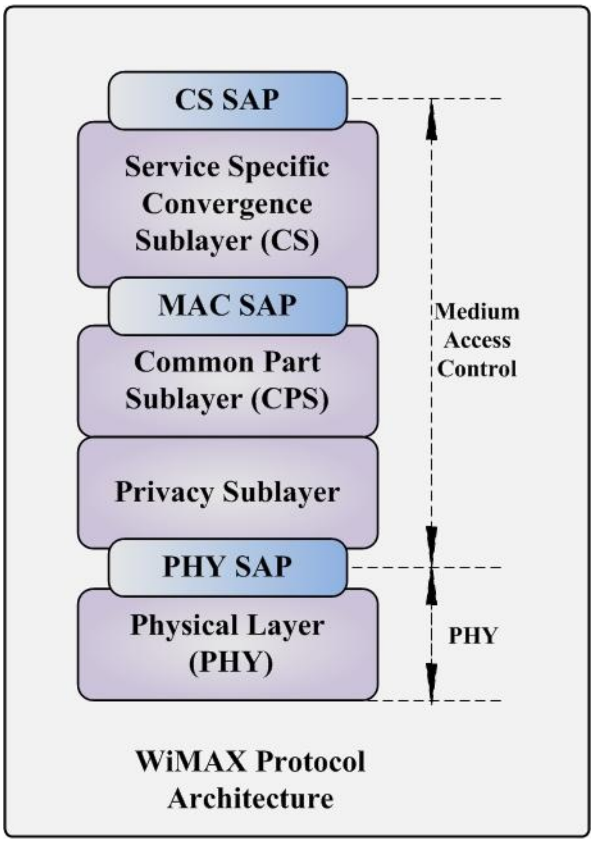


Figure 5 – WiMAX and LTE protocol architectures

The Service Flow (SF) concept in WiMAX is a MAC layer connection that supports connection oriented QoS mechanisms which enable end-to-end QoS control. The QoS parameters are negotiated dynamically or statically through MAC messages and provide scheduling and transmission ordering over the air interface. WiMAX QoS mechanism supports various SF types: rtPS (Real Time Polling Service), UGS (Unsolicited Grant Service), ErtPS (Extended Real Time Polling Service), nrtPS (Non Real Time Polling Service) or BE (Best Effort Service) [25].

In WiMAX, a service is associated to every application that runs over the network, by using an SF identifier (SFID) and all data packets transmitted by that application through the network will be marked with this specific SFID in order to assure a certain QoS level. Besides the SFID, the data packet generated by the application will also be identified by a connection identifier (CID), which is assigned by the network when the application establishes a connection. Using these two identifiers, SFID and CID, the data packets can be sorted and placed in the correct queues (Figure 6).

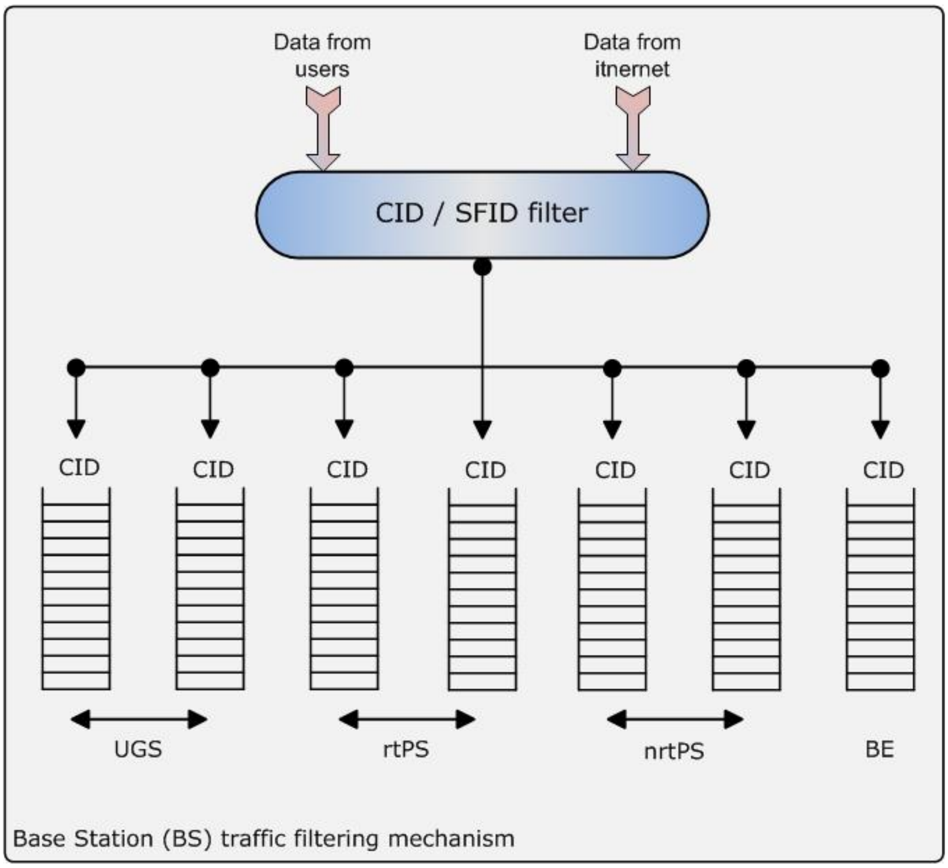


Figure 6 –Traffic flow classification in WiMAX

During the last years, the evolution of applications was extremely rapid, but most importantly, it was connected with that of the computing devices on which they run. In these conditions, IP services provision anytime and anywhere became very challenging and it is perceived by the mobile operators as a great opportunity for boosting the average revenue per unit. But the further success of IP services deployment requires true mobile broadband IP connectivity on a global scale. For accomplishing this request, two technologies stand out with the aim of providing voice, data, video and multimedia services on mobile devices at high speeds and cheap rates: WiMAX and LTE (3GPP Long Term Evolution). The economical aspects regarding these two technologies are comparable, considering that both are still under development [26], [27] but based on the industrial trend and on some major advantages of the LTE technology [28], like backwards compatibility with other 3GPP standards, the author of this thesis considers that LTE has the edge and that it will become the technology of choice for the majority of mobile network operators.

2.2 LTE Networks

LTE technology evolved from UMTS/HSDPA cellular technology to meet current used demands of high data rates and increased mobility. The LTE radio access is based on OFDM technique and supports different carrier frequency bandwidths (1.4-20 MHz) in both frequency-division duplex (FDD) and time-division duplex (TDD) modes [29]. The use of SC-FDMA in the uplink reduces Peak-to-Average Power Ratio (PAPR) compared to OFDMA, increasing the battery life and the usage time on the UEs. In downlink, peak data rates go from 100 Mbps to 326.4 Mbps, depending on the modulation type and antenna configuration used. LTE aims at providing IP backbone services, flexible spectrum, lower power consumption and simple network architecture with open interfaces.

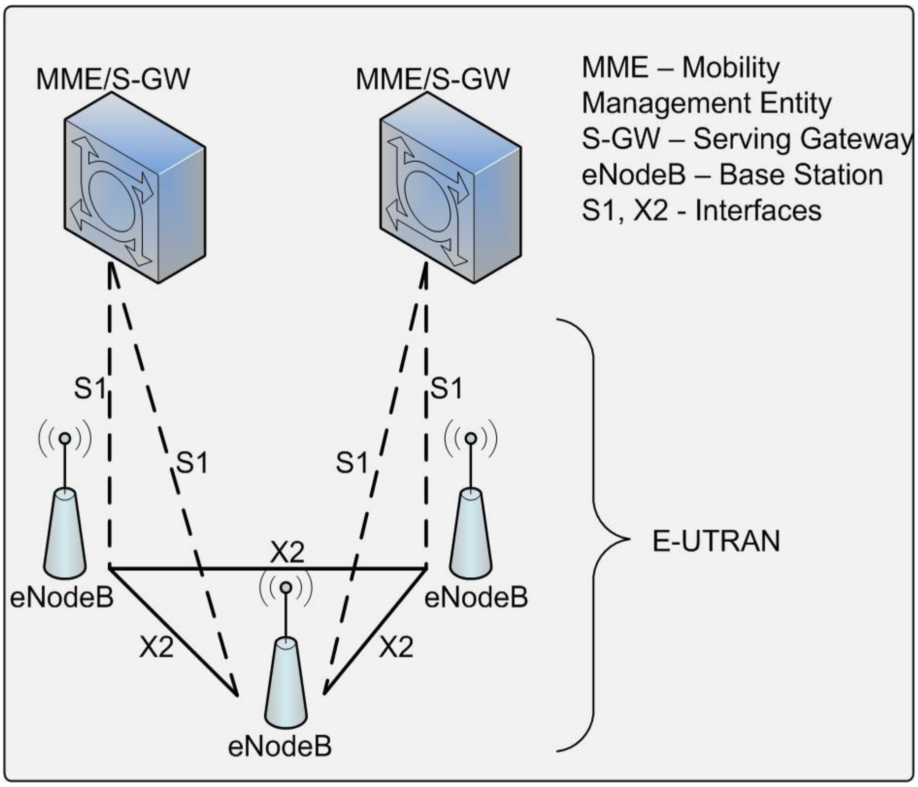


Figure 7 - LTE Overall Architecture

The LTE architecture can be seen as a two-node architecture because only two nodes are involved in the communication between the user equipment and the core network. These two nodes are the base station (eNodeB) and the serving gateway (S-GW) in the user plane (U-plane) and the mobility management entity (MME) in the control plane (C-plane), respectively [30]. Through this architecture, LTE offers the operators the possibility for gradual deployment over their existing networks, assuring service continuity by allowing handover to and from any older 3GPP or 3GPP2 network [26].

LTE architecture is composed of Core Network (CN) and Access Network (AN), where CN corresponds to the Evolved Packet Core (EPC) and AN refers to E-UTRAN. The CN and AN together correspond to Evolved Packet System (EPS). EPS connects the users to Packet Data Network (PDN) by IP address in order to access the internet and services like Voice over IP (VoIP).

The overall network architecture is shown in Figure 7 [31], [32].

MME is the control plane entity within EPS supporting the following functions: inter CN node signaling for mobility between 3GPP access networks, S-GW selection, roaming, authentication, bearer management functions and NAS (Non Access Stratum) signaling. Serving Gateway is the gateway which terminates the interface towards E-UTRAN. For each user associated with the EPS, at a given point in time, there is a single Serving GW that is responsible for transferring user IP packets, lawful interception and mobility anchor for inter-eNodeB handover and for inter-3GPP mobility.

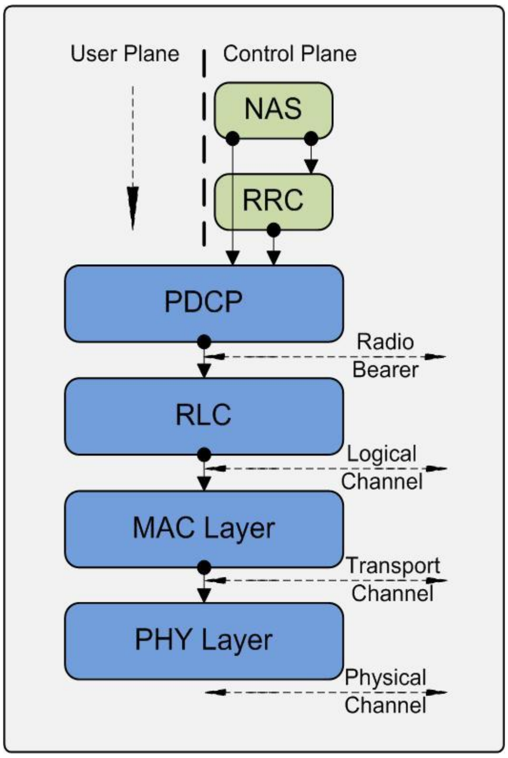


Figure 8 - LTE protocol architecture

LTE protocol architecture depicted in Figure 8 is similar to WiMAX architecture, except it uses the first three layers of OSI model. Even if both technologies rely on OFDM modulation, allowing them to support very high peak rates and their performances are comparable, the architecture differs and the question arising from this difference is what technology should be used by the operators for upgrade based on CAPEX (Capital expenditures) and OPEX (Operating expenditure). LTE architecture was designed in such way that the operators interested in it, will be able to deploy it over their existing infrastructure with a minimum of changes and investments, and this may qualify it as the first choice based on deploying and day-to-day costs.

The radio link specific protocols, including radio link control (RLC) and medium access control (MAC) protocols are terminated in the eNodeB. The packet data convergence protocol (PDCP) layer, responsible of IP header compression along with ciphering, is also located in the eNodeB. The radio resource connection (RRC) is the highest protocol in the control plane on the radio side allowing the exchange of signaling messages between eNodeB and UE, and to forward signaling messages coming from the core network, called NAS (Non-Access Stratum) messages.The NAS layer performs authentication, security control and idle mode mobility and paging.

The data arriving at PDCP is grouped in the sa called PDCP SDUs, to which PDCP attaches header information to form PDCP PDUs which are sent to the RLC for further processing. The RLC performs the transfer of data towards MAC layer in 3 modes: transparent mode (real-time services), unacknowledged mode (signaling) and acknowledged mode (non real-time services). The MAC layer is converting MSDUs to MPDUs by adding the MAC header and is mapping logical channels to transport channels. The MPDUs are then sent to the physical layer which performs encoding/decoding of data and organizes the MPDUs in transport blocks [33].

2.2.1 LTE Physical Layer

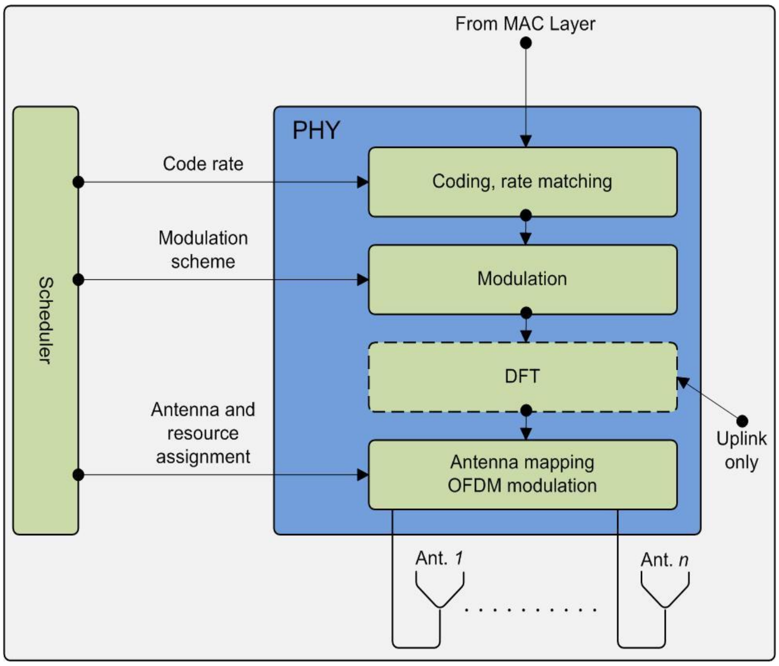


Figure 9 - Components of the LTE Physical layer

The main objectives for the LTE physical layer (Figure 9) are to increase peak data rates up to about 360Mb/s in DL and 75 Mb/s in UL within a 20 MHz spectrum and a 4x4 MIMO configuration, to increase cell edge bit rates, to reduce user and control plane latency to less than 5 ms [34], to support interactive real-time services such as video/audio conferencing or multiplayer gaming, and to provide mobility for speeds up to 350 km/h.

The physical layer is able to facilitate the technology to coexist with other standards by implementing a scalable bandwidth (1.25/2.5/5/10/20MHz). Also the spectral efficiency is improved, allowing the operators to increase number of accommodated customers within their spectrum allocation, with a reduced cost of delivery per bit and an improved cost and power consumption.

The fulfillment of all LTE objectives presented above determined the choice of air interface technology. The spectrum requirements, data rates and performances can be achieved using a multiple access technology called Orthogonal Frequency Division Multiplexing (OFDM) for the downlink. For the uplink the decision was to use Single Carrier-Frequency Division Multiple Access (SC-FDMA) with dynamic bandwidth to reduce power consumption for the user terminal.

2.2.1.1 OFDM In LTE

For LTE, OFDM divides the available frequency bandwidth into small subcarriers spaced at 15 kHz, modulating every subcarrier using one of the available modulation formats: QPSK, 16-QAM, or 64-QAM. OFDMA is used to allocate each user the required bandwidth for their transmission, excluding the unassigned subcarriers, while also reducing the interference and the power needed to complete the data exchange. In Figure 10 can be observed that in case OFDM is used for a data transmission, the entire bandwidth is used by a single user for a defined period. In case OFDMA is employed, multiple users are able to share the bandwidth width at any point in time:

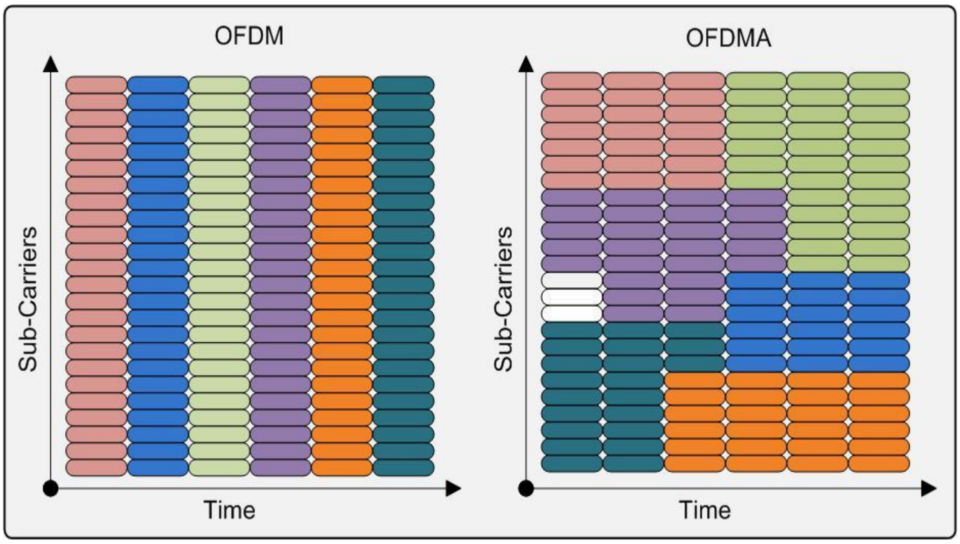


Figure 10 - OFDM vs. OFDMA

For the UL channel LTE uses a version of OFDM, called SC-FDMA, where the data is spread across multiple subcarriers. The main advantage of SC-FDMA is that it has a low PAPR compared to OFDM, thus reducing the battery power consumption. This digital modulation scheme also requires a simpler amplifier design, improving the cell-edge performance.

The OFDMA parameters used in LTE are presented in Table 2. For a 5 MHz bandwidth, there are 512 subcarriers of15 kHz each, whereas the total band is 7.68 MHz, which is larger than the 5 MHz band. But only 301 subcarriers are used (pilot, DC and data), the other are just used as guard subcarriers. In this case there are 301 subcarriers of 15 kHz each, summing to a 4.515 MHz band (<5MHz).

Table 2 - OFDMA parameters for LTE

Spectrum allocation	***1.4 MHz***	***3 MHz***	***5 MHz***	***10 MHz***	***15 MHz***	***20 MHz***
Subcarrier spacing	15 kHz					
Sampling frequency [MHz]	1.92 (1/2x3.84)	3.84	7.68 (2x3.84)	15.36 (4x3.84)	23.04 (6x3.84)	30.72 (8x3.84)
Number of subcarriers	128	256	512	1024	1536	2048
Number of useful subcarriers	75 (76)	150 (151)	300 (301)	600 (601)	900 (901)	1200 (1201)

The OFDM symbol useful duration depends on the subcarrier bandwidth and it is calculated using the formula below:

$$T_{sym} = {}^{1}/_{subcarrier\ BW} = {}^{1}/_{15\ KHz} = 66.6\ \mu s$$

To avoid ISI, a guard interval is inserted between two consecutive symbols. The guard interval is then filled with the Cyclic Prefix (CP) which is a copy of fixed number of the last samples at the start of the symbol. The time structure of an OFDM symbol is presented in Figure 11.

In LTE, 2 CPs are defined:

- Long Cyclic Prefix: 16,67 μs
- Short Cyclic Prefix: 4.69 μs

In case a Long CP is used, the total duration of an OFDM symbol will be:

$$T_{sym} + CP = 66.6\ \mu s + 16.67\ \mu s = 83.33\ \mu s$$

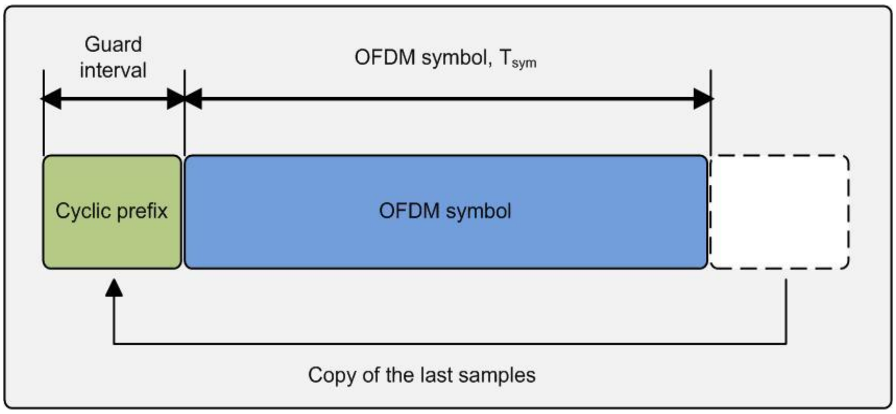


Figure 11 - OFDM symbol time structure

2.2.1.2 Physical Resource Structure

Physical resources in the radio interface are organized into radio frames. Two radio frame types are supported: Type 1, used in Frequency Division Duplexing (FDD) and Type 2 used in Time Division Duplexing (TDD). Type 1 frame is 10 ms long and consists of 10 consecutive sub-frames of length $T_{sub\text{-}frame}$ = 1ms. Each sub-frame divides into 2 slots of 0.5ms long. Figure 12 depicts the Type 1 radio frame structure:

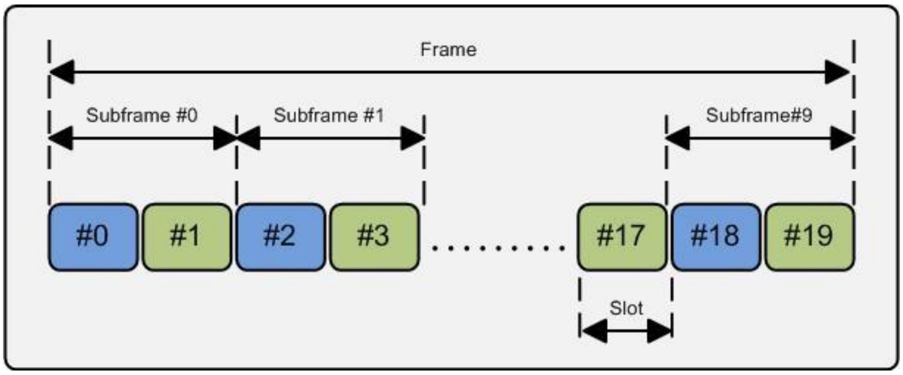


Figure 12 - Type 1 radio frame structure

In LTE, the smallest modulation unit is the Resource Element (one 15 kHz subcarrier by one symbol), but the minimum resource unit a scheduler can allocate to a user is called Physical Resource Block (PRB), which corresponds to 12 consecutive subcarriers by six or seven symbols (depending on the length of the CP used – long/short) (Figure 13).

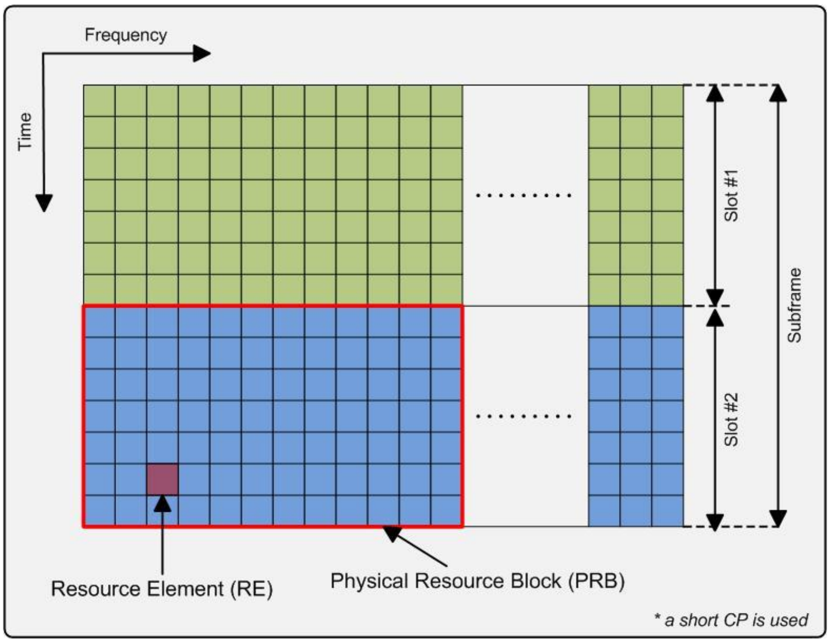


Figure 13 – Sub-frame structure

2.2.1.3 Adaptive Modulation And Coding (AMC)

Adaptive Modulation and Coding refers to the ability of the network to dynamically select the appropriate modulation type and coding rate based on the current channel conditions reported by the User Equipment (UE), in order to improve the QoS of the delivered data. In LTE, the modulation type can be one of the following: QPSK, 16-QAM and 64-QAM. If QPSK is used, each of the four symbols carries 2 bits of information. In 16-QAM there are 4 bits of information per symbol while in 64-QAM, one symbol has 6 bits of information. 16-QAM and 64-QAM modulations are very sensitive to poor channel conditions compared to QPSK because of the small differences between symbols in the constellation (Figure 14).

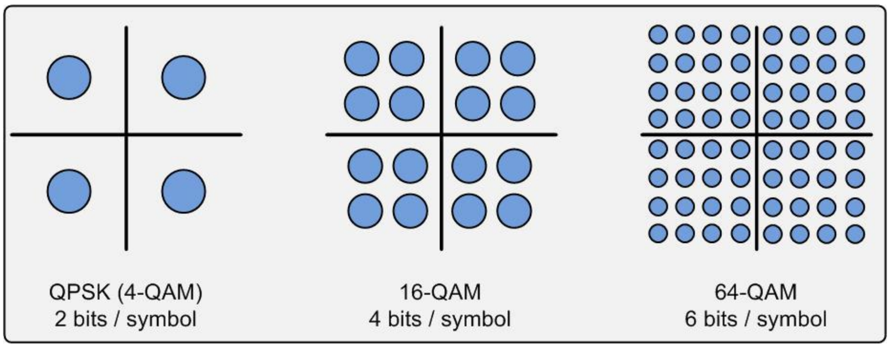


Figure 14 - Ideal constellations for the modulation types available in LTE

The coding scheme is used to determine the amount of redundant bits that are added to the useful data for increasing the reliability of the transmission. The coding rate *R* is the ratio between the useful data length and the transmitted (coded) data. Considering this, if the link quality is good there will be less redundancy compared to a bad quality link. Also, in good RF conditions, a high-order modulation scheme can be used, increasing the number of bits per symbol that can be delivered.

2.2.1.4 Multiple Antenna Systems

Multiple antenna systems are being considered in all next generation cellular standards, including LTE and WiMAX, for their ability to increase capacity or to provide spatial diversity. The key benefits of multiple antenna systems are the increased capacity, diversity, data rates and efficiency when compared to single antenna systems. LTE uses multiple antenna techniques and wider spectrum to provide data rates in the entire cell coverage area. The advanced antenna techniques used by LTE are Beam-forming, Spatial Division Multiple Access (SDMA) and MIMO. The antenna configuration supported by LTE downlink is (2x2) and (4x4) having 2 or 4 antennas at eNodeB and 2 or 4 antennas at UE. The uplink of LTE supports 2x2 MIMO having 2 antennas at UE as well as at eNodeB. Table 3 describes the MIMO antenna configurations used by LTE.

Table 3 - Multiple antenna schemes in LTE

Tx data streams	***Multiple antenna scheme***	***Gain***	***Benefits***
One	Transmit diversity	Diversity gain	Link robustness Coverage
	Beam forming	Power gain	Coverage Capacity
Multiple	Spatial multiplexing	Capacity gain	Spectral efficiency Data rates

2.2.1.5 Physical Channels And Physical Signals

The physical layer in LTE uses physical channels and physical signals. The physical channels are physical resources that carry data or information from the MAC layer. The physical signals are also physical resources that supports the functions of the physical layer, but do not carry any information from the MAC layer.

For the downlink, we have the following physical channels and signals:

- Physical channels
  - Physical Downlink Shared Channel (PDSCH) – user data from MAC
  - Physical Broadcast Channel (PBCH) – broadcast data from MAC
  - Physical Multicast Channel (PMCH) – multicast data from MAC

- Physical Downlink Control Channel (PDCCH) – signaling for PDSCH and PUSCH
- Physical Control Format Indicator Channel (PCFICH) – indicates the number of PFDM symbols used for control signaling in the current subframe
- Physical Hybrid ARQ Indicator Channel (PHICH) – transmits acknowledgements for uplink data

➢ Physical signals
- Reference signals to support coherent demodulation in downlink
- Synchronization signals to be used in cell-search procedure

Table 4 - Modulation schemes for LTE UL and DL physical channels

Direction	***Channel***	***Modulation scheme***
Downlink	PDSCH	QPSK, 16-QAM, 64-QAM
	PMCH	QPSK, 16-QAM, 64-QAM
	PBCH	QPSK
	PCFICH	QPSK
	PDCCH	QPSK
	PHICH	BPSK
Uplink	PUSCH	QPSK, 16-QAM, 64-QAM
	PUCCH	BPSK, QPSK
	PRACH	*u*th Root Zadoff-Chu

For the uplink, the following physical channels and signals are used:

➢ Physical channels
- Physical Uplink Shared Channel (PUSCH) – user data from MAC
- Physical Random Access Channel (PRACH) – transmits information necessary to obtain scheduling grants and timing synchronization for asynchronous random access
- Physical Uplink Control Channel (PUCCH) – sends downlink CQI information to the eNodeB, ACK/NACK for downlink transmissions and scheduling requests

➢ Physical signals
- Reference signals to support coherent demodulation in uplink
- Reference signals for uplink channel sounding – to obtain channel quality for the entire bandwidth for each user

Table 4 summarizes the modulation schemes allowed for LTE uplink and downlink [30].

2.2.2 Media Access Control (MAC) Sub-Layer

The medium access control (MAC) sub-layer [35] provides hybrid ARQ and is responsible for the functionality that is required for medium access, such as scheduling operation and random access.

Figure 15 presents the components of MAC layer:

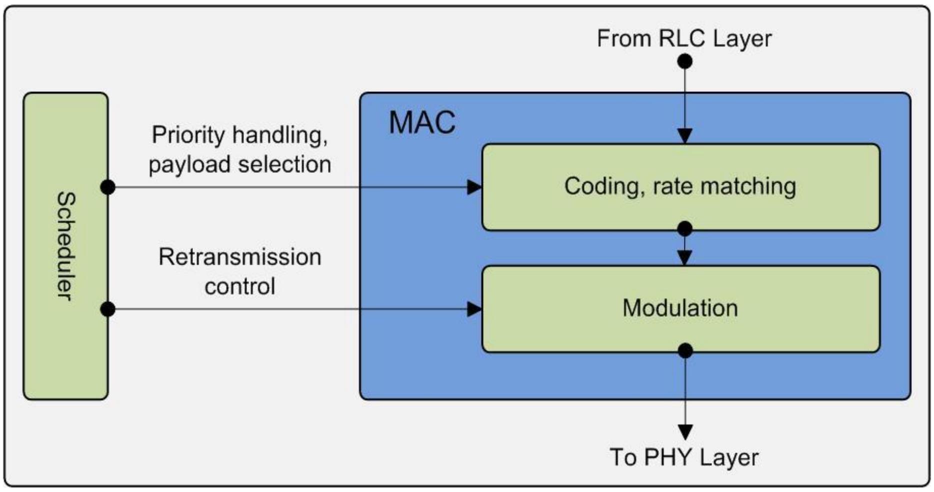


Figure 15 - Components of the LTE MAC layer

Like in any wireless communication system, occasional transmission errors are expected to occur due to channel noise, channel fading, interference etc. LTE was design to stop the error propagation to higher layers, instead to perform a retransmission of the corrupted data block if the service has a high sensitivity to packet loss or to drop the erroneous blocks if the service QoS permits it.

2.2.2.1 Hybrid ARQ

MAC sub-layer is responsible of retransmitting the corrupted transport blocks in order to correct most of the transmission errors. This is done using the hybrid ARQ mechanism, which is similar to the solution implemented for HSDPA [36]. The protocol uses multiple stop-and-wait HARQ processes, the functionality being comparable to that of a window-based selective repeat protocol (it allows continuous transmission which cannot be achieved using a single stop-and-wait phase). Even if the complexity in implementation is higher for HARQ than for the traditional ARQ strategy, HARQ gains in terms of simplicity, control overhead and delay by using a single-bit HARQ feedback ACK/NACK with a well defined relation in time with the transmitted data. Two types of retransmission strategies are implemented:

- Chase combining (the retransmission is identical with the original transmission)

- Incremental redundancy (only a code word is sent, containing systematic and parity bits used for error correction)

In case of chase combining, the integrity of a transport block is checked by calculating the CRC (cyclic redundancy check) and comparing it to the CRC sequence. If there are any decoding errors, a retransmission request is generated and the transmitter will send the same transport block. The second transport block is stored and its integrity is again checked using the CRC sequence. If there are errors, then the combining process uses the two received transport block affected by errors to increase the probability of successful decoding. Figure 16 presents the chase combining HARQ.

If incremental redundancy HARQ is used in case a transport block is erroneous, a transmission request is sent to the transmitter. The transmitter will supply to the decoder with additional parity bits which are combined with the original transport block. This process is repeated until the block is correctly decoded or until the retransmission limit is reached. This type of HARQ has better performances than chase combining but it requires a larger buffer size and the implementation is more complex.

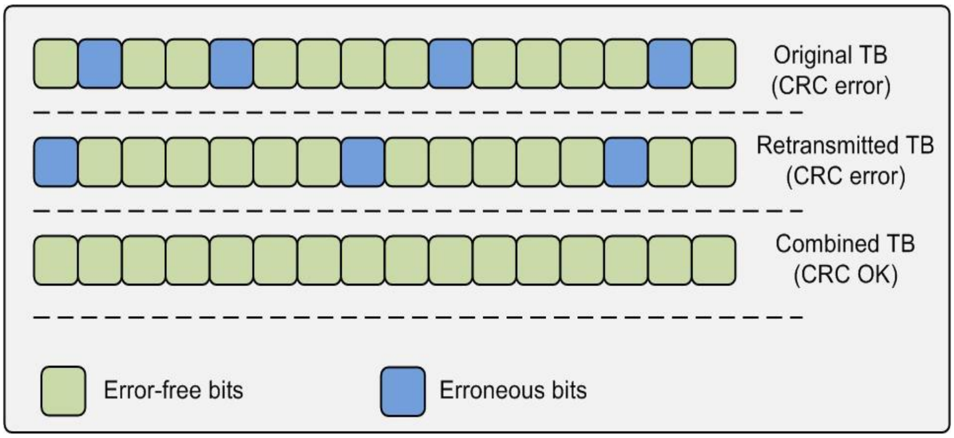


Figure 16 - Chase combining HARQ

2.2.2.2 MAC Scheduler

The scheduler, located in the eNodeB, determines dynamically each 1 ms (each TTI or Sub-frame), which UEs are scheduled to transmit/receive data on the Uplink/Downlink shared channel and also what resources should be used. If one considers that different data flows have different QoS requirements, these requirements cannot be guaranteed using a round-robin scheduling policy because each user has unique network conditions and the average link capacity is not the same. Hence, in order to adequately allocate radio resources to LTE users, advanced scheduling algorithms had to be developed that take into account important aspects like QoS requirements, instantaneous channel conditions, UE capabilities, pending retransmissions etc.

In order to select the suited adaptive modulation and coding scheme, the scheduler needs measurement reports in both downlink and uplink. In the uplink the eNodeB is able to measure the signal quality, because the UEs transmit their data towards eNodeB. Using this measurements, the eNodeB is able to select the modulation and coding scheme and the transmit power. In the downlink, some feedback is required from the UEs so that eNodeB can adapt to the channels' conditions. UEs are reporting periodically the measurements reflecting the instantaneous channel quality of a group of resource blocks. These measurements are built into indicators called Channel Quality Indicators (CQI) and can be used by the eNodeB for the following purposes:

- Selection of modulation and coding scheme
- Time/frequency selective scheduling
- Interference management
- Transmission power control for physical channels.

Table 5 presents the mapping between available CQI index values and the modulation scheme used, as defined in [37]:

Table 5 - CQI index - Modulation scheme relations for DL channels

CQI Index	***Modulation scheme***	***Coding rate x 1024***	***Efficiency****
0	not used		
1	QPSK	78	0.1523
2	QPSK	120	0.2344
3	QPSK	193	0.3770
4	QPSK	308	0.6016
5	QPSK	449	0.8770
6	QPSK	602	1.1758
7	16QAM	378	1.4766
8	16QAM	490	1.9141
9	16QAM	615	2.4063
10	64QAM	466	2.7305
11	64QAM	567	3.3223
12	64QAM	666	3.9023
13	64QAM	772	4.5234
14	64QAM	873	5.1152
15	64QAM	948	5.5547

** Efficiency = coding rate x nb of bits per symbol*

Figure 17 and Figure 18 are describing the simplified reporting mechanisms used in uplink and downlink.

CQI reports are considered to be sufficient feedback for data transmissions where only one antenna is used. In case advanced antenna techniques are utilized, in order to support them the scheduler must consider other two parameters, PMI (Precoding Matrix Indication) and RI (Rank Indicator).

PMI is an array that describes how a UE (User Equipment) would prefer to receive a certain modulated symbol that is transmitted on multiple antennas from the eNB (e.g. it can specify by how many degrees the symbol transmitted on antenna 2 should be out of phase compared to the symbol from antenna 1). [30]

The RI refers to the number of antennas a UE can be able to discern in an eNB transmission. If a UE is close to the eNB, it will probably be capable to distinguish all four of them and it will report this. The eNB will then place different versions of the modulation symbol on all four antennas, with consideration to the feedback received in the PMI report relative to their phase.

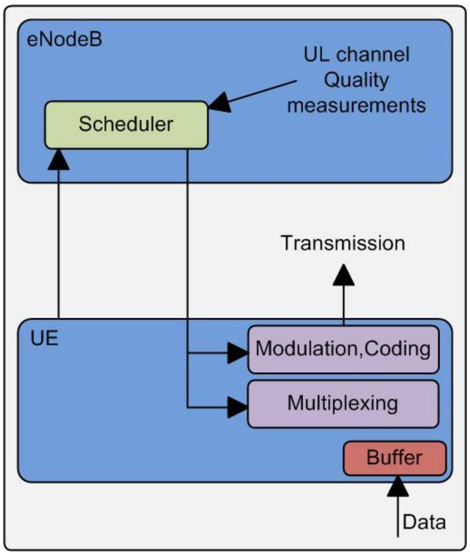


Figure 17 - UL reporting mechanism

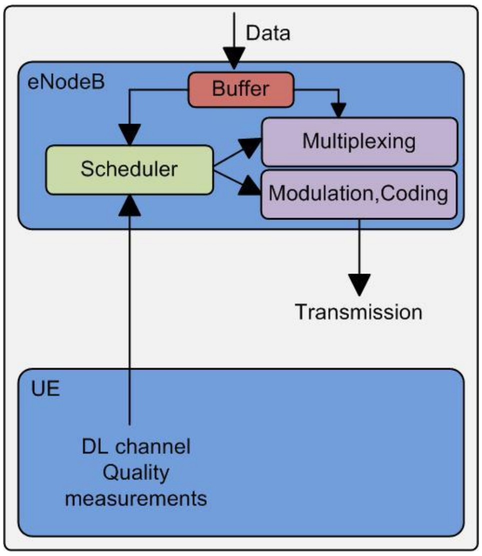


Figure 18 - DL reporting mechanism

Another important aspect that the eNB scheduler is considering is the UE category, which describes the capabilities of the equipment (Table 6) [38].

Table 6 - LTE UE capabilities

	UE Category	***Max nb of DL-SCH transport block bits received within a TTI***	***Peak Data Rate***	***Max nb of supported layers for spatial mux in DL***
Downlink Performance	Cat. 1	10040	10 Mbps	1
	Cat. 2	50000	50 Mbps	2
	Cat. 3	100000	100 Mbps	2
	Cat. 4	150112	150 Mbps	2
	Cat. 5	300064	300 Mbps	4

	UE Category	Max nb of DL-SCH transport block bits received within a TTI	Peak Data Rate	Support for 64QAM in UL
Uplink Performance	Cat. 1	5032	5 Mbps	No
	Cat. 2	25008	25 Mbps	No
	Cat. 3	50000	50 Mbps	No
	Cat. 4	50000	50 Mbps	No
	Cat. 5	75056	75 Mbps	Yes

In order to achieve a higher system performance in LTE, scheduling must also be aware of the buffer status reports, the QoS parameters for the UEs in the coverage area and of the HARQ processes. Figure 19 presents the factors used by the eNB scheduler to decide about what users to be scheduled, what type of allocation should be used (Type 0, 1 or 2) and what strategy for the time validity of the allocation is more suited (semi-persistent or non-persistent).

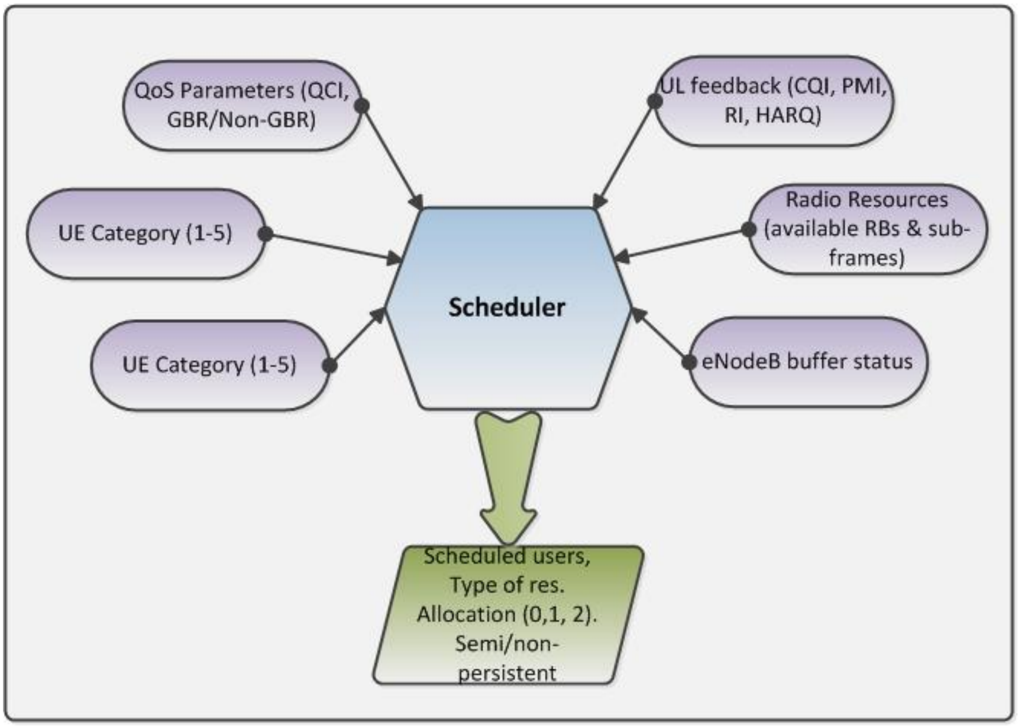


Figure 19 – Factors influencing the scheduling process

2.2.3 QoS Provisioning In LTE Networks

LTE technology implements dedicated mechanisms for QoS control to be able to deal with the network congestions, limited radio spectrum, limited backhaul bandwidth and to manage efficiently the variety of services requested by users (guaranteed bit rate for VoIP, high speed throughput for video, etc). From the operators' perspective, the standard allows the implementation of Policy Control and Charging functions that can be used to improve the efficiency of the network resources used, to ensure customer satisfaction, to comply with certain laws or regulations in different geographical areas, to offer superior performances for the users that pay for them or to provide attractive offers that can lead to a market share gain.

A Service Data Flow (SDF) in an application level data packet flow between a UE and one or more devices in an external Packed Data Network (PDN) and it is the level at which LTE defines the QoS parameters (Figure 20).

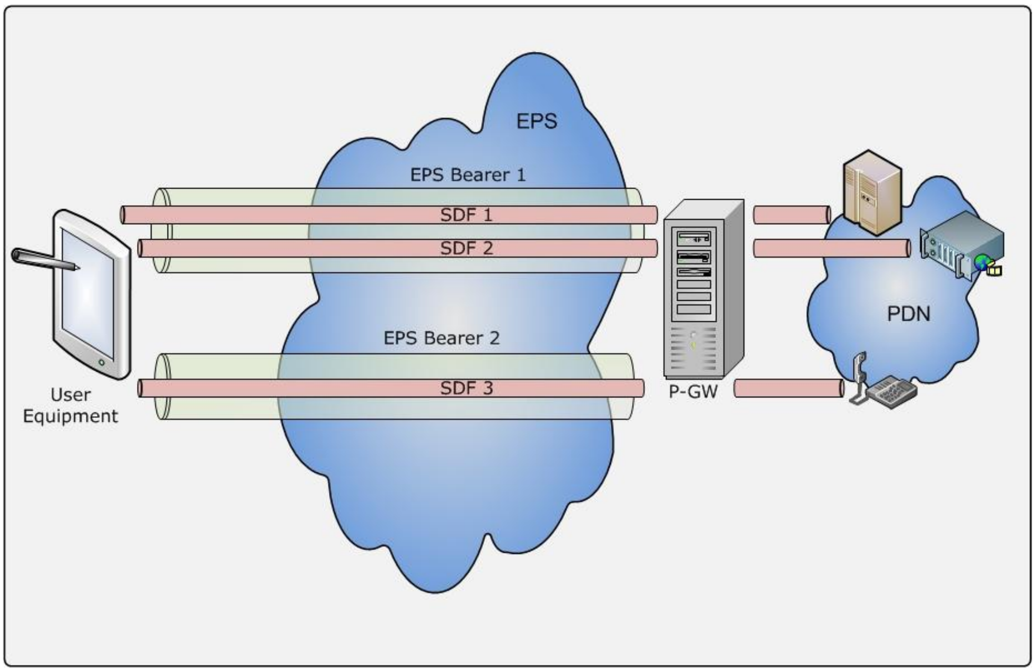


Figure 20 – SDFs mapping on different bearers

If a user wants to access a home video streaming service, the QoS parameters can set up the connection with a specific packet delay, by assigning a relative priority in comparison to other services and possibly a guaranteed bit rate. If the same user requests another streaming session from an internet service, the quality requirements of this new connection can be the same as the previous ones or they can be different. For the user to experience the same quality for both streaming sessions, the network needs to aggregate them and treat the two connections that are running simultaneously as a single flow. In order to make this concept possible, a logical channel named EPS bearer is used. The EPS bearer is the level at which QoS policies are enforced in the network [39].

After the UE establishes a signaling radio connection with the eNB, it performs the initial attach procedure, followed by the MME selection (performed by the eNB). The UE can now start the Authentication procedure with the MME, which in case the procedure is successful proceeds and selects the S-GW and the P-GW. At this step, the UE has always-on IP connectivity, provided by the default EPS bearer established during the network attachment procedures. The EPS bearer is composed of three virtual traffic paths, as depicted in Figure 21:

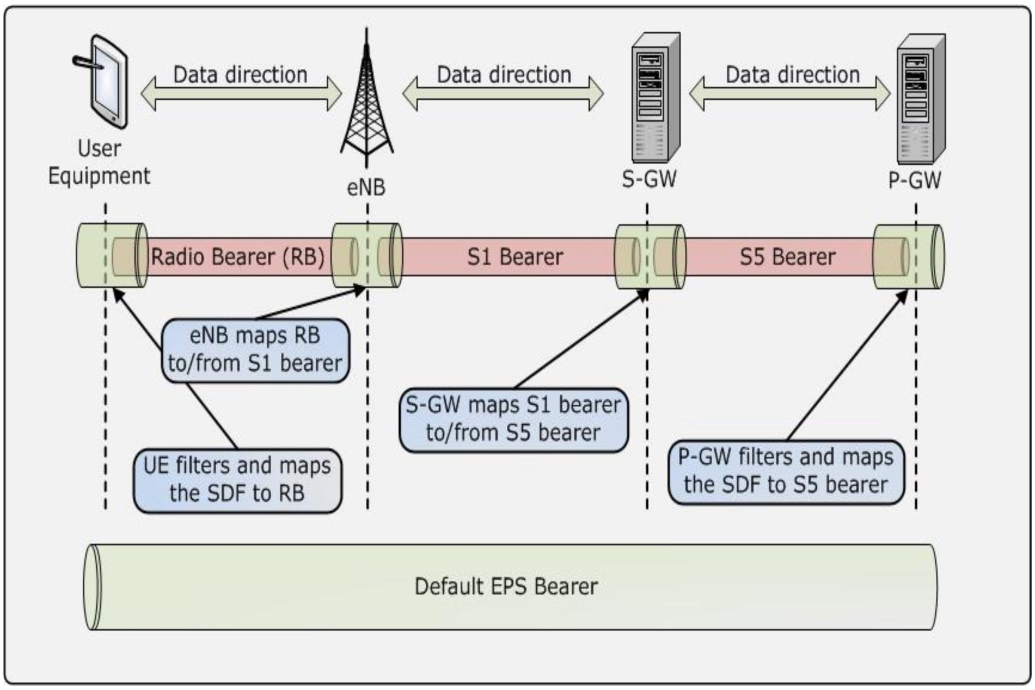


Figure 21 – Default EPS bearer anatomy

Every user that is connected to the LTE network has established a default virtual GTP (GPRS Tunneling Protocol) tunnel between him and the P-GW that is able to provide an always-on IP connectivity, delivering traffic on a "best effort" basis. This default EPS bearer should provide enough resources for applications like e-mail, file sharing, ftp, progressive video, etc. If a user demands an application that requires a specific QoS treatment, then a new virtual tunnel will be created, the dedicated EPS bearer, able to deliver the specific traffic with respect to the established QoS parameters (Figure 22).

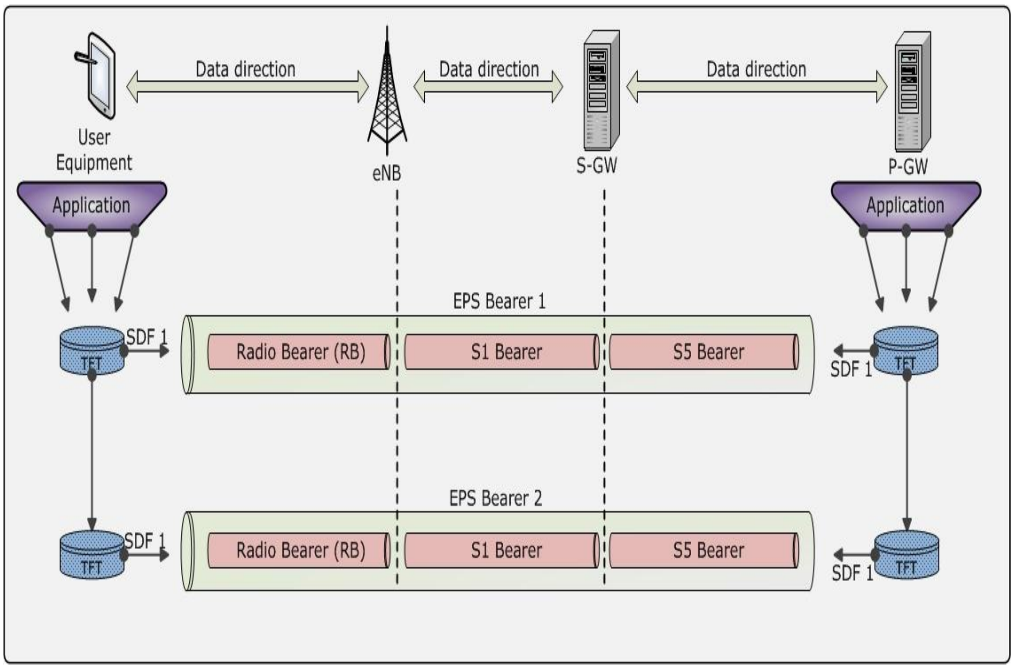


Figure 22 – Traffic Flow Templates in LTE networks

In order to map the users' service traffic flows on the associated bearer, the standard specifications are proposing the use of traffic filters, able to differentiate between data packets that require specific QoS treatment. These traffic filters called Traffic Flow Templates (TFT) are implemented for both UL and DL directions and are used to recognize a specific SDF, to aggregate it with other SFDs that have the same QoS requirements and to map it on the bearer with the suited QoS parameters. While the UL TFT just maps the uplink data packets to the appropriate Radio Bearer, the DL TFT has to use the IP network layer and transport layer header information to classify and detect the correct SDF for the traffic arriving from an external PDN [40].

The key QoS parameters associated with an EPS bearer, as presented in Figure 23, are:

- Bearer type: Guaranteed Bit Rate (GBR) or Non-GBR bearers; for GBR bearers, the guaranteed bit rate and the maximum bit rate (MBR) must be specified, while for Non-GBR bearers only the aggregate maximum bit rate (AMBR) will be specified.
- Allocation and Retention Priority (ARP): this parameter is used in congestion situations, when not all users or their requests can be accommodated.
- The QoS Class Identifier (QCI): it is used to define the general class of the service and it is associated to a certain priority, a specific delay and packet loss values. There are currently 9 QCI values defined in the standard (Table 7) [41], used in the process of determining the resource scheduling or rate shaping.

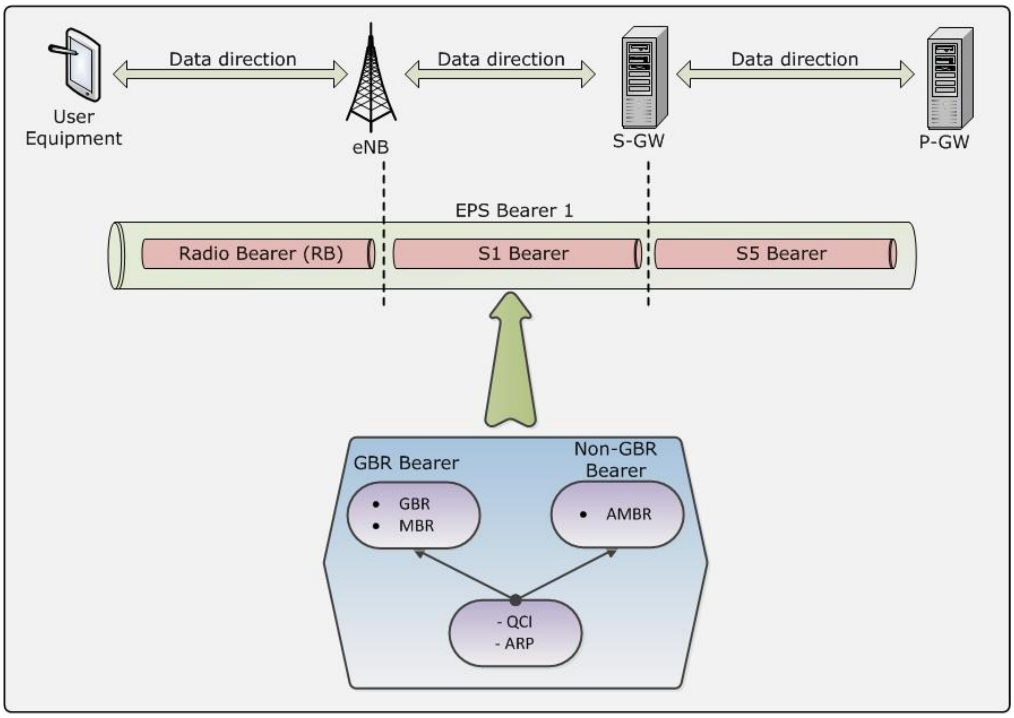


Figure 23 – QoS parameters of an EPS bearer

Table 7 - Standardized QCI characteristics

QCI	***Resource Type***	***Services (example)***	***Packet Delay***	***Packet Error***	***Priority***
1	GBR	VoIP	100 ms	10^{-2}	2
2		Live video streaming	150 ms	10^{-3}	4
3		Real-Time gaming	50 ms	10^{-3}	3
4		Buffered video	300 ms	10^{-6}	5
5	Non-GBR	IMS signaling	100 ms	10^{-6}	1
6		Voice, video, interactive gaming	300 ms	10^{-6}	6
7		Buffered streaming, TCP based apps (web, mail ftp transfer, etc)	100 ms	10^{-3}	7
8			300 ms	10^{-6}	8
9					9

3. IEEE 802.11 WIRELESS LAN NETWORKS

First section of this chapter presents the general characteristics of the IEEE 802.11 Wireless LAN networks, followed by a detailed description of the physical and MAC layers in sections two and three, where the QoS aspects of this technology are also discussed.

3.1 Introduction

In the author's vision, achieving the anytime - anywhere IP connectivity desiderate, one other network should be considered, besides 3GPP LTE. Taking into account that the number of available IEEE 802.11 hot-spots increased exponentially during the last years all over the world, offering high speed data rates at very low costs, it is fair to assume that IEEE 802.11 networks [42] will become the worldwide accepted technology for wireless short-range internet access, similar to what the GSM ETSI standard is for long-range mobile voice communications. The success of IEEE 802.11 has motivated many researchers to improve its performances through sustained efforts and brilliant ideas. Next sub-sections are offering an overview of the standard, focusing on detailing the physical and MAC layers.

A WLAN (Wireless Local Area Network) connects two or more devices using a wireless distribution method (most common are OFDM – 802.11g – and spread-spectrum – 802.11b). It is usually providing a wireless connection through an access point to the wider internet for a limited number of users, giving them the possibility to move around within a local coverage area and still be connected to the network.

A typical wireless router using 802.11g with a stock antenna might offer a range of 35 m indoors and 100 m outdoors. The new 802.11n however, can exceed that range by more than two times. This type of networks have become popular in both home and public environments due to the ease of installation and the increase of mobile devices equipped with WNICs (Wireless Network Interface Cards). The content delivered via wireless networks is usually rich media-based and often puts significant pressure on the existing networks in order to support high quality of delivery.

A classification of available wireless networks based on coverage can be the one presented in Figure 24:

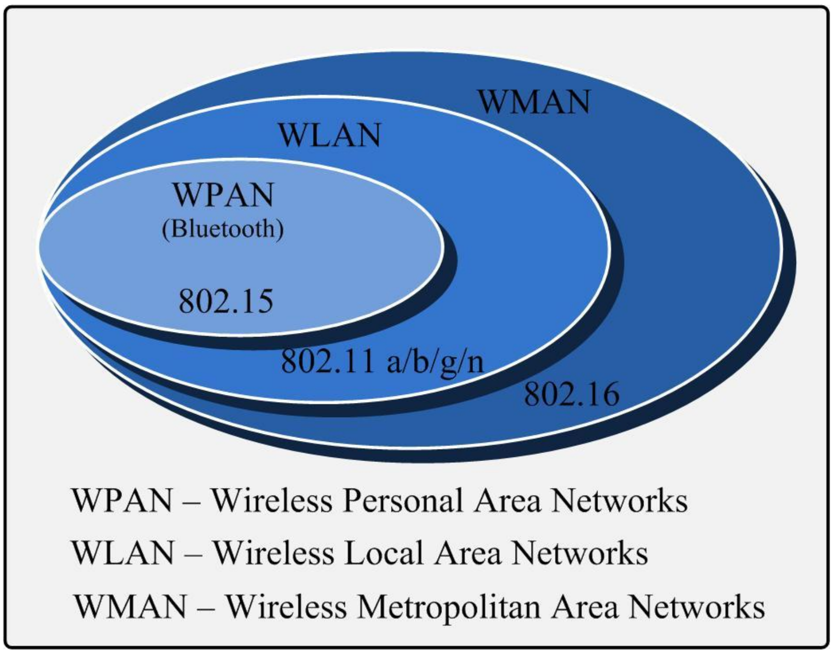


Figure 24 - Classification on wireless networks based on coverage

The IEEE 802.11 standard has three basic topologies for a network [43]:

- Independent Basic Service Set (IBSS)
- Basic Service Set (BSS)
- Extended Service Set (ESS)

In case of an IBSS there is no base and no one gives permission to talk. This type of network is also known as a peer-to-peer mode (Figure 25 (a)), where it is allowed for devices to directly communicate with each other. This method is typically used by two computers so that they can connect to each other to form a network. BSS is a cellular topology, where a cell is composed of an access point (AP) and all the stations connected. In this specific case, the communication between two stations is realized through the AP, which is usually connected to an Ethernet network. This type of network is also known as infrastructure mode (Figure 25 (b)). For the ESS there are a number of interconnected APs using an infrastructure called Distribution System (DS) (e.g. Ethernet).

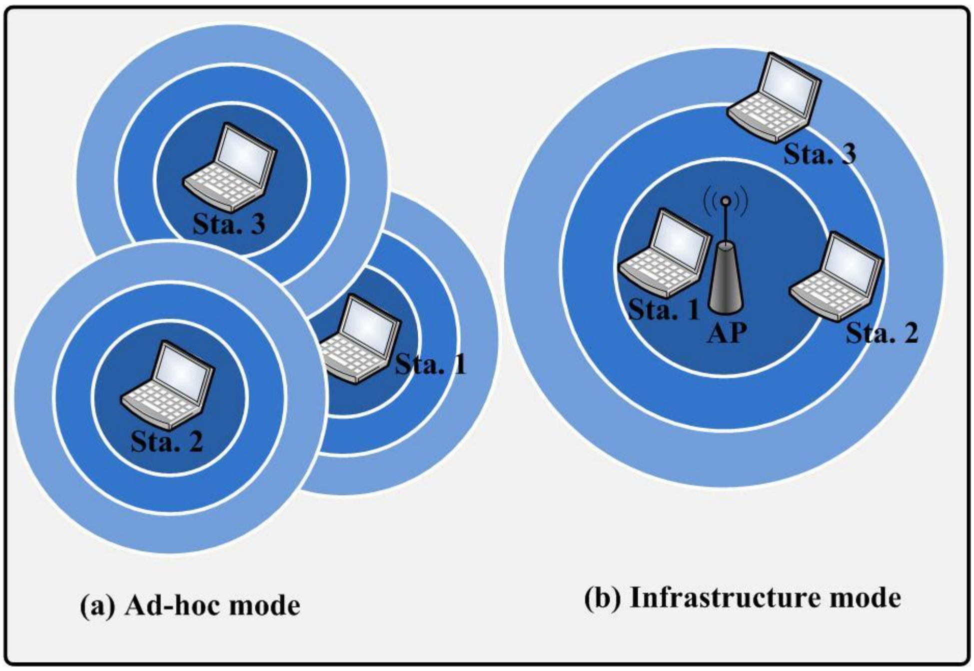


Figure 25 - 802.11 modes of operation

The most popular standards from 802.11 family are those defined by the 802.11b and 802.11g protocols, which are amendments to the original standard [43]. They both use the 2.4 GHz frequency band and this is the reason why 802.11b and g equipment can occasionally suffer interference from microwave ovens, cordless telephones or Bluetooth devices. Table 8 presents the parameters of 802.11 network standards:

Table 8 - 802.11 network standards

802.11 Protocol	***Release***	***Freq (GHz)***	***Bandwidth (MHz)***	***MIMO streams***	***Modulation***	***Indoor range***	***Outdoor range***
Legacy	Jun 1997	2.4	20	1	DSSS, FHSS	20	100
a	Sep 1999	5	20	1	OFDM	35	120
b	Sep 1999	2.4	20	1	DSSS	38	140
g	Jun 2003	2.4	20	1	OFDM, DSSS	70	140
n	Oct 2009	2.4 / 5	40	4	OFDM	70	250

The 802.11 standards are focused on the bottom two layers from the protocol stack for WLANs, as presented in Figure 26.

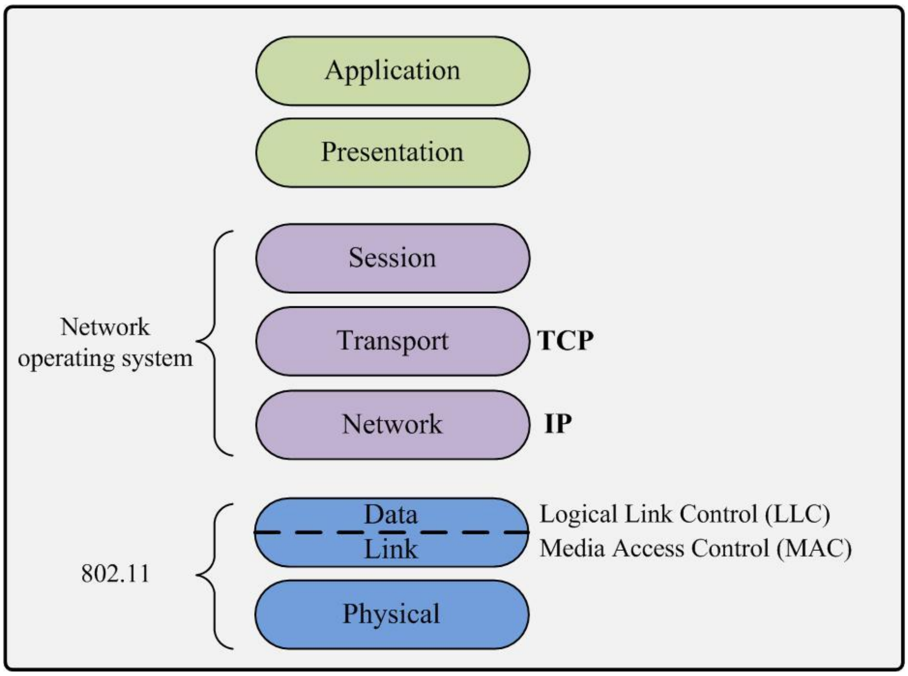


Figure 26 - WLAN protocol stack

3.2 Physical Layer

The IEEE 802.11 in its original form defines three types of PHYs: Frequency Hopping Spread Spectrum (FHSS), Direct Sequence Spread Spectrum (DSSS) and Infrared (IR). With FHSS, communicating stations operate on a common frequency channel only for a short period of time, before hopping to another frequency. The frequency channels used for hopping are known by all stations in the group based on a pseudo-random list of frequencies. In contrast to FHSS, in DSSS all stations operate on the same center frequency. Both FHSS and DSSS in 802.11 are not CDMA because all stations operate with the same code sequence. The objective of the spread spectrum approach is to perform a balanced distribution of radio emissions over broader bandwidths in the spectrum in order to facilitate spectral co-existence with other radio systems (Bluetooth) [44]. In case of newer physical layers using OFDM, spread spectrum is not needed because of the flat shape of the transmitted signals. The IR physical layer is not commercially successful, but may become a solution for future residential in-house applications, where users can prefer light over radio waves.

Orthogonal Frequency Division Multiplexing is the transmission scheme used by 801.11a, 802.11 g and 802.11n. In 802.11a, one OFDM symbol (52 carriers using 16.6 MHz) has a duration of 4 µs and the system can dynamically select a coding and modulation scheme based on current QoS requirements and radio channel conditions. From the 52 carriers of a symbol, 48 are used to transport user data while 4 are used for pilot symbols.

Table 9 [43] summarizes numerical values for the main parameters of the 802.11a OFDM transmission system, which are identical for most of the 802.11g, operating in the 2.4 GHz frequency band.

Table 9 - OFDM Parameters of IEEE 802.11a

Parameter	***Value***
Sampling rate 1/T	20 MHz
OFDM block duration T_b	65*T = 3.2 μs
Guard interval duration T_g	16*T = 0.8 μs
OFDM symbol duration $T_b' = T_b + T_g$	80*T = 4 μs
Number of data sub-carriers	48
Number of pilot sub-carriers	4
Sub-carrier spacing D_f	$1/T_b$ = 0.3125 MHz
Spacing between the outmost sub-carriers	$(N_{total} - 1) * D_f$ = 15.9375 MHz

3.3 Media Access Control (MAC) Sub-Layer

The IEEE 802.11 MAC layer specifies two coordination functions, i.e., the Distributed Coordination Function (DCF) for traffic without QoS, known as asynchronous services, and the Point Coordination Function (PCF) for traffic with QoS requirements, known as synchronous services [42]. The mandatory Distributed Coordinator Function (DCF) is based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) while the optional Point Coordinator Function is based on a pooling mechanism.

3.3.1 Distributed Coordination Function

In wireless communications, a transmitter cannot detect collision at a receiver while transmitting. To overcome this problem, the 802.11 defines collision avoidance mechanisms to reduce the probability of such unwanted problems. Before starting a transmission, the station performs the so called backoff procedure: stations having an MSDU to deliver need to sense the channel for a random time duration after detecting the channel being idle for the minimum duration Distributed Inter-Frame Space (DIFS), which is 34 μs for 802.11a. If the channel remains idle after this period of time, the station will initiate its transmission [45]. The duration of the random time period is a multiple of the slot duration SlotTime. Each station has a Contention Window (CW) which is used to determine the number of SlotTime intervals it has to wait before initiating a transmission. For the first transmission attempt the CW is set to the minimum value CW_{min}. It is doubled for every unsuccessful transmission attempt up to a maximum value CW_{max} as depicted in Figure 27. If a successful transmission is achieved, the CW is reset to the CW_{min} value. During this backoff procedure, the backoff timer is decremented for each time slot that the medium remains idle. If the medium becomes busy during this period, the timer is paused and it is resumed once the medium is sensed idle for a time interval of DIFS. The station is allowed to transmit once the backoff timer reaches zero [46].

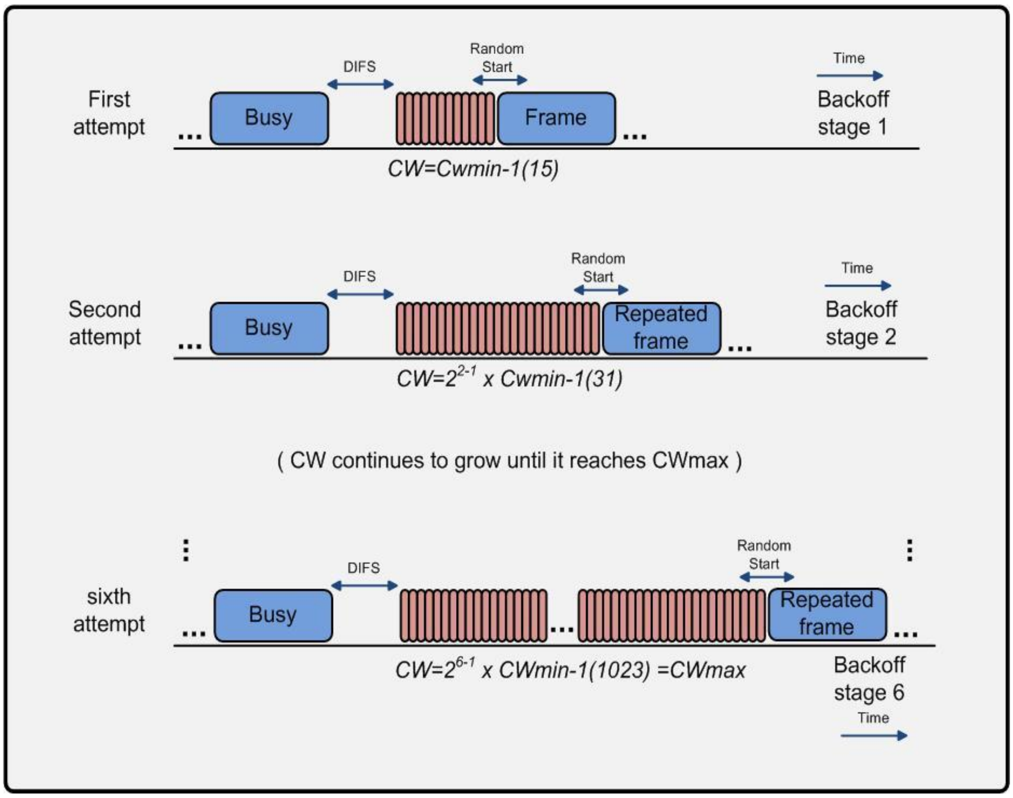


Figure 27 - Backoff procedure for collision avoidance in IEEE 802.11a

A positive acknowledgment frame (ACK) is used to inform the sender that the frame has been successfully received. A receiver returns an ACK frame after a Short Inter-Frame Space (SIFS). If a sender does not receive an ACK within $ACK_{timeout}$, it assumes the packet has been lost due to collision or erroneous frame and reschedules the transmission by running the backoff procedure again. After each successful transmission, another random backoff procedure is performed by the transmitting station even if there is no MSDU to be delivered. This is called a "post-backoff" procedure and is aimed to ensure that all frames waiting for transmission will be delivered with backoff. Any MSDU arrived in the station's queue from a higher layer can be transmitted without a delay if all following are true: the transmission queue is empty, the post-backoff procedure is finished and the channel is idle for at least DIFS. By using this procedure when the system is not highly loaded, a shorter delivery delay can be obtained.

To reduce the duration necessary for a transmission, data frames (MSDUs) can be segmented into smaller MPDUs if their length exceeds a certain threshold.

This process is called fragmentation and it is used to limit the probability of long MSDUs colliding and being retransmitted. The main advantage of fragmentation is that the data to be transmitted in case of an erroneous transmission is less than the original MSDU. It also increases the probability of successful MSDU transmission in cases where the radio channel characteristics cause more errors for longer frames than for shorter frames. One important aspect related to fragmentation is that only MSDUs involved in a unicast communication can be fragmented. Broadcast or multicast frames cannot be fragmented even if their length exceeds the threshold [43], [47].

3.3.2. Hidden Station Problem

An important problem that can occur in wireless communication systems that use carrier sensing arises when a station is able to receive frames coming from two different sources but those transmitting stations cannot detect each other. This problem is referred in the literature as the hidden station problem [47]. If two stations that are both within the range of the same AP cannot detect each other, they may sense the channel as idle, and can start initiating a transmission. In these circumstances, when transmitting stations are not aware of the other hidden stations, a collision is taking place and severely interfered frames are detected at stations that are aware of the all the hidden stations. In order to reduce the impact of hidden stations transmitting simultaneously, IEEE 802.11 standard specifies the exchange of Request-To-Send/Clear-To-Send (RTS/CTS) frames.

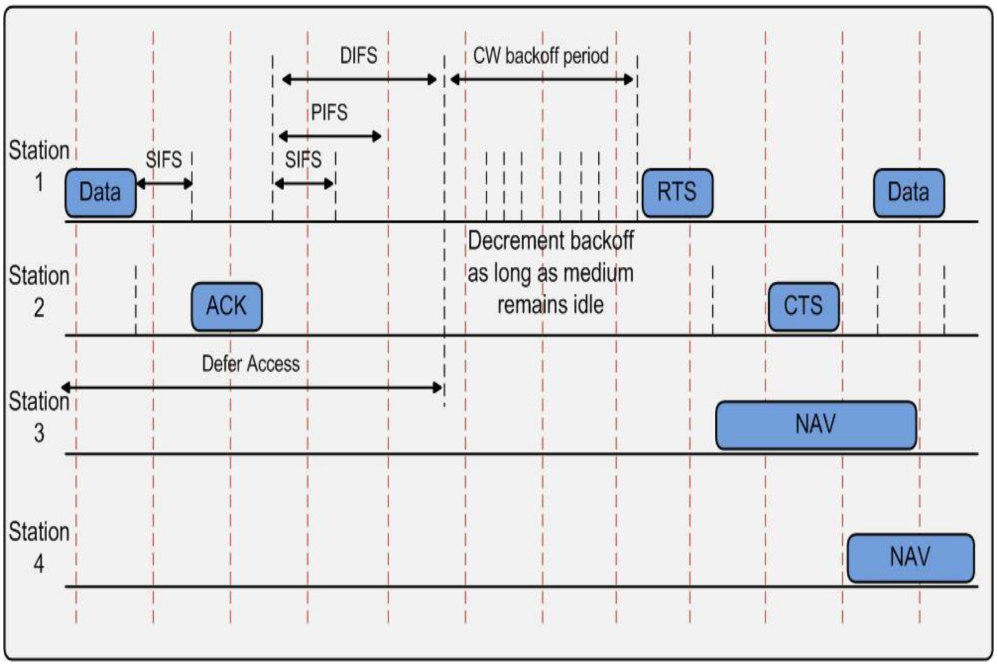


Figure 28 - RTS/CTS mechanism

If the conditions are favorable for a data exchange, the receiving station will send back a CTS frame spaced by SIFS. RTS/CTS procedure is used together with CSMA/CA to overcome the problems discussed above. A station that has a pending frame for transmission first performs the CSMA/CA procedure followed by RTS/CTS, detailed in Figure 28 [43].

The transmitting station sends a RTS broadcast to all nodes within its carrier sense range. Like this, all nodes that are receiving the RTS broadcast will be informed to not access the medium for the duration time specified by the Network Allocation Vector (NAV) field in the RTS frame. The receiver of the data frame will respond to the RTS with a CTS, which is also received by all stations within its range. The stations will not access the medium for the duration time specified by the NAV field in the CTS frame. The transmitting station can now proceed with the transmission of the data frame. Even if the RTS/CTS procedure reduces the number of collisions it also increases the overhead required to transmit a packet.

3.3.3 Point Coordination Function

PCF is an optional MAC access technique that resides in the AP, being located exactly above the DCF in the MAC architecture. It is used to provide support for time-sensitive services by using a Point coordinator (PC), that can control and allow prioritized user access to the wireless medium. When the PCF is used in a BSS topology, the access medium is divided into Contention Free Period (CFP) and Contention Period (CP) intervals called super-frames. While DCF is used to access the channel during CP intervals, the PCF is used to access the channel during the CFP intervals. During CFP, the AP will interrogate stations about the existence of any pending frames and will deliver those frames if necessary. The AP's pooling mechanism will continue until the CFP ends, at which point a CFP-End control frame is transmitted by the AP to signal the end of the CFP [47].

The PCF access technique is also known to have a sum of problems, from which the most important are the unpredictable beacon delays and the unknown transmission durations of the interrogated stations.

The unpredictable beacon delays may happen when the beacon cannot be transmitted at Target Beacon Transmission Time (TBTT) until the medium is considered idle for a time interval equal or larger than PCIF, even if the beacon was scheduled as the next frame. Another problem is the unknown transmission duration of polled stations. One station that was polled by Point Coordinator has the permission to transmit an MSDU of an arbitrary length that may be fragmented. Applying different modulation and coding schemes can affect the delivery duration of the selected MSDU which is no more under the control of the PC. This can reduce the QoS that is provided to other stations that are polled during the rest of the CFP.

3.3.4 MAC Enhancements With IEEE802.11e

The enhancements that were added to the 802.11 MAC layer, due to the drastic increase in the real-time traffic (streaming media, VoIP, network gaming etc) had led to a extension to the standard, known as 802.11e (IEEE, 2005b) [48], that allows service differentiation of various traffic flows within a wireless network. The IEEE 802.11e introduces a new MAC Layer function called Hybrid Coordination Function (HCF) for QoS support. The HCF defines two medium access mechanisms: contention-based channel access and controlled channel access. The contention-based channel access is provided using Enhanced Distributed Channel Access (EDCA) while the controlled channel access is provided using HCF Controlled Channel Access (HCCA). EDCA is used only in the CP while HCCA is used in both CP and CFP. Both HCCA and EDCA use the 802.11 MAC frame format of the DCF to transmit user data on the radio channel (DATA/ACK frame exchange sequence with optional RTS/CTS). Figure 29 illustrates the main elements of the 802.11e MAC architecture in the context of 802.11:

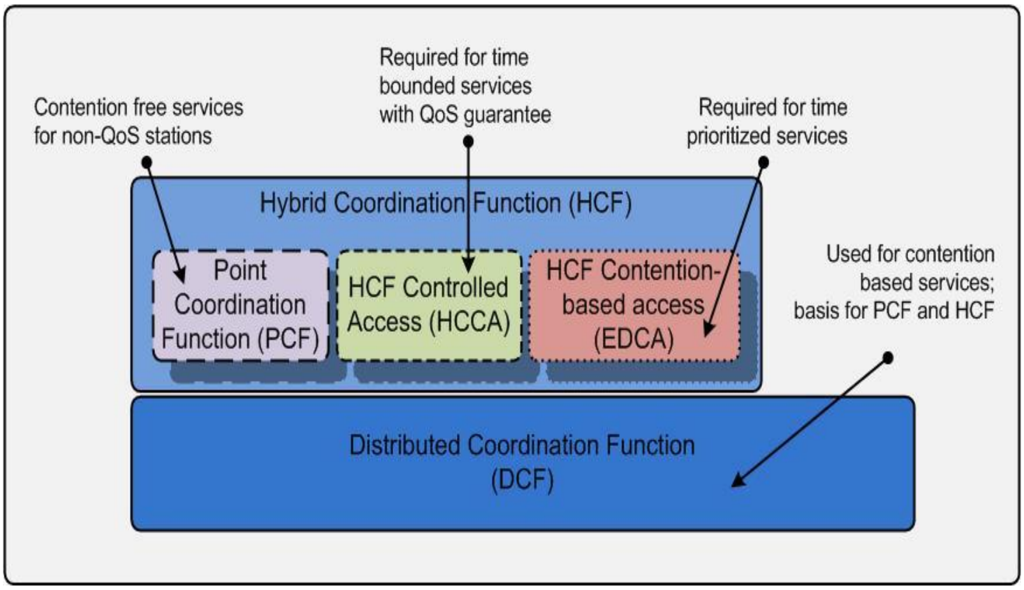


Figure 29 - MAC architecture of 802.11e

Enhanced Distributed Channel Access [43] is designed to provide prioritized QoS by enhancing the contention based DCF mechanism outlined above. This prioritization is achieved by associating a priority level with every packet entering the IEEE 802.11e MAC. These user level priorities are known as Traffic Categories (TC). EDCA also introduces four First-in, First-out (FIFO) queues at the MAC layer called Access Category (AC). Packets arriving at the MAC layer are filtered into their corresponding ACs in accordance with the IEEE 802.1D bridging protocol. The four ACs are labeled according to their target application: AC_VO (voice), AC_VI (video), AC_BE (best effort) and AC_BK (background). The filtering mechanism for arriving packets is shown in Figure 30.

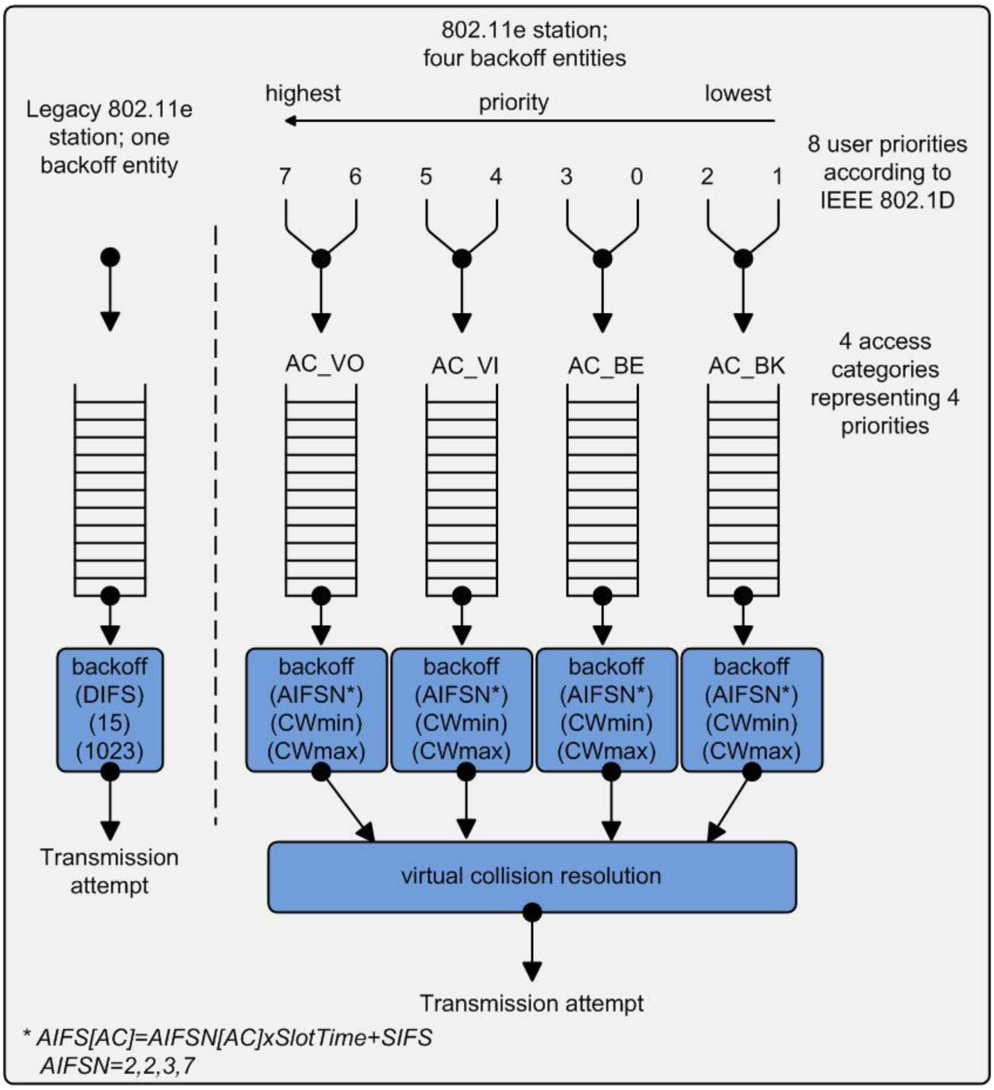


Figure 30 - Legacy 802.11 station and 802.11e station with 4 ACs

In addition to these new coordination functions, the HCF also introduces the concept of Transmission Opportunity (TxOP), which refers to a time duration during which a station is allowed to transmit a burst of data frames. Each backoff entity within a station independently competes for a TxOP. It starts to decrease the backoff counter after it detects that the medium is idle for a certain duration, defined by the Arbitration Interframe Space (AIFS[AC]) instead of DIFS, which is used by legacy stations. The AIFS[AC] is greater of equal to DIFS, and can be enlarged per access category with the help of the (AIFSN[AC]) [47]. The Arbitration Interframe Space Number defines the duration of AIFS[AC] according to:

$$AIFS[AC] = AIFSN[AC] \times SlotTime + SIFS,\ AIFSN[AC] \geq 2$$

AIFSN[AC] is selected by the HC such that the earliest access time of EDCA stations is DIFS. Basically, an access category uses AIFS[AC], CW_{min}[AC] and CW_{max}[AC] instead of the DCF parameters (DIFS, CW_{min} and CW_{max}) for the contention process. These parameters are chosen to allow a quicker medium access for higher priority traffic. The smaller their values, the shorter the channel access delays, and the higher the capacity share for a given traffic condition. The CW range increases exponentially after each failed transmission attempt and resets after each successful transmission. The size of the CW_i [AC] in backoff stage *i* is:

$$CW_i\ [AC]=min\ [\ 2i(CW_{min}[AC]+1)-1,\ CW_{max}[AC]\ ]$$

The contention window value, which is defined per AC as part of the EDCA parameter set, never exceeds CW_{max}[AC]. Small value for CW_{max}[AC] translates into high medium access priority, but this also mean an increase of the collision probability. The default set of EDCA parameters are presented in Table 10:

Table 10 - Default EDCA parameter set

AC	***CW_{min}***	***CW_{max}***	***AIFSN***
AC_BK	aCW_{min}	aCW_{max}	7
AC_BE	aCW_{min}	aCW_{max}	3
AC_VI	$(aCW_{min}+1)2-1$	aCW_{min}	2
AC_VO	$(aCW_{min}+1)/4-1$	$(aCW_{min}+1)/2-1$	2

The HCF Controlled Channel Access is similar to IEEE 802.11 PCF, except there is no division between CFPs and CPs. A QoS Hybrid Coordinator (HC) controls the access to the medium when in HCF Controlled Channel Access mode. If the network operates in infrastructure mode, the HC is represented by the AP, while in ad-hoc mode the HC is represented by the HC station, which has high priority access to the medium in order to initiate a Controlled Access Phases (CAPs). The CAP is the time period in which a HC can initiate a downlink frame transfer with a station or poll a station for pending frames. A more detailed description of the operating mode when in HCCA can be found in [43], [49].

4. MULTIMEDIA STREAMING OVER WIRELESS NETWORKS

The first part of this chapter focuses on the aspects that concern the multimedia streaming, trying to explain the QoE concept for multimedia traffic. In the second section, various network and transport layer protocols used to deliver multimedia content are presented, while section three describes the most common multimedia streaming algorithms. The last section presents some options for resource scheduling algorithms used in both IEEE 802.11 wireless LAN and LTE networks.

4.1 QoE Concept For Multimedia Traffic

The area of adaptive streaming in wireless environments has been continuously improved during the last years due to the efforts of many researchers who imagined, developed and implemented various solutions that aimed to reduce the network load while increasing the link utilization, to maintain a minimum quality level of the service provided, etc.

Users in wireless networks usually have a subjective perception about the delivered service quality that can be influenced by a number of factors: environment factors, delivery of a service above a minimum required level, fulfilling the applications' requirements in terms of bandwidth, loss or delay, etc. The pressure on the service providers is constantly increasing in a very competitive market, as they are aware that assuring a satisfactory end-user quality of experience (QoE) can mean their survival. In order to better understand the user experience concept, all factors that influence the user perception must be carefully analyzed. Figure 31 presents these factors:

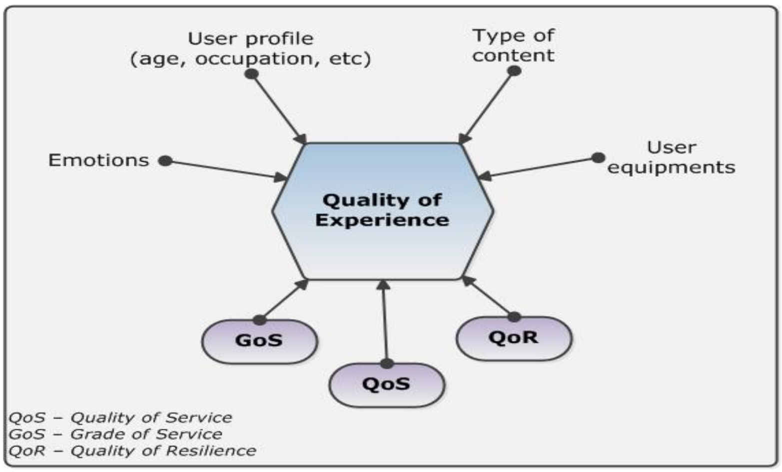


Figure 31 – Factors influencing QoE

It is a very difficult task, if not impossible, for a network operator to control all these factors in order to offer an acceptable quality for all services that a user might request. But in the context of particular applications it is possible to find a mapping between the network intrinsic characteristics (QoS, GoS and QoR) and the user satisfaction, as QoE is strongly dependent on them.

The **Grade of Service** (GoS) was first used in circuit switched telephone networks [50] to describe the performances of connection setup, connection release and connection maintenance. Currently, GoS is used in all types of networks, with or without admission control [51], to establish the performance of the connection setup or release indicators by evaluating parameters like connection setup delay, authentication delay, service rejection, etc.

Quality of Resilience (QoR) [52] is used to describe the availability and retainability of a network, analyzing parameters like service outages or service drops. Because users have a different perception about service drops, about the mean time to reconnect or about the unplanned outages and because QoE can be easily influenced by these problems, network operators establish very low targets for them, usually under 0.5%, by implementing complex recovery mechanisms able to satisfy these demands.

Adaptation as a counter measure for congestion has been an interesting research topic since the 80's [53]. In order to design a suitable algorithm for multimedia stream adaptation, it is necessary to consider a good congestion control algorithm that is able to avoid the situations when the network is ready to collapse due to congestion. One such algorithm should optimize the most important network parameters – throughput, loss, delay, jitter – for multimedia traffic. These parameters and their values are providing important information about the service quality, known as the **Quality of Service** or QoS.

Throughput is the network parameter that measures the average rate of successful message delivery over a communication channel. An important aspect that it is taken into consideration by the researchers is the maximum throughput. Maximum throughput can be separated into four different categories: maximum theoretical throughput, peak measured throughput, maximum achievable throughput and maximum sustained throughput. Maximum theoretical throughput is represented by the maximum amount of data that can be theoretically transported through a communications channel. It is closely related to the channel capacity and together with maximum achievable throughput, is considered generally in the design phase of a communication system. The maximum sustained throughput it is considered to be the most accurate indicator of system performance, because it is the throughput value averaged over a long period of time. The peak measured throughput, like the maximum sustained throughput, it is a real measured parameter and it gives the throughput of a system or a communication channel averaged over a short period of time. The parameter is important in communication systems that rely on burst data transmissions. As a whole, throughput is an important factor that can influence the multimedia streaming process by modifying the QoS and in consequence the QoE. One system that has a large available bandwidth will offer a high QoE because the throughput achieved by the multimedia application it is better than the one that can be achieved in a system with limited resources. But besides the available bandwidth, there are other factors that affect the throughput: the analog limitations, the total number of users using the network resources, the protocol limitations etc.

Packet loss is the phenomenon that occurs into an IP-based communication network when packets of data traveling across the network fail to reach their destination. This can happen because of some factors like signal degradation, network congestion, faulty network equipment, etc. In case of network congestion, the available link capacity is lower than the combined data rates of the incoming streams and the router buffers will overflow, resulting in dropped packets. In wireless networks, the main factors that conduct to packet losses are the signal degradation and the interferences. Loosing or dropping packets can lead to performance issues especially for streaming applications, degrading the user QoE. Even if the video streaming or services like VoIP can cope with small packet loss rates, exceeding these rates will dramatically decrease the service performances and the user perceived QoE.

The network delay is an important parameter of a telecommunication network and it specifies the duration necessary for a bit of data to travel from the source point to the destination point. Generally, in any communications network there are four sources for the delay:

- processing delay: introduced by the routers while they process the packet header
- queuing delay: introduced by the packets' waiting period in routers queues
- transmission delay: introduced by the mechanisms used to push the data bits of each packet into the network
- propagation delay: introduced by the distance between the source and the target point

Packet jitter is another important characteristic of computer networks and it is used to measure the variation in time of packet latency across network. An ideal network will have a constant latency and as a consequence no jitter. But in real networks jitter is present and it is expressed as the average deviation from the network mean latency. Many experts consider that the term jitter is incorrectly used for describing the above problem, and that the correct term is packet delay variation or PDV. PDV is a very important QoS factor in communications networks, especially when the applications that are using the network are real time applications. In case of multimedia streaming, the solution to remove PDV is to select a properly sized play-out buffer at the receiver which can cause just a delay before starting the media playback. For interactive real time applications like VoIP, PDV can cause serious performance degradation and in consequence a very bad QoE. To prevent this from happening, QoS-enabled networks are needed in order to provide a quality channel with a reduced PDV.

The QoE experienced by a user is in direct connection with the network performances and specific features but there is no simple mapping between these parameters and the perceived quality. However, data delivery with appropriate QoS, QoR and GoS levels is a "must do" in order to be able to achieve a high QoE. As the user experience is also affected by other factors (Figure 31), one can expect the users to experience in different ways a service provided with the same network parameters.

4.2 Network, Transport And Upper Layer Protocols

4.2.1 Internet Protocol (IP)

All of today's multimedia applications are using an IP-based network layer, with all the other layers depending on the applications requirements. Considering the wireless streaming process, IP is a very important component that provides mandatory services to the layers above in the TCP/IP architecture. IP is a connectionless and unreliable protocol that delivers the data packets using the best-effort strategy, able to offer the following:

- It defines the basic unit for data transfer over TCP/IP networks, and the exact format of all data in such network is specified
- It realizes the routing function, choosing the path for data transmission
- Besides routing and packet formatting information, it also contains a set of rules that characterize the way a station or a gateway should process the received packets, why and when error messages can be generated and under what conditions the packets can be dropped.

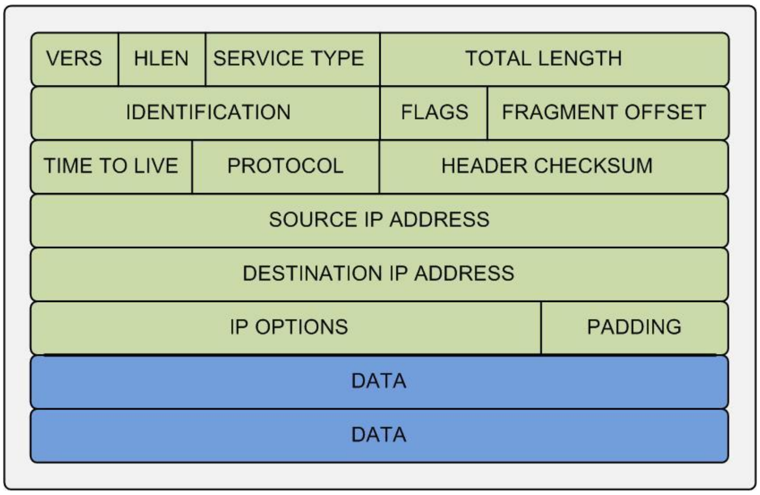


Figure 32 - IP header structure

The IP packet, like any other packet sent over the network, has a header and a payload. The header (Figure 32) contains [54]:

- IP version (VERS)
- Headers length (HLEN)
- The service type (SERVICE TYPE)
- Total length of the packet (TOTAL LENGTH)
- Packet identification number (IDENTIFICATION)
- Fields used in the fragmentation process (FLAGS, FRAGMENT OFFSET)
- Time to live for the packet (TIME TO LIVE)
- The protocol (PROTOCOL)

- Header checksum
- Source IP address
- Destination IP address
- Bits for IP options (IP OPTIOINS)

The IP dimension, opposite to the physical networks frames, is not limited by any hardware characteristic. For IP version 4 (IPv4) [54], the maximum length of a packet can be 2^{16} bytes. An IP packet that is delivered over the network from one station to another has to be encapsulated into a network frame (Figure 33). The ideal case is when the entire IP packet fits into a single network frame, increasing the transmission efficiency. To make the transmission as efficient as possible, Maximum Transfer Unit (MTU) term was introduced, so that any IP packet will fit into the network frame (e.g. Ethernet MTU = 1500, proNET10 MTU = 2044). If the MTU is small, inefficient data transfers will happen over a network that can deliver larger frames.

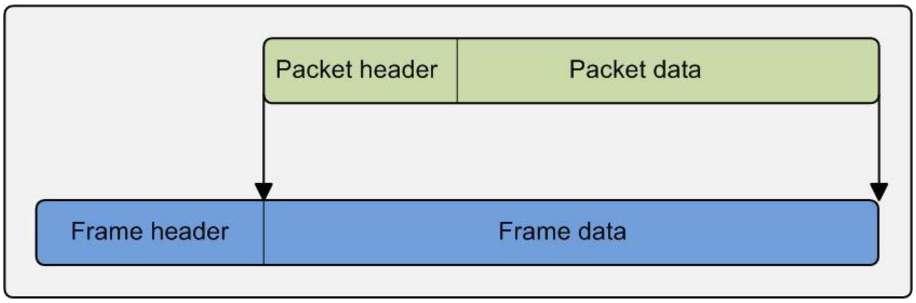


Figure 33 - IP packet encapsulation

There are currently two IP versions: Internet Protocol version 4 (IPv4) and Internet Protocol version 6 (IPv6) [55]. As compared toIPv4, which can support 2^{32} = 4.294.967.296 addresses, IPv6 has a larger address space, providing 2^{128} = 340.282.366.920.938.463.463.374.607.431.768.211.456 addresses. IP addressing has a critical importance in any IP network because it uniquely identifies all the network nodes (stations and routers).

4.2.2 UDP (User Datagram Protocol)

User Datagram Protocol (UDP) [56] is one of the main transport protocols in IP stack. It is suitable for transporting delay sensitive data (e.g. voice, video) which have real time characteristics because it does not provide any reliability in terms of ordered delivery and duplicate protection, and no means of congestion control. It is meant to use datagrams (data in small fragmented packets) in a packet-switched interconnected network with a minimum protocol mechanism. UDP provides application multiplexing (via port numbers) and integrity verification (via checksum) of the header and payload.

The UDP header consists of only 4 fields:

- Source port Number – this field identifies the port to reply if needed. If this field it is not used, its value should be zero.

- Checksum – this field is used to check the UDP packet for transmission errors.
- Destination Port Number – this field identifies the receiver port. If the destination of the packets is the source host, then the port number will be a well known port number while if the destination is the client, then the port number will be temporarily assigned.
- Length – this field is used to specify the length in bytes of the entire UDP packet. The minimum length is 8 bytes while the maximum theoretical length can be up to 65.535 bytes.

First two fields are optional in case IPv4 is used while in IPv6 only the first one is optional. Figure 34 presents the UDP packet format:

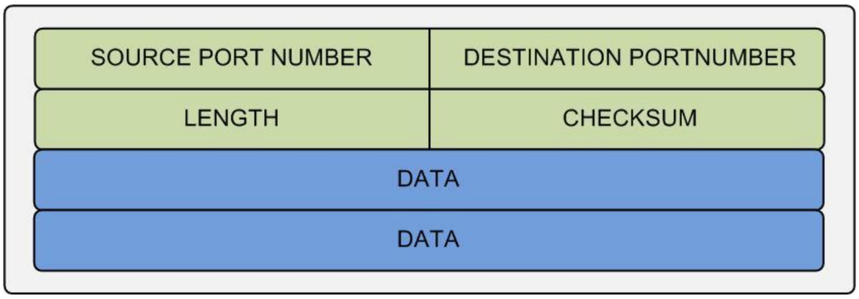


Figure 34 - UDP packet structure

Generally, applications using UDP do not require reliability mechanisms. In these conditions, some network-based mechanisms and elements (e.g. routers using packet queuing and dropping techniques) are required to reduce the potential congestion that can appear due to excessive traffic generated by UDP applications (streaming media, video gaming, voice over IP).

4.2.3 TCP (Transport Control Protocol)

The Transmission Control Protocol [57], [58] is one of the two components of the Internet Protocol Suite and together with IP protocol is building the TCP/IP suite. It is the most widely used Transport layer protocol on the IP protocol stack and it guarantees ordered and reliable delivery of data packets at the cost of delay. TCP is considered to be a reliable, connection-oriented, congestion control byte stream service and all major internet applications like World Wide Web, e-mail or file transfer are using it.

TCP was designed to provide a communication service between an application program and the IP. If one application program sends a large amount of data across the network using IP, a request to TCP to handle the IP details is sent by the application. This way, breaking the data into small IP-sized packets is no longer necessary.

The TCP segment header contains 11 fields, of which only one is optional:

- Source port – field used to identify the sending port number
- Destination port – field used to identify the receiving port number

- Sequence number - field used to identify the correct sequence number (initial sequence number if SYN flag is set, accumulated sequence otherwise)
- Acknowledgement number – field used to identify the next sequence number
- Data offset – field used to specify the TCP header size (minimum 20 bytes and maximum 60 bytes)
- Reserved – field that should be set to zero
- Flags – field used to set all 8 flags from TCP header: CWR, ECE, URG, ACK, PSH, RST, SYN and FIN
- Window – field used to specify the size of the receive window
- Checksum – field used to check the TCP segment for transmission errors
- Urgent pointer – field used to indicate the last urgent data byte, if the URG flag is set
- Options – optional field used to set some TCP options like selective acknowledgement, window scale, TCP alternative checksum request etc.

The format of a TCP segment is illustrated in Figure 35:

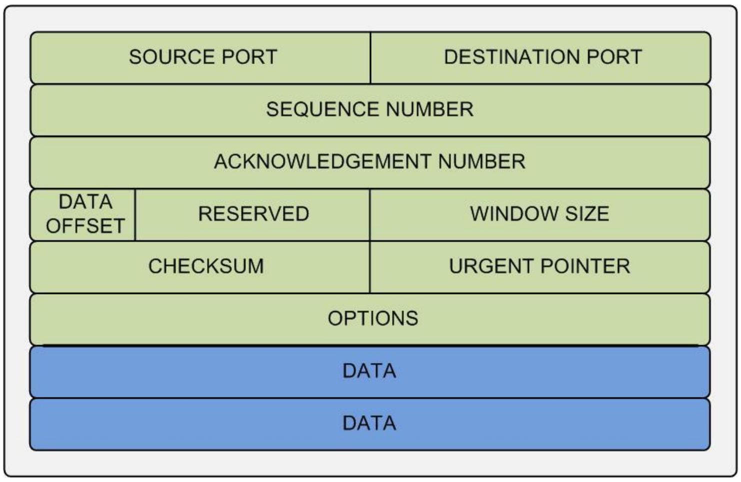


Figure 35 - TCP segment structure

Extensive research has been done to optimize the TCP for wireless network because originally the protocol was designed for wired networks. Normally, in a wired network, any loss of packets is the result of congestion and in consequence the congestion window size is reduced. But since in wireless networks the loss of packets can appear from other reasons than congestion (fading, shadowing or handover), reducing the window size will have as a result the underutilization of the radio link. The proposed solutions can be end-to-end [59], link layer solutions [60] or proxy based solutions [61].

4.2.4 RTP (Real-Time Transfer Protocol)/RTCP (Real-Time Transport Control Protocol)

Real-time Transport Protocol [62] was developed by the Audio-Video Transport Working Group of the IETF (Internet Engineering Task Force and it defines a standardized packet format for delivering audio and video over the internet. RTP is an upper layer protocol which provides services for end-to-end delivery of data that has real time characteristics which generally runs over UDP. The RTP architecture is depicted in Figure 36:

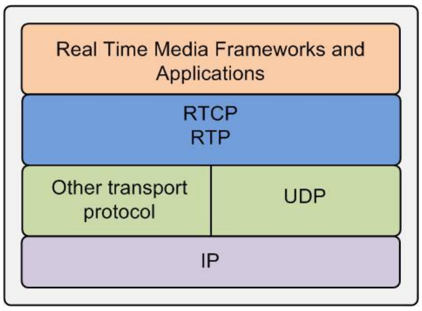


Figure 36 - RTP architecture

For real-time multimedia applications that require on-time delivery of data but that can cope with some packet loss, RTP can provide the means to compensate the jitter and to detect the senders' sending sequence using sequence numbers. The protocol can be used for both unicast and multicast sessions and it is regarded as the primary standard for audio/video transport in IP networks. The header format for RTP is presented in Figure 37 and it contains the following fields:

- Ver. – field used to indicate the version of the protocol
- P – Padding field, used to indicate if there are any extra padding bytes at the end of the packet
- X – Extension field, used to indicate the presence of an extension header
- CC – CSRC count field, that contains the number of CSRC identifiers
- M – Marker field, used to indicate the relevance of the current data for the application
- PT – Payload Type field, used to indicate the format of the payload and the way it should be interpreted by the receiving application
- Sequence Number – field used to indicate the senders sending sequence
- Timestamp – field used to synchronize the receiver so that it can play back the received samples at the right interval
- SSRC identifier – field used to uniquely identify the source of a stream
- CSRC identifiers – field used identify the contributing sources for a stream that is generated by multiple sources
- Extension header – optional field used to indicate the length of the extension

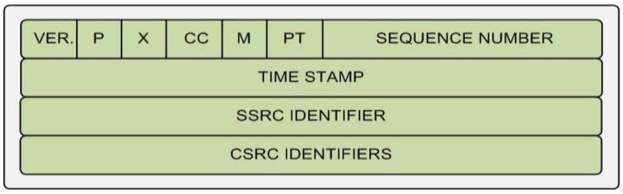


Figure 37 - RTC header

According to the standards' specifications, RTP uses two so called sub-protocols, one for the data transfer and one for feedback and synchronization between media streams [63]. The data transfer protocol has the role to transfer real-time multimedia data along with additional information for synchronization and for packet loss detection. The sub-protocol responsible with QoS feedback and synchronization of different media streams is called Real Time Control Protocol and its traffic is around 5% when compared to RTP [63]. RTCP is not providing flow encryption or authentication methods but it offers information about packet counts, jitter, round trip delay, lost packets count etc. The primary function of RTCP is to provide feedback on the QoS in media distribution by sending periodically relevant information to participants in a streaming multimedia session. This type of feedback information can be used by the source for transmission fault detection or for adaptive media encoding. RTCP also provides canonical end-point identifiers (CNAME) to all session participants. CNAME is used to establish a unique identification of end-points and for third-party monitoring. In order to limit the protocol traffic when reports are sent to multiple participants involved in a multicast session, dynamic control of the frequency of report transmission is done by session bandwidth management function.

The above functionality of the protocol is possible using five types of RTCP packets:

- Sender Report (SR): used to report transmission and reception statistics for RTP packets by the active senders. This packet is important for synchronization when both audio and video streams are transmitted simultaneously
- Receiver Report (RR): this type of packets are sent by passive participants in order to get statistical information
- Source Description (SDES): used to send the CNAME identifier to all users involved in a session
- End of Participation (BYE): used by the source to notify that is about to leave the session
- Application-specific Message (APP): used to design application-specific extensions to the RTCP protocol

RTP/AVP (Audio Visual Profile) is a profile used to interpret the generic fields of RTP and to associate RTCP for audio and video conference with minimum control. This profile also defines a payload type for audio and video data.

RTP/AVPF (Audio Visual Profile with Feedback) profile is an extension of RTP/AVP that allows more relaxation in feedback timing. In this profile, the receiver feedback can be used more effectively by implementing fast adaptation and repair mechanisms as it is possible to immediately send a feedback to report a particular event. Three types of feedback messages are defined in this profile:

- Transport layer feedback messages: are independent of codecs or applications and are offering general purpose feedback information
- Payload specific feedback messages: carries specific information for a particular payload type
- Application specific feedback messages: carries specific information for a particular application

4.2.5 Explicit Congestion Notification For Video Traffic In LTE

The Explicit Congestion Notification (ECN) mechanism was first proposed in 1999 [64] as an experimental protocol for TCP. From 2001, ECN was defined as an extension to the IP and to TCP in RFC 3168 [58]. Generally, dropping packets in a TCP network is an indication of congestion. In order to increase the QoS of delay sensitive services like audio or video streaming in TCP networks, the use of ECN will allow for an end-to-end notification of congestion without dropping any packets. ECN works by setting a mark in the header of an IP packet to signal the congestion. This way, the packet will not be dropped by the ECN-aware router and the receiver will offer feedback to the sender about congestion in the network. Using the ECN mechanism implies that both end-nodes and all network elements are ECN-aware, otherwise some network equipments may just drop the packets with the ECN bits set instead of ignoring it.

ECN mechanism uses two bits from the IP header (less significant bits from SERVICE TYPE field) to send congestion information (Figure 38). Depending on the value of the two bits, we have four different code points:

- 00: indicates that the packet is not using ECN
- 01 and 10: indicates that the transport protocols are ECN capable
- 11: set by the routers to inform the senders and receivers about congestion

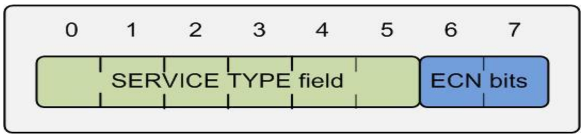


Figure 38 - SERVICE TYPE field from IP header with the 2 less significant bits used for ECN

ECN support for TCP traffic being already standardized, there is an increased research interest around ECN support for UDP but nothing is yet standardized. One solution offered in [65] is to keep the UDP protocol intact while adding some minor changes to the RTP header in order to signal the congestion to the participants.

To make ECN effective, Active Queue Management (AQM) policy has to be used. The benefits obtained by using ECN are a reduced number of dropped packets, a reduced latency and a reduced jitter. Some problems can appear when ECN with AQM is used in highly loaded networks because the packets are never dropped and the load on the network is increased. To overcome this disadvantage, specific AQM implementations are needed, that will rather drop packets than mark them in high loaded network conditions.

4.3 Multimedia Streaming Algorithms For Wired And Wireless Networks

Important research efforts have been put in investigating how to provide better QoS and a guaranteed bandwidth for multimedia traffic. One approach used to improve the multimedia streaming performances is the **end-to-end approach** [59] which assumes that all the intelligence should be at the end points while the network is dumb. The role of the network in this approach is just to offer the minimal service set needed to transport data packets from one end point to another. Figure 39 is illustrating the end-to-end approach:

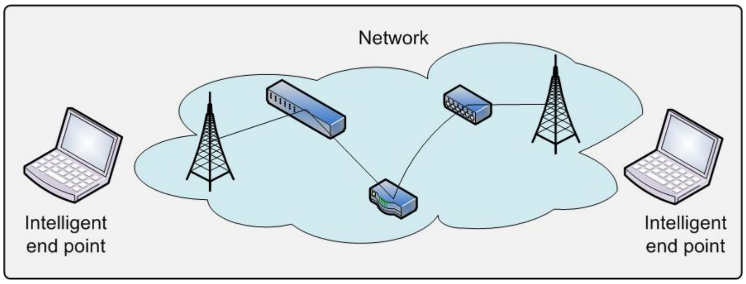


Figure 39 - End-to-end approach

The two most popular protocols for data transfer over IP networks that follow an end-to-end approach are TCP and UDP. The use of these protocols for streaming applications is therefore an expected decision, but neither one of them are suited for video streaming. This is because UDP protocol offers too less support while TCP protocol offers too much. TCP's characteristic throughput fluctuations have a negative effect on video quality because the video streaming process requires continuous bandwidth availability and controlled end-to-end delays. Even if TCP is a very reliable protocol and very robust to congestion in wired networks, the retransmission mechanism activated by the NACK feedback when packets are dropped, causes unacceptable pauses in media playback while the streaming application is waiting for dropped packets to be retransmitted. The use of large receiver side buffers [66] can help overcoming this problem but it introduces a new one, big startup delays. Unlike TCP, the unreliable UDP protocol is not offering real support for congestion avoidance, but is providing the necessary requirements to build application specific mechanisms able to implement rate control mechanisms for a smoother throughput variation as compared to TCP, while maintaining the friendliness.

Streaming Media Congestion Control (SMCC) [67] protocol is offering an end-to-end approach to rate and congestion control. The protocol estimates the bottleneck bandwidth share of a connection using algorithms similar to those introduced in [68]. The sender transmits data packets to the receiver and each time the receiver identifies a lost packet, it sends a negative acknowledgement feedback

(NACK) to sender. The sender is informed about the lost packet and can initiate a retransmission if the dropped packet can be delivered before the event horizon is reached. The feedback packet sent by the receiver contains also information about the current bandwidth (Bandwidth Share Estimate or BSE) that is used by the sender to adjust the transmission rate based on current network conditions. Like in TCP, SMCC increases its sending rate by one packet per RTT until it is informed that packets are lost. When a NACK feedback is received, the sending rate is set to BSE and after one RTT, if there are no packets lost, the sender will start increasing linearly the sending rate with one packet per RTT. Test results are showing that SMCC has a good behavior in environments where that are subject of random packet loss.

Another congestion control algorithm that uses the end-to-end approach is the **Rate Adaptation Protocol** (RAP). Unlike SMCC, in RAP every packet of received data is acknowledged by the receiver with an ACK feedback sent to the sender. Based on this feedback packet, the sender can estimate the RTT and can detect any losses. RAP protocol was compared with TCP in [69] and it was proven that the performances obtained are lower than those of the TCP protocol. RAP performances can be improved if queuing routers are added in the network.

Because TCP protocol has built in its own congestion control mechanism, the traffic generated will be generous in nature and it will make way for other types of traffic if the network is under congestion conditions. But if the other traffic flows do not address the generosity of TCP, then TCP will tend to use as much network bandwidth as possible. The idea of TCP friendly traffic is introduced first time in [70], where **TCP Friendly Rate Control** (TFRC) protocol is proposed. TFRC is an equation based rate control algorithm that trades off the benefits between UDP and TCP like approaches. The source uses an equation which is a function of RTT, packet size (*s*) and packet loss probability (*p*) in order to determine the sending rate:

$$X = \frac{s}{RTT\sqrt{\frac{2bp}{3}} + 12 \text{ x } RTT \text{ x } p\sqrt{\frac{3bp}{8}}(1 + 32p^2)}$$

TFRC algorithm was designed for best-effort unicast multimedia traffic while being fair when sharing resources with TCP flows. The fact that TFRC throughput variation over time is low recommends it for streaming media applications. But TFRC performances in congestion control, when used for streaming applications in wireless environments are affected by the specific characteristics of the radio interfaces. In wired networks, packet loss is the result of congestion in the network, while in wireless networks packet loss is mainly generated by the propagation problems through the physical channel. Because TFRC was designed for wired networks, it assumes that all losses are generated by network congestion. As a consequence, it cannot distinguish the propagation losses, treating them as congestion losses.

An adapted version of TRF to wireless environment is proposed in [71]. **TFRC Wireless** (TFRC-W) has the same rate equation as TFRC but the packet loss probability is considering also the random losses introduced by the wireless environment. These losses are recognized using the Loss Discrimination Algorithm (LDA) and they are not considered when the packet loss probability p is computed. LDA uses the RTT measurements in order to determine if a loss is generated by congestion or by the wireless medium. If the RTT measured is high, this indicates congestion and therefore the cause for the loss of packets during this period is less likely to be the wireless environment.

The **MULTFRC** mechanism proposed in [72] is designed to support streaming media applications over wireless networks. The idea was born from the work that was investigating the use of multiple concurrent TCP connections for streaming. In order to acquire more bandwidth, multiple TCP connections are opened, increasing the competition with the other data flows. Considering the fact that fairness between TCP friendly applications is based on their individual connections, the use of multiple TCP connections can increase the throughput that an individual application can achieve. This solution uses the existing network infrastructure but it requires a complex scheduling algorithm able to deliver relevant data chunks in a timely manner. The use of a utility function and a discovery mechanism is also required for mapping the device characteristics into a relevant number of streams.

Another algorithm based on TFRC that can provide a solid support for QoS of multimedia applications is proposed in [73]. **TCP Friendly Rate Control with Compensation** (TFRCC) uses the same rate equation as TFRC but it is QoS aware, considering the QoS requirements of the application. If the transmission rate calculated based on actual RTT, packet size and packet loss probability is lower than the threshold needed to satisfy the applications' QoS, the algorithm allows a temporary adjustment of the rate, increasing it until it satisfy the requirements of the application. These temporary adjustments are breaking the TCP friendliness but in long term the TCP friendliness is maintained.

Content-aware TCP-Friendly Congestion Control for Multimedia Transmission [74] is another solution proposed by the research community based on the TFRC algorithm. The novelty introduced by this solution is the new mechanism that is able to optimally prioritize the multimedia classes for the transmission, besides adapting the congestion window size in relation to the measured packet loss rate. While other solutions adapt the user's sending rate based only on the actual network conditions, this algorithm also uses the application characteristics in the adaptation process. Among interest factors, the distortion impacts, the delay constraints and the dependencies between multimedia packets of different classes are considered. In order to maximize the quality of the multimedia stream over a longer period of time, the algorithm uses finite-horizon Markov process (FHMP), being able to improve the PSNR with more than 3dB, when compared to the conventional TCP congestion control approaches [75].

Video Transport Protocol (VTP) [76] is an algorithm designed for real time streaming over wireless networks. Using the same technique like TCP, VTP is continuously monitoring the network parameters until congestion is detected. Because it uses the LDA algorithm, VTP can distinguish the congestion and wireless medium losses, thing that makes it suitable for use in a wireless environment. After a congestion loss is detected, the mechanism does not decrease the sending rate like TCP. Instead, the rate is lowered to the last throughput value that was successfully received. This rate called Achieved Rate (AR) is used to avoid the severe rate reduction that has a negative impact on video perceived quality. In long term, the VTP is able to maintain the same average throughput as a TCP connection, but without the fluctuations that are affecting the streaming process by maintaining a reduced rate for a longer period of time.

Another multimedia adaptation method proposed for video streaming in heterogeneous networks is **Dynamic Adaptive Streaming over HTTP** (DASH), enabling the streaming of high-quality multimedia content over wired and wireless networks using conventional http web servers. Also known under the name of MPEG-DASH, the adaptive bitrate streaming solution that became an international standard in November 2011 [77] operates by partitioning the multimedia files into a number of segments and delivering them to the clients using http. Each segment resulted after partition is described by a set of information that includes the URL, the video resolution and bit rate. Because it is expected that this technology will become very popular, 3GPP adopted it and developed a version called 3GP-DASH [78]. This solution is able to handle various service types like on-demand or live, while providing a rich feature set: adaptive bitrate switching, insertion of ads or multiple language support. In general, streaming solutions like RTSP are state oriented, meaning that once a user-server connection is established, the streaming server will continuously sends the media towards the client as a stream of data packets over UDP or TCP. As HTTP is a stateless protocol, the http servers will respond to any client demands by sending the requested data and afterwards closing the connection. A method of pseudo-streaming over http is the progressive download, but as it is used on large scale, this option brings some disadvantages through the fact that it is not able to support live streaming sessions and also it is not able to avoid the resource wasting when a user decides to stop using the content but it doesn't terminate its progressive download session. The disadvantages of progressive download strategy are eliminated when DASH method is used, because in this case the client has to make use of the http GET method in order to obtain another segment of its streaming session. In this way, the client has total control over the multimedia streaming session, being able to manage the on-time requests or changing other attributes like bit rate.

The research community noticed the potential of this new technology and already started to develop and propose solutions for obtaining an optimal streaming strategy for DASH. As presented in [79], one possibility is to model the bitrate adaptation as a Markov Decision Process (MDP) with the target to optimize the final user perceived quality of experience. Characteristics like video streaming smoothness, average quality or interruptions are considered in order to obtain an optimal MPD solution each time the user has to make a request for a new segment. Because the MPD is able to address the random network conditions using a relatively simple approach, the solution is well suited for handheld devices which have a limited energy supply. The results are showing that this method is feasible and even more it is able to obtain good results, improving the user QoE in comparison with other existing approaches.

Battery and Stream-Aware Adaptive Multimedia Delivery mechanism (BaSe-AMy) [80] is a multimedia adaptation algorithm that aims at maximizing the mobile device's battery life by adapting the bit rate of the streamed video file, while keeping a high level of QoS. In this case, the multimedia adaptation process is triggered when the algorithm detects that the battery life of the device drops below a certain threshold or when the loss rate of the received stream is below a predefined value. When the threshold is reached, the adaptation mechanism selects another version of the same file, coded at a lower bit rate. It is shown that by monitoring the loss rate, the stream playback duration and the energy consumption, this algorithm is capable of improving the video quality by 4%, while increasing the battery performance of the considered mobile device by more than 18%, in comparison with the situation when no adaptation is performed.

Adaptive multimedia streaming is an interesting topic for researchers also when the wireless network considered is a cellular network. These types of networks are starting to show their limits in terms of high data rates delivery, as the number of intelligent terminals available in the market is exponentially increasing. Users' demands are putting a significant pressure on the network, especially if we consider that many of the popular applications available are requiring a high data rate multimedia transfer (video conferencing, live multimedia streaming, etc). One solution that addresses this problem is presented in [81], where a new method for multimedia transmission in two-hop cellular networks is proposed. The **Dynamic Time Slot Partitioning** (DTSP) algorithm is designed to enhance the TDMA scheme, presenting an efficient time slot allocation method, based on statistical multiplexing that is able to increase the available bandwidth resources by dividing the pre-defined time slots into a number of "minislots". Using the statistical multiplexing, the DTSP algorithm differentiates itself from the conventional time division multiplexing (TDM) through the possibility to arbitrarily divide the bandwidth into a number of channels and the assignment of time slots that often appear to be random for a user, depending on the amount of data to be sent or received. The results in terms of system capacity obtained by DTSP in downlink mode using adaptive modulation show an improvement of up to 41%, compared to other existing solutions.

Another end-to-end approach adaptive mechanism with very good results when multimedia traffic is streamed over a network is the **Quality-Oriented Adaptation Scheme** (QOAS) [8]. This algorithm, unlike TFRC, TFRCP, LDA or RAP, is taking into account the end-user perceived quality and is using this parameter in the rate adjustment process. This rate-adaptive scheme is a unicast multimedia steaming solution able to maximize the QoE using client and server-located components. On the client side there is a QoE parameters monitor named Quality of Delivery Grading Scheme (QoDGS) that constantly computes the quality of delivery scores of the received multimedia stream and sends it as feedback information to the server. On the server side, the Server Arbitration Scheme (SAS) is used to analyze the feedback from the client. Based on this analysis, SAS proposes adjustment decisions meant to improve the user QoE in the current network conditions. For the adaptation process, the server uses a number of states, each of them assigned to a different stream quality. The functionality of this algorithm is exemplified in [82] with five different server states corresponding to five different stream quality versions of the same multimedia instance. The difference between the five versions is given by stream parameters like resolution, frame rate or color depth. After one multimedia streaming session is initiated, the server will dynamically adjust the sending rate based on the QoDGS feedback. If congestion is detected in the network, QoDGS will report a decrease in the end-user perceived quality and it will switch to a lower quality version of the same stream, reducing the amount of data sent through the network. The QOAS adaptation principle is illustrated in Figure 40.

The simulation results are showing that compared to TFRCP, LDA+ and a non-adaptive algorithm, QOAS has a lower packet loss rate and in consequence a higher end-user perceived quality, then the number of clients is kept constant. Another major advantage of QOAS is the decreased rate of control feedback messages. While TFRCP uses acknowledgement for each successfully received packet and LDA+ uses RTCP packets for feedback, QOAS uses only a QoDscore at a time, reducing the bandwidth usage to a very low value, around 0.1%. QOAS algorithm was also tested in a IEEE 802.11 wireless environment [83] and the results out-performs those obtained when other adaptation techniques like TFRC or LDA+ are used. The number of simultaneously served clients while maintaining the same perceived quality is higher if QOAS scheme is used and also the total throughput is increased.

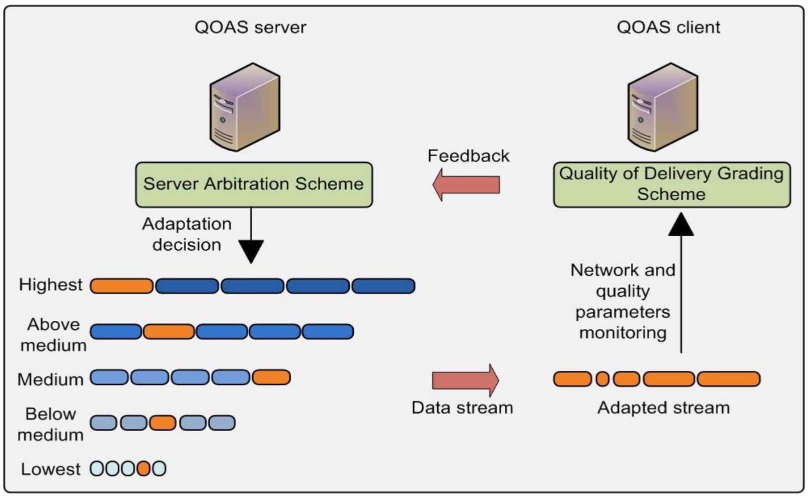


Figure 40 - QOAS adaptation principle

The second approach to provide better QoS and a guaranteed bandwidth for multimedia traffic is the **network centric approach**. Unlike the end-to-end approach, where the intelligence resided in the end nodes, here the intelligence is built into the network elements. Figure 41 illustrates the network centric approach.

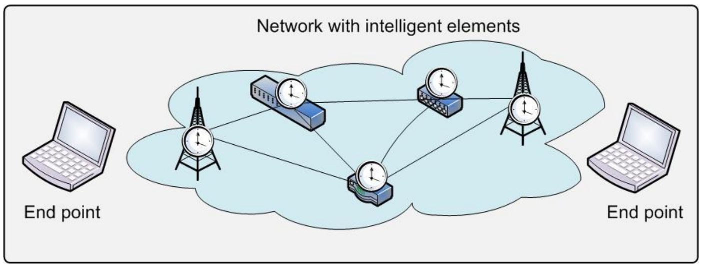


Figure 41 - Network-centric approach

One of the first architecture models used to achieve a certain QoS level was proposed by the IETF. The **Integrated Services** (IntServ) [84] model uses the Resource Reservation Protocol (RSVP) for a per-flow signaling in order to provide the required QoS. In this model, for every router in the network it is required to have the IntServ implemented, which means major modifications, to make possible the resource reservation for each data flow that needs QoS guarantees. The routers between the sender and the receiver have to decide if they can handle the requested reservation and if not, they must notify the received through a specific message. In case they decide that the reservation can be handled, they must deliver the required traffic.

The drawbacks of this model are the high complexity and the scalability problems that may appear when new elements are added to the network. This is the reason why this solution was used only in small network and it was never implemented on a larger scale.

To overcome the issues introduced by the IntServ model, IETF proposed another solution that uses a simple scalable mechanism to provide the required QoS guarantees for the data flows that are requiring a certain QoS level [85]. In contrast with IntServ, the **Differentiated Services** (DiffServ) is not managing the data based on traffic flows, it uses instead simple traffic classes. Instead of differentiating the traffic based on individual flows, DiffServ classifies each data packet, pacing it in the required traffic class. In order to signal that a data packet needs preferential treatment, the SERVICE TYPE field from the IP header is used. This model does not specified which traffic classes have higher priority, instead leaves this for the operator. In this model, the network is a group of routers that are set by an administrator to form the administrative domain that uses a predefined list of forwarding rules.

Basically, DiffServ is just a mechanism that is able to decide what data packets should be dropped or delayed, when the link is overloaded. But this means that the link is already close to its limit, so instead of integrating a complex solution like DiffServ, the operator might as well extend the network capacity.

More recent, other mechanisms based on the network-centric approach were developed to improve the QoS parameters, **Cross-Layer techniques** being one of them [86]. By using Cross-Layer optimization techniques, the natural feedback path between layers, as defined by the OSI or TCP/IP model, is broken. Here, the signaling between layers can be performed in different ways, improving the latency or to compensating the overload from certain routing paths.

Some disadvantages that arise from the fact that this is a mechanism that avoids the waterfall concept implemented by the OSI model are the impossibility for a layer approach by the designer, the system stability or the implementation dependencies, as detailed in [87].

4.4 Resource Scheduling In Wireless Networks

Scheduling is the process that dynamically allocates the physical resources among the users based on a well defined set of rules called scheduling algorithms. Another technique closely related to the scheduling process is the link adaptation and it refers to the adequate selection of the modulation and coding scheme (MCS) to be used. The selection of MCS and the good functionality of the scheduling process are based on feedback data transporting network and user-related information. Depending on the wireless technology used, different scheduling algorithms have been developed in order to overcome the problems generated by the unpredictable radio interface. In IEEE 802.11 environment, the original scheduling mechanisms provided fair scheduling for best effort traffic. Using this type of scheduling algorithms without any kind of data flow prioritization, the results of multimedia streaming in terms of end user perceived quality are poor. IEEE 802.11e is offering a solution by introducing the priority based scheduling via Enhanced Distribution Coordinator Function (EDCF) that offers certain QoS guarantees for multimedia applications. Designing a competitive scheduling algorithm for wireless networks needs to take into account the limited bandwidth available and the distributed nature of the radio channels. An overview of the scheduling mechanisms for WLAN is presented in [88] while a detailed description of available scheduling algorithms for LTE downlink can be found in [89].

The Distributed Weighted Fair Queue (DWFQ) [90] used for WLANs adapts the contention window size depending on the difference between the actual and the expected throughput. To create a proportional bandwidth distribution and to allocate bandwidth for each data flow according to their queues weights, DWQF uses the CW mechanism of the IEEE 802.11 MAC DCF. Depending on the window size, a flow will receive more or less bandwidth (e.g. if the CW is small, the achieved throughput will be high). Using this strategy, two algorithms are actually proposed by the authors: in the first one, if the actual throughput is higher than it is expected, the CW size will be decreased in order to give priority to that flow; in the second one, the CW size is adjusted after comparing the flows' requirements.

[91] proposes the **Persistent Factor DCF**, which assigns a persistent factor P to each traffic flow. The value of the persistent factor depends on the traffic class priority in the way that for a high priority traffic class P has a small value while for a low priority class, P has a larger value. During the backoff period, a random number r is assigned to each time slot and the transmission of a flow in a time slot can only start if the condition $r > P$ is satisfied.

Distributed Fair Scheduling (DFS) wireless scheduling technique is introduced in [92] as an extension to the Self-Clocked Fair Queuing [93]. The advantage of this method is that both prioritization and fairness are taken into account in the scheduling process. This way, a station will always perform a back-off for every packet to be transmitted. The back-off period is a function of packet size and a prioritization parameter that can be seen as the stations' weight. A high station weight means that the priority for traffic generated by that station will be high. The fairness of the technique comes from the packet size consideration in computing the back-off interval. The data flows consisting of small packets (e.g. VoIP traffic) will be sent more often.

Another scheduling method that adapts the ARQ limit dynamically is detailed in [94]. The **Content Aware Adaptive Rate** scheduling mechanism modifies dynamically the ARQ limit according to the carried content. This algorithm is specially adapted for video transmission because it considers the packets' priority based on its position within the Group of Pictures (GOP). If a packet contains the I-frames, the mechanism will try really hard to retransmit it, while if a packet contains the B frames, the retransmission effort will be low.

As Long Term Evolution technology evolves and important operators in telecom world announced their interest for LTE, researchers are starting to develop algorithms capable of improving the network delivery. Their work concerns both the uplink and downlink, considering multiple solutions for implementing scheduling algorithms in different traffic conditions, considering multiclass flows. The scheduling methods are looking for improving the system capacity in terms of number of QoS flows that can be supported and also for reducing the resource utilization. Reference [89] divides the work done in this area into two categories, based on the type of traffic the scheduler was designed for: scheduling for elastic (non-real-time flows) [95] and scheduling for real-time flows [96]. LTE schedulers can also be classified based on their awareness parameter(s) into channel-aware schedulers [97], queue-aware schedulers and queue and channel-aware schedulers [89].

Regarding the Uplink schedulers, a lot of work has been done. A performance comparison on control-less scheduling policies for Voice over IP (VoIP) in LTE UL was conducted in [98] and it was proven that semi-persistent scheduling obtains better performances that group scheduling when no group interactions occurs for group scheduling. In [99], the authors suggested an opportunistic scheduling algorithm based on the gradient algorithm called **Heuristic Localized Gradient Algorithm** (HLGA) that allocates resource blocks to users while maintaining the allocation constraint and considers retransmissions requests. Channel-aware scheduling algorithms for SC-FDMA are proposed in [97] in local and wide area scenarios. The first two, **First Maximum Expansion** (FMA) and **Recursive Maximum Expansion** (RME), represent simple solutions for localized allocation of the resource blocks, whereas the third algorithm, **Minimum Area-Difference to the Envelope** (MADE), is more complex but performs closer to the optimal combinatorial solution.

The QoS aspects of the LTE OFDMA Downlink are influenced by a large number of factors – channel conditions, resource allocation policies, available resources, delay sensitive/insensitive traffic, etc – and therefore new means were needed to enhance QoS beyond what the default IP service provided. This problem is addressed in [89], where a new scheduler for LTE downlink is proposed. The performance of this scheduler is analyzed using multiclass traffic and the results are indicating that a channel- and queue-aware scheduler is a good choice for LTE DL. The work in [100] showed that strict prioritization for session initiation protocol (SIP) packets over other packets – voice and data – can lead to better overall performances. References [101] and [102] are analyzing the packet scheduling of mixed traffic in LTE DL. The results in both are showing that it is necessary to perform service differentiation and prioritization of delay-critical traffic as VoIP traffic, especially when in combination with delay-insensitive traffic like web surfing or TCP download. Some of the basic LTE scheduling algorithms relevant to this thesis work are described in the following paragraphs.

The **Round Robin** (RR) scheduler is a very simple scheduler that allows users having data to transmit to take turns without considering the channel quality information. RR can be considered a fair scheduler because every user has the same amount of time and the same radio resources at his disposal for transmitting the queued data packets. But in terms of spectral efficiency, this algorithm performs very poor because it is not considering the instantaneous channel conditions in the scheduling process. The **Frequency Division Multiplexing** (FDM) scheduler is a particular case of the RR scheduler where all users are scheduled each TTI and benefits form an equal share of frequency resources. The performances of this scheduler are similar to those obtained by RR.

Maximum Throughput (MT) scheduler has a high spectral efficiency but it is not a fair algorithm because it gives an advantage to the users that have the best channel conditions at a given time. This is happening because the scheduled user for a TTI is selected based only on the instantaneous channel conditions.

One algorithm that realizes a tradeoff between spectral efficiency and fairness is the **Proportional Fair** (PF) scheduler. Both the instantaneous channel conditions and the users' past average throughput are considered in the scheduling process, offering the same throughput for each user: $M_n=d_n/r_n$, where d_n is the instantaneous supported throughput and r_n is the past average throughput. Thus, the users who have the best channel quality relative to their average channel quality will get scheduled.

5. PROPOSED ALGORITHM FOR MULTIMEDIA DELIVERY

This thesis chapter describes the proposed multimedia streaming algorithm, Dynamic Quality-Oriented Adaptive Scheme (DQOAS). The chapter starts by presenting the necessity of the proposed algorithm, followed by the description of the e-learning system where the algorithm was first introduced with good results. The last part describes the main components of the adaptation mechanism and gives an exhaustive presentation of the DQOAS functioning principle

5.1 Introduction

The general understanding of an adaptation process that involves a sender and a receiver is that it will use an adaptive behavior that will allow the sender to control and change its sending rate depending on the availability of network resources. One definition for adaptation is given in [103]:

> *"Adaptation is a technique for monitoring network utilization and manipulating transmission or forwarding rates for data frames to keep traffic levels from overwhelming the network medium"*

But in order to design a competitive adaptation mechanism for congestion avoidance in wireless environments, one must offer a solution for both network and user-related problems. Considering this, a newly proposed algorithm must be capable to offer QoS guarantees and a satisfactory end-user perceived quality for the multimedia content being delivered. An analysis of the most common network problems that can affect the multimedia streaming process is done in chapter 2, where parameters like delay, throughput, packet loss rate or jitter are discussed, along with some solutions already proposed by researchers. The difficulty in designing a performant algorithm appears when one has to analyze the user-related parameters and more important, when one has to find a way to describe and quantify how these parameters are impacting the multimedia streaming process. The problem arises from the fact that the behavior and the expectations of an end-user (or a group of end-users) has highly dynamic characteristics and it is difficult to define a set of parameters and rules able to describe the continuously changing user environment.

Considering the definition above, an adaptive process is complete if just the network-related problems are addressed. Contrary to this conclusion, the author of this thesis has the opinion that a complete adaptive process needs to consider the network-related problems, but also the end-user preferences and requirements, because both network and user-specific requirements can be finally translated into end-user satisfaction level. If a user declares himself satisfied by his experience in the network, one can conclude that the adaptation process applied for that specific user was successful. But as the actual wireless networks are able to accommodate more than one user, a new variable must be introduced in the problem: multiple users treatment. So in order to design a competitive adaptation mechanism, all (or the majority of) users must finish the multimedia streaming session and be satisfied by its quality.

Before starting to implement the algorithm, three questions had to be answered:

I. Who is responsible with the adaptation process?
II. How is the adaptation performed?
III. When should the adaptation be done?

For the first question, three solutions are available in the literature: server side adaptation, client side adaptation and in-network adaptation. The solution chosen in this thesis uses the first option, server side adaptation using client feedback. The multimedia adaptation process is performed on the server side by dynamically modifying the data stream transmission rate, using as input data the network conditions and user preferences. Server side adaptation method was chosen because it is considered to be suitable for heterogeneous environments like wireless networks, where the clients may have different medium access technique compared to the server, and can experience a wide range of congestion scenarios. One other big advantage offered by server side adaptation is that it provides more scalability compared with the other two options.

In order to answer the second question, four different solutions have been analyzed:

- Adaptive Increase and Multiplicative Decrease (AIMD) [104], where the sender, based on the feedback messages, will either decrease the sending rate with a constant multiplication value or increase it with a constant additive value:

$$R_{(i+1)} = R_i + \alpha,\ if\ C_i = 0 \text{ or}$$
$$R_{(i+1)} = R_i \times (1-\beta),\ if\ C_i \neq 0 \quad (1),$$

  where C_i represents the congestion notification and R_i is the sending rate at time t_i. α is the increase constant and β is the decrease constant.
- Adaptive Increase and loss Proportional Decrease (AIPD), where the sender adjusts its sending rate proportional to a fraction of packet loss rate:

$$R_{(i+1)} = R_i + \alpha,\ if\ f_i = 0$$
$$R_{(i+1)} = R_i\ x\ \beta x f_i, if\ f_i > 0 \quad (2),$$

  where f_i is the loss fraction.
- Adaptive Increase and Adaptive Decrease (AIAD), where the sending rate decreases and increases in a non-linear form. The modality in which the transmission rate is modified depends on one or more variables that can be network and user dependent. The equations used for this approach are variations of the equations (1) and (2), from AIMD and AIPD.

- Equation based Rate Adaptation, where the transmission rate is adjusted based on a particular equation. One algorithm that uses this approach is the TFRC, described in chapter 2. This approach is considered to be media unfriendly and it is not suitable for environments where the error rates are high.

For the algorithm proposed in this thesis, the approach used is the Adaptive Increase and Adaptive Decrease method because it is the only one that can use both network and user-related parameters in the adaptation process.

The answer to the last question was decided after two possible solutions were discussed: first option available is when the adaptation mechanism is not triggered by the congestions present in the network because it starts adapting the transmission rate before any of these congestions can occur; the second option available is to trigger the adaptation mechanism only when a congestion is detected in the network. The first approach is well suited for delay and packet loss sensitive applications but it has a high complexity in implementation. This is one of the reasons why the proposed algorithm uses the second method presented here, based on constant probing the network channels and user parameters. In order to overcome the problem that can arise due to delayed responses to network congestion that can degrade the QoS, the algorithm uses a "fast decrease, slow increase" mechanism, ensuring a fast reaction during bad network conditions, allowing them to improve and stabilize before a quality increase.

5.2 DQOAS Architecture

The proposed algorithm is named Dynamic Quality-Oriented Adaptation Scheme (DQOAS). DQOAS is a user-oriented adaptation mechanism that dynamically adapts the multimedia streams sent over the network, based on user preferences and actual network conditions [105]. By employing this algorithm, it is expected to obtain an increased number of users that can be simultaneously accommodated, while maintaining the perceived quality levels at or above their individual "acceptable quality" thresholds. This goal is achieved by using dynamic modification of the adaptation policy during the multimedia delivery process, or in other words, by increasing / decreasing the granularity of the adaptive data stream for certain users, based on their preferences and on the network parameters.

DQOAS is designed to perform an optimal multimedia content management according to three indicators: end-user profile, user preferences and network conditions. The dynamic adaptation performed by the proposed algorithm has the final goal to support a higher end-user quality of experience while increasing the total number of simultaneously active users. The e-learning system that was first used to develop and test DQOAS algorithm is named Performance-Aware Multimedia-based Adaptive Hypermedia (PAMAH). This e-learning system was proposed and developed by the researchers from the Performance Engineering Laboratory, Dublin City University and the design of DQOAS algorithm was done as a part of this project.

5.2.1 PAMAH e-Learning Adaptive System

Performance-Aware Multimedia-based Adaptive Hypermedia system was designed to enhance the classic Adaptive Hypermedia System architecture by including performance and QoS aspects in relation to adaptive multimedia content delivery. PAMAH supports the delivery of high quality personalized educational content to e-learners via heterogeneous networks. Its goal is to optimize users' QoE and their learning outcomes by automatically adapting content and navigational support based on user interests, knowledge levels and current network delivery conditions.

The problem of delivering personalized multimedia content presents a significant challenge. If the case of a single user is considered, the process of delivering multimedia content towards that user can be seen as an attempt to offer the highest data quality, with respect to the actual network conditions and the user's operational environment. However, multimedia content is comprised of a number of components each with different performance characteristics. The performance trade-off between the various components yields an interesting optimization problem. In addition, multimedia content may be paralleled by equivalent content such as diagrams or text (see Figure 42).

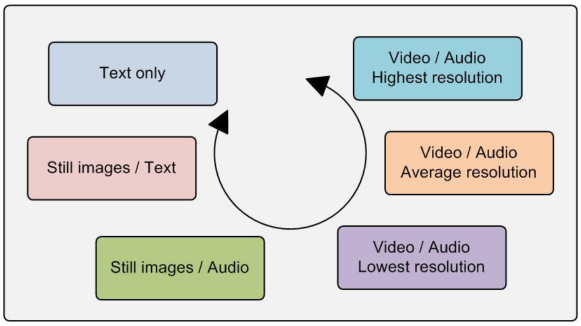


Figure 42 - Content adaptation in PAMAH

Therefore the decision of what content to deliver may be governed not only by what is feasible from a QoS standpoint, but also by the learning style of the individual user. It is likely that some users would prefer the static content (just as some people would rather read a book than watch a video). In addition, within a session, the user's preferred content might change (they might want the video the first time they go through the material, but afterwards prefer just to refer to a diagram to reinforce their learning). The problem is thus not simply one of maximizing the QoS of multimedia content delivery, but also requires individual (and possibly time-varying) user preferences to be taken into account.

PAMAH framework is build based on five main components: Domain Model (DM), User Model (UM), Experience Model (EM), Adaptation Model (AM) and Performance Model (PM). Each model has attached its own database, used to store information specific to that model: D Db, U Db, Exp Db, A Db and P Db.

- The Domain Model is designed to store the educational content, being organized in a hierarchical structure of concepts, amongst which logical relationships exist. A concept can be identified as a section of text, an audio file, a video clip, etc. All together with a set of rules, they form the educational units.

- The User Model is built and maintained by the system. Here some user-related parameters are assessed: user knowledge levels, preferences, goals. The multimedia delivery considers these parameters in the adaptation process, improving user quality of experience.
- The Experience Model is specially designed to assess user quality of experience, by storing information about user preferences of media or activities, learning goals, interaction preferences, preferred type of feedback, etc. Some of the stored information are behavior characteristics (behavior trackers) and others are experience dependent (experience trackers).
- The role of the Adaptation Model is to decide on personalization and performance adaptations to the content, based on the information gathered by the System Engine from PM and UM.
- The Performance Model performs real time monitoring for different factors that influence network delivery (e.g. throughput, RTT, QoS metrics, etc).

Figure 43 presents the architectural framework for the PAMAH system:

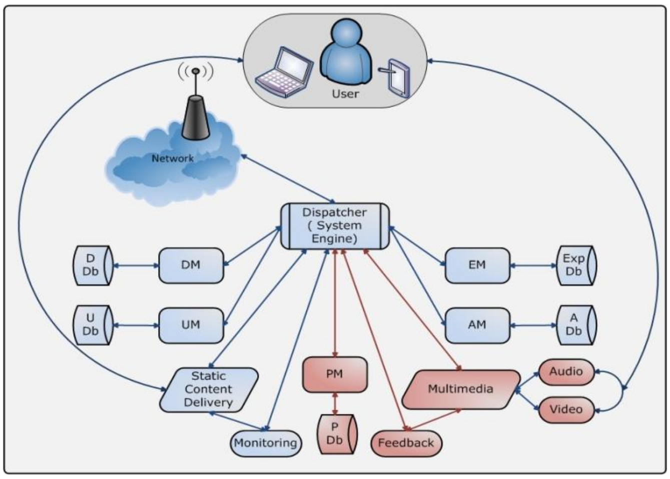


Figure 43 -PAMAH system – block-level architecture

All these models are controlled by and interconnected through the Dispatcher, or the System Engine. The Dispatcher uses information stored in different models and dynamically builds a list of rules which will be used further for both static content delivery and multimedia delivery. This novel PAMAH architecture is meant to improve the final outcome of the e-learning process by using dynamic adaptation techniques that consider both network-related and user-related factors.

Multimedia Delivery block implements the DQOAS adaptation algorithm presented above. The Feedback Module has the role to create a report that summarizes the delivery conditions and the user feedback for the delivered stream and sends it to the Dispatcher which is extracting the important parts and forwards them to the concerned modules. The network-related information is forwarded to the PM which is checking the personal database for any resembling patterns. After placing the current parameters in a certain category, the Performance Module generates a report for the Dispatcher. Using the report received from the Performance Module and also the reports from the other modules, the System Engine builds a list of rules that is sent to Multimedia Delivery block, as the current input for the adaptation algorithm. As the e-learning process itself is highly dynamic, DQOAS algorithm is designed to keep up with the continuously changing conditions and parameters involved in the e-learning process and to improve end-user quality of experience by adjusting dynamically its adaptation policy.

5.2.2 DQOAS Principles

Dynamic Quality-Oriented Adaptation Scheme algorithm is a network and user-aware adaptation mechanism that was designed and developed based on the need to increase the users' quality of experience during a multimedia streaming session, and in the same time to enable a higher number of simultaneously connected clients.

Considering the PAMAH architecture depicted in Figure 43, one role of the Adaptation Model is to perform real-time bandwidth estimation for each user involved in the learning process. Based on the estimation performed by the AM and on the feedback received from the other modules, System Engine decides if a certain video stream rate can be delivered or not to a specific user. In case the System Engine decides that a streaming session can be initiated (or continued) for a specific user, it builds a list of user-specific rules and forwards this list to the Multimedia Delivery block. Based on the user requirements found in the list of rules, Multimedia Delivery is able to compute the user-specific streaming parameters and the session starts.

Unlike Quality-Oriented Adaptation Scheme [8], which is a static adaptation algorithm that uses some pre-defined potential bitrate adaptive levels, DQOAS algorithm is highly dynamic, continuously adapting the multimedia bitrate levels in order to maximize the delivery process. DQOAS was designed to improve QOAS, by introducing the QoE expectation level parameter in the adaptation process. Like this, the algorithm will be able to apply differentiated treatment to specific users, the adaptive process not being based only on network parameters, but also on the user requirements and expectations in terms of quality of experience.

Figure 44 presents the component blocks and the architectural framework used to build the Dynamic QOAS algorithm that aims to improve the user QoE during mobile and wireless e-learning, deployed on the PAMAH concept structure detailed in section above.

The adaptation algorithm is located on the server side, but it needs also a component that runs on the client side. This client side component is highly important for the algorithm's functionality because it is designed to offer feedback for both user and network parameters.

Three functional blocks are used to implement the adaptation mechanism part situated on the client side: the Decoding & Playing module, the QoE estimator module and the Feedback module. The role of Decoding & Playing module is to decode and play the adapted video stream received from the server in real time. Connected to this block is a dedicated module that is designed to perform accurate estimations about the end-user perceived quality of the received multimedia stream. Finally, the Feedback module has the role of monitoring the delivery parameters (such as loss rate, delay and jitter) and also of assessing the quality of delivery.

Based on the observations and measurements performed, it builds and sends short real time reports to the server. The reports are offering all network-related and user-specific parameters needed for the adaptation algorithm located on the server side. On receiving the report, the System Engine will update the list of rules and parameters with new user specific information for all concerned user, updating also the relevant system databases (A Db and U Db). Then, the list containing user-related information and session-specific information will be fed to the N-Level Builder module that will decide if a new adaptation cycle must be initiated. This list offers information about user-related parameters (e.g. QoE expectation levels for users) and also about session-specific data (e.g. users to be considered for a streaming session, video metadata, etc).

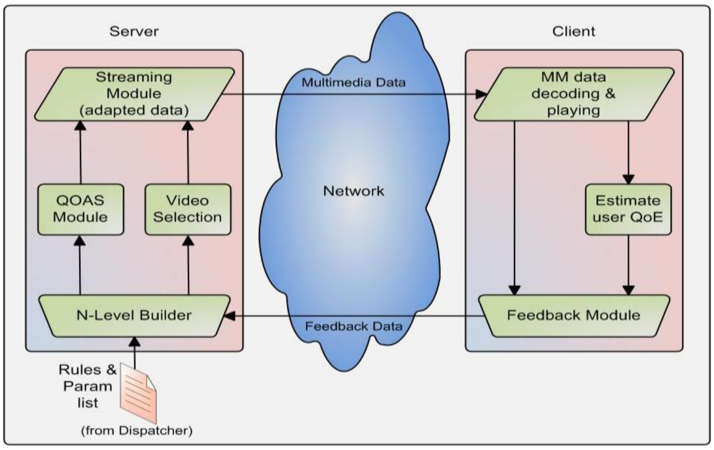


Figure 44 - DQOAS adaptation algorithm

The list of rules and parameters has an important role in the adaptation process, because based on the information found in it, the algorithm will decide what adaptation strategy should be adopted. If the data found in the list is not updated in real time or if the report arrives late at the N-Level Builder, there is a possibility that the algorithm will make its adaptation decision based on old data that do not reflect the actual network conditions or actual user preferences. To overcome this problem, each report must have a timestamp that will be checked by the N-Level Builder. Both parts of the adaptation mechanism (on client and on server side) are synchronized, so when the timestamp will be analyzed, an accurate decision can be taken related to the report relevance. If the report is considered "old", then it will be discarded and the settings will not change, if not, the adaptation process can start. The feedback message send by the client contains a quality coefficient Q, which is computed based on throughput, delay and loss using the following formula: $Q = (w_1 \times Q_{throughput}) + (w_2 \times Q_{delay}) + (w_3 \times Q_{loss})$, where the weights *w1, w2 and w3* were determined to be 0.4, 0.3 and 0.3 after extensive tuning of the algorithm. $Q_{throughput}$ is the ratio between the actual throughput sent by the server and the throughput experienced by user, the Q_{delay} represents the report between the packet delay and the expected delay while Q_{loss} is the ratio between the actual packet loss and the expected packet loss.

On the server side, DQOAS adaptation algorithm has three important functioning routines: Initial Level Building, Dynamic Update of Quality Levels and Adaptation Cycle. Every routine has specific trigger mechanisms and can run individually, depending on the conditions at a given time. The three routines are detailed below:

1. *Initial Level Building*

The Initial Level Building routine is called every time there is a new user in the wireless network that sends a request for multimedia streaming towards the server. Since this new user never requested a multimedia stream, hence never required any adaptation, the QoE Estimation Module has no history about the user preferences because it was never triggered to establish the minimum QoE expectation level for it. In this case, the level building will not be done dynamically because the algorithm does not know the minimum QoE level. The level building is done statically by employing directly the QOAS Module.

The newly detected user will be logged in the list of rules and parameters with the minimum QoE expectation level set to 0. After the list is transmitted to N-Level Builder Module, it is analyzed and the new user will be discovered. At this point, N-Level Builder will send a level request to QOAS Module that will create the static levels just for this user.

For a new user U_k the QOAS algorithm uses M potential levels of quality: L_1 L_2 … L_j … L_M. L_1 represents a low bitrate (the lowest bitrate level used by the other users) while L_M represents the maximum video bitrate available. As it was detailed in section 4.3, QOAS algorithm obtains good results when five granularity levels are used. Same number of initial levels is also used here, where the minimum level is L_1 and the maximum level is L_M as defined above, while the three intermediary levels are equally spaced in the interval between L_1 and L_M. The video transmission will start with the lowest level available, L_1, and depending on the feedback received, the algorithm may decide to increase this level. Even if L_1 is the lowest QoE level used for the users involved in the streaming session and it is possible that this level is not suited for the new user, the algorithm uses this bitrate because it is more likely to find faster the available conditions that will allow a quality increase.

After the streaming session starts, the system tries to estimate the QoE expectation level for the new user in order to prepare the start of the second step, the update of the quality levels based on the new user-specific parameter. After QoE estimation, one of the inputs in the rules list corresponding to this user indicates the video-quality level which represents the acceptable quality expectation level associated to this user, below which he will terminate its multimedia streaming session. Considering this level, DQOAS algorithm can start building another set of levels, this time having updated limits, other than the static levels: the minimum level is represented by the user minimum QoE expected level and the maximum level is the maximum video level, L_M.

2. *Dynamic Update of Quality Levels*

N-Level Builder is the module that has the role to make the decision if the dynamic quality level update procedure should be initiated for a certain user. This decision is based on the information found in the list of rules and parameters received from the Dispatcher, and also on the information received from the

Feedback Module. The dynamic quality level update procedure is triggered for every new user as soon as its minimum accepted quality level is reported to the server. For users with an active streaming session, this process can be triggered by changes in the monitored network parameters or by an update on their accepted quality level. Knowing the minimum accepted level $L_1 = M_i$ and the maximum accepted level as L_M for user U_i, the levels are dynamically built using the following formula: $L_n = Q \times L_{(n-1)}$, where L_n represents the n-th level and Q is the quality coefficient reported by the client. If the calculated value of L_n exceeds the maximum video bitrate, then this level will become the maximum level and its value will be L_M.

If the QoE Estimator detected that for user U_k, who already had passes one or more adaptation cycles a new QoE expectation level should be set, it will determine this level and set it to be M_{k_new}. Once M_{k_new} is determined and the parameters and rules list is updated, the N-Level Builder Module proceeds and rebuilds the adaptation levels for user U_k. Every time the user QoE expectation level is changing, the N-Level Builder will initiate the procedure to dynamically update the user levels. The user QoE expected level found in the rules and parameters is unique for each user and represents the minimum video bitrate that will be transmitted for that user.

While a new user has a fixed number of levels, a user who has a QoE expectation level assigned in the list, will have a variable number of levels, N. This value can vary depending on the network conditions, bandwidth estimation, user QoE expectation level and the network load. By using the user-specific levels of quality, the algorithm is able to obtain a very fine adjustment control of the video quality. Carefully manipulating the multimedia stream quality, by lowering or increasing the quality should improve the overall end-user perceived QoE, since it was demonstrated that viewers prefer a controlled reduction in the quality of the streamed multimedia content rather than random losses of a high quality multimedia stream [106].

3. Adaptation Cycle

Considering a wireless network where P users $U_1\ U_2\ U_3 \ldots U_k \ldots U_{P-1}\ U_P$ have an active e-learning session, there is a QoE expectation level that specifies the expected quality for every user: $M_1\ M_2\ M_3 \ldots M_k \ldots M_{P-1}\ M_P$. For any active user U_k, the algorithm builds the dynamic levels by dividing the amount of bandwidth between M_k and L_M levels into intervals. The resulted intervals are not equal, as the division is highly dependent on the network QoS parameters which are considered in real-time. If the network is loaded and most of the resources are allocated, the algorithm will build most of the levels close to M_k, in order to avoid any congestion. The congestion can happen when DQOAS algorithm will try to increase the streamed multimedia quality for that user, by increasing the transmission bitrate with one level. If the levels are having a small granularity and are close to the QoE expectation level of that user, the possibility of obtaining resources for a level increase is higher. On the other hand, if network resources are available and the congestions in the network are missing, then the levels will have a larger granularity in order to facilitate a faster quality increase.

At this step, the QOAS module is tuned on these new adapted quality levels obtained for user U_k and sets the delivered media at the established bitrate. The Streaming Module will take over and will send the selected data according to the input obtained from the QOAS module. If user U_k is tuned on level L_n, the server will receive feedback messages containing the quality coefficient, Q. Two consecutive coefficients are compared and based on the result, a decision to increase or decrease the quality level is taken by the algorithm. A level increase will be decided if the last coefficient is greater than the previous otherwise the level will be decreased. In case when a user is receiving the multimedia stream at the lowest bitrate available, L_1, and the feedback reports are suggesting another quality decrease, the algorithm will keep sending the traffic flow to the user at the minimum rate until a number of feedback messages are received. This is done because during this period, the quality level can be reduced for other users, releasing some radio resources and allowing this user to receive a stream at or above its minimum expected quality level. If these messages are still indicating a quality decrease after this waiting period, then the algorithm will terminate the current session for the user, freeing the resources.

N-Level Builder module applies the described adaptation mechanism whenever media-rich content fragments transmitted over the network need to adapt their bit rate to match the continuously changing network parameters and user requirements. The process of delivering such media rich fragments generally consumes significant network resources and lasts over a longer period of time, when the delivery conditions will most likely vary. In this case, random losses have a greater impact on the end-user perceived quality than a controlled reduction in quality.

Figure 45 presents using pseudo-code the proposed algorithm:

```
SERVER: Adaptation Module implementation (pseudo-code)

Begin

Is_Feedback (received package)
{       if (received package) is fbk then
                Analyze_Report (received package)
        else
                exit;
}

Analyze_Report
{
        Extract_Interest_Factors
        Estimate_User_Quality
        Network_Conditions
        Network_Infrastructure
        Decide_Rate
}

End

CLIENT: Feedback Module implementation (pseudo-code)

Begin

Is_Multimedia (received package)
{       if (received package) is Mm then
                Report (received package)
        else
                exit;
}

Report (received package)
{
        Calculate_Interest_Factors
        Build_Report
        Deliver_Report
}

End
```

Figure 45 - DQOAS algorithm – pseudo-code representation

One important function performed on the client side is the computation of interest factors. The parameters considered by the algorithm in calculating these factors are the encoded stream bit rate, the packet loss and packet delay, which are constantly monitored during the streaming session. The client regularly collects samples of these parameters and based on the length of the client monitoring window it averages these values and then computes the interest factors: $Q_{throughput}$, Q_{loss} and Q_{delay}. In LTE environment, one sample is collected every TTI while the monitoring window duration is set to 10 TTIs. The interest factors used are defined as follows: $Q_{throughput}$ is the ratio between the actual throughput sent by the server and the throughput experienced by user, Q_{delay} represents the report between the packet delay and the expected delay while Q_{loss} is the ratio between the actual packet loss and the expected packet loss. The obtained values are then forwarded to the server as a feedback packet, each time a monitoring window expires.

On server side, first action after receiving the feedback message from the client is to determine if these packets are not delayed. This is done by analyzing the time stamp existent in each feedback message and comparing it with the time stamp of the last correct message received. After this step, the algorithm extracts the interest factors and proceeds in estimating the user perceived quality. Because DQOAS is an extension of QOAS algorithm, the same quality metric is used, the no-reference moving picture quality metric. The formula that defines this metric is given below:

$$Quality = Q_0 + \mathcal{X}_Q * \left(\frac{\bar{R}}{X_R}\right)^{-\frac{1}{\xi_R}} + \mathcal{X}_L * \bar{R} * PLR$$

The parameter Q_0 is determined to have a value always around 5, corresponding to the highest quality. $\mathcal{X}_Q$, $\mathcal{X}_R$, $\mathcal{X}_L$ and ξ_R are related to the encoding complexity of the set of frames considered, while $\bar{R}$ represents the mean bit rate of the stream. PLR, or Packet Loss Rate, is defined as the number of packets lost per time unit divided by the number of packets transmitted during the same time unit. The parameters' values for the quality metric used in the current implementation of the algorithm are the average values determined in [106] for a complex video stream: $\mathcal{X}_Q$ = -0.045, $\mathcal{X}_R$ = 124.7, $\mathcal{X}_L$ = -33.9, ξ_R = 1.12 and Q_0 = 5.2.

While the user perceived quality is computed based on every feedback report, the quality coefficient Q used in the rate decision process is calculated by averaging the received interest factors over the length of the server monitoring window. In LTE, the server receives a feedback message every 10 TTIs and uses a window length of 10 frames (100 TTIs). The quality coefficient Q is computed using the following formula: $Q = (w_1 \times Q_{throughput}) + (w_2 \times Q_{delay}) + (w_3 \times Q_{loss})$, where the weights w_1, w_2 and w_3 were determined to be 0.4, 0.3 and 0.3 after extensive tuning of the algorithm.

Because the data packets simulating a video streaming session used in the test scenarios are dummy packets, the quality metric values are not directly considered in the rate decision process. They are used instead as an extra confirmation regarding the decision to increase or decrease the transmission bit rate, after two consecutive quality coefficients are compared. If the last quality coefficient is at least 5% larger/smaller than the previous and if the computed quality metrics values are indicating an improvement/degradation, the adaptation algorithm will decide to increase/decrease the transmission bit rate. If the quality metrics values are suggesting an opposite action, the algorithm will wait for another quality coefficient to be computer and then will take the rate adaptation decision based only on their last two values.

DQOAS design allows the proposed adaptation algorithm to perform a dynamic manipulation of the transmitted information bit rate, to suit the delivery conditions and user interests. For example, after a period of increased traffic load on the network, when the stream was adapted and set to be delivered at low bit rates, if the algorithm detects any improvements in network conditions, a report will signal this to the N-Level Builder module, who will start increasing the bit rate on a step-by-step basis, according to the user-specific levels, improving therefore the user-perceived quality. Another scenario where DQOAS proves its utility is when the network conditions are degrading due to an increase in the background traffic or to new users requesting network resources. DQOAS algorithm will rebuild the dynamic levels for every user with a smaller granularity, and as a consequence the transmission level considered might decrease, avoiding congestions.

6. DQOAS RESULTS IN IEEE 802.11 WIRELESS LAN NETWORKS

In this chapter, simulation results for DQOAS in IEEE 802.11 wireless LAN environment are presented and discussed. The chapter starts with a brief description of the simulation environment used, including the simulation settings. Results of different simulations for every experiment performed are compared and discussed in separate sections.

6.1 Network Model And Test Setup For IEEE 802.11 Networks

The proposed algorithm was tested in IEEE 802.11 wireless environment using Network Simulator (NS) with NOAH (No Ad-Hoc) patch installed [107]. NS is an open source discrete event network simulator that can simulate a wide range of protocols in both wired and wireless environments. First released in 1995, NS is an advanced version of the Realistic and Large network simulator (REAL), first developed in 1988. The current NS development is supported through the Defense Advanced Research Projects Agency (DARPA) by the Measurement and Analysis for Network (SAMAN) and through National Science Foundation (NSF) by the Collaborative Simulation for Education and Research project (CONSER). Because of its open source nature, NS allowed other researchers and research groups to include their contributions, like the wireless code from Carnegie Mellon University (CMU) Monarch Project and Sun Microsystems.

NS is an object oriented simulator, written in C++, with an OTcl interpreter as frontend. NS uses two languages because simulator has two different kinds of things it needs to do. On one hand, detailed simulations of protocols require a systems programming language which can efficiently manipulate bytes, packet headers, and implement algorithms that run over large data sets. For these tasks run-time speed is important and turn-around time (run simulation, find bug, fix bug, recompile, re-run) is less important.

On the other hand, a large part of network research involves slightly varying parameters or configurations, or quickly exploring a number of scenarios. In these cases, iteration time (change the model and re-run) is more important. Since configuration runs once (at the beginning of the simulation), run-time of this part of the task is less important.

NS meets both of these needs with two languages, C++ and OTcl. C++ is fast to run but slower to change, making it suitable for detailed protocol implementation. OTcl runs much slower but can be changed very quickly (and interactively), making it ideal for simulation configuration. NS (via tclcl) provides glue to make objects and variables appear on both languages. The NS architecture is presented in Figure 46:

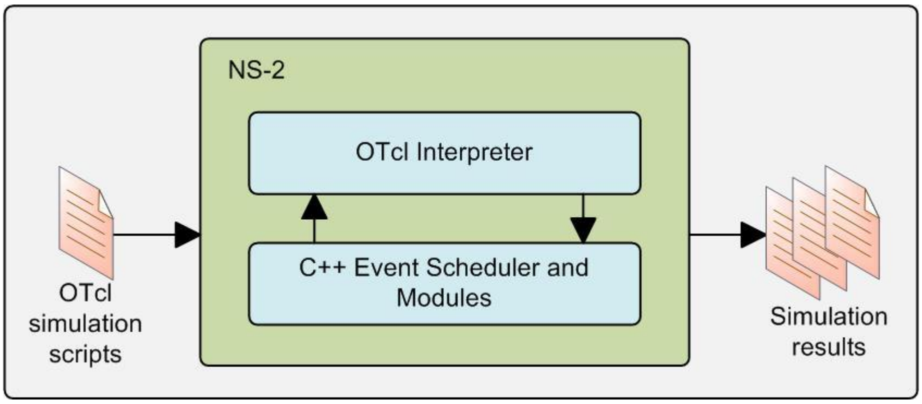


Figure 46 - NS-2 architecture

In order to simulate an infrastructure based WLAN topology, the NOAH patch was used. NOAH implements the direct wireless routing between base stations and mobile devices.

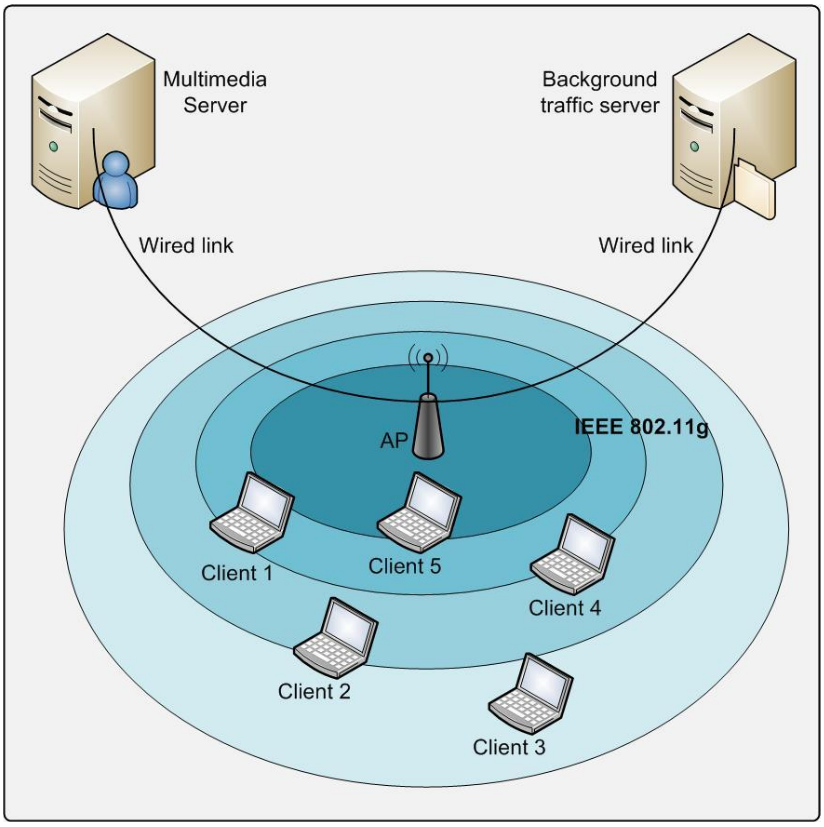


Figure 47 - General test architecture

The general network architecture used in all simulation scenarios is presented in Figure 47. One streaming server and one Background traffic server are connected through a wired link to the access point AP. Five clients are accessing the wireless medium using IEEE 802.11g standard, requesting data from both serves. The clients have a static position or they can follow a mobility pattern inside the coverage area of the AP, depending in the test scenario. They can also attach or detach to the network at any time, requesting or freeing the radio resources. The parameters used in all simulation scenarios are presented in the Table 11.

Each node (client) has specific characteristics like mobility pattern, maximum achievable throughput and one parameter defining the expected quality level of an incoming video stream.

Table 11 - NS-2 simulation parameters

Parameter	***Value***
Data Rate	54 Mbps
CW_{min}	15
CW_{max}	1023
Slot Time	9 μs
SIFS	16 μs
Short Retry Limit	7
Long Retry Limit	4

The wired link between the AP and the two servers (Multimedia streaming server and Background Traffic server) is considered to be a bottleneck link with a delay of 2 ms and a fixed value of 3.85 Mbps, that is slightly greater than the sum of minimum QoE expectation levels for every user, presented in Table 12.

Table 12 - User-specific thresholds for video quality

Client	***Minimum accepted quality level***
Client 1	0.3Mbps
Client 2	0.6 Mbps
Client 3	1.0Mbps
Client 4	0.8 Mbps
Client 5	0.45 Mbps

For all simulation scenarios that were used to test DQOAS algorithm in a wireless LAN environment, other selected algorithms were tested: a non-adaptive solution [80], the TFRC algorithm and the QOAS algorithm. The results obtained by DQOAS are compared with those obtained by the other delivery mechanisms and the performance is assessed. In the proposed test scenarios, one media-rich content stream is delivered to the five users in different mobility and network load conditions.

6.2 DQOAS Results

The performances of the proposed adaptation algorithm were analyzed and tested using an open source discrete event network simulator called Network Simulator. Using this software, the author was able to simulate the IEEE 802.11 environment and to perform four different experiments, each with duration of 60 seconds. During the first two experiments all five users are static, while the last two experiments are employing user mobility. Background traffic is setup to run only in experiment 2 and 4. For the experiments that are using mobility, the movement pattern used is the following: during the first 10 seconds of the simulations the users have a static position, after which they start moving away from the AP with a speed of 1.1 m/s. First four users are randomly attaching to the network within first 6 seconds of the simulation, while the fifth user is always attaching after 7.5 seconds.

To evaluate the proposed algorithm, the same test scenarios were conducted for a non-adaptive streaming solution and for two adaptive streaming methods: TFRC and QOAS. The non-adaptive delivery mechanism was tested with three different transmission rate settings: first non-adaptive streaming version uses the transmission rate set to the maximum video bit rate available (1.5 Mbps), the second version uses a transmission rate set to a medium video bit rate (0.9 Mbps), while the last version uses a transmission rate that is set to the lowest video bit rate available (0.3 Mbps). The TFRC model used is the one included in the current version of the NS-2 simulator and it adjusts the transmission rate to match the expected throughput of a TCP stream in similar conditions. The analyzed throughput is considered the received throughput by every user.

6.2.1 Exp. 1: Static Users With No Background Traffic

During the 60 seconds of this first experiment, all five users are allocated with a set of static coordinates that determine a set of five unique positions inside the range of the AP, specially chosen so that the received signal strength is enough to satisfy their minimum QoE requirements presented in Table 12. No background traffic is used in this simulation.

The multimedia throughput obtained for each user when the delivery solution used was the QOAS adaptive mechanism is depicted in Figure 48. It can be observed that the algorithm adapts each of the transmitted multimedia streams, improving the spectral efficiency and offering similar delivery conditions to all five users. But because users usually have different quality expectations, this solution may prove to be unsatisfactory when the QoE is considered.

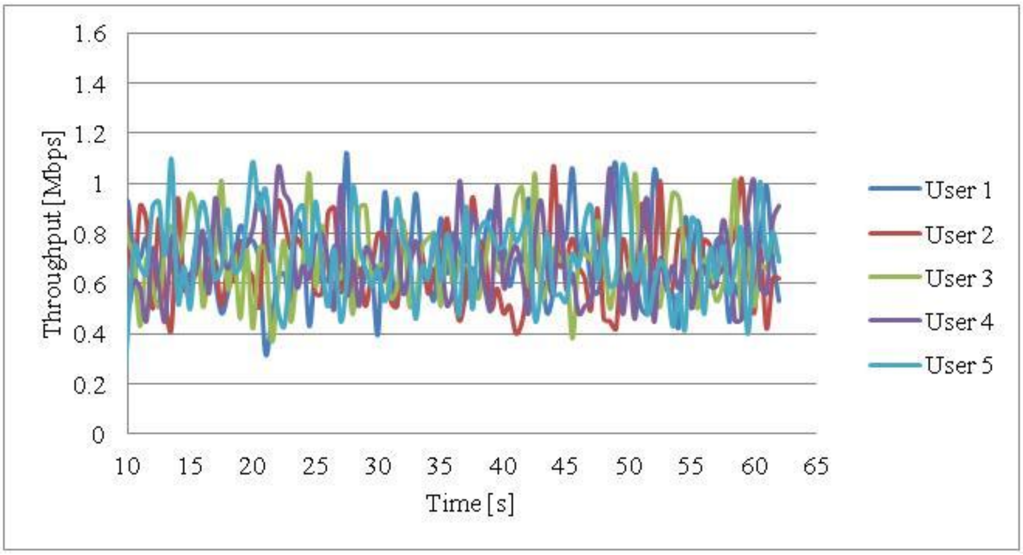


Figure 48 - Multimedia throughput when QOAS adaptive method is used

The results obtained when the non-adaptive method was used with all three transmission rates available (1.5, 0.9 and 0.3 Mbps) are presented in Table 13. In case the transmission rate is set to the maximum bit rate level, only the first two users that requested the multimedia stream (in this particular case *User 1* and *User 2*) can be served at or above the expected quality level. If the transmission quality is set to the medium value, the number of users that receive the stream at that level increases to 4, while the fifth user has a continuous buffering process. In this situation the non-adaptive algorithm is using the available bandwidth in an inefficient manner, because even if the BW available is enough to accommodate all five users' needs, one user is not able to receive the requested multimedia stream. When the transmission rate is set to its minimum value, all five users are receiving the multimedia stream, but the user satisfaction in this case is low, because only User1 is receiving the multimedia stream above his minimum expected level.

Table 13 - Results obtained with the non-adaptive solution

Transm. rate	***User 1 throughput [Mbps]***	***User 2 throughput [Mbps]***	***User 3 throughput [Mbps]***	***User 4 throughput [Mbps]***	***User 5 throughput [Mbps]***
1.5 Mbps	1.5	1.5	0.0	0.0	0.0
0.9 Mbps	0.900	0.900	0.900	0.900	0.0
0.3 Mbps	0.300	0.300	0.300	0.300	0.300

The results obtained when DQOAS algorithm is employed for multimedia transmission are presented in Figure 49. It is easily observed that the algorithm is able to differentiate between users, based on the information related to their minimum expected quality level.

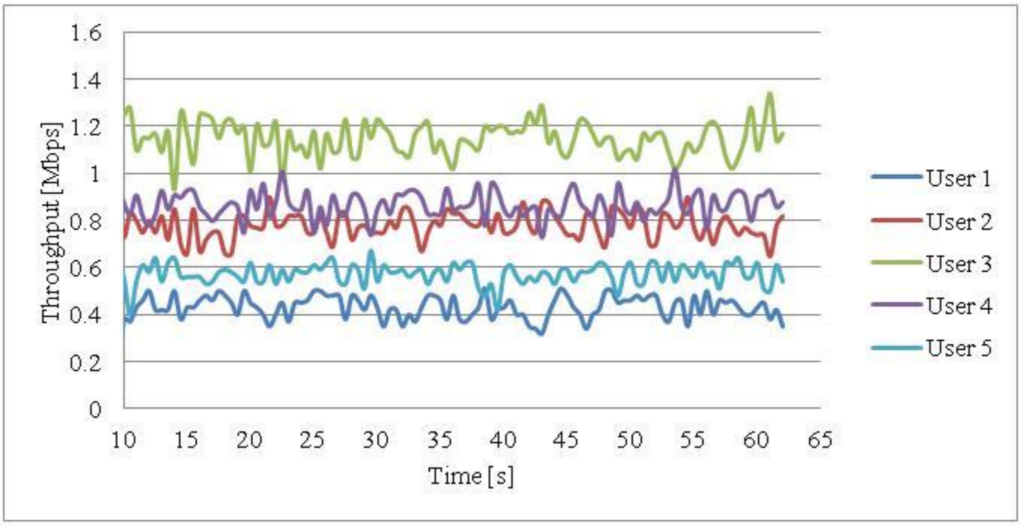


Figure 49 - Multimedia throughput when DQOAS algorithm is used

The oscillations present in the throughput are representing tries performed by DQOAS to improve the stream quality, using a step increase for certain streams. The algorithm is pushing to improve the quality of the delivered stream because it detects that there are still some network resources available and by this it tries to maximize the bandwidth usage.

Considering that no background traffic is present in the network, the total throughput obtained by summing the individual user throughput can offer an idea about how efficient was used the available bandwidth by each tested algorithm (Figure 50).

When the TFRC algorithm is used, the total throughput obtained is comparable with the QOAS algorithm throughput, but the TFRC loss rate is bigger, thus reducing the perceived quality of the multimedia stream. DQOAS algorithm manages to obtain a high bandwidth usage because of its small granularity adaptation levels used for the delivered stream and because it is constantly trying to increase the received quality for each user.

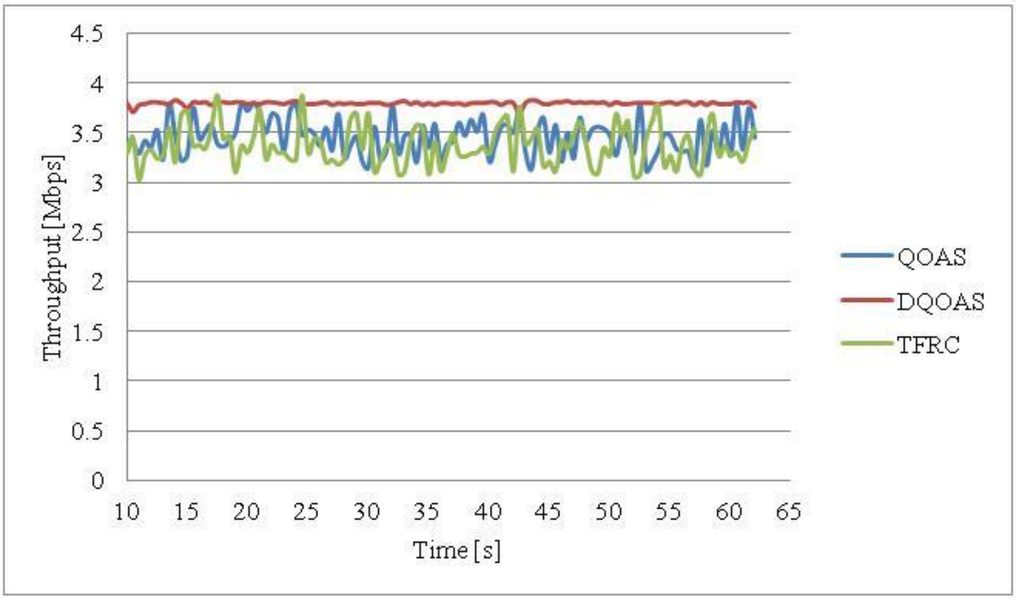


Figure 50 - Total throughput for the three adaptive solutions used

The loss rate experienced when the adaptive algorithms were used in the context of the first experiment is presented in Table 14.

Table 14 - Loss rate experienced when adaptive algorithms are used

Adaptive algorithm	***Loss rate [%]***
QOAS	0.67
TFRC	3.49
DQOAS	0.72

The average throughput obtained by each user during the last 50 seconds of the simulation, when different adaptive algorithms are used, is presented in Table 15. When compared to the minimum accepted quality level for every user, it can be observed that QOAS and TFRC algorithms are able to deliver the expected quality for 3 users, while DQOAS method is able to satisfy all five users in the conditions of this experiment.

Table 15 - Average throughput per user

Min. accepted level	***0.300***	***0.600***	***1.0***	***0.800***	***0.450***
Adaptive algorithm	User 1 throughput [Mbps]	User 2 throughput [Mbps]	User 3 throughput [Mbps]	User 4 throughput [Mbps]	User 5 throughput [Mbps]
DQOAS	0.431	0.786	1.176	0.879	0.573
QOAS	0.766	0.726	0.712	0.697	0.694
TFRC	0.768	0.760	0.762	0.748	0.762

6.2.2 Exp. 2: Static Users With Background Traffic

In the second simulation, all five users were assigned with the exact same static positions like in the first experiment. The difference between the two experiments comes from the fact that now each user has a second wireless connection enabled. This connection is established between the background traffic server and every user. The background traffic will feed the user with a data stream who's bit rate varies between 0.10 Mbps and 0.11 Mbps.

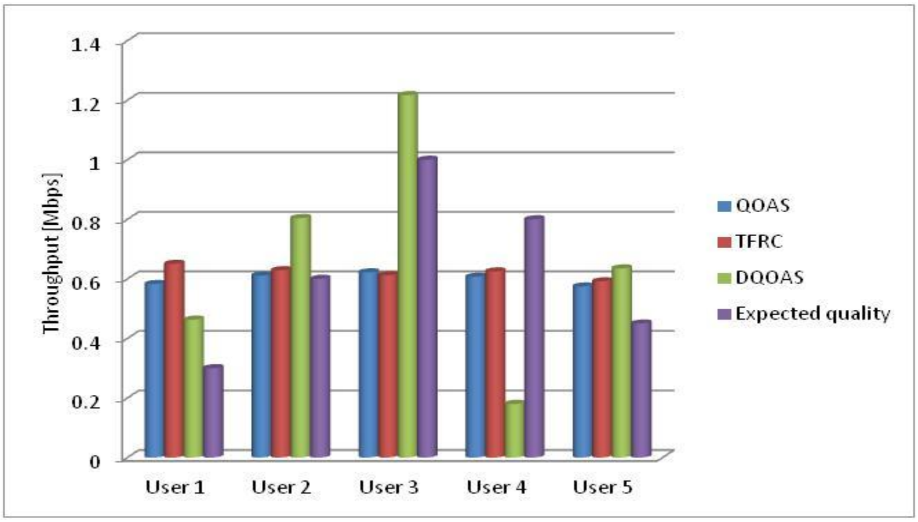


Figure 51 - Average user throughput when QOAS, TFRC and DQOAS algorithms are used

Figure 51 presents the average throughput obtained by each user exclusively on their connection with the multimedia server when QOAS, TFRC and DQOAS delivery algorithms were used.

User throughput evolution in time that is obtained using all three algorithms deployed in this scenario is similar with the one obtained during the first experiment.

During this simulation experiment, a particular scenario is taking place. Because of the congestion that appears in the network due to the presence of the background traffic, the adaptation algorithms are forced to reduce rate at which they are delivering the multimedia stream to the users. In these conditions, TFRC and QOAS algorithms keep sending the adapted data stream to all five users that are requesting the multimedia traffic, while DQOASs' adaptation mechanism will decide to terminate the session for one of the users, in this case User 4. This decision is taken after analyzing both user requirements and the network conditions, and observing that in the current network conditions it is impossible to deliver a satisfactory multimedia stream quality. By doing so, it will free those radio resources that will become available for the other users. Figure 52 is illustrating the throughput variation for the five users in the case described above, when DQOAS algorithm is used.

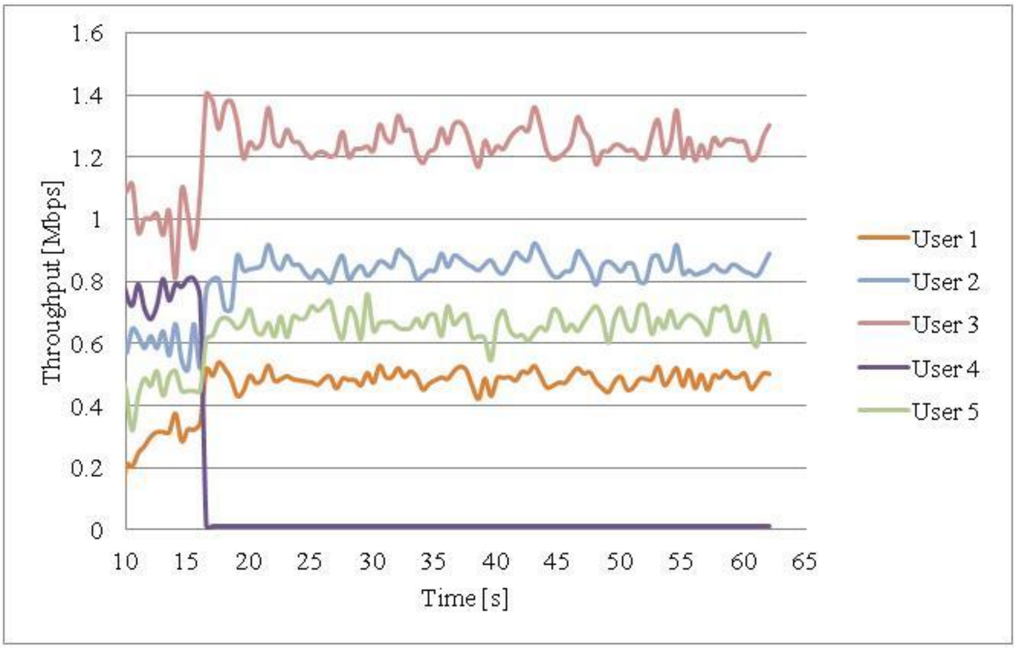


Figure 52 - Multimedia throughput when DQOAS algorithm is used

In case the non-adaptive solution is used, the result obtained when the lowest transmission rate is used is the same with the one obtained in the first experiment, because there are enough radio resources remaining free for the background traffic delivery. If the transmission rate is set to its maximum available bit rate, 1.5 Mbps, only two users will receive the requested multimedia stream, exactly like in the first test scenario. The difference from the first simulation appears when the transmission rate is set to 0.9 Mbps, because the presence of the background traffic will decrease the number of simultaneously satisfied users to only three.

The loss rate for the adaptive algorithms in the context of the first experiment is presented in Table 16.

Table 16 - Loss rate experienced when adaptive algorithms are used

Adaptive algorithm	***Loss rate [%]***
QOAS	0.71
TFRC	3.14
DQOAS	0.86

The total instant multimedia throughput per all five users, obtained by the adaptive algorithms when background traffic in on is presented in Table 17. For the background traffic, the total instant throughput was determined to be 0.54 Mbps in all scenarios of this experiment.

Analyzing Table 17, one can observe that the instant throughput obtained when DQOAS algorithm was used, is the closest to the bottleneck link limit (3.85 Mbps) of all three. This is an indication that DQOAS algorithm is well designed to maximize the bandwidth usage, even when other types of traffic that cannot be manipulated are present in the network.

Table 17 - Total instant throughput

Adaptive algorithm	***Instant multimedia throughput [Mbps]***	***Instant throughput (multimedia+bckground) [Mbps]***
QOAS	2.99	3.53
TRFC	3.10	3.64
DQOAS	3.30	3.84

6.2.3 Exp. 3: Mobility With No Background Traffic

The third experiment performed to test the performances of the adaptation algorithm proposed in this thesis introduces the mobility factor. The initial position of all 5 users is near the Access Point. They will start their movement towards the edge of the coverage area with a constant speed of 1.1 m/s before the first 10 seconds of the simulation had passed. The second connection for background traffic is not active during this experiment, so compared to the first proposed experiment the streaming algorithms used have to deal with the users' mobility. Figure 53 and Figure 54 represent the multimedia throughput obtained by all five users, considering the hypothetical throughput of IEEE 802.11g standard reported to distance when TFRC and DQOAS algorithms are deployed.

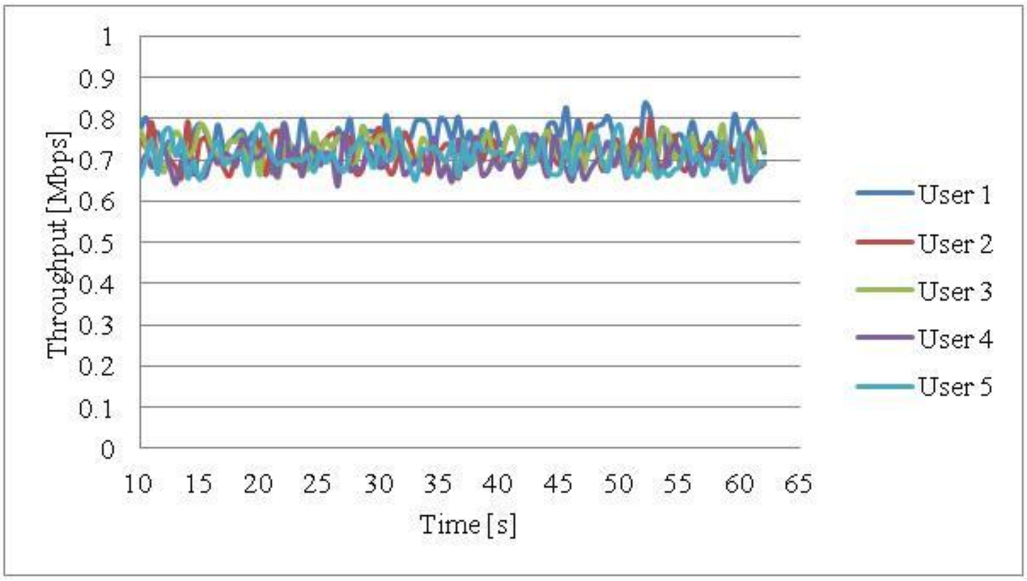


Figure 53 - Multimedia throughput when TFRC algorithm is used

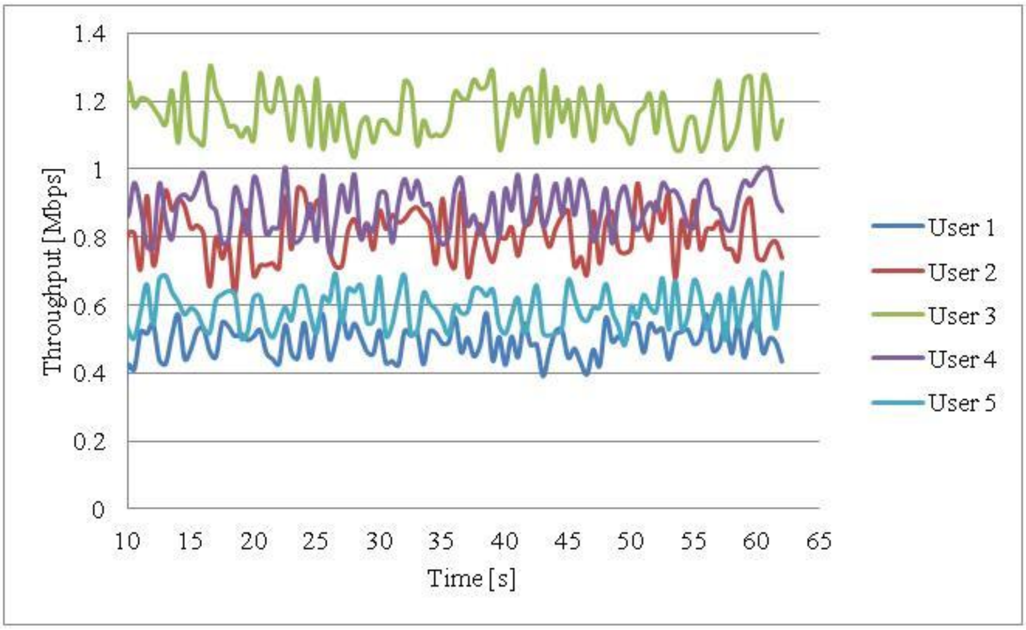


Figure 54 - Multimedia throughput when DQOAS algorithm is used

Table 18 presents the total instantaneous throughput and the loss obtained after the simulation scenario was run for all three adaptive algorithms. Compared with the first scenario, QOAS loss rate degraded while DQOAS improved it by 0.35, absolute value. This translates into a better user mobility management done by DQOAS algorithm because of its dynamic adaptation process.

Table 18 - Throughput and loss for all adaptive algorithms

Adaptive algorithm	***Total instant throughput [Mbps]***	***Loss rate [%]***
TFRC	3.39	1.35
QOAS	3.59	0.72
DQOAS	3.78	0.37

The non-adaptive algorithm results are identical with those obtained in the first experiment. The maximum number of satisfied users is four and it is obtained when the transmission rate is set to the medium quality, 0.900 Mbps.

Table 19 presents the average user throughput obtained when the adaptive algorithms are used. The conclusion from the first experiment, that DQOAS algorithm is able to deliver the multimedia stream over the minimum required quality for every user, while QOAS and TFRC are offering the desired QoE only for three users, can also be drawn here.

Table 19 - Average throughput per user

Min. accepted level	***0.300***	***0.600***	***1.0***	***0.800***	***0.450***
Adaptive algorithm	User 1 throughput [Mbps]	User 2 throughput [Mbps]	User 3 throughput [Mbps]	User 4 throughput [Mbps]	User 5 throughput [Mbps]
DQOAS	0.452	0.771	1.146	0.861	0.552
TFRC	0.766	0.726	0.712	0.697	0.694
QOAS	0.714	0.695	0.663	0.664	0.660

6.2.4 Exp. 4: Mobility With Background Traffic

During this simulation setup, the users are keeping the mobility pattern detailed in the previous experiment but they also activate their second connection used for downloading background traffic with data rates between 0.10 Mbps and 0.11 Mbps.

The effect of the background traffic connection on the user perceived quality of the multimedia stream and also on the loss rate is presented in Table 20.

Table 20 - Throughput and loss for all adaptive algorithms

Adaptive algorithm	***Total throughput [Mbps]***	***Loss rate [%]***
TFRC	3.24	1.93
QOAS	3.21	0.88
DQOAS	3.27	1.05

Figure 55 presents the multimedia throughput obtained in this scenario for every user. It can be observed that TFRC and QOAS adaptive algorithms deliver the media rich stream above the expected quality level for three users, while DQOAS outperforms them and manages to deliver a satisfied multimedia quality for four out of five users.

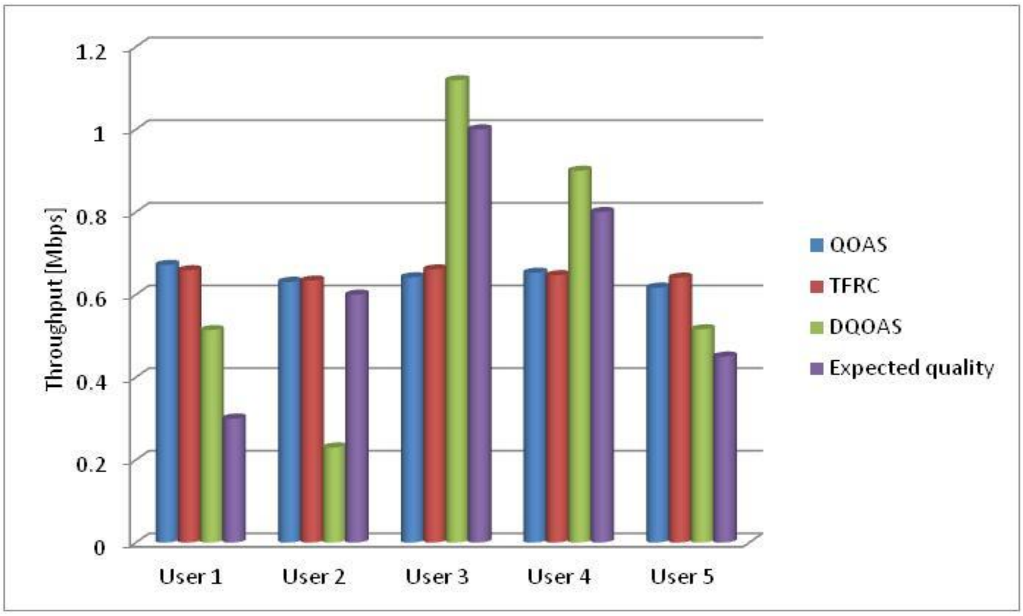


Figure 55 - Average user throughput when QOAS, TFRC and DQOAS algorithms are used

The presence of additional the background traffic on the bottleneck link between the servers and the AP determines DQOAS algorithm to terminate one of the connections because of the limited resources. After it tries to deliver the multimedia streams above the minimum satisfaction level for each user, it decides based on the feedback reports received that not all users' requirements can be satisfied and proceeds to end the streaming session for user 2 (Figure 56).

The results obtained when the non-adaptive method is used are similar with those obtained in the second experiment, where the presence of the background traffic was affecting the number of served users when the transmission rate is set to medium quality (0.900 Mbps), decreasing it to only three.

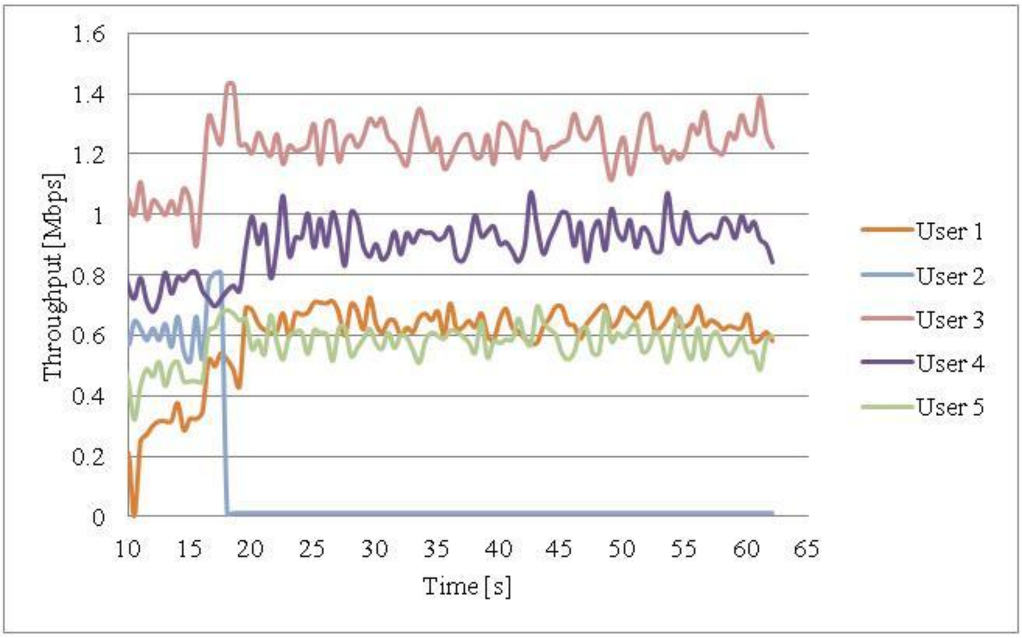


Figure 56 - Multimedia throughput when DQOAS algorithm is used

6.3 Conclusions

The results presented in these simulations are indicating the fact that DQOAS algorithm is able to manage both static and mobile users, obtaining good results in comparison with the other tested algorithms. Being an extension of QOAS, the proposed algorithm is keeping the advantages of the first (very good loss rate, improved link utilization) while increasing the number of simultaneously satisfied users through its dynamic stream granularity adaptation [105].

Because it was designed to differentiate the users based on their minimum accepted quality level, DQOAS is not applying an even distribution of radio resources among the users. It allocates the resources in a prioritized manner in order to increase the efficiency of the streaming process and to serve as many users as possible in the same time. Given the adaptive granularity of a stream that is continuously built by the algorithm based on the received feedback, the loss rate is kept low and the link utilization is increased.

Simulation scenarios two and four are illustrating the case when DQOAS algorithm is not able to offer the minimum requirements to all users because of the lack of radio resources. While TFRC and QOAS are reducing the stream bitrate of other users in order to increase the transmission rate for the user with the poorest quality, the proposed method decides to terminate the streaming session for that user and to reallocate those resources among the other users. Using this mechanism, the number of users that are receiving a multimedia stream above their minimum expectations is increased without the risk of wasting some of the valuable resources for users that do not have the proper radio channel conditions for receiving the video stream in a satisfactory manner [108]. If this procedure is used in PAMAH context, the poorest session will still be terminated but a new multimedia session will be initiated for transmitting data with a reduced multimedia content (audio or still images) in order to avoid the congestions and to still offer the possibility to that user to continue its learning process.

Table 21 gives an overview of the four experiments in terms of link utilization and simultaneously satisfied users.

Table 21 - Link utilization and satisfied users (per experiment)

Adaptive algorithm	***Experiment 1***		***Experiment 2***		***Experiment 3***		***Experiment 4***	
	Link util. [%]	Users satisfied	Link util. [%]	Users satisfied	Link util. [%]	Users satisfied	Link util. [%]	Users satisfied
DQOAS	99	5	99	4	98	5	98	4
QOAS	93	3	90	3	93	3	96	3
TFRC	98	3	93	3	88	3	97	3

Considering all the aspects discussed above, DQOAS algorithm can be a very good solution for multimedia delivery over wireless LANs when the user preferences are known by the application. The reduced loss rate, the high link utilization and the total number of simultaneously served users recommends the use of DQOAS by applications that are generating different levels of media content (video, audio, still images) [108].

7. DQOAS RESULTS IN 3GPP LTE NETWORKS

In this chapter, simulation results for DQOAS in 3GPP LTE environment are presented and discussed. The chapter starts with a brief description of the simulation environment used, including the simulation settings. Results of different simulations performed are compared and discussed in sections 5.2.1 and 5.2.2, along with the traffic prioritization scheme used. Section 5.2.3 describes the new mapping scheme proposed for applications that are generating multiple traffic flows, in order to improve DQOAS performances, followed by the simulations to test the implemented method. Last section presents another set of simulations and their results, when three different data flows from the same application are present for each user.

7.1 Network Model And Test Setup For 3GPP LTE Networks

The results obtained in IEEE 802.11 wireless LAN environment by the proposed multimedia delivery mechanism – DQOAS –, designed to improve the end users' QoE during a multimedia experience, are showing that a dynamic adaptation policy based on user preferences and network conditions is improving significantly the end-user perceived quality. Also, the total number of simultaneous served users is increased, as well as the link utilization [105], [108], [109]. Taking into account the good results of DQOAS and the user-oriented approach, the author considered to be of further interest to analyze the behavior of this algorithm over wide coverage wireless networks, like Long Term Evolution networks.

The proposed algorithm was tested using the LTE System Level Simulator [110], capable of simulating LTE SISO (Single Input Single Output) and MIMO (Multiple Input and Multiple Output) networks using TxDiv (Transmission Diversity) or OLSM (Open Loop Spatial Multiplexing) transmission modes. The Physical layer model is based on the post-equalization Signal to Interference and Noise Ratio (SINR), offering pre-calculated fading parameters and so reducing computational complexity at run-time. The schematic block diagram of the simulator is presented in Figure 57, and like other system-level simulators, the core part consists of a link measurement model [111] and a link performance model [112]. The link measurement model abstracts the measured link quality used for link adaptation and resource allocation. The link performance model determines the link Block Error Ratio (BLER) at reduced complexity.

The simulator is implemented using Matlab software, allowing the addition of new functionalities and algorithms. The DQOAS functioning principle proposed in section 3.2 was implemented in order to test the utility of an adaptation algorithm deployed over LTE networks. Considering that LTE has a performant QoS provisioning mechanism, the simulation results obtained when DQOAS is used in conjunction with different schedulers are compared with the results obtained when no adaptation algorithm is employed, leaving just the LTE QoS architecture to manage the multimedia flows.

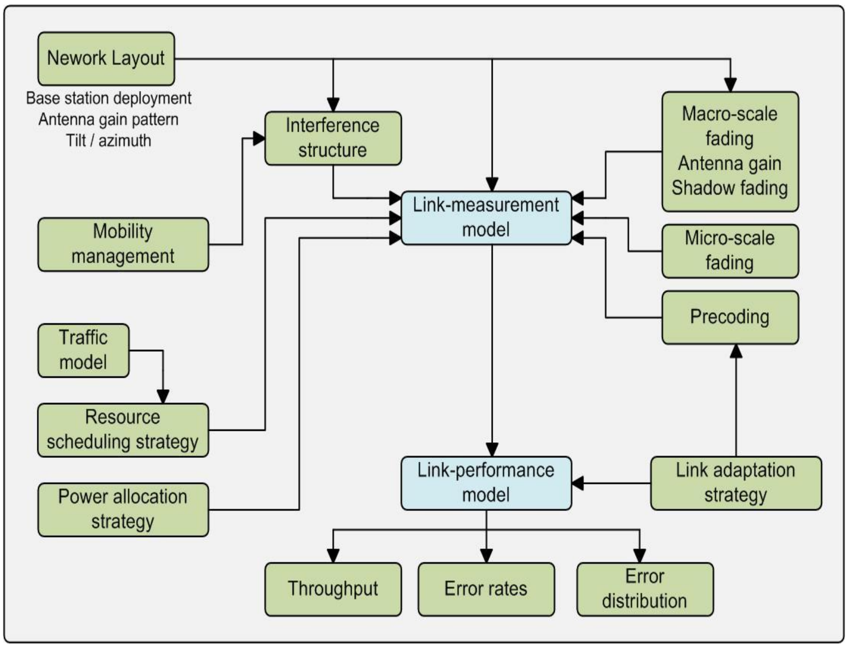


Figure 57 – LTE System Level Simulator architecture

The LTE network parameters used in all simulation scenarios are presented in the Table 22.

Table 22 - LTE parameters used for running simulation scenarios

Parameter	***Value***
Frequency	2.0 GHz
Bandwidth	5 MHz
Thermal noise density	-174 dBm/Hz
Receiver noise figure	9 dB
nTX x nRX	2 x 2
TTI length	1e-3 s
Simulation length	50000 TTIs
Subcarrier averaging algorithm	EESM
UE speed	5 Km/h

The extended network map that was used in the simulation scenarios is presented in Figure 58 and consists of 7 eNodeBs, each with 10 User Equipments attached, that move randomly through the map with a constant speed of 5 km/h. For some experiments, the number of eNBs or UEs attached to each of them was reduced for a better evaluation.

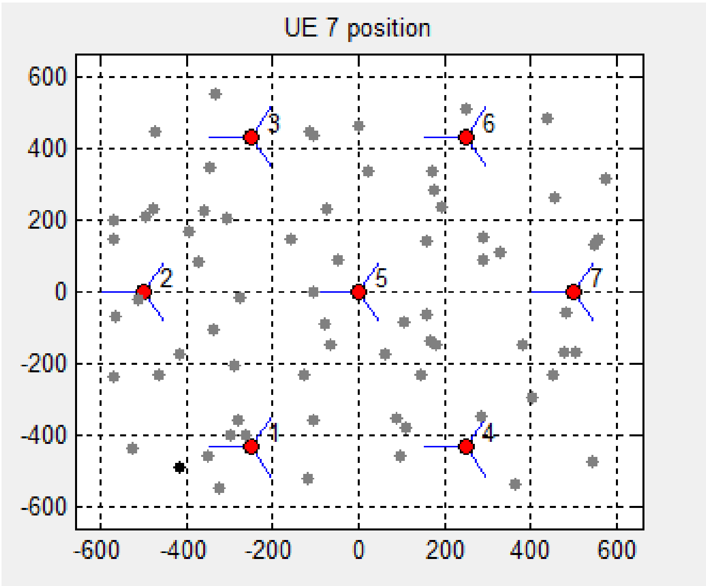


Figure 58 - LTE network map used in test scenarios

7.2 Experiments And Results

DQOAS algorithm was tested in 3GPP LTE environment by performing different experiments, each lasting 50 seconds. Based on the results obtained in different simulation scenarios, further improvements to the delivery scheme were added, increasing the performances. In the test setup used, every user can receive one or more data streams, from which at least one has rich media content, while the other streams can simulate a web browsing connection or a VoIP session that is being used simultaneously with the rich media content streaming. One representative user was chosen in order to analyze the throughput and BLER in each particular scenario, considering three different downlink schedulers: Round Robin, Maximum Throughput and Proportional Fair. The overall assessment of the results was done at the end of every simulation scenario.

7.2.1 Initial DQOAS Assessment In LTE Environment

The experiments in this section were performed in order to determine if DQOAS algorithm is suited to be used in 4G LTE networks, where the multimedia delivery process is enforced by the QoS mechanisms provided by the technology. In these experiments, the users are able to receive one or two multimedia streams, both having the same priority and the same QoS requirements. In order to ease the implementation, the minimum bitrate level required for the multimedia stream was set to 0.500 Mbps for all users. For evaluation, the LTE QoS mechanism results are compared against the results when DQOAS algorithm is employed for multimedia delivery using two simulation scenarios:

- one data flow per user when LTE mechanism and DQOAS algorithm are used
- two similar data flows per user when the LTE mechanism and DQOAS algorithm are used

7.2.1.1 Exp. 1: One Data Flow Per User

For this experiment, the traffic flows are transmitted using the LTE delivery mechanism or the proposed algorithm. One media-rich data flow is delivered to every user, having set a minimum required quality level at a bitrate of 500 kbps. The simulations are run using three different schedulers (RR, MT and PF) and the results obtained are compared in all cases.

Figure 59 and Figure 60 present the throughput (blue line) and BLER (green line) evolution for User 7 when the Proportional Fair scheduler is used, in conjunction with the LTE delivery mechanism and DQOAS algorithm respectively. In both cases, the data stream is delivered to User 7 in good quality conditions, with an average bitrate value of 1.487 Mbps in the first case and 1.654 Mbps in the second. The BLER values experienced by this user are 1.9% and 1.8% respectively. In case Maximum Throughput scheduler is used, the average throughput and BLER values obtained for User 7 in the first situation are 1.612Mbps and 2.4% while in the second situation the user is receiving the stream with an average throughput of 1.595Mbps and a BLER of 2.1%. For Round Robin scheduler, the values obtained are similar with those when the PF is employed: 1.490Mbps average throughput and 2.0% BLER for the first case and 1.492Mbps average throughput and 2.0% BLER when DQOAS mechanism is used.

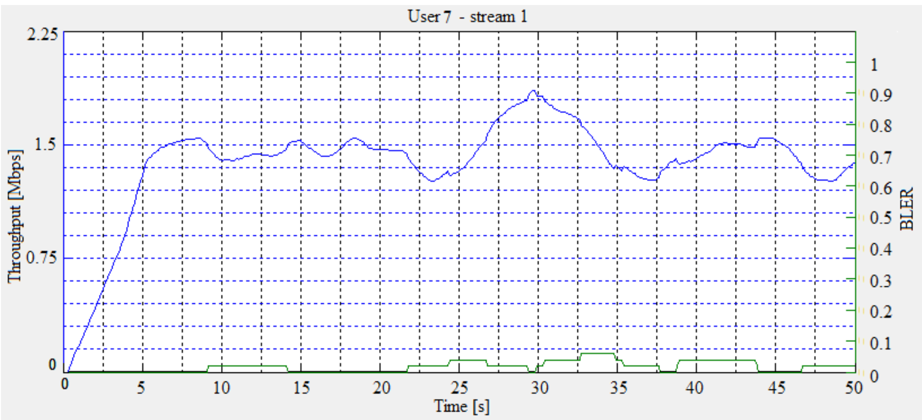


Figure 59 – Throughput and BLER for User 7 when PF scheduler and LTE delivery mechanism are used

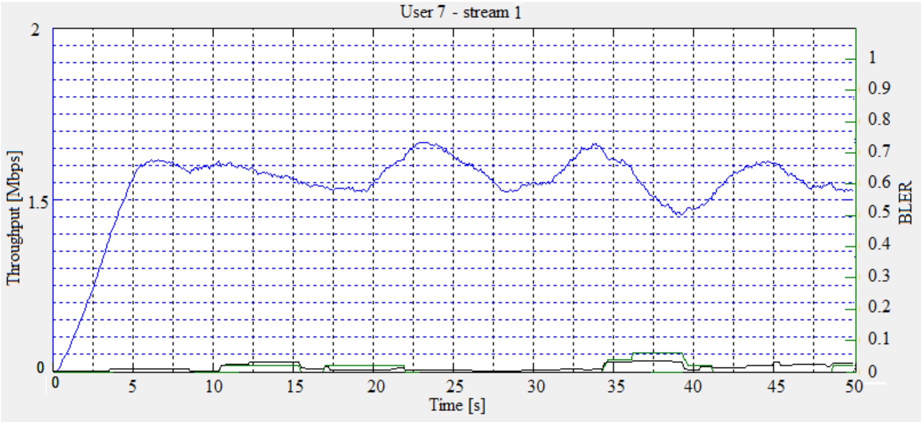


Figure 60 - Throughput and BLER for User 7 when PF scheduler and DQOAS algorithm are used

By sending only one data stream per user, no major congestions occur in the network and the differences in terms of user satisfaction between the proposed algorithm and the LTE delivery mechanism are small, as shown in section 7.2.1.3. In order to perform a better performance evaluation of the proposed algorithm in comparison with the LTE delivery scheme, a new set of experiments in which the network congestion is increased had to be conducted.

7.2.1.2 Exp. 2: Two Similar Data Flows Per User

For these test scenarios, in addition to the conditions presented in the previous experiment, a new UDP stream is delivered to every user in the network. Like the first stream, the second one has the same minimum accepted quality level, 0,500 Mbps. When DQOAS algorithm is used, the two streams received by a user are managed individually and no adaptation is done in order to accommodate both streams, but based on the fairness of the algorithm it is expected that it will be able to satisfy more users, compared with the LTE delivery mechanism.

Figure 61 present the throughput and BLER evolution for User 12 when the Round Robin scheduler is used together with the LTE delivery mechanism. It can be observed that both streams are experiencing high BLER values, their throughput being similar since they have the same priority and are treated equally by the scheduler.

In Figure 62, the throughput and BLER variations are presented for the same user when DQOAS algorithm is used. It this case, each stream bitrate is individually controlled by the algorithm based on the feedback about network conditions and on the required quality level. This is the reason why the user is not experiencing the same throughput and BLER for the two streams. In the case depicted in Figure 62, the algorithm decides to terminate one of the sessions for User 12 because it is not able to deliver this stream above the required level.

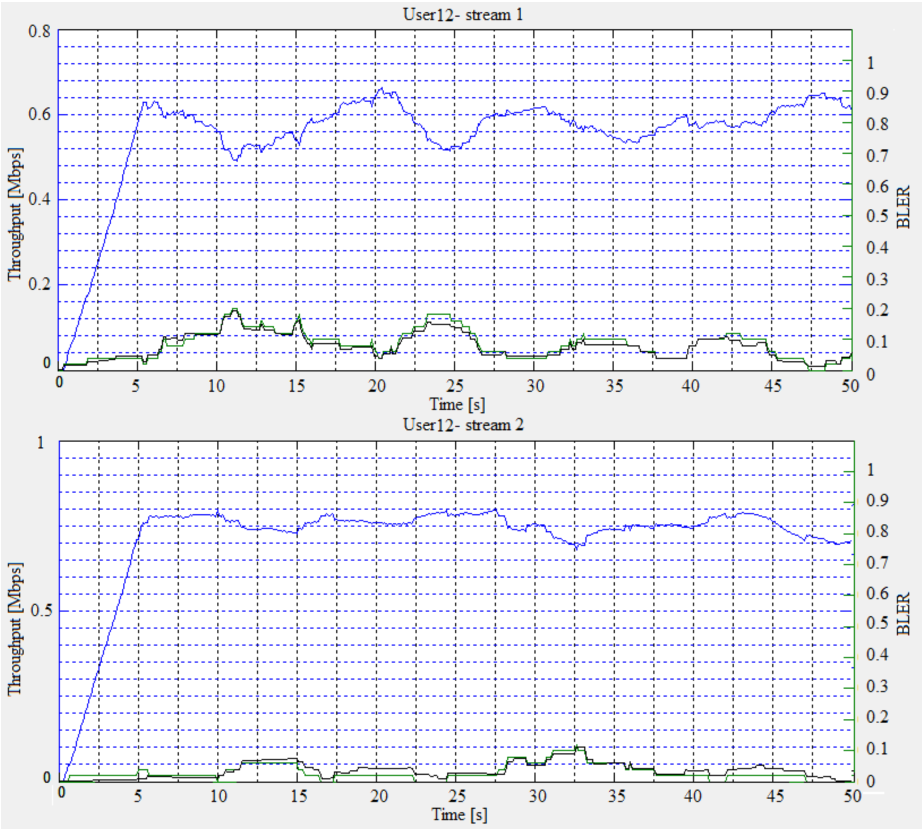


Figure 61 - Throughput and BLER for User 12 when RR scheduler and LTE delivery mechanism are used

The average throughput for user 12 in the first case is 0.588 Mbps for the first stream and 0.755 for the second, while in the last scenario the average throughput for the first stream is 0.669 Mbps, the second stream being terminated. In terms of BLER, the average values obtained in the first case are 6.9% for stream 1 and 6.0% for stream 2. For the second test, the average block error rate of the first stream is 3.2%.

In case PF scheduler is used, the average throughput values obtained in the first situation are 0.607 Mbps for the first stream and 0.649 for the second, while BLER was 6.8% for both streams. In the second simulation, the user is receiving the first stream with an average throughput of 0.721Mbps and a BLER of 6.8%. The second stream is not terminated in this case, and the user is experiencing a throughput of 0.673Mbps and a BLER of 6.5%. For Maximum Throughput scheduler, the values obtained in the first case are 0.714Mbps average throughput and 8.1% BLER for the first stream and 0.703 Mbps and 8.4% for the second one. In the second test performed, one of the data streams is terminated by the algorithm, the other one having an average throughput of 0.718Mbps and a BLER of 6.0%.

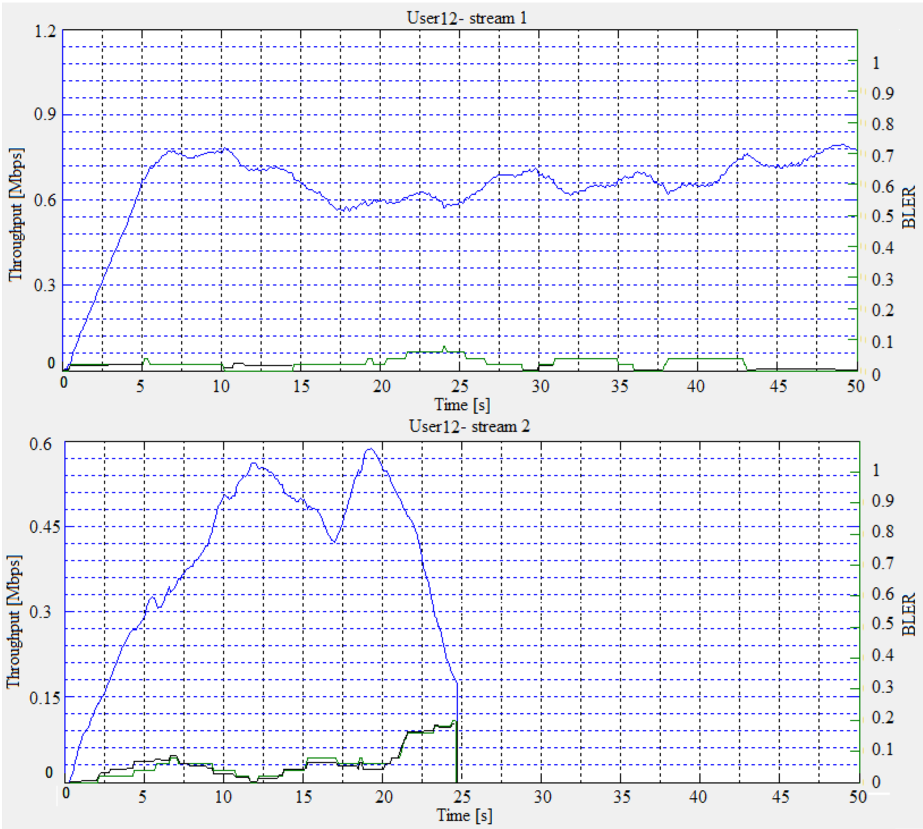


Figure 62 - Throughput and BLER for User 12 when RR scheduler and DQOAS algorithm are used

7.2.1.3 Conclusions

Experiment 1

The average throughput and BLER values normalized with the number of active users obtained during the first experiment are presented in Table 23. BLER is defined as the ratio of the number of erroneous blocks received to the total number of blocks sent, where an erroneous block is defined as a Transport Block for which the cyclic redundancy check (CRC) is wrong. Looking at the BLER values obtained during the first set of experiments, the best results are experienced when the proposed algorithm is used, because DQOAS reacts to any increase in BLER by reducing the bitrate of the multimedia stream.

In terms of user satisfaction, the results obtained for the first experiment and presented in Table 24 are suggesting that DQOAS algorithm can be used with good results in 3GPP LTE networks increasing the number of simultaneously satisfied users, being able to adapt the multimedia stream based on the network conditions while considering the user preferences.

Table 23 - Throughput and BLER average values when different schedulers are used

		PF scheduler	***RR scheduler***	***MT scheduler***
Throughput [Mbps]	LTE delivery mechanism	1.503	1.491	1.595
	DQOAS algorithm	1.542	1.512	1.616
BLER [%]	LTE delivery mechanism	2.1	2.2	2.2
	DQOAS algorithm	2.0	2.1	2.2

Table 24 - The percentage of satisfied users

	The percentage of satisfied users when the scheduler used is		
	PF	**RR**	**MT**
LTE delivery mechanism	74	71	67
DQOAS algorithm	79	73	70

The 5% increase (3 users) in the total number of satisfied users is not an impressive increase, but considering the conditions of this experiment (reduced congestion) it gives confidence that the algorithm can perform better when congestions are present in the network.

The results obtained during these simulations are also suggesting that the Proportional Fair scheduler performs better compared with Round Robin and Maximum Throughput, being able to deliver the data stream with a satisfactory quality to the highest number of users in both cases considered.

Experiment 2

The overall results obtained during this experiment are presented in Table 25. Looking at the user satisfaction figures, the results obtained are suggesting that by using DQOAS algorithm it is possible to increase the total number of simultaneously satisfied users, mainly because the traffic management process considers both user-related and network related parameters. A user is considered to be satisfied if it receives the data flows at an average bitrate at or above their minimum quality level, without any drops below.

BLER values are also reduced, because the adaptation algorithm considers this parameter when it decides to increase or reduce the delivered stream bitrate. In terms of throughput, the values obtained when DQOAS is used are higher compared to those obtained when the LTE delivery mechanism is used. This is possible because DQOAS cam decide to terminate some data sessions for certain users, freeing this way the radio resources that were used inefficiently.

Like in the first set of tests, the results are showing that PF scheduler obtains better performances, compared to Round Robin and Maximum Throughput, delivering the data stream above the quality threshold to the highest number of users.

Table 25 - Throughput and BLER average values when different schedulers are used

<table>
<tr><th colspan="3"></th><th>PF scheduler</th><th>RR scheduler</th><th>MT scheduler</th></tr>
<tr><td rowspan="7">LTE delivery mechanism</td><td rowspan="3">Stream 1</td><td>Throughput [Mbps]</td><td>0.633</td><td>0.624</td><td>0.691</td></tr>
<tr><td>BLER [%]</td><td>7.9</td><td>8.4</td><td>8.6</td></tr>
<tr><td>Satisfied users [%]</td><td>37</td><td>37</td><td>23</td></tr>
<tr><td rowspan="3">Stream 2</td><td>Throughput [Mbps]</td><td>0.628</td><td>0.618</td><td>0.690</td></tr>
<tr><td>BLER [%]</td><td>7.9</td><td>8.3</td><td>8.5</td></tr>
<tr><td>Satisfied users [%]</td><td>37</td><td>37</td><td>23</td></tr>
<tr><td colspan="2">Users satisfied by both streams [%]</td><td>37</td><td>37</td><td>23</td></tr>
<tr><td rowspan="7">DQOAS algorithm</td><td rowspan="3">Stream 1</td><td>Throughput [Mbps]*</td><td>0.705</td><td>0.689</td><td>0.736</td></tr>
<tr><td>BLER [%]*</td><td>6.6</td><td>6.9</td><td>7.8</td></tr>
<tr><td>Satisfied users [%]</td><td>50</td><td>42</td><td>31</td></tr>
<tr><td rowspan="3">Stream 2</td><td>Throughput [Mbps]*</td><td>0.647</td><td>0.652</td><td>0.738</td></tr>
<tr><td>BLER [%]*</td><td>6.5</td><td>7.0</td><td>7.6</td></tr>
<tr><td>Satisfied users [%]</td><td>44</td><td>44</td><td>35</td></tr>
<tr><td colspan="2">Users satisfied by both streams [%]</td><td>41</td><td>39</td><td>30</td></tr>
</table>

**the users for which the data stream was droped are not considered*

7.2.2 DQOAS Assessment When An Application Is Generating A Traffic Mix

The first set of experiments demonstrated that DQOAS algorithm is able to improve the multimedia delivery performances when one or multiple data flows of the same type are delivered to a user. Considering that today's' applications are able to generate multiple data streams that can be classified with different priorities, the next step was to test and analyze the algorithms' performances when a user receives two streams from the same application, that belong to different classes with different probabilities of being scheduled.

Because the current implementation of the LTE simulator has no mean to prioritize traffic that belong to different classes, a method for traffic prioritization was designed based on the work presented in [102].

Considering an application that generates traffic belonging to two different service classes, classified in different QoS queues, and a RR scheduler, the equation describing the *i*-th user satisfaction according to [102] is:

$$\frac{\left(f_1 + \frac{\alpha}{\rho} f_2\right) \cdot T \cdot \left[\frac{N}{n}\right] \cdot \Delta}{(T + d^{max}) \cdot \beta} \leq \frac{1}{1-\varepsilon}, \quad (1)$$

f_1 and f_2 represent the average packet transmission ratio, ρ is the priority of the first service over the second, T is the time interval in which the transmission takes place, N represents the maximum cell load that satisfies the quality criteria for user i, n denotes the number of scheduled users at every Transmission Time Interval (TTI), Δ is TTI length, d^{max} is the maximum scheduling delay and ε is the maximum ratio of delayed and loss packets with which the service quality perceived by the user is still satisfactory. If s_1 and s_2 are the average packet sizes of the two services, and s_i^{max} is the average amount of data that can be transmitted to user i in a single transport block, then $\alpha = s_2/s_1$ and $\beta = s_i^{max}/s_i$.

Adopting this model and considering that the first stream has a priority ρ over the second service, the amount of data the scheduler will consider when computing the capacity for the first stream will be $f_1 + \frac{\alpha}{\rho} f_2$. In this thesis, it was considered that the average packet sizes of the two services, s_1 and s_2 are equal ($\alpha = s_2/s_1 = 1$) and the priority of the first service - a live streaming session (QCI2) - over the second one - a TCP based app (QCI8) – is ρ = 2, with regard to the service priorities presented in Table 7.

To test this proposed situation, two different streams are sent to each user: the first stream is a multimedia stream and the second is a TCP-based flow. The minimum required level for the multimedia stream (first stream) was set to 0.500 Mbps for all users while for the second stream, the minimum accepted level is set to 0.250 Mbps. The results obtained when the LTE QoS mechanism is used are compared with the results collected when DQOAS algorithm is employed for multimedia delivery. Following two simulation scenarios were run:

- two different streams for each user when the LTE QoS mechanism is used
- two different streams for each user when DQOAS algorithm is used for multimedia delivery

7.2.2.1 Exp. 1: Two Different Streams For Each User When The LTE QoS Mechanism Is Used

In this experiment, the proposed LTE prioritization scheme for service flows is used, along with the LTE delivery mechanism. Two streams with different priorities are delivered to every user; first one is delivering multimedia traffic and the second one is a TCP download. The results obtained when PF and MT schedulers are used are illustrated in Figures 63 and Figure 64. Each figure presents both the throughput and the BLER for each stream received by User 9.

Figure 63 - Throughput and BLER for User 9 when PF scheduler is used

The average throughput of stream 1 when PF scheduler is used is 1.392 Mbps, satisfying the minimum requirements for User 9. For the second stream, the average throughput calculated shows a value of 0.252 Mbps. Even if the average throughput is above the required satisfaction rate (0.250 Mbps), there is a period longer than 15 seconds, when the experienced throughput is below the requested quality. In terms of user perceived quality, these seconds can determine the user to terminate the session and by doing so, perceiving the overall quality of the application as unsatisfactory because on the second stream.

Using the Maximum Throughput Scheduler in the context of this experiment, the average throughput value obtained for the first stream is 1.581 Mbps and 0.204 Mbps for the second. In this case, the average throughput experienced by the user for the second stream is below the minimum required level, and as a consequence the overall user satisfaction will decrease. The results when Round Robin scheduler is used – 1.361 Mbps for the first stream and 0.328 Mbps for the second – are suggesting that both streams have an average throughput over the required levels, without any big variations in time.

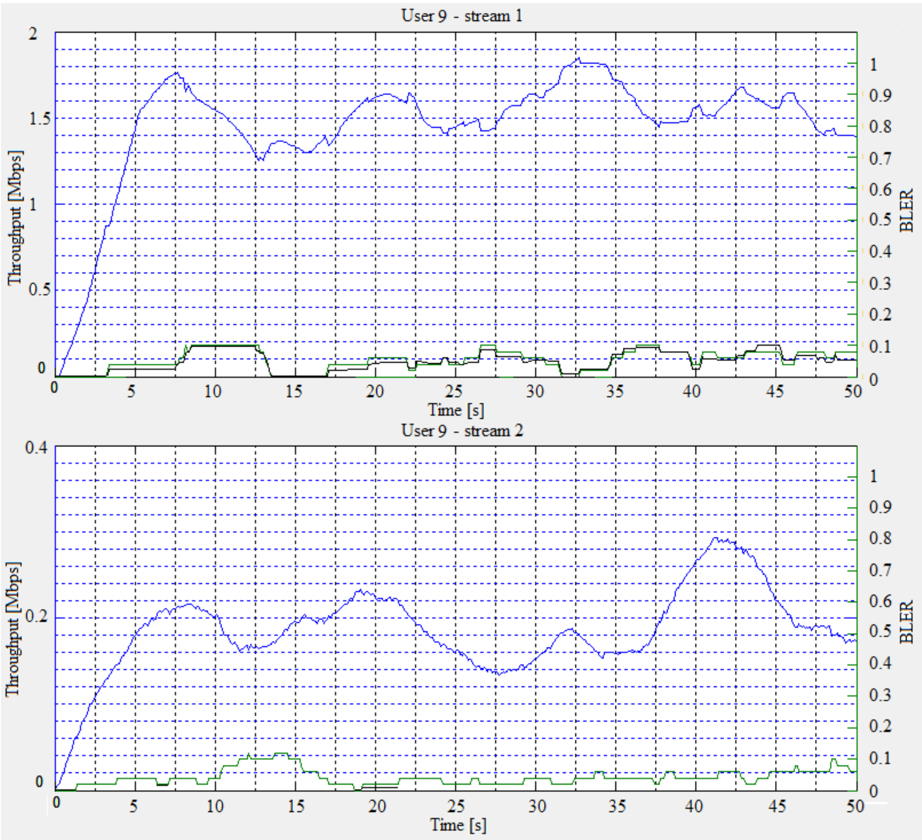


Figure 64 - Throughput and BLER for User 9 when MT scheduler is used

The BLER average values obtained by User 9 during this experiment are presented in Table 26. It can be observed that the best results are experienced by this user when the scheduler used is the Proportional Fair. The PF scheduler is also obtaining a better overall performance compared to Round Robin and Maximum Throughput schedulers, as shown is section 7.2.2.3.

Table 26 – User 9 BLER values when different schedulers are used

	PF scheduler [%]	***RR scheduler [%]***	***MT scheduler [%]***
Stream 1	5.6	6.6	7.9
Stream 2	5.1	5.1	5.3

7.2.2.2 Exp. 2: Two Different Streams Per User When DQOAS Algorithm Is Used For Multimedia Delivery

The second experiment performed is using the proposed flow prioritization mechanism in conjunction with the delivery algorithm presented in this thesis. All three schedulers were tested in order to analyze their performances and to outline the eventual improvements DQOAS brings in identical delivery conditions when compared with the LTE delivery mechanism. User 9 throughput variation in time for the two streams is presented in Figure 65 and 66 for PF and RR schedulers.

The average throughput value for the first stream obtained when the PF scheduler is used is 1.421 Mbps, which is greater than the one obtained in similar conditions in the context of the first experiment. The second stream average throughput value, 0.305 Mbps, is also greater than the one obtained when LTE QoS mechanism is used, and it never drops below 0.260 Mbps, which is above the minimum required quality (0.250 Mbps). This is an important aspect, because the chances User 9 will terminate the session due to a long period of unsatisfactory quality are reduced.

The average throughput simulation values when RR scheduler is used are 1.021 Mbps for the first stream and 0.437 Mbps for the second. Compared with the result from the first experiment, the throughput value of the first stream is lower, but still above the user satisfaction level. For the second steam, the average value is slightly higher than the one obtained during the first experiment and the throughput does not drop below the 0.250 Mbps limit.

If the Maximum Throughput scheduler is used, the results are showing that DQOAS method performs better in terms of final user satisfaction, compared with the first case. If during the first experiment, the second stream was not received in acceptable quality conditions, having an average throughput of just 0.204 Mbps, in the second setup the average value is 0.458 Mbps, without any drops below 0.290 Mbps. The first stream was also above the accepted quality level, averaging to a total of 1.463 Mbps.

The BLER average values experienced by User 9 during this simulation are presented in Table 27. It can be noticed that the best BLER values for the adapted stream (stream 1) are obtained when Proportional Fair scheduler is employed. This scheduler has the best performances because its allocation strategy is based on a tradeoff between spectral efficiency and fairness. The same results are obtained also in the overall analysis, when all scheduled users are considered (section 7.2.2.3).

Table 27 – User 9 BLER values when different schedulers are used

	PF scheduler [%]	RR scheduler [%]	MT scheduler [%]
Stream 1	2.1	3.2	4.1
Stream 2	4.3	3.7	3.2

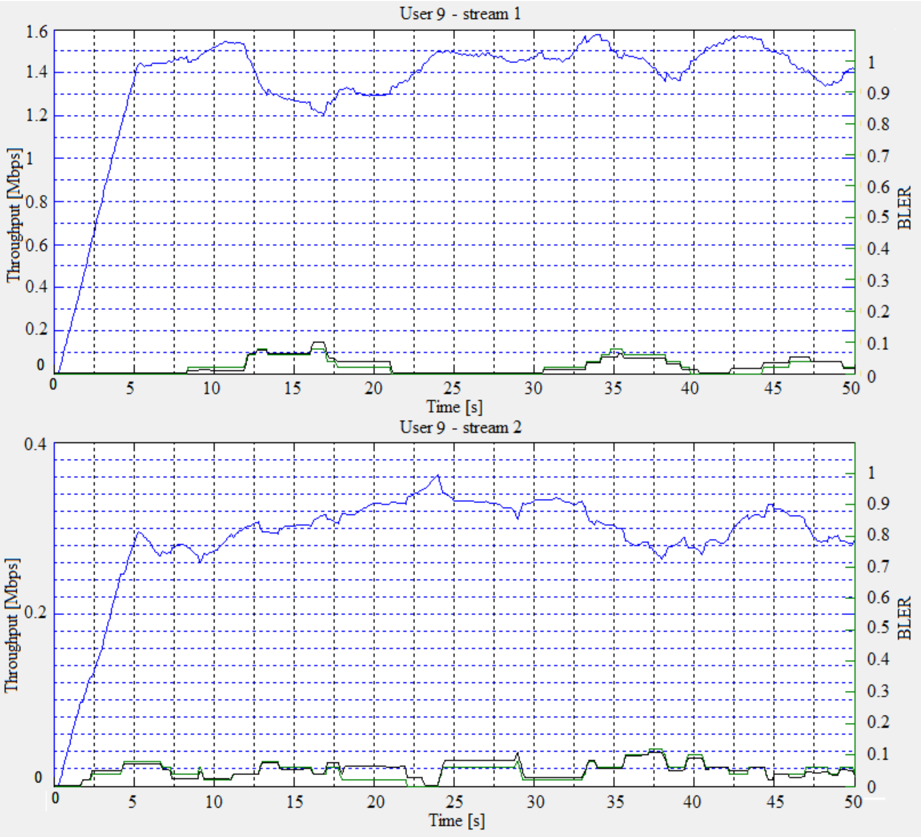


Figure 65 - Throughput and BLER for User 9 when PF scheduler is used

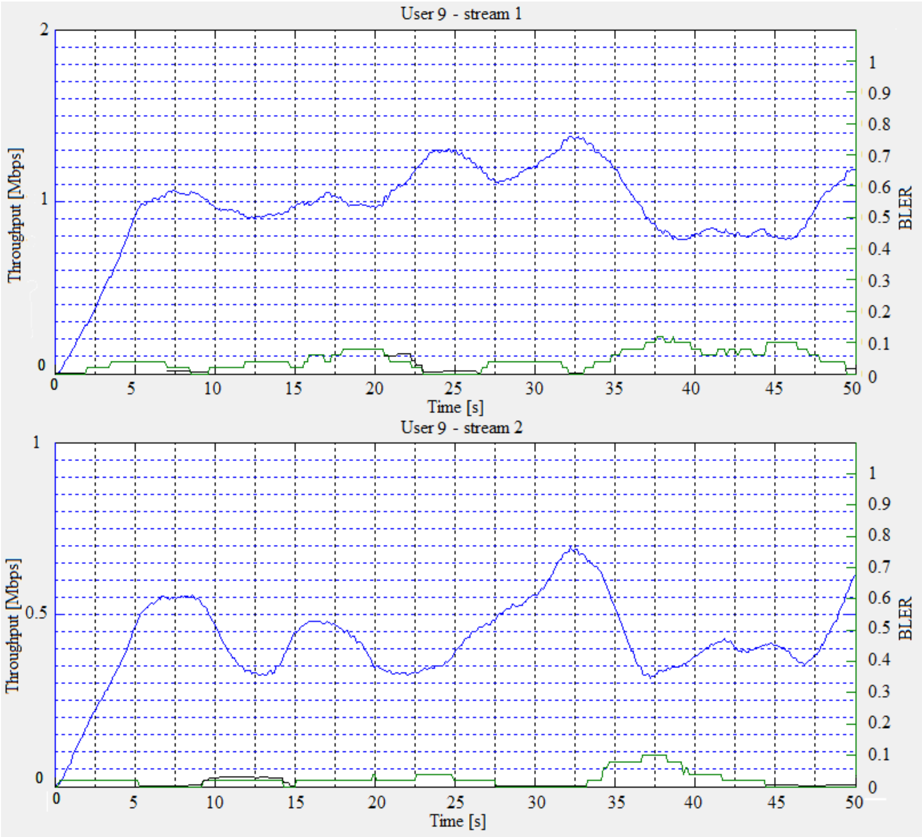


Figure 66 - Throughput and BLER for User 9 when RR scheduler is used

7.2.2.3 Conclusions

The average overall throughput and BLER value per user during the two experiments are presented in Table 28 and Table 29.

Table 28 - Throughput and BLER average values in context of the first experiment

		PF scheduler	***RR scheduler***	***MT scheduler***
Throughput [Mbps]	Stream 1	1.385	1.366	1.596
	Stream 2	0.487	0.433	0.396
BLER [%]	Stream 1	5.1	6.5	6.9
	Stream 2	6.4	6.9	6.2

Table 29 - Throughput and BLER average values in context of the second experiment

		PF scheduler	**RR scheduler**	**MT scheduler**
Throughput [Mbps]	Stream 1	1.411	1.168	1.534
	Stream 2	0.395	0.447	0.363
BLER [%]	Stream 1	3.2	3.9	4.4
	Stream 2	4.1	4.0	4.3

Looking at the overall values per user obtained, one can conclude that using the DQOAS mechanism for multimedia delivery over LTE networks generally improves the transmission quality, compared to the LTE QoS-based delivery mechanism, reducing the BLER of the transmitted streams, while maintaining them above the minimum bitrate level required for a satisfactory experience [113].

Table 30 gives an overview of the total satisfied users by both streams, considering all users.

Table 30 - The percentage of satisfied users per experiment

Experiment number	***% of satisfied users when the scheduler used is***		
	PF	**RR**	**MT**
Scenario 1	48	46	23
Scenario 2	51	46	31

As expected, the results obtained in both simulation scenarios are suggesting that the Proportional Fair scheduler obtains better performances compared with Round Robin and Maximum Throughput, being able to deliver the two streams with a satisfactory quality to the highest number of users [114].

Even if 3% increase in the total number of satisfied users is not an impressive increase, it is important to mention that this value was calculated considering that all users are using the same application simultaneously while moving inside the network map. Some of these users are likely to find themselves at the edge of their cell, where radio resources are limited and a good quality reception of the two streams in parallel is impossible.

The results presented in this section are suggesting that DQOAS algorithm can be used with good results in 3GPP LTE networks when a service prioritization scheme is used to differentiate the traffic flows received by a user, improving the BLER average values while maintaining the end-user perceived quality at or above their expectation level, for the highest number of users.

7.2.3 QoS Parameters Mapping Scheme For Optimizing DQOAS In Case Of Traffic Mix

Prioritizing the services types has very good results when the traffic sources are independent. A flow with a higher priority will have a significant capacity gain with the cost of a small capacity loss of the second service. But when the traffic source is the same for the different flows, the user might achieve a higher per-application QoE if both flows have the same priority (ρ = 1).

Following these assumptions, the second traffic type (with a lower priority) will have the same QoS class as the streaming traffic, as conceptually illustrated in Figure 67. The advantage of being in the same QoS class is that the queue specific sorting algorithm will consider both flows with the same priority during the resource allocation process. In this way, the data packets coming from the same application, for two different services, will have the same queuing delay and the same chance of being scheduled, improving the QoE of the application as a whole [114], [115].

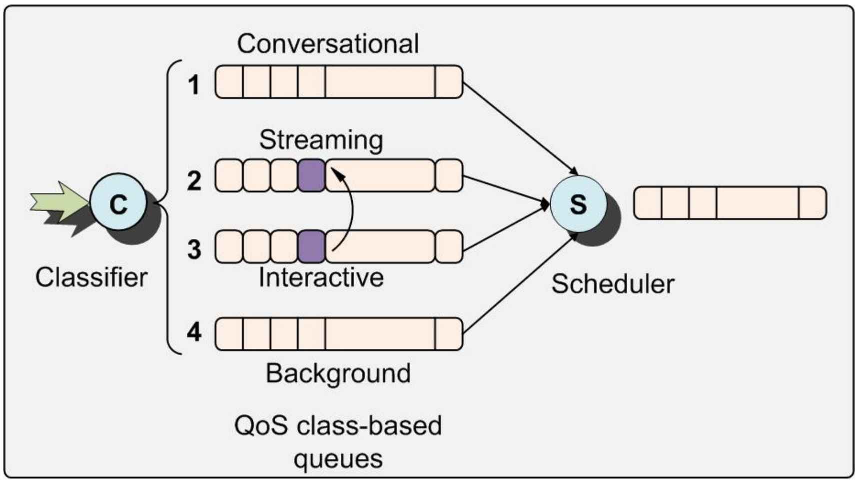


Figure 67 - Proposed framework for QoS class change for the low priority flow

In these conditions, DQOAS can update the quality levels for the multimedia stream based on user preferences, on instantaneous channel conditions and on the resource allocation scheme, with a minimum impact on the second service.

To test this solution, one new simulation scenario was created. The experiment uses the new mapping scheme that modifies the flow priority settings for streams generated by the same application. The new approach presented above is created to optimize the use of DQOAS over LTE network in case of an application that generates data streams with different priorities, increasing the total number of simultaneously per-application satisfied users.

7.2.3.1 DQOAS Results When The New Prioritization Scheme Is Used

Like in the previous experiments, three schedulers were used in order to perform an analysis of the proposed solution. All users are receiving two different data streams with the same minimum required levels like before: the minimum required level for the multimedia stream (first stream) is a bitrate of 0.500 Mbps, while for the second stream the level is set to 0.250 Mbps.

For the first scheduler used, Proportional Fair, the simulations are showing an average throughput of 0.76 Mbps for the first stream and 0.43 Mbps for the second one, both big enough to satisfy User 4 quality requirements. In the second experiment from section 5.2.2, DQOAS cannot manage in an optimal way the two streams with different priorities because the scheduler is overriding it, considering the flows' priorities in the resource allocation process. In this experiment the two streams are having the same priority and in consequence they receive the same treatment from the scheduler. This gives DQOAS the opportunity to perform an optimal adaptation for the first stream without influencing the second one, increasing the overall QoE of the application.

Figure 68 illustrates the results obtained when Proportional Fair scheduling method is used, while Figure 69 presents the simulation results obtained with RR scheduler.

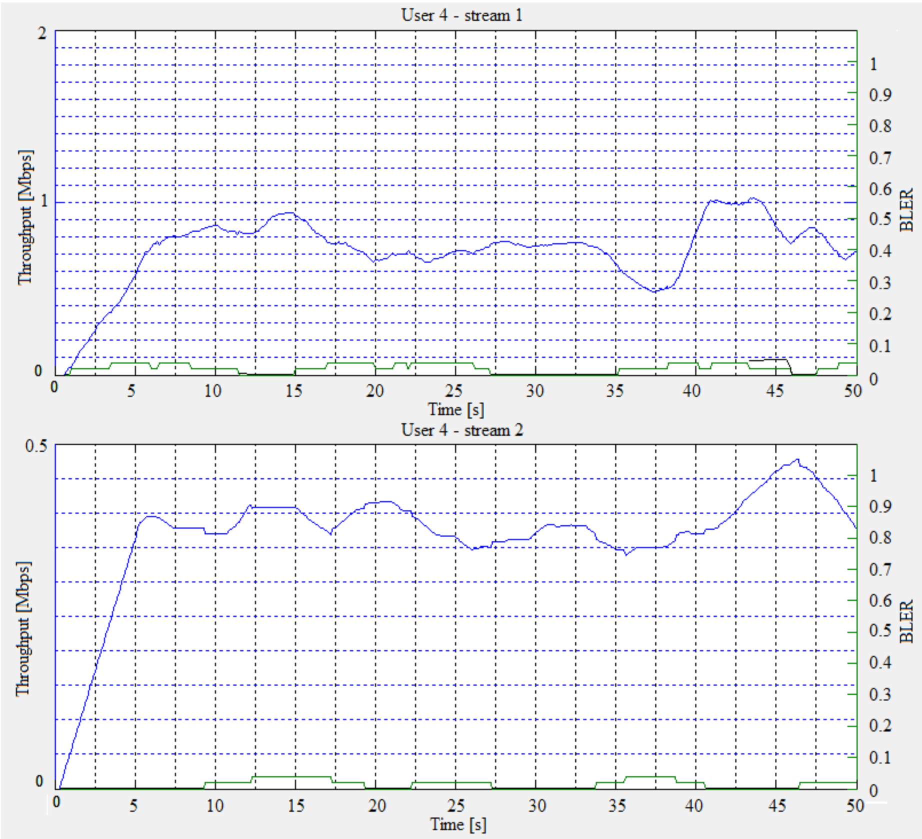


Figure 68 - Throughput and BLER for User 4 when PF scheduler is used

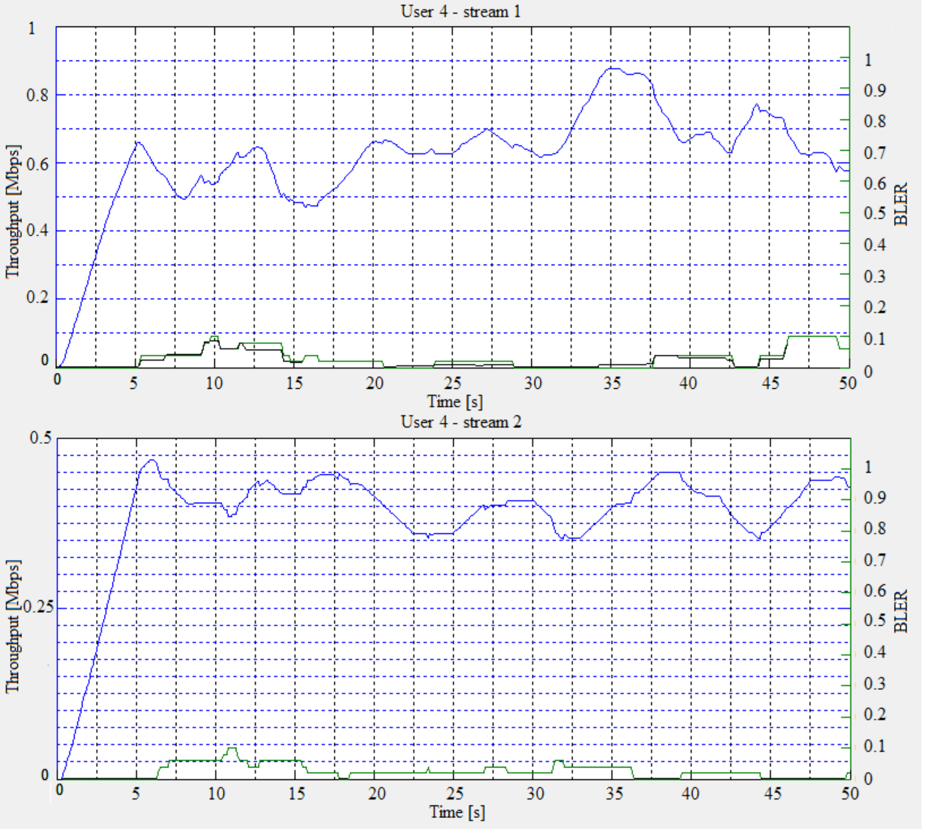


Figure 69 - Throughput and BLER for User 4 when RR scheduler is used

Throughput average values obtained when RR scheduler is used are 0.642 Mbps for the first stream and 0.411 Mbps for the second one. These values are again enough for User 4 satisfaction, even if compared with the two simulations in section 5.2.2, the throughput for the first stream is much lower.

Maximum Throughput scheduler does not performs as well as RR and PF, but is keeping the streams' average throughput over the minimum accepted level. The second stream is dropping for a period of 2.5 seconds below these levels, so the user might consider the overall performance as unsatisfactory in this case. The values obtained when MT scheduler is used are 0.668 Mbps for stream 1 and 0.318 Mbps for stream 2. Figure 70 describes the throughput variation in the latest case.

Figure 71 presents the BLER values of the two streams for User 4 when all three schedulers are used. It can be observed that the best BLER results registered for the first stream are obtained when the scheduler used is Proportional Fair, confirming the results from previous simulations.

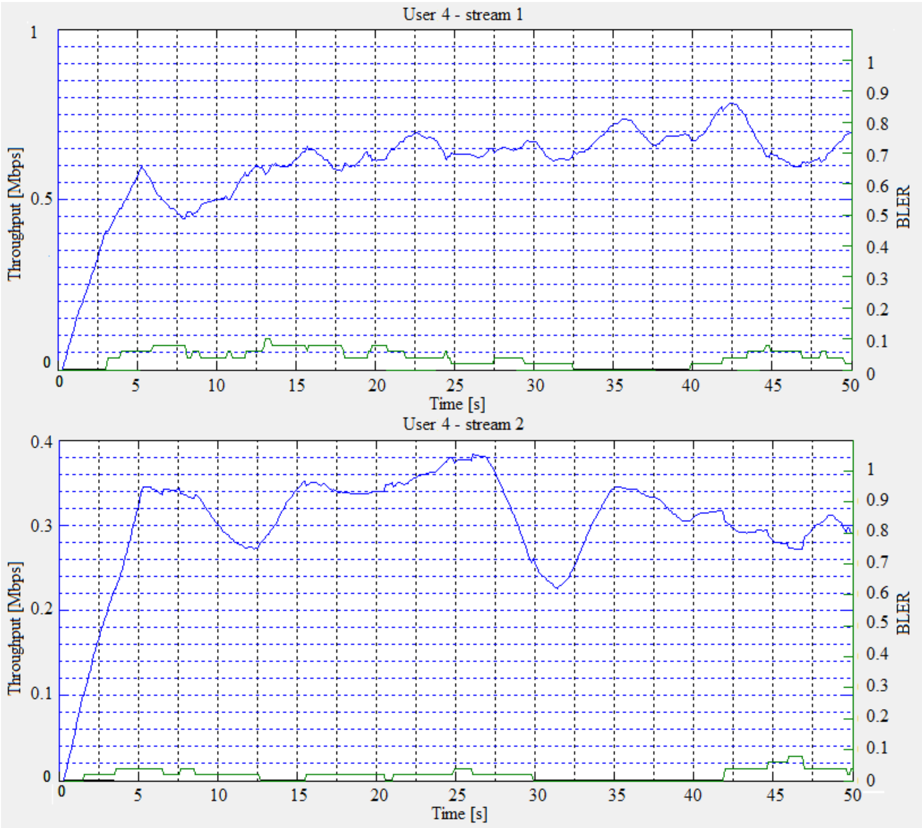
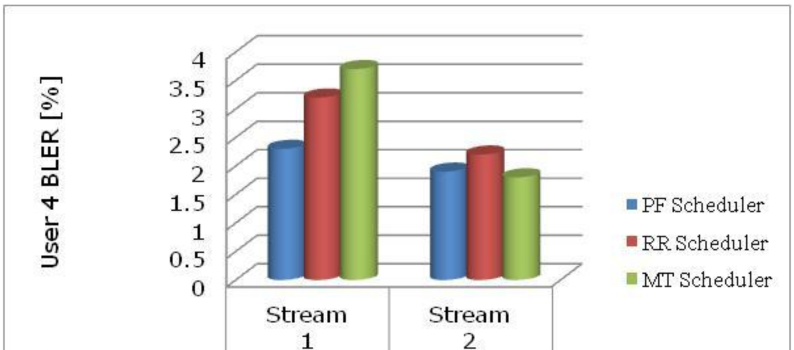


Figure 70 - Throughput and BLER for User 4 when MT scheduler is used



	Stream 1	Stream 2
PF Scheduler	2.3	1.9
RR Scheduler	3.2	2.2
MT Scheduler	3.7	1.8

Figure 71- BLER values when different schedulers are used

7.2.3.2 Conclusions

The overall throughput and BLER values obtained during this experiment are suggesting that DQOAS algorithm can be used with good results in 3GPP LTE networks if the traffic prioritization scheme is adapted in order to give the same priority for two different streams.

Compared with the two experiments presented in section 7.2.2, the BLER values obtained in this experiment are lower, which means that the user perceived quality is improved, as it can be observed from Figure 72. If DQOAS is used in conjunction with the proposed prioritization scheme, the number of users considered to be satisfied by the offered services is increased with 12% compared to the LTE delivery scheme (48%) and with 9% compared to DQOAS used with the original prioritization scheme (51%) [114], [115].

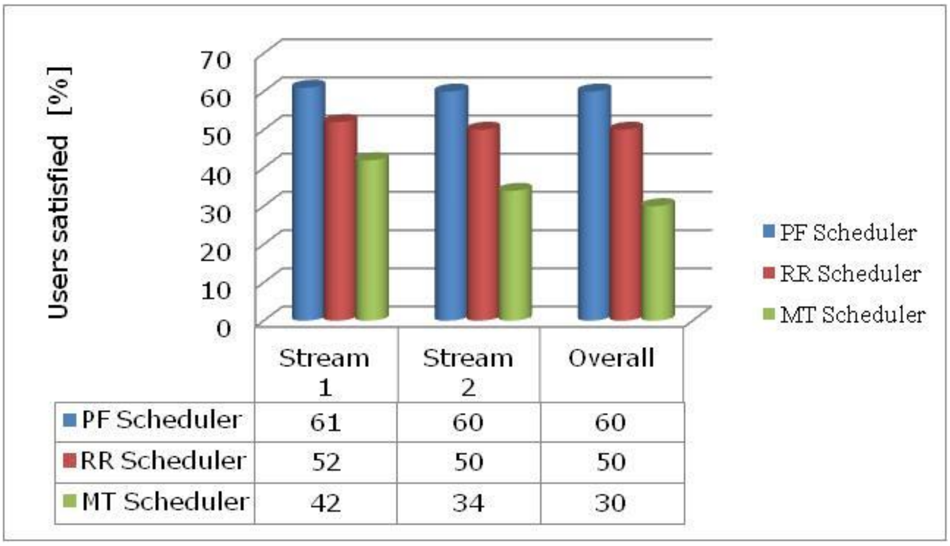


Figure 72- Satisfied users when different schedulers are used

One other aspect that can be noticed in Figure 72 is that like in previous scenarios, Proportional Fair scheduler performs better in terms of user satisfaction than Round Robin or Maximum Throughput [116].

7.2.4 Further Testing

Another set of tests were performed to further investigate the performances of the proposed adaptation mechanism in the situation where a third stream is delivered, in addition to the multimedia stream and TCP download [117]. This third stream is considered to be a VoIP connection. These new tests were conducted because the number of applications that generate these three different data flows are becoming more and more popular (e-Learning applications, live-conferencing, etc) and the situations where a user can consider the application experience as not satisfactory because one data flow is unsatisfactory might increase.

Each traffic type is managed differently by the downlink scheduler: VoIP data is scheduled using a semi-persistent strategy, while media-rich content and TCP-based traffic are scheduled dynamically, as depicted for exemplification in Figure 73. Using the proposed mapping scheme for multimedia streaming and TCP-based traffic, the scheduler will have to manage only two data flows with different characteristics instead of three. Considering the data-flows coming from the same application, DQOAS can update the quality levels for the multimedia stream based on user preferences, on instantaneous channel conditions and on the resource allocation scheme with a minimum impact on the VoIP traffic as it employs a semi-persistent scheduling policy.

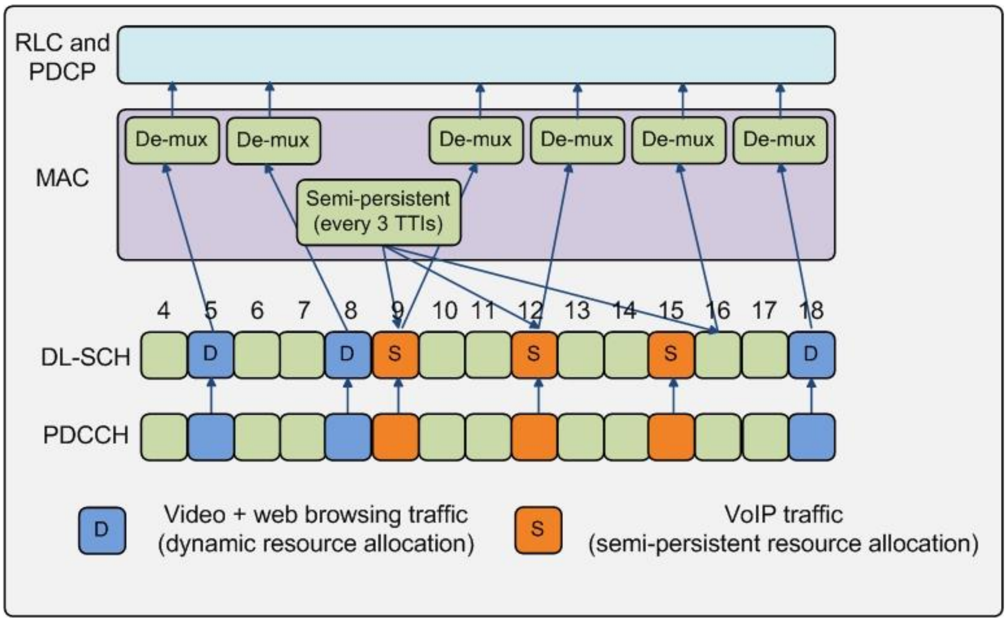


Figure 73 – Semi-persistent and dynamic scheduling

7.2.4.1 Three Different Data Flows Per User

A number of two simulation scenarios were considered for testing the performance of the proposed adaptation mechanism. In each simulation scenario, three streams with different priorities are considered for every user in order to simulate a traffic mix (VoIP, video streaming and web browsing data).

In the first scenario, the LTE delivery mechanism is tested using the proposed mapping scheme, in which video and web browsing data streams are considered to have the same priority with the same weight in the schedulers' queue while VoIP traffic employs a semi-persistent scheduler. Second scenario is using DQOAS as delivery algorithm in conjunction with the new mapping scheme. For both scenarios, three schedulers are considered: Maximum Throughput scheduler, Round Robin scheduler and Proportional Fair scheduler.

The minimum required level for the multimedia stream (stream 1) was set to 0.500 Mbps for all users, while for the second stream the minimum accepted level is set to 0.250 Mbps. It is considered that by using a semi-persistent scheduler, VoIP traffic is transmitted to all users at or above their quality expectations and no further analyzes were performed.

For each test scenario, mean BLER and throughput were analyzed for all users and also for one single user, User 2 (Table 31).

Table 31 – User 2 throughput and BLER average values when different schedulers are used

		PF scheduler		***RR scheduler***		***MT scheduler***	
		Scenario 1	*Scenario 2*	*Scenario 1*	*Scenario 2*	*Scenario 1*	*Scenario 2*
Throughput [Mbps]	Stream 1	0.653	0.674	0.526	0.680	0.691	0.717
	Stream 2	0.438	0.409	0.422	0.418	0.446	0.423
BLER [%]	Stream 1	1.8	1.7	1.7	1.8	2.3	2.1
	Stream 2	1.9	1.6	1.7	1.7	2.0	2.1

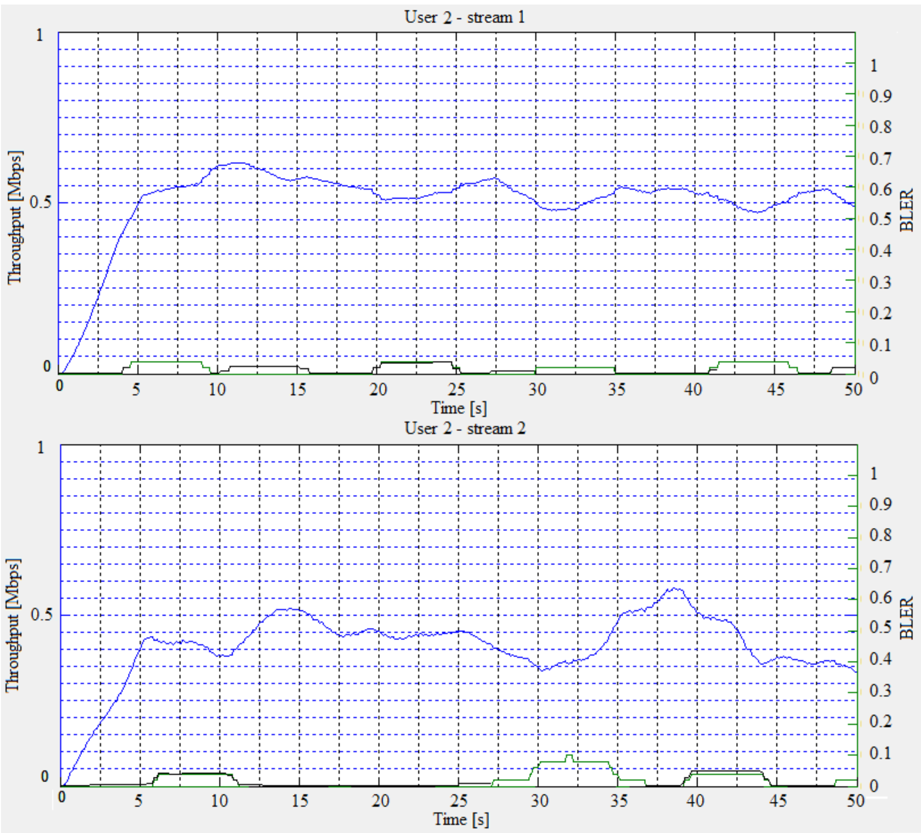


Figure 74 - Throughput and BLER for User 2(RR scheduler and LTE QoS mechanism)

Figure 74 presents the throughput and BLER experienced by User 2 for the first two streams, when it requests three different traffic types from the same application. The LTE QoS mechanism and the proposed mapping scheme are used in this case. In Figure 75, the same data is plotted when the new mapping scheme is deployed, together with DQOAS adaptive algorithm.

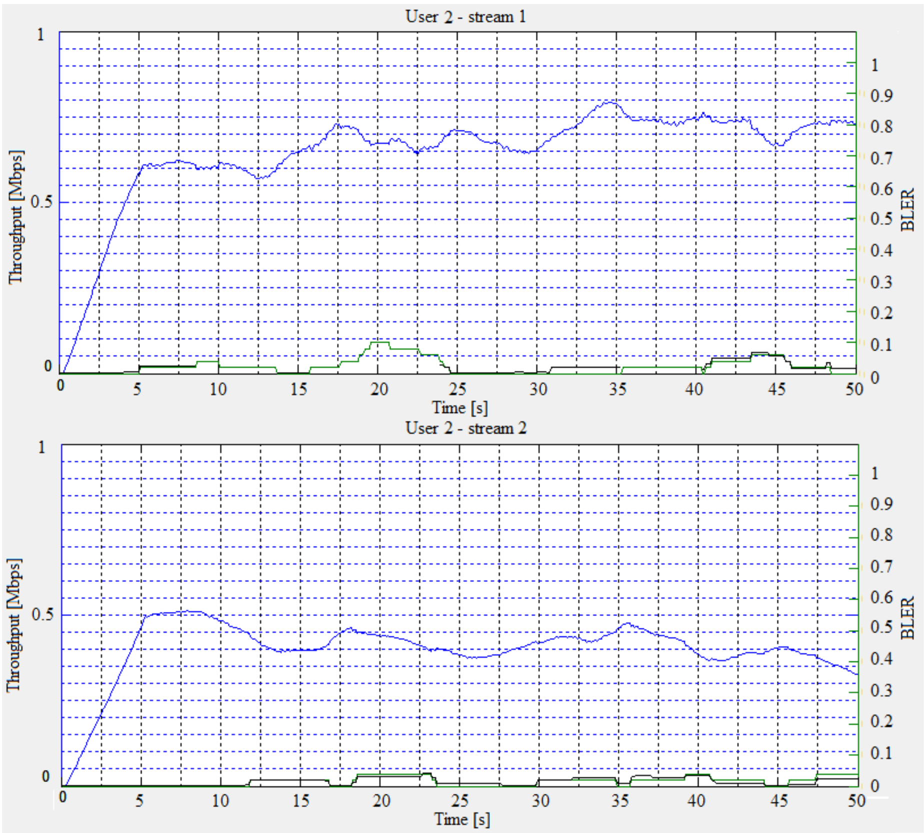


Figure 75 - Throughput and BLER for User 2 when Round Robin scheduler and DQOAS algorithm are used

It can be observed from Table 31 that the BLER values are reduced slightly in the second scenario for most cases, while maintaining an acceptable throughput value for both streams.

7.2.4.2 Conclusions

In terms of per-application satisfied users, it can be observed from Table 32 that again, Proportional Fair scheduler obtained better results than the other schedulers. The results obtained by DQOAS in this case are lower than the ones obtained in section 7.2.3 because in this case the scheduler performs a semi-persistent resources allocation for the third stream and by doing so, is reducing the available bandwidth that the other two streams can use. Comparing the user satisfaction rates obtained when LTE delivery mechanism and DQOAS algorithm are used, it can be noticed that also in this case, DQOAS algorithm outperforms the standard LTE mechanism [117].

Table 32 - The percentage of satisfied users when all three schedulers are considered

	The percentage of satisfied users when the scheduler used is:		
	PF	**RR**	**MT**
LTE Delivery mechanism	50	48	25
DQOAS algorithm	57	52	26

8. CONCLUSIONS AND FUTURE WORKS

This chapter summarizes the work done in this thesis, presenting different findings and conclusions. Further research works that can be performed based on this thesis are presented in the end.

8.1 Conclusions

During the last years, research efforts in the multimedia delivery field were concentrated on developing new algorithms able to improve the end-user quality of experience when a wireless technology is used as a communication environment. Offering a good end-user perceived quality is a pressing problem as an increasing number of users are accessing multimedia content from mobile devices via wireless networks. Complex applications, like e-learning or live conferencing, involves using multiple processes – web browsing, video and audio streaming, interactive voice calls and ftp background traffic – that are generating a rich traffic mix. Managing the traffic mix flows is an even more difficult problem, especially when it is delivered over wireless networks. This is because wireless technologies have limited radio resources and are highly susceptible of being affected by environmental factors, traffic load, number of clients and their mobility patterns. In a wireless medium, the radio resources fairly shared among the connected devices are used to deliver the required content. If one is to consider the adaptation process just between the delivery server and the device, than many adaptation algorithms may be able to offer very good performances. But in order to increase the end-used perceived quality, one has to take into account another factor, which is the users' individual set of characteristics. The fact that users have subjective opinions on the quality of a multimedia application can be used to increase their QoE by setting a minimum quality threshold below which the connection is considered to be undesired. Like this, the use of precious radio resources can be optimized in order to simultaneously satisfy an increased number of users.

In this thesis the above discussed aspects of a multimedia delivery were taken into consideration, with the goal to evaluate if the performances of an adaptation process can be improved in terms of end-user perceived quality. In the process of achieving this goal a new user-oriented adaptive algorithm based on QOAS was designed and developed to address the user satisfaction problem. Simulations have been carried out with different adaptation schemes to compare the performances and benefits of the DQOAS mechanism. The simulation results are showing that using a dynamic stream granularity with a minimum threshold for the transmission rate, improves the overall quality of the multimedia delivery process, increasing the total number of satisfied users and the link utilization, by influencing the radio resources allocation. The algorithm's decision to terminate a session when this is considered to be unsatisfying proves to be good, because it allows the scheduler to reallocate those resources to other users, maintaining their delivery rate over the minimum accepted level.

The good results obtained by the algorithm in IEEE 802.11 wireless environment, motivated the research about the utility of the newly proposed algorithm in another wireless environment, LTE. The study shows that DQOAS algorithm can obtain good results in terms of application perceived quality, when the considered application generates one or multiple streams. The results can be further improved by using a new QoS parameters mapping scheme.

Because the scheduler is one important component of the LTE QoS mechanism, some scheduling strategies were also analyzed and the results show that the Proportional Fair algorithm obtains better performances than Round Robin or Maximum Throughput.

8.2 Contributions

This thesis has the following contributions:

- *proposes a new user-oriented delivery algorithm able to improve the final users' QoE, the link utilization and the number of satisfied users by dynamically building the stream granularity based on current network conditions and on user preferences*

- *evaluates the algorithm's performances in a simulated IEEE 802.11 wireless environment, using both static and mobile scenarios, with and without background traffic*

- *evaluates the algorithm's performances using a simulated 3GPP LTE wireless environment, when one, two or three data flows are delivered simultaneously to each user*

- *uses a priority scheme to prioritize the different data flows in case a traffic mix is sent to a user and test the proposed solution*

- *proposes a new QoS mapping scheme for data flow prioritization in LTE in order to increase the performances obtained by DQOAS algorithm when an application is generating multiple traffic flows with different priorities*

- *demonstrates that the Proportional Fair scheduler used in LTE technology gives better results in all simulation scenarios, compared to Round Robin and Maximum Throughput schedulers.*

8.3 Future Works

As this thesis was only concerned with the downlink aspect of LTE, a natural extension of this work is to enlarge the testing grounds by including a corresponding study for the uplink, where the data transmission and the resource allocation are implemented differently.

Another interesting direction could be to investigate the scenarios in which users have both WLAN and LTE technologies available for multimedia streaming. The processes of inter-technology cooperation and network selection decision should be taken into consideration for a multi-technology adaptive algorithm.

Bibliography

[1] J. Kurose, D. Towsley, R. Koodli and J. Padhye, "A model based TCP Friendly rate control Protocol", in Proc. NOSSDAV, 1999.

[2] D. Sisalem and A. Wolisz, "LDA + TCP-friendly adaptation: A measurement and comparison study," Proc. NOSSDAV, 2000.

[3] L. Qiong and M. van der Schaar, "Providing adaptive QoS to layered video over wireless local area networks through real-time retry limit adaptation", IEEE Trans. Multimedia, vol. 6, no. 2, April 2004.

[4] K. Chen, K. Nahrstedt and S. H. Shah, "Dynamic bandwidth management for single-hop ad hoc wireless networks", IEEE Int'l Conf. on Pervasive Computing and Communications in Dallas-Fort Worth, USA, March 2003.

[5] K.-A. Cha, "Content complexity adaptation for MPEG-4 audio-visual scene", IEEE Trans. Cons. Elec., vol. 50, no. 2, pp. 760–765, May 2004.

[6] L. Murphy, P. Perry and N. Cranley, "Optimum adaptation trajectories for streamed multimedia", ACM Multimedia Systems Journal, New York: Springer-Verlag, 2005.

[7] V. H. Muntean – master thesis, "Adaptive multimedia streaming control algorithm in wireless LANs and 4G networks", Dublin City University, Ireland, 2010.

[8] G.-M. Muntean, P. Perry and L. Murphy, "A new adaptive multimedia streaming system for all-IP multi-service networks", IEEE Transactions on Broadcasting, vol.50, no.1, March 2004.

[9] G.-M. Muntean, P. Perry and L. Murphy, "Subjective Assessment of the Quality-Oriented Adaptive Scheme", IEEE Transactions on Broadcasting, vol. 51, no. 3, September 2005.

[10] Robert V. Bruce, "Alexander Graham Bell and the Conquest of Solitude", Little, Brown & Company, Canada, Toronto, ISBN 0-316-11251-8, 1973.

[11] 3GPP TS 05.01: "Technical Specification Group GSM/EDGE Radio Access Network; Physical layer on the radio path", Release 1999, 2004.

[12] 3GPP TS 05.02: "Technical Specification Group GSM/EDGE Radio Access Network; Multiplexing and multiple access on the radio path", Release 1999, 2003

[13] 3GPP TS 05.05: "Technical Specification Group GSM/EDGE Radio Access Network; Radio transmission and reception", Release 1999, 2005.

[14] R. Koodli and M. Puuskari, "Supporting packet-data QoS in next-generation cellular networks", IEEE Communications Magazine, vol. 39, no. 2, February 2001.

[15] ETSI TS 123.107: "Quality of Service (QoS) Concept and Architecture (Release 5)", Release 5, 2002.

[16] ETSI TS 150.059: "Enhanced Data rates for GSM Evolution (EDGE)", Release 4, 2001.

[17] 3GPP TS 23.107: "Technical Specification Group Services and System Aspects; Quality of Service (QOS) concept and architecture," Release 9, December 2009.

[18] 3GPP2 C.S0024-A "cdma2000 High Rate Packet Data Air Interface Specification", Version 3, 2006.

[19] 3GPP2 C.S0046-0 "3G Multimedia Streaming Services", Version 1, 2006.

[20] 3GPP2 SC.R2005-002-0 "System Release Guide for the UMB-1 Release of the cdma2000 System Specifications", Version 1, 2008.

[21] J. G. Andrews, A. Ghosh and R. Muhamed, "Fundamentals of WiMAX – Understanding Broadband Wireless Networks", Prentice Hall PTR, 2007.

[22] http://www.wimaxforum.org, "Mobile WiMAX-Part I: A Technical Overview and Performance Evaluation" on WiMAX Forum, August 2006.

[23] IEEE P802.16 Rev3: "Air Interface for Broadband Wireless Access Systems", 2009.

[24] L. Nuaymi, "WiMAX: Technology for Broadband wireless Access", John Wiley&Sons, ISBN 978-0470028087, 2007.

[25] M. Alasti, B. Neekzad, J. Hui and R. Vannithamby, "Quality of Service in WiMAX and LTE Networks", IEEE Communications Magazine, vol.48, no. 5, May 2010.

[26] S. Z. Asif, "WiMAX Developments in the Middle East and Africa", IEEE Communications Magazine, vol. 47, no. 2, February 2009.

[27] Darren McQueen, "The Momentum Behind LTE Adoption", IEEE Communications Magazine, vol. 47, no. 2, February 2009.

[28] V. H. Muntean and M. Otesteanu, "WiMAX versus LTE; An Overview of technical aspects for Next Generation Networks technologies", International Symposium on Electronics and Telecommunications, November 2010.

[29] K. Bogineni, R. Ludwig et co., "LTE Part I: Core Network – Guest Editorial", IEEE Communications Magazine, vol. 47, no. 2, February 2009.

[30] B. Furht and S. Ahson, "Long Term Evolution – 3GPP LTE Radio and Cellular Technology". CRC Press, Taylor & Francis Group, LLC, 2009.

[31] 3GPP TS 36.300 v8.7: "Technical Specification Group Radio Access Network, Rel. 8," 3GPP LTE, Technical Report 2008.

[32] I. Toufik and M. Baker S. Sesia, "LTE – The UMTS Long Term Evolution: From Theory to Practice", John Wiley & Sons, Ltd., 2009.

[33] A. Larmo, M. Lindstrom et co., "The LTE Link-Layer Design", IEEE Communications Magazine, vol. 47, no. 4, April 2009.

[34] D. Kliazovich, F. Granelli, S. Redana and N. Riato, "Cross-Layer Error Control Optimization in 3G LTE", IEEE Global Telecommunication Conference, Trento, 2007.

[35] 3GPP TS 36.321: "E-UTRA Medium Access Control (MAC) Protocol Specifications", Release 8, 3GPP LTE, Technical Report 2008.

[36] E. Dahlman, "3G Evolution: HSPA and LTE for Mobile Broadband", 2nd edition, Academin Press, 2008.

[37] ETSI TS 136.213: "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures", Release 8, 2009.

[38] 3GPP TS 36.306: "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) User Equipment (UE) radio access capabilities", Release 8, 2007.

[39] Hannes Ekström, "QoS Control in the 3GPP Evolved Packet System", IEEE Communications Magazine, vol. 47, no. 2, February 2009.

[40] S. Sesia, I. Toufik and M. Baker, "LTE – The UMTS Long Term Evolution: From Theory to Practice", John Wiley&Sons, ISBN (ebook) 978-0470978511, 2011.

[41] 3GPP TS 23.203: "Technical Specification Group Services and System Aspects; Policy and charging control architecture", Release 8, 2010.

[42] IEEE Technical Report: "IEEE 802.11 Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specification", 2000.

[43] S. Mangold, L. Berlemann and B. H. Walke, "IEEE 802 Wireless Systems: Protocols, Multi-hop Mesh/Relaying, Performance and Spectrum Coexistence", John Wiley & Sons, ISBN 978-0470014394, 2007.

[44] http://www.bluetooth.org, "Specification of the Bluetooth System", Version 1.1, 2001.

[45] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function", IEEE Journal on Selected Area in Comm., vol.18, no.3, 2000.

[46] A. Ksentini, A. Nafaa, A. Gueroui, M. Naimi, "Determinist Contention Window Algorithm for IEEE 802.11", IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC 2005), vol.4, 2005.

[47] Q. Ni , L. Romdhani and T. Turletti, "A Survey of QoS Enhancements for IEEE 802.11 Wireless LAN", Journal of Wireless Communications and Mobile Computing, vol.4, no. 5, Wiley, 2004.

[48] IEEE, Technical Report, "IEEE 802.11e: Medium Access Control (MAC) Quality of Service Enhancements", 2005.

[49] Q. Ni, "Performance Analysis and Enhancements for IEEE 802.11e Wireless Networks", IEEE Network, vol.19, no.4, 2005.

[50] ITU-T Recommendation E.771: "Network Grade of Service parameters and target values for circuit-switched land mobile services", 1994.

[51] ITU-T Recommendation Y.2001: "Next Generation Networks – Frameworks and functional architecture models", 2004.

[52] P. Cholda, A. Mykkeltveit, B. E. Helvik, O. J. Wittner and A. Jajszczyk, "A survey of resilience differentiation frameworks in communication networks", IEEE Communications Surveys & Tutorials, vol.9, no.4, 2007.

[53] IETF Technical Report RFC 896 by (J. Nagle): "Congestion Control in IP/TCP Internetworks", 1984.

[54] IETF Technical Report RFC 791: "Internet Protocol", 1981.

[55] IETF, Technical Report RFC 2460 (by S. Deering and R. Hinden): "Internet protocol version 6 (ipv6) specifications", 1998.

[56] IETF Technical Report RFC 768 (by J. Postel): "User datagram protocol", 1980.

[57] IETF Technical Report RFC 793: "Transmission control protocol", 1981.

[58] IETF Technical Report RFC 3168: "The addition of explicit congestion notification (ecn) to ip", 2001.

[59] D. P. Reed, D. D. Clark and J. H. Saltzer, "End-to-end arguments in system design", ACM Transactions on Computer Systems, vol. 2, 1984.

[60] J. Qhang, M. E. Aydin and J. Yang, "Optimization of WCDMA radio networks with consideration of link-level performance factors", International Journal of Mobile Networks Design and Innovation, vol. 2, no. 1, 2007.

[61] H. Yu, M. Handley and D. Estrin R. Rejaie, "Multimedia proxy caching for quality adaptive streaming applications in the internet", Proc. 19th Annual Joint Conference IEEE Computer and Communications Societies 2000 (INFOCOM'00), Tel-Aviv, Israel, March 2000.

[62] IETF Technical Report RFC 3550 (by H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson): "Rtp: A transport protocol for real-time applications", 2003.

[63] Colin Perkins, "RTP Audo and Video for the Internet", Pearson Education, Inc, Boston, ISBN 978-0321833624, 2003.

[64] IETF Technical Report RFC 2481 (by K. Ramakrishnan and S. Floyd): "A proposal to add Explicit Congestion Notification (ECN) to IP", 1999.

[65] F. Hultin, "Congestion Notification and rate adaptation for real-time services in all-IP radio networks", Lulea University of Technology, Master Thesis.

[66] R. Rejaie, M. Handley and D. Estrin, "RAP: An end-to-end rate-based congestion control mechanism for realtime streams in the internet", IEEE Computer and Communications Society (INFOCOM), vol. 3, March 1999.

[67] N. Aboobaker, D. Chanady, M. Gerla and M. Sanadidi, "Streaming media congestion control using bandwidth estimation", IEEE International Conference on Management of Multimedia and Mobile Networks, October 2002.

[68] S. Mascolo, C. Casetti, M. Gerla, M. Y. Sanadidi and R. Wang, "TCP westwood: Bandwidth estimation for enhanced transport over wireless", Mobile Computing and Networking, 2001.

[69] F. Menta and G. Schembra, "Effcient design of red routers for tcp/rap fairness optimization", IEEE International Conference on Communications, vol. 4, 2002.

[70] IETF Technical Report RFC 3448 (by M. Handley, S. Floyd, J. Padhye and J. Midmer): "TCP friendly rate control (TFRC): Protocol specification", 2003.

[71] S. Cen, P. C. Cosman and G. M. Voelker, "End-to-end differentiation of congestion and wireless losses", IEEE/ACM Transactions on Networking, vol. 11, no. 5, 2003.

[72] M. Chen and A. Zakhor, "Multiple tfrc connections based rate control for wireless networks", IEEE Transactions on Multimedia, vol. 8, no. 5, October 2006.

[73] P. Zhu, W. Zeng and C. Li, "Joint design of source rate control and qos-aware congestion control for video streaming over the internet", IEEE Transaction on Multimedia, 2007.

[74] H.-P. Shiang and M. van der Schaar, "Content-aware TCP-friendly congestion control for multimedia transmission", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, May 2011.

[75] H.-P. Shiang and M. van der Schaar, "A Quality-Centric TCP-Friendly Congestion Control for Multimedia Transmission", IEEE Transactions on Multimedia, vol. 14, no. 3, June 2012.

[76] G. Yang, M. Gerla and M. Sanadidi, "Adaptive video streaming in the presense of wireless errors", Management Multimedia Networks & Services, October 2004.

[77] ISO/IEC 23009-1:2012, "Information technology -- Dynamic adaptive streaming over HTTP (DASH) -- Part 1: Media presentation description and segment formats", April 2012.

[78] Thomas Stockhammer, "Dynamic Adaptive Streaming over HTTP – Standards and Design Principles", ACM Multimedia Systems 2011, San Jose, California, USA, February 2011.

[79] S. Xiang, L. Cai and J. Pan, "Adaptive scalable video streaming in wireless networks", ACM Multimedia Systems 2012, Chapel Hill, North Carolina, USA, February 2012.

[80] M. Kennedy, H. Venkataraman and G.-M. Muntean, "Battery and Stream-Aware Adaptive Multimedia Delivery for wireless devices", IEEE 35th Conference on Local Computer Networks (LCN), Denver, Colorado, USA, October 2010.

[81] H. Venkataraman and G.-M. Muntean, "Dynamic Time Slot Partitioning for Multimedia Transmission in Two-Hop Cellular Networks", IEEE Transactions on Mobile Computing, vol. 10, no. 5, April 2011.

[82] G.-M. Muntean, "Efficient delivery of multimedia streams over broadband networks using QOAS", IEEE Transactions on Broadcasting, vol. 52, no. 2, June 2006.

[83] G.-M. Muntean and N. Cranley, "Resource efficient quality-oriented wireless broadcasting of adaptive multimedia content", IEEE Transactions on Broadcasting, vol. 53, no. 1, March 2007.

[84] IETF Technical Report RFC 1633 (by R. Braden, D. Clark and S. Shenker): "Integrated services in the internet architecture: An overview", 1994.

[85] IETF Technical Report RFC 2475 (by S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang and W. Weiss): "An architecture for differentiated services", 1998.

[86] F. Granelli and D. Kliazovich, "Cross-Layering for Performance Improvement in Multi-Hop Wireless Networks", Proceedings of the 8th International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN 2005), Trento Univ., Italy, 2005.

[87] V. Kawadia and P. R. Kumar, "A cautionary perspective on cross-layer design", IEEE Wireless Communications, vol.12, no.1, 2005.

[88] H. Fattah and C. Leung, "An overview of scheduling algorithms in wireless multimedia networks", IEEE Wireless Communications, vol.9, no.5, 2002.

[89] B. Sadiq, R. Madan and A. Sampath, "Downlink schedulilng for multiclass traffic in LTE", EURASIP Journal on Wireless Communications and Networking, vol.2009, July 2009.

[90] A. Banchs and X. Perez, "Distributed weighted fair queueing", International Communications Conference, vol. 5, April 2002.

[91] G. Ye and J. Hou, "An analytical model for service differentiation in IEEE 802.11", IEEE International Conference on Communications, vol.5, 2003.

[92] N. Vaidya, P. Bahl and S. Gupa, "Distributed fair fcheduling in a wireless LAN", Annual International Conference on Mobile Computing and Networking, August 2000.

[93] S. Golestani, "A self-clocked fair queueing scheme for broadband applications", IEEE Infocom, 1994.

[94] M.-H. Lu, P. Steenkiste and T. Chen, "Video streaming over 802.11 WLAN with content-aware adaptive retry limit", IEEE International Conference on Multimedia & Expo, July 2005.

[95] K. Seong, M. Mohseni and J. M. Cioffi, "Optimal resource allocation for OFDMA downlink systems", Proceedings of the IEEE International Symposium on Information Theory (ISIT '06), Seattle, USA, July 2006.

[96] N. Chen and S. Jordan, "Downlink scheduling with probabilistic guarantees on short-term average throughputs", Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '08), Las Vegas, USA, April 2008.

[97] L. A. de Temino, G. Berardinelli et co., "Channel-Aware Scheduling algorithms for SC-FDMA in LTE Uplink", IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2008), Cannes, France, September 2008.

[98] H. Wang and D. Jiang, "Performance Comparison of Control-less Scheduling Policies for VoIP in LTE UL", IEEE Wireless Communications and Networking Conference (WCNC 2008), Las Vegas, USA, April 2008.

[99] M. Al-Rawi, R. Jantti, J. Torsner and M. Sagfors, "On the Performance of Heuristic Opportunistic Scheduling in the Uplink of 3G LTE Networks", IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2008), Cannes, France, September 2008.

[100] M. Wernersson, S. Wanstedt and P. Synnergren, "Effects of QoS scheduling strategies on performance of mixed services over LTE", Proceedings of the 18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '07), Athens, Greece, September 2007.

[101] J. Puttonen, N. Kolehmainen et co., "Mixed Traffic Packet Scheduling in UTRAN Long Term Evolution Downlink", IEEE 19th International Symposium on

Personal, Indoor and Mobile Radio Communications (PIMRC 2008), Cannes, France, September 2008.

[102] I. Siomina and S. Wanstedt, "The Impact of QoS Support on the End User Satisfaction in LTE Networks with Mixed Traffic", IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2008), Cannes, France, September 2008.

[103] *www.netchico.com/support/glossary/c.html*, August 2007.

[104] D.-M. Chiu and R. Jain, "Analysis of the increase and decrease algorithms for congestion avoidance", Computer Networks, Journal of Computer Networks and ISDN Systems, 1989.

[105] V. H. Muntean and G.-M. Muntean, "A novel adaptive multimedia delivery algorithm for increasing user quality of experience during wireless and mobile e-learning", IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, BMSB '09, 2009.

[106] O. Verscheure, P. Frossard, and M. Hamdi, "User-oriented QoS analysis in MPEG-2 video delivery," Journal of Real-Time Imaging, vol. 5, no. 5, October 1999.

[107] NOAH NS-extension, *http://icapeople.epfl.ch/widmer/uwb/ns-2/noah/*

[108] V. H. Muntean, M. Otesteanu and G.-M. Muntean "QoS-oriented Multimedia Delivery over 4G Wireless Networks: Dynamic Quality-Oriented Adaptation Scheme - a user-oriented adaptation mechanism", LAP LAMBERT Academic Publishing, ISBN 978-3846545911, November 2011.

[109] V. H. Muntean and M. Otesteanu, "QoS-oriented Multimedia Delivery Algorithm for Next Generation Wireless Networks", Workshop "Interdisciplinabilitatea si managementul cercetarii – Prezentarea rezultatelor obtinute de doctoranzi", POSDRU/88/1.5/S/50783, Timisoara, Romania, 24-25 November 2011.

[110] J. C. Ikuno, M. Wrulich and M. Rupp, "System level simulator of LTE networks", Proceedings of IEEE 71st Vehicular Technology Conference, Taipei, Taiwan, May 2010.

[111] M. Wrulich, S. Eder, I. Viering and M. Rupp, "Efficient link-to-system level model for MIMO HSDPA", Proc. of the 4th IEEE Broadband Wireless Access Workshop, 2008.

[112] K. Brueninghaus, D. Astelyet co., “Link performance models for system level simulations of broadband radio access systems”, IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2005), September, 2005.

[113] V. H. Muntean and M. Otesteanu, “QoE-oriented multimedia delivery algorithm for e-learning in next generation wireless networks”, Proceedings of the 8th International Scientific Conference "eLearning and Software for Education", Bucharest, Romania, April 26 - 27, 2012.

[114] V. H. Muntean, M. Otesteanu and G.-M. Muntean, “DQOAS Performances for Traffic Mix Delivery over LTE Networks Using a New QoS Parameter Mapping Scheme”, Scientific Bulletin of “Politehnica” University of Timisoara, Romania – Transactions on Automatic Control and Computer Science, ISSN 1224-600x, Ed. “Politehnica”, Timisoara, Romania, vol.55,No. 3, September 2010, pp 161-170.

[115] V. H. Muntean, M. Otesteanu and G.-M. Muntean, “QoS parameters mapping for the e-learning traffic mix in LTE networks”, International Joint Conference on Computational Cybernetics and Technical Informatics (ICCC-CONTI), Timisoara, Romania, May 2010.

[116] V. H. Muntean and M. Otesteanu, “Techniques for improving the overall QoE for applications”, Workshop “Cercetari doctorale in domeniul tehnic”, POSDRU/88/1.5/S/50783, Craiova, Romania, February 2011.

[117] V. H. Muntean and M. Otesteanu, “Performance evaluation of DQOAS algorithm in case of applications generating VoIP and video streaming when a new QoS prioritization scheme for LTE is used”, IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Nuremberg, Germany, 8-10 June 2011.