Adaptive multimedia streaming control algorithm in wireless LANs and 4G networks

Vasile Horia Muntean
B. Eng.

Thesis presented for the degree of
Masters in Engineering
to the
Faculty of Electronic Engineering
Dublin City University
Ireland

Research Supervisor
Dr. Gabriel-Miro Muntean
August 2010

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading o the award of Masters of Engineering is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: ___________________________

Vasile Horia Muntean

ID No. : 58119965

Date: ___________________________

Acknowledgements

First, I want to express my heartily gratitude to my supervisor, Dr. Gabriel-Miro Muntean, whose guidance and continuous technical and moral support proved to be crucial to the realization of this thesis. Also, 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, I would like to thank to my dear friends Arghir, Ramona and Andreea, to my sister Ada and to a special person, Rocs, who all offered me their support all this time.

Abstract

E-learning has become an important service offered over the Internet. Lately many users are accessing learning content via wireless networks and 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. Additionally, to support this for mobile e-learning over wireless LANs becomes even more difficult. If we want to increase the end-used perceived quality, we have to take into account 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 a new user-oriented adaptive algorithm based on QOAS was designed and developed in order 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

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 multiple streams. These results can be improved by using a new QoS parameters mapping scheme able to modify the streams' priority and thus allowing the algorithms decisions to not be overridden by the systems' scheduler.

List of publications

1) V. H. Muntean, G.-M. Muntean – "A novel adaptive multimedia delivery algorithm for increasing user quality of experience during wireless and mobile e-learning", in IEEE International Symposium on Broadband Multimedia Systems and Broadcasting 2009, Bilbao, Spain.

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

List of figures

List of Tables

INDEX

3GPP - 3^{rd} 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
TBTT - Target Beacon Transmission Time
TC - Traffic Categories
TCP - Transport Control Protocol
TDD - Time-Division Duplex
TFRC - TCP Friendly Rate Control

TFRC - TCP-Friendly Rate Control Protocol
TFRCC - TCP Friendly Rate Control with Compensation
TFRC-W - TFRC Wireless
TTI - Transmission Time Interval
TxD - Transmission Diversity
TxOP - Transmission Opportunity
UDP - User Datagram Protocol
UE - User Equipment
UL - Uplink
UM - User Model
UMTS - Universal Mobile Telecommunications Systems
VoIP - Voice over IP
VTP - Video Transport Protocol
WiMAX - Worldwide Interoperability for Microwave Access
WLAN - Wireless LAN
WMAN - Wireless Metropolitan Area
WNIC - Wireless Network Interface Cards
WPAN - Wireless Personal Area Network

Table of Contents

1. Introduction

This chapter starts with a brief introduction of this thesis to the reader followed by the problem statement, the goals and the scope of this thesis. The chapter ends by presenting the contributions brought in the area of quality-oriented adaptive wireless multimedia streaming followed by the thesis structure.

1.1 Overview

Multimedia content is being delivered via many types of networks to viewers 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. This quality of experience influences in turn the quality of any application. It is important to realise 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 LANs, as wireless technologies offer lower bandwidth and their service is highly affected by environmental factors, traffic load and number of clients, as well as their location and mobility patterns.

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) [1]. PAMAHs' goal is to optimise 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.

First part of this thesis introduces a novel solution for adaptive multimedia streaming over IEEE 802.11 Wireless LANs – Dynamic Quality-Oriented Adaptation Scheme (DQOAS) – with focus on mobile e-learning. The proposed algorithm concentrates on performing an optimal multimedia content management in accordance with end-user learning profile, knowledge level, goals, preferences as well as network conditions. The algorithm supports higher end-user quality of experience during the learning process by considering important factors that influence user perceived quality, adaptive multimedia delivery aspects and the main factors affecting learning process.

Taking into account the good results of DQOAS and its user-oriented approach in IEEE 802.11 wireless LANs, it is of further interest to analyze the behavior of this algorithm over other wireless networks, like Next Generation Networks (NGNs), and in this thesis Long Term Evolution (LTE) is the chosen technology as the next wide coverage wireless network. In order for DQOAS to perform optimally when used over LTE networks, it is necessary to use a new QoS parameters mapping scheme that changes the original data flow prioritization. The proposed scheme together with the results obtained when DQOAS algorithm is used over LTE are presented in the second part of this thesis.

1.2 Problem statement

E-learning has become an important service offered over the Internet. Lately many users are accessing learning content via wireless networks and 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. Additionally, to support this for mobile e-learning over wireless LANs becomes even more difficult. 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) [2] and the enhanced Loss-Delay Adaptation Algorithm (LDA+) [3] but they are mainly designed to offer good network-related results when streaming multimedia over wired networks. Using these algorithms over wireless networks returns medium/poor multimedia quality and does not include end-user perceived quality in the adaptation process. Also, a variety of solutions have been proposed for streaming scalable multimedia content over wireless networks [4] or wireless ad-hoc networks [5]. Among these are adaptive algorithms that operate at the level of layers [5] or objects [6], fine-granular scalability schemes [7] and perception-based approaches [8] are most suitable for further research considering the results obtained. However none of these algorithms has considered user preferences and interests as well as network delivery conditions together.

Consider a typical IEEE 802.11g 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 fair 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. Figure1 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, let's assume that each has a minimum required throughput 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. As a consequence it is 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. 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 used resources for the other users. There will 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.

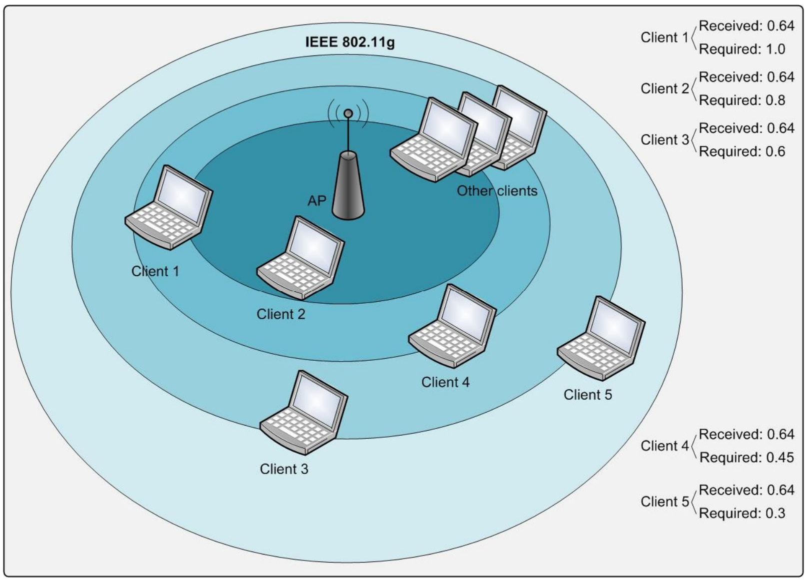


Figure 1 - Hypothetical streaming scenario (the received and required rates are expressed in Mbps)

In a LTE network, some problems can appear when one application is generating more than one stream for every user. If those streams are having different priorities, it is possible that the scenario presented above for IEEE 802.11 WLANs to affect the perceived quality of one of the streams. The solution presented for WLANs will be efficient only for streams that have the same priority and in consequence, the overall quality of one application generating two streams with different priorities can be perceived by the user as unsatisfactory if only one stream is received in good quality conditions. To overcome this problem, a new prioritization scheme must be used in order to optimize the proposed 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). In this conditions, the dynamically adaptation solution proposed above can manage the resource allocation for all streams, improving the final user satisfaction.

1.3 Contributions of this thesis

This thesis proposes the Dynamic Quality-Oriented Adaptation Scheme (DQOAS), a user-oriented adaptation mechanism that improves the quality of experience while also enabling a higher number of simultaneously connected clients to communicate in a IEEE 802.11 network. DQOAS algorithm enhances the Quality-Oriented Adaptation Scheme (QOAS) [9], [10] by adding different user QoE expectation levels in the adaptation process. Consequently, DQOAS will perform a differentiate treatment of the users, adaptive process 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.

In order to improve the DQOAS performances when one application is generating multiple streams with different priorities in a LTE network, a new QoS parameters mapping scheme for data flows is proposed. This prioritization scheme is modifying the stream priorities to that of the stream with the highest priority, give one priority class for the entire application generating these streams. Used in conjunction with DQOAS, this new scheme offers a mapping alternative for LTE QoS parameters in case on an e-learning traffic mix, in order to obtain an optimal end-user QoE.

1.4 Thesis structure

In this thesis, chapter 2 presents the background study and literature review. It gives an overview of the studied technologies, their advantages and disadvantages and reviews different methods proposed by the research community, trying to identify the problems concerning the stated problem.

Chapter 3 introduces and describes the design and the functioning stages of the proposed adaptation mechanism. It also presents the proposed solution for LTE QoS parameters mapping used to optimize the algorithms' performances.

Chapter 4 presents the simulation environment along with the test setup used. The simulation scenarios are detailed and the results obtained are discussed.

Chapter 5 summarizes the research work and presents further work.

2. Background

First two sections of this chapter are presenting in detail the architecture, the physical and MAC layer of the two wireless technologies used in this thesis: IEEE 802.11 wireless LANs and 3GPP LTE networks. In the third section various network and transport layer protocols used to deliver multimedia content are discussed while section 2.4 is offering an overview of the research efforts in multimedia streaming.

2.1 Wireless LAN Networks

IEEE 802.11 [11] standardized WLAN technology has became today the worldwide accepted technology for wireless short-range internet access, similar to what the GSM (Global System for Mobile Communications) 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 three sub-sections are offering an overview of the standard, focusing on detailing the physical and MAC layers.

2.1.1 Introduction

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 (Figure 2) to the wider internet for a limited number of users. This gives users 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 have 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.

Figure 2 – Wireless access using a laptop and an AP

A classification of available wireless networks based on coverage can be the one presented in Figure 3:

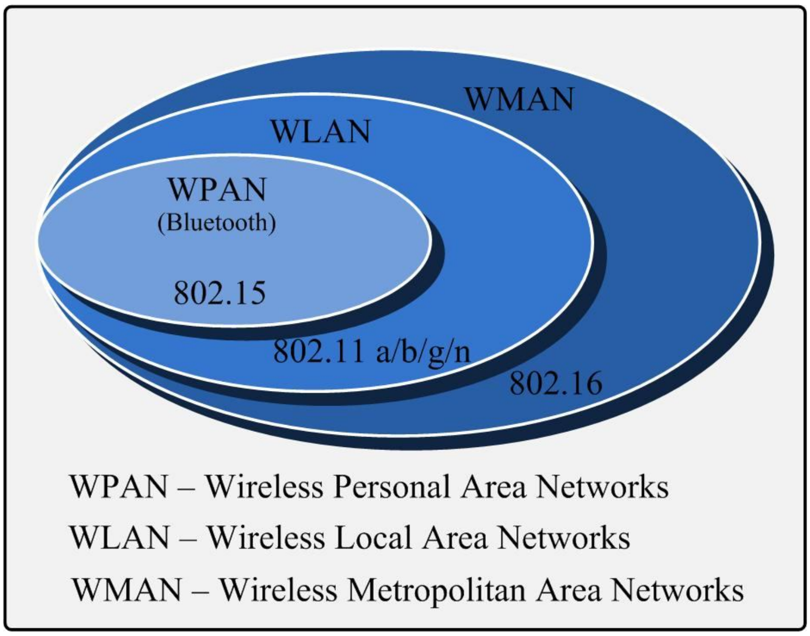


Figure 3 - Classification on wireless networks based on coverage

The IEEE 802.11 standard has three basic topologies for a network [12]:

- 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 4 (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 a Ethernet network. This type of network is also known as infrastructure mode (Figure 4 (b)). For the ESS there are a number of interconnected APs using an infrastructure called Distribution System (DS) (e.g. Ethernet).

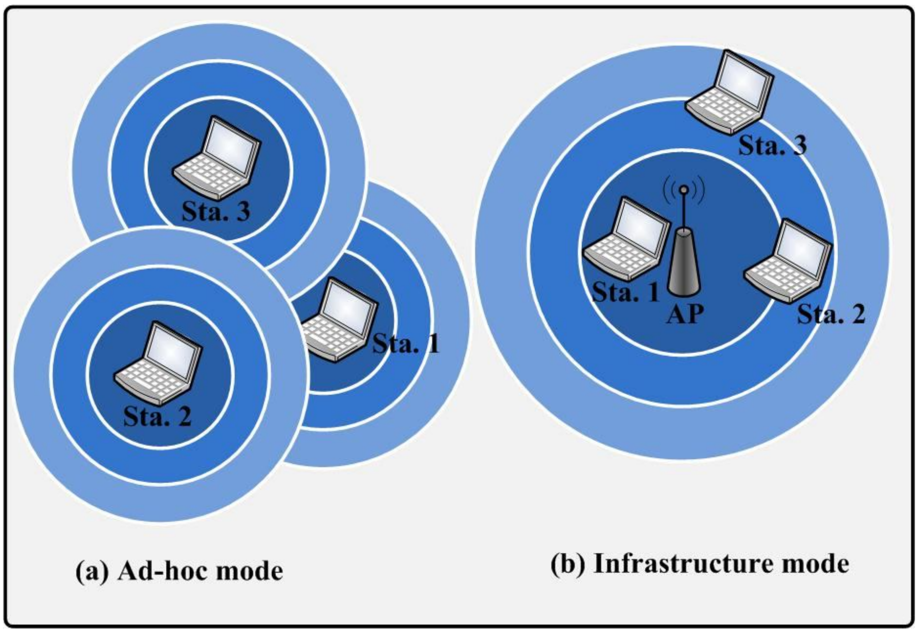


Figure 4 - 802.11 modes of operation

From 802.11 family, the most popular standards are those defined by the 802.11b and 802.11g protocols, which are amendments to the original standard [12]. 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 1 presents the parameters of 802.11 network standards:

Table 1 - 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 5.

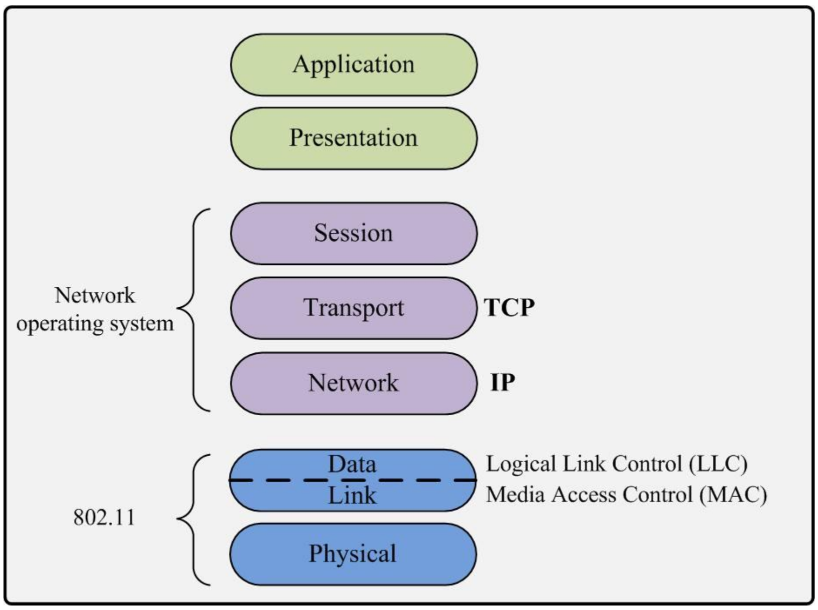


Figure 5 - WLAN protocol stack

2.1.2 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). In case of newer physical payers 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.

Table 2 - OFDM Parameters of IEEE 802.11a [12]

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

Orthogonal Frequency Division Multiplexing (OFDM) 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 used data while 4 are used for pilot symbols.

Table 2 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.

2.1.3 Media Access Control (MAC) Sublayer

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 [11]. The mandatory Distributed Coordinator Function (DCF) is based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) and the optional Point Coordinator Function is based on a pooling mechanism.

Distributed Coordination Function

In wireless communication, 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 needs 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. 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 *CWmin*. It is doubled for every unsuccessful transmission attempt up to a maximum value *CWmax* as depicted in Figure 6. If a successful transmission is achieved the CW is reset to the *CWmin* 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. 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.

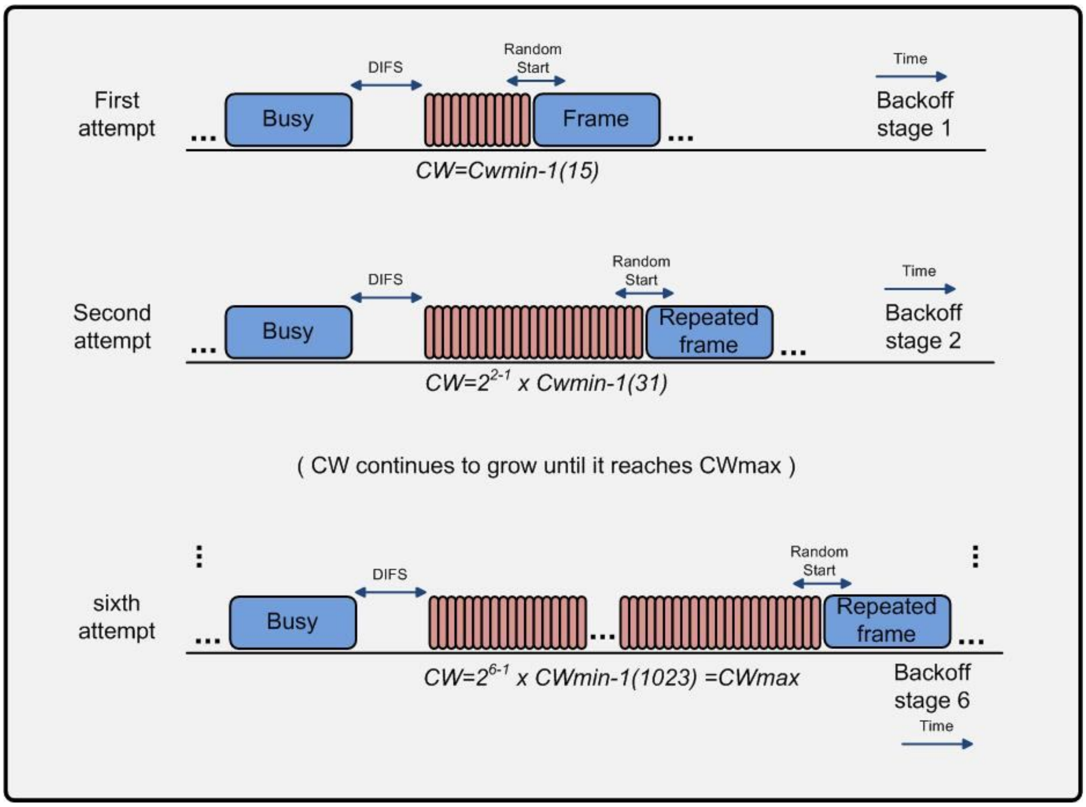


Figure 6 - Backoff procedure for collision avoidance in IEEE 802.11a

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 *ACKtimeout*, 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" because the procedure is performed after the transmission is acknowledged. The aim of the post-backoff is to ensure that any frame waiting to be transmitted, will be delivered with backoff. An MSDU arriving in the station's queue from a higher layer can be transmitted without a delay if the transmission queue is empty, the post-backoff procedure is finished and the channel is idle for a time duration equal to at least DIFS. Using this procedure in case of a lightly loaded system a shorter delivery delay can be obtained.

To reduce the duration the channel is occupied by a transmission, data frames (MSDUs) can be partitioned 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 benefit of fragmentation is that in case of an erroneous transmission, there is less data to be retransmitted. It also increases the probability of successful MSDU transmission in cases where the radio channel characteristics cause errors for longer frames than for shorter frames. It is important to mention 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. The receiver is performing the process called defragmentation by recombining the received MPDUs into a single MSDU.

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. 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. The decision to use RTS/CTS it is made locally by the receiving station. Before transmitting a data frame, a station may transmit a short RTS frame. 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 7. The transmitting station sends a RTS broadcast to all nodes within its carrier sense range. 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.

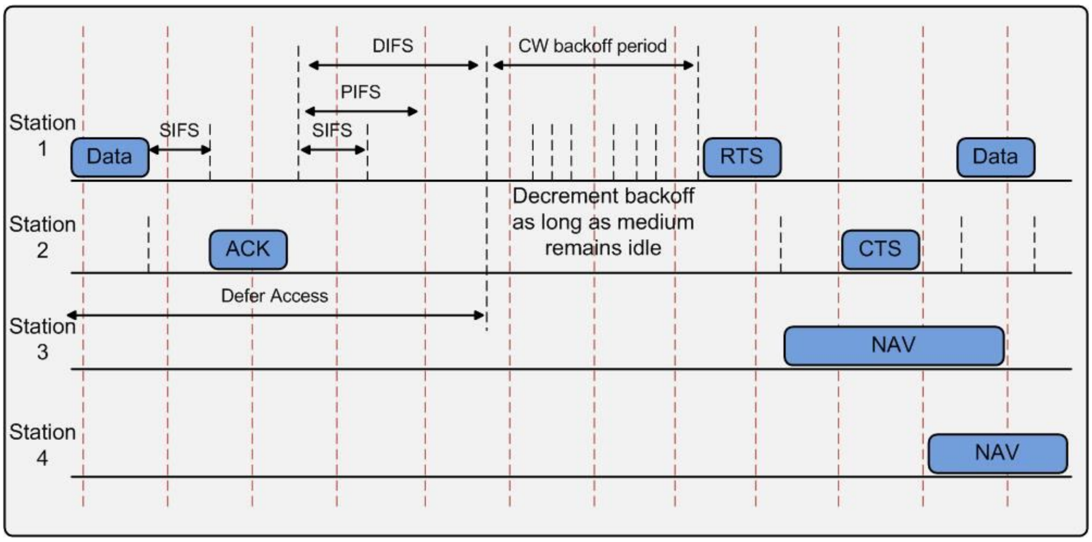


Figure 7 - RTS/CTS mechanism

Point Coordination Function

PCF is an optional MAC access mechanism that provides support for time-sensitive services by allowing stations to have priority access to the wireless medium, coordinated using a Point Coordinator (PC) which resides in the AP if in infrastructure mode. When a Basic Service Set is using the PCF, the medium is divided into Contention Free Period (CFP) and Contention Period (CP) intervals called superframes. If DCF is used to access the channel during CP intervals, PCF is used to access the channel during the CFP intervals. During the CFP the AP will poll stations for pending frames and deliver any pending downstream frames. The APs' 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.

The PCF also comes with a sum of problems, which include the unpredictable beacon delays and unknown transmission durations of the polled stations. At Target Beacon Transmission Time (TBTT), even if the PC schedules the beacon as the next frame to be transmitted, the beacon can only be transmitted when the medium has been idle for at least PIFS. Depending on the state of the wireless medium at TBTT, the beacon can be delayed from TBTT, which determines the delay of the transmission for time-bounded MSDUs that must be delivered in the CFP. This is a weak spot because it introduces time delays in each CFP. Another problem is the unknown transmission duration of polled stations. One station that was polled by PC 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 reduces the QoS that is provided to other stations that are polled during the rest of the CFP.

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) [13], 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 8 illustrates the main elements of the 802.11e MAC architecture in the context of 802.11:

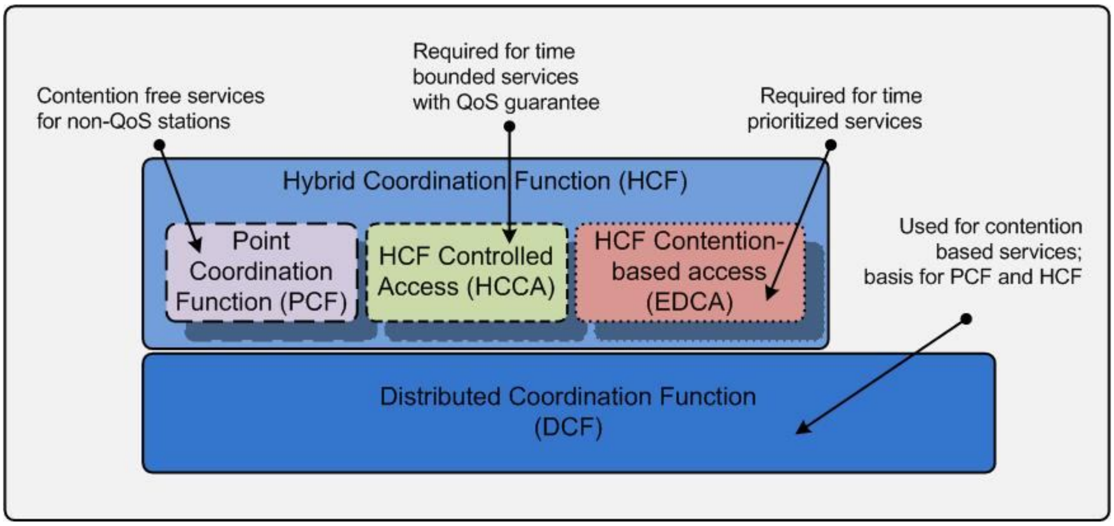


Figure 8 - MAC architecture of 802.11e [12]

Enhanced Distributed Channel Access [12] 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 9.

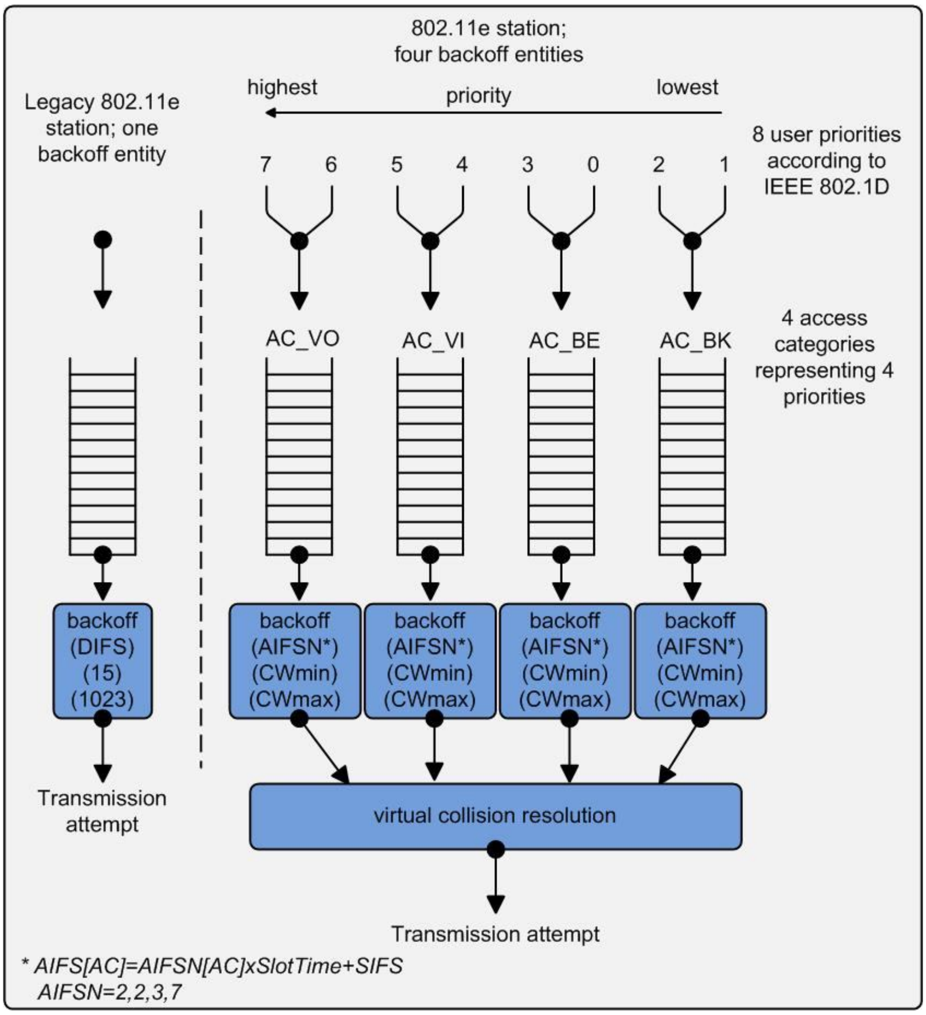


Figure 9 - Legacy 802.11 station and 802.11e station with 4 ACs

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 contends for a TxOP. It starts downcounting the backoff counter after detecting the medium being idle for a duration defined by the Arbitration Interframe Space (*AIFS*[*AC*]) instead of DIFS, which is used by legacy stations. The *AIFS*[*AC*] is at least DIFS, and can be enlarged per AC with the help of the Arbitration Interframe Space Number (*AIFSN*[*AC*]). The *AIFSN*[*AC*] defines the duration of AIFS[AC] according to:

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

AIFSN[*AC*] should be selected by the HC such that the earliest access time of EDCA stations is DIFS, equivalent to legacy 802.11. Basically, an AC uses AIFS[AC], CWmin[AC] and CWmax[AC] instead of the DCF parameters DIFS, CWmin, and CWmax for the contention process to transmit a frame. These parameters are chosen to allow higher priority traffic gain access to the medium quicker than lower priority traffic. The smaller the values of CWmin[AC], CWmax[AC], and AIFS[AC], the shorter the channel access delays, and consequently the higher capacity share for a given traffic condition. The CW range increases exponentially after each failed transmission attempt and reset after each successful transmission. The size of the CW_i *[AC]* in backoff stage *i* is:

$$CW_i\ [AC]=min[2^i(CWmin[AC]+1)-1, CWmax[AC]]$$

The contention window never exceeds the value of *CWmax*[*AC*]. This parameter is defined per AC as part of the EDCA parameter set. The smaller the *CWmax*[*AC*], the higher the medium access priority. However, a small *CWmax*[*AC*] may increase the collision probability. The EDCA parameters of each backoff entity are defined by the HC. Table 3 presents the default set of EDCA parameters:

Table 3 - Default EDCA parameter set

AC	CWmin	CWmax	AIFSN
AC_BK	aCWmin	aCWmax	7
AC_BE	aCWmin	aCWmax	3
AC_VI	(aCWmin+1)2-1	aCWmin	2
AC_VO	(aCWmin+1)/4-1	(aCWmin+1)/2-1	2

The HCF Controlled Channel Access is similar to IEEE 802.11 PCF, with the difference that there is no division between CFPs and CPs. A QoS Hybrid Coordinator (HC) controls the access to the medium when in HCCA mode. Depending on the operating mode of the network (infrastructure or ad-hoc), the HC is represented by the AP or by the station. In ad-hoc mode, the HC station has a higher priority access to the medium than other stations, allowing it 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. The transfer can be initiated only after a Priority Inter-Frame Spacing (PIFS) that will allow HC to gain the priority access to the medium. A more detailed description of the operating mode when in HCCA can be found in [12].

2.2 3GPP Technologies

In the last few years, we have witnessed an explosion of IP connectivity demand, yielding rapid development of the corresponding technologies in the wireless access network domain. IP services provision anytime and anywhere becomes very challenging and is seen by the mobile operators as a major opportunity for boosting the average revenue per unit. The further success of IP services deployment requires true mobile broadband IP connectivity on a global scale. For accomplishing this request, two technologies emerged with the aim of providing voice, data, video and multimedia services on mobile devices at high speeds and cheap rates: WiMAX (Worldwide Interoperability for Microwave Access) and LTE (3GPP Long Term Evolution). The economical aspects regarding these two technologies are comparable, considering that both are new and still under development [14], [15]. Important companies are developing new equipments, with improved performances that are starting to be available in the open market. In this thesis, the technology of choice is LTE, based on the industrial trend and on some advantages of this technology, detailed in [16].

2.2.1 Introduction

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 [17]. The use of SC-FDMA in the uplink reduces Peak-to-Average Power Ratio 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.

LTE architecture can be seen as a two-node architecture because only two nodes are involved 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 and the mobility management entity (MME) in the control plane, respectively [18]. Through this architecture, LTE offers smooth integration and handover to and from existing 3GPP and 3GPP2 networks, ensuring that operators can deploy LTE in a gradual manner using their existing legacy networks for service continuity [15]. 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 10 [19], [20].

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.

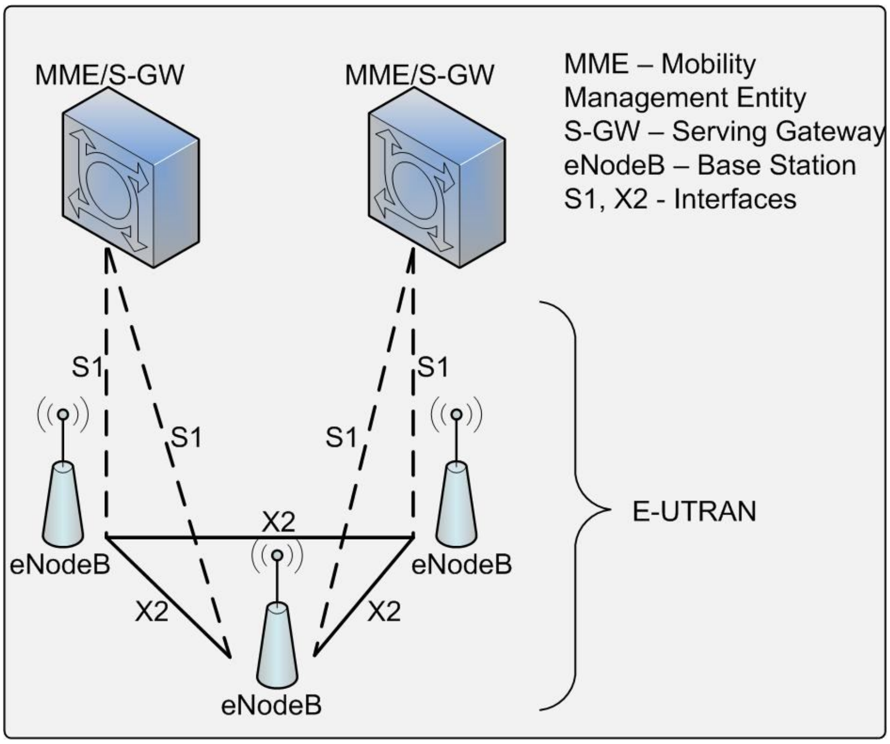


Figure 10 - LTE Overall Architecture

LTE protocol architecture depicted in Figure 11 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 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.

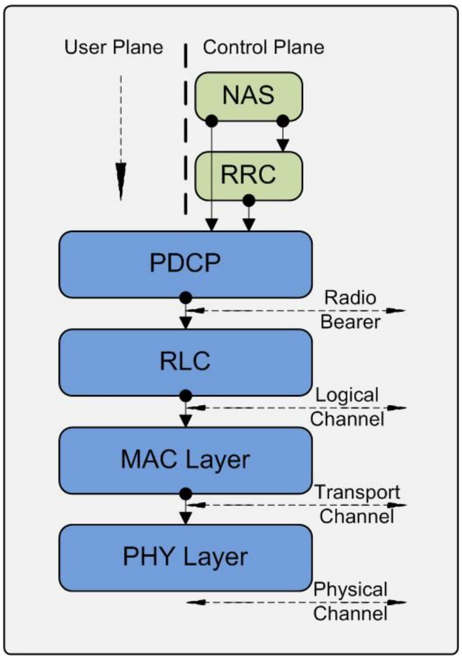


Figure 11 - LTE protocol architecture

The NAS layer performs authentication, security control and idle mode mobility and paging. At PDCP, the data is called PDCP SDUs. PDCP attaches header information with PDCP SDUs, 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 [21].

2.2.2 Physical Layer

The objectives of LTE physical layer are to significantly increase peak data rates up to 100Mb/s in downlink and 50 Mb/s in uplink within a 20 MHz spectrum - leading to a spectrum efficiency of 5Mb/s - , to increase cell edge bit rates, to reduce user and control plane latency to less than 10 ms and less than 100 ms respectively [22], to provide interactive real-time services such as high quality video/audio conferencing and multiplayer gaming, to support mobility for up to 350 km/h and to reduce the operation costs. It also provides a scalable bandwidth (1.25/2.5/5/10/20MHz) in order to allow the technology to coexist with other existing standards. The spectrum efficiency is improved 2 to 4 times compared to Release 6 HSPA in order to permit operators to accommodate an increased number of customers within their existing and future spectrum allocation, with a reduced cost of delivery per bit and an acceptable cost and power consumption. Figure 12 presents the components of PHY layer:

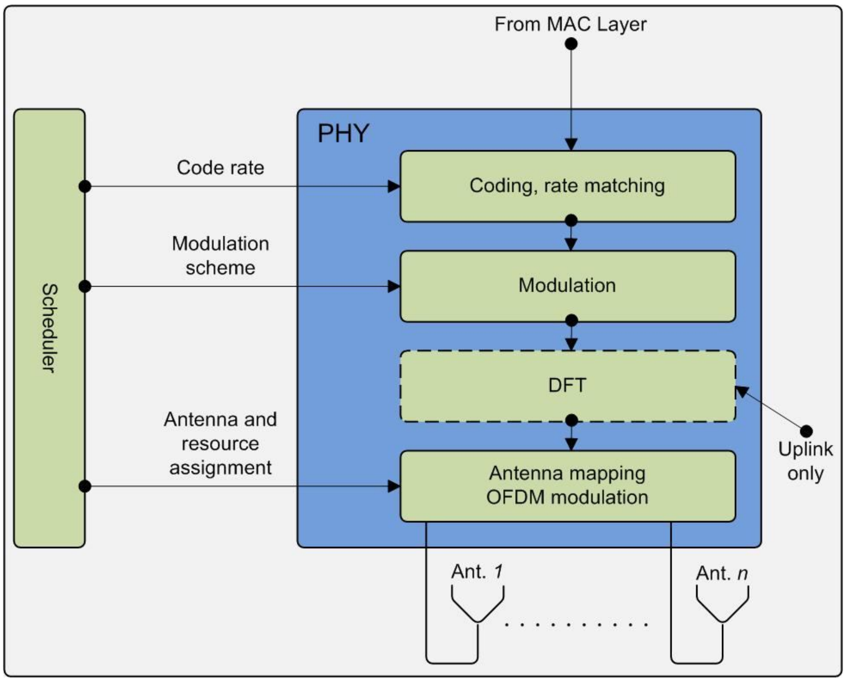


Figure 12 - Components of the LTE Physical layer

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.

OFDM in LTE

For LTE, OFDM splits the carrier frequency bandwidth into many small subcarriers spaced at 15 kHz, and then modulates each individual subcarrier using the QPSK, 16-QAM, or 64-QAM digital modulation formats. OFDMA assigns each user the bandwidth needed for their transmission. Unassigned subcarriers are off, thus reducing power consumption and interference. OFDMA uses OFDM; however, it is the scheduling and assignment of resources that makes OFDMA distinctive. The OFDM diagram in Figure 13 below shows that the entire bandwidth belongs to a single user for a period. In the OFDMA diagram, multiple users are sharing the bandwidth at each point in time:

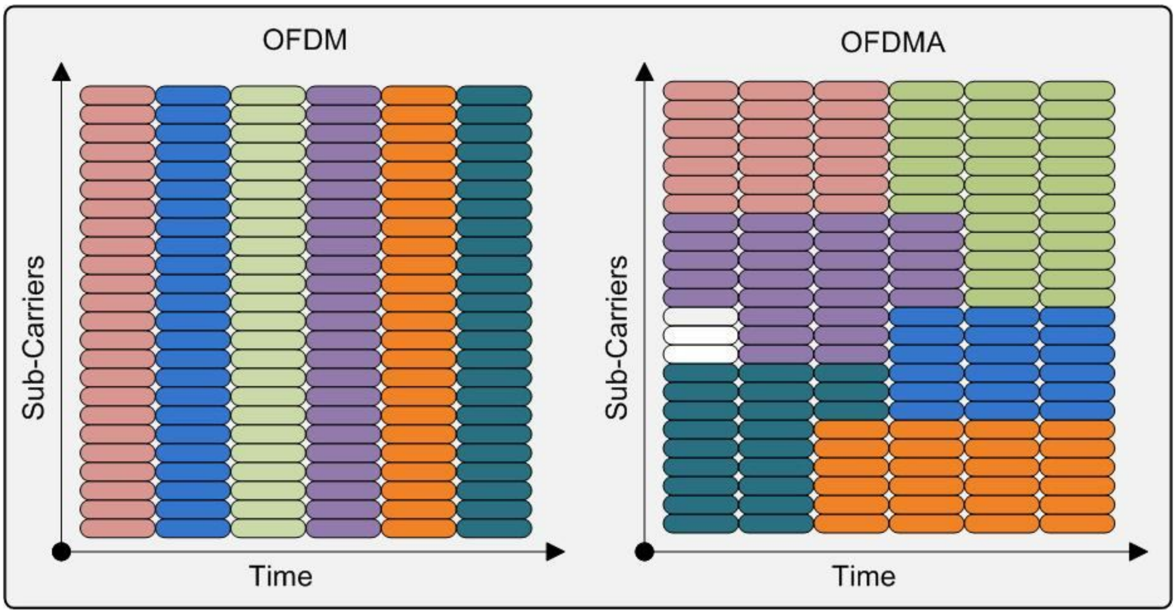


Figure 13 - OFDM vs. OFDMA

In the uplink, LTE uses a pre-coded version of OFDM called SC-FDMA. SC-FDMA has a lower PAPR (Peak-to-Average Power Ratio) than OFDM. This lower PAPR reduces battery power consumption, requires a simpler amplifier design and improves uplink coverage and cell-edge performance. In SCFDMA, data spreads across multiple subcarriers, unlike OFDMA where each subcarrier transports unique data. The need for a complex receiver makes SC-FDMA unacceptable for the downlink. Table 4 presents the OFDMA parameters used in LTE:

Table 4 – 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)

For the 5 MHz band, 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).

The OFDM symbol useful duration depends on the subcarrier bandwidth:

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

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 14:

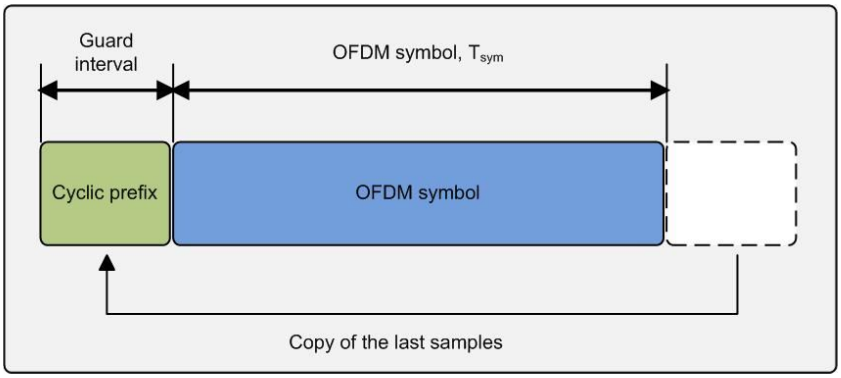


Figure 14 - OFDM symbol time structure

In LTE, 2 CPs are defined:

- Long CP: 16,67 µs
- Short CP: 4.69 µs

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

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

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 subframes of length $T_{subframe}$ = 1ms. Each subframe divides into 2 slots of 0.5ms long. Figure 15 depicts the Type 1 radio frame structure:

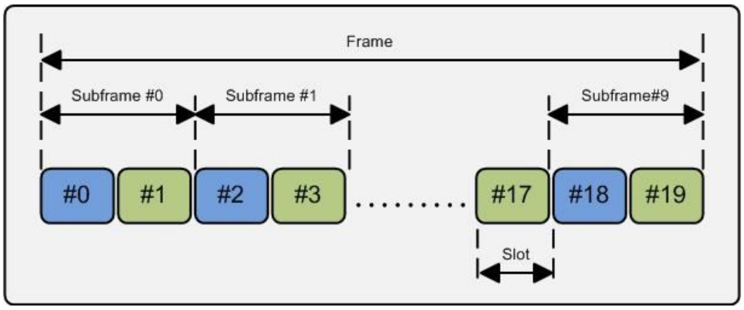


Figure 15 - Type 1 radio frame structure

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 16).

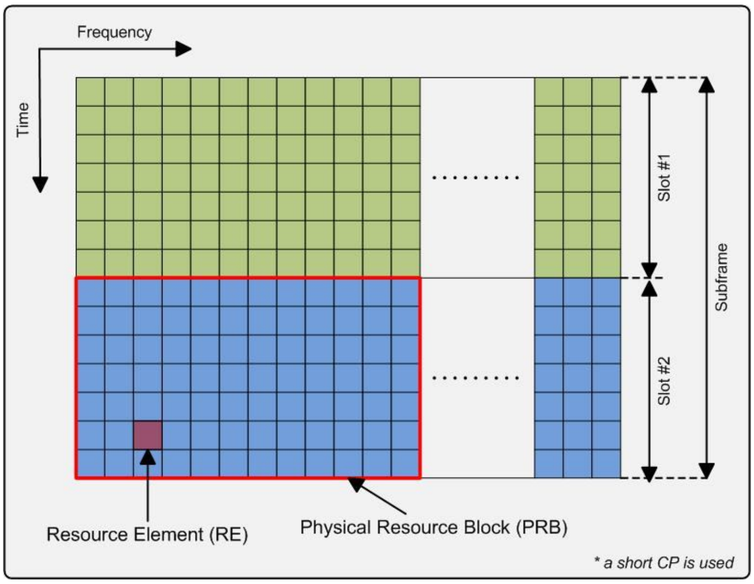


Figure 16 - Subframe structure

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 17).

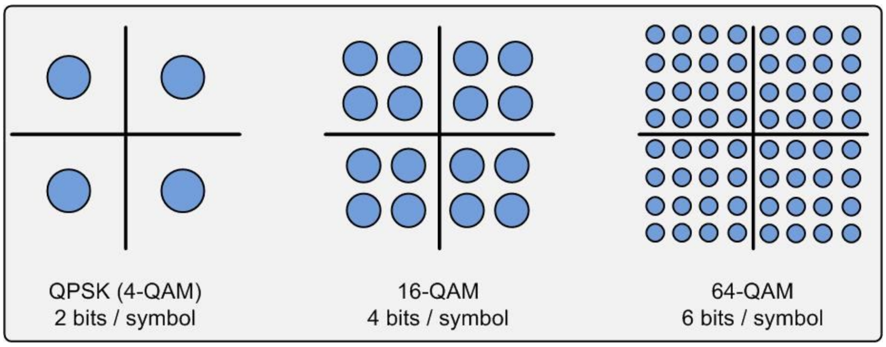


Figure 17 - Ideal constellations for the modulation types available in LTE

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.

Multiple antenna systems

Multiple antenna systems are being considered in all next generation cellular standards, including LTE and WiMAX, to increase capacity or to provide spatial diversity. The technologies being considered are MIMO, both Spatial Multiplexing and Space-Time/Space-Frequency Block Coding, and Beamforming. 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 Beamforming, Spatial Division Multiple Access (SDMA) and MIMO. The antenna configuration supported by LTE DL is (2x2) and (4x4) having 2 or 4 antennas at eNodeB and 2 or 4 antennas at UE. The UL of LTE supports 2x2 MIMO having 2 antennas at UE as well as at eNodeB. Table 5 describes the MIMO antenna configurations used by LTE.

Table 5 - 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

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

- o Physical Broadcast Channel (PBCH) – broadcast data from MAC
- o Physical Multicast Channel (PMCH) – multicast data from MAC
- o Physical Downlink Control Channel (PDCCH) – signaling for PDSCH and PUSCH
- o Physical Control Format Indicator Channel (PCFICH) – indicates the number of PFDM symbols used for control signaling in the current subframe
- o Physical Hybrid ARQ Indicator Channel (PHICH) – transmits acknowledgements for uplink data

➢ Physical signals

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

Table 6 - Modulation schemes for LTE UL and DL physical channels [18]

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

- o Physical Uplink Shared Channel (PUSCH) – user data from MAC
- o 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 6 summarizes the modulation schemes allowed for LTE uplink and downlink.

2.2.3 Media Access Control (MAC) Sublayer

The medium access control (MAC) sublayer [23] provides hybrid ARQ and is responsible for the functionality that is required for medium access, such as scheduling operation and random access.

Figure 18 presents the components of MAC layer:

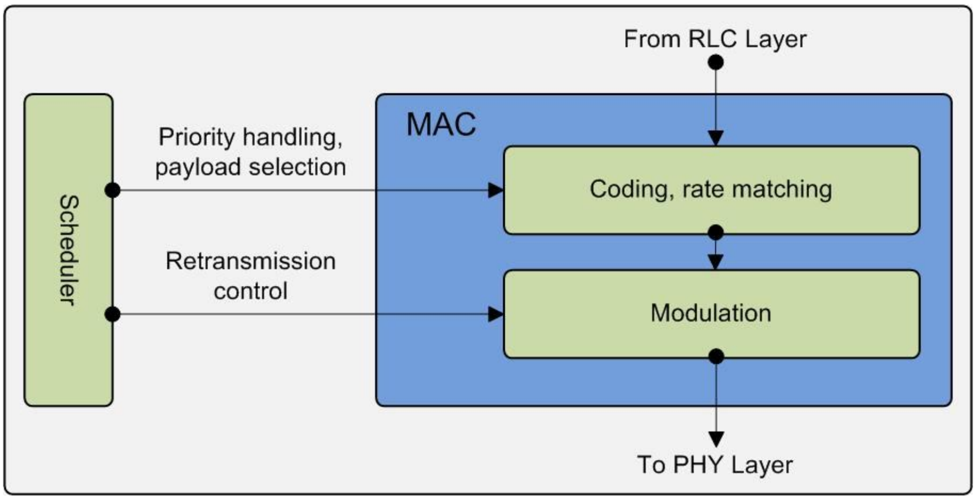


Figure 18 - Components of the LTE MAC layer

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. It will perform a retransmission of the corrupted data block if the service has a high sensitivity to packet loss or it will drop the erroneous blocks if the service QoS permits it.

Hybrid ARQ

MAC sublayer 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 [24]. 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 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, the combining process uses the two received transport block affected by errors to increase the probability of successful decoding. Figure 19 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.

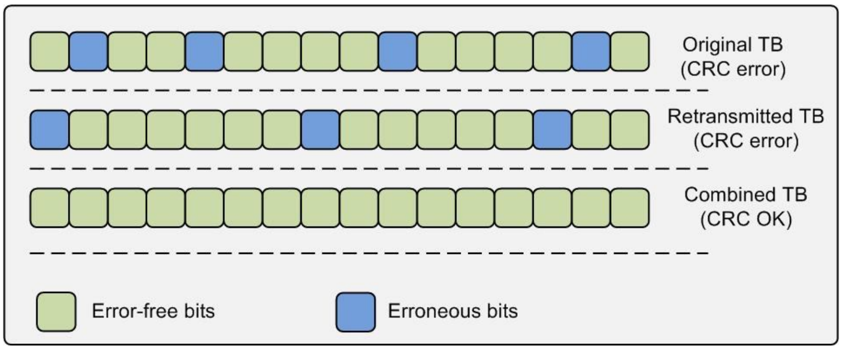


Figure 19 - Chase combining HARQ

MAC Scheduler

The scheduler, located in the eNodeB, determines dynamically each 1 ms (each Subframe), which UEs are scheduled to transmit/receive data on the Uplink/Downlink shared channel and also what resources should be used. If we consider 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.

Figure 20 and Figure 21 are describing the reporting mechanisms used in uplink and downlink.

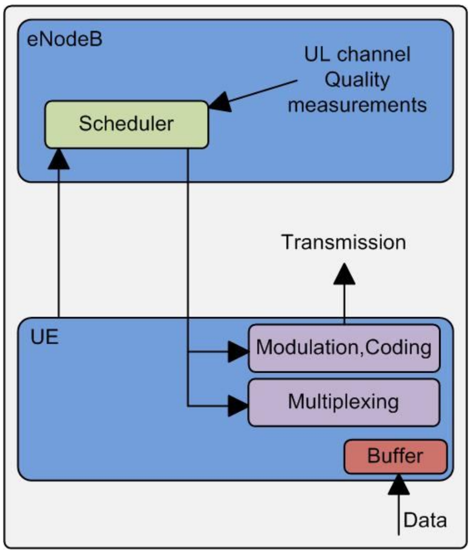


Figure 20 - UL reporting mechanism

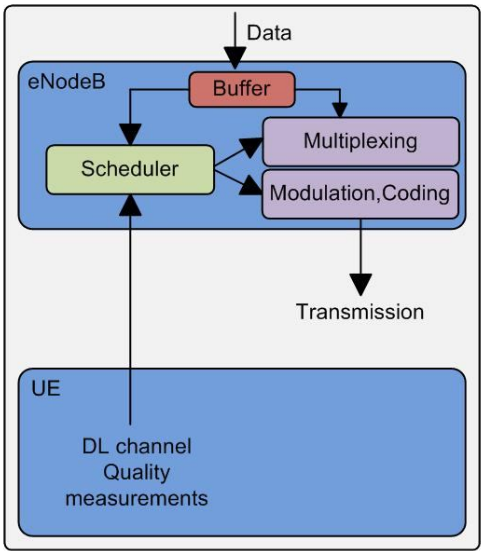


Figure 21 - DL reporting mechanism

In order to achieve a higher system performance in LTE, scheduling must be integrated with link adaptation and HARQ processes.

2.3 Network and Transport Layer Protocols

2.3.1 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 not a connectionless and unreliable protocol that delivers the data packets using the best-effort strategy. IP protocol is offering 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.

The IP packet, like any other packet sent over the network, has a header and a payload. The header (Figure 22) contains:

- 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)

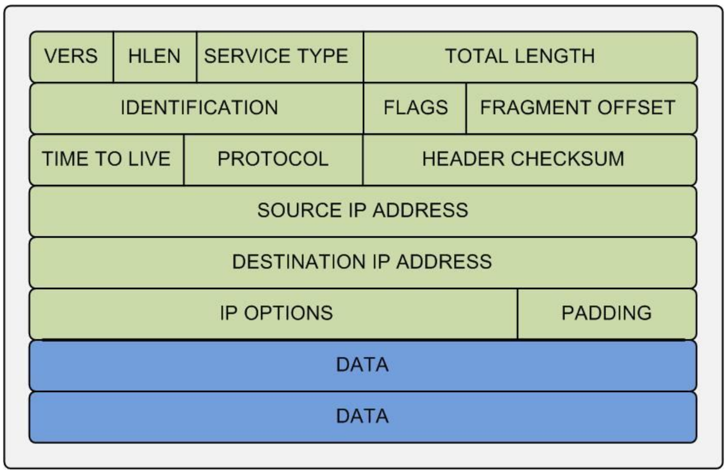


Figure 22 - IP header structure

The IP dimension, opposite to the physical networks frames, is not limited by any hardware characteristic. For IP version 4 (IPv4) [25], the maximum length of al packet can be 2^{16}. An IP packet that is delivered over the network from one station to another has to be encapsulated into a network frame (Figure 23). 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, then inefficient data transfers will happen over the networks that can deliver larger frames.

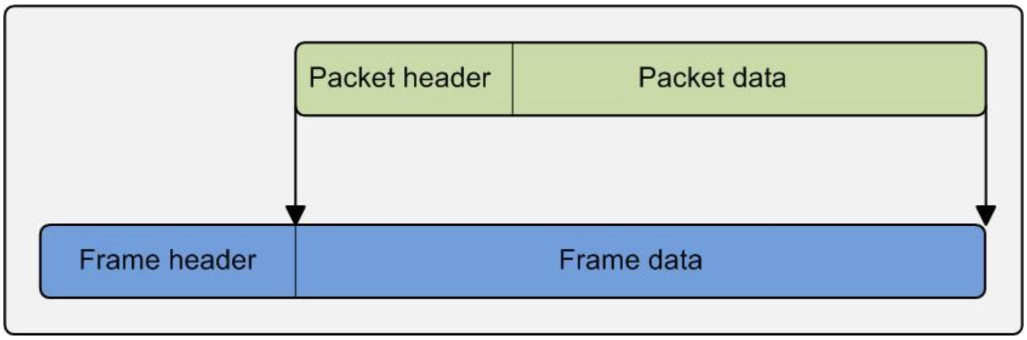


Figure 23 - IP packet encapsulation

There are currently two IP versions: Internet Protocol version 4 (IPv4) and Internet Protocol version 6 (IPv6) [26]. As compared to IPv4, 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).

2.3.2 UDP (User Datagram Protocol)

User Datagram Protocol (UDP) [27] is one of the main transport protocols in the 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 does not provide any 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 24 presents the UDP packet format:

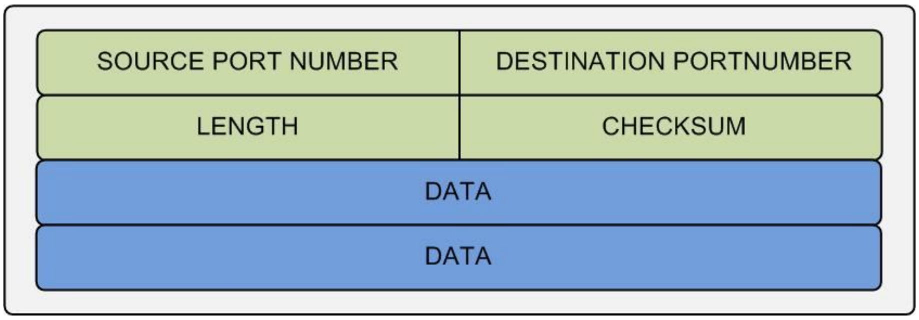


Figure 24 - UDP packet structure

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).

2.3.3 TCP (Transport Control Protocol)

The Transmission Control Protocol [28], [29] 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

o Sequence number - field used to identify the correct sequence number (initial sequence number if SYN flag is set, accumulated sequence otherwise)

o Acknowledgement number – field used to identify the next sequence number

o Data offset – field used to specify the TCP header size (minimum 20 bytes and maximum 60 bytes)

o Reserved – field that should be set to zero

o Flags – field used to set all 8 flags from TCP header: CWR, ECE, URG, ACK, PSH, RST, SYN and FIN

o Window – field used to specify the size of the receive window

o Checksum – field used to check the TCP segment for transmission errors

o Urgent pointer – field used to indicate the last urgent data byte, if the URG flag is set

o 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 25:

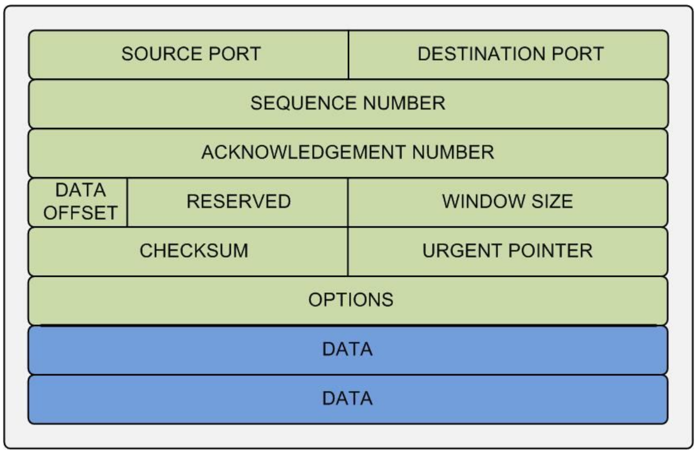


Figure 25 - TCP segment structure

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 [30], link layer solutions [31] or proxy based solutions [32].

2.3.4 RTP (Real-Time Transfer Protocol)/RTCP (Real-Time transport Control Protocol)

Real-time Transport Protocol [33] it was developed by the Audio-Video Transport Working Group of the IETF (Internet Engineering Task Forceand 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 26:

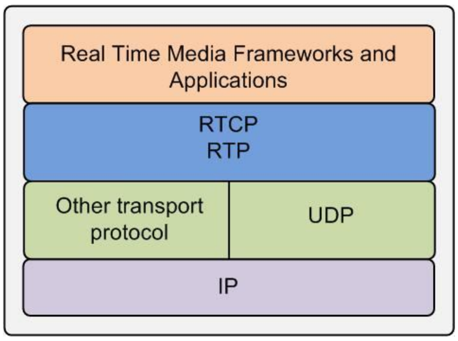


Figure 26 - RTP architecture

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 27 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

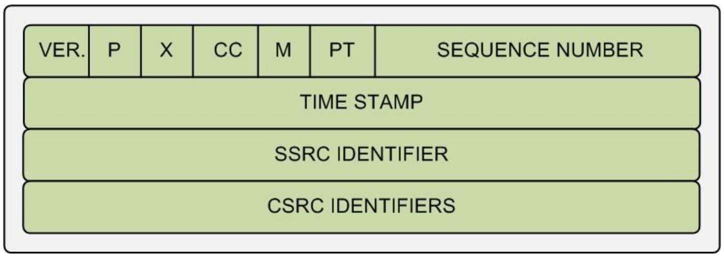


Figure 27 - RTC header

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 [34]. 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 [34]. 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

2.3.5 ECN (Explicit Congestion Notification) for video traffic in LTE

The Explicit Congestion Notification (ECN) mechanism was first proposed in 1999 [35] as an experimental protocol for TCP. From 2001, ECN was defined as an extension to the IP and to TCP in RFC 3168 [29]. 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 28). Depending on the value of the two bits, we have four different codepoints:

- 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

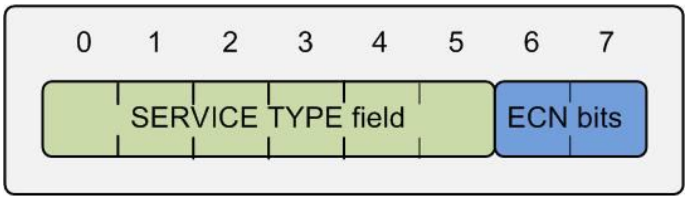


Figure 28 - SERVICE TYPE field from IP header with the 2 less significant bits used for ECN

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 [36] 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.

2.4 Related work

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 improve the quality of perception while keeping the same QoS level etc. There are solutions for every layer in the OSI or TCP/IP model and depending on their approach they can be categorized as end-to-end or network centric. The most important network and multimedia parameters are taken into consideration and different algorithms are optimized for specific networks with specific conditions.

2.4.1 Multimedia characteristics

Adaptation as a counter measure for congestion has been an interesting research topic since the 80's [37]. 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. Besides QoS, another important factor is the perceived quality, as expressed by the user. Quality of Experience or QoE can be seen as the overall result of the QoS on the multimedia stream, as perceived by the user.

Throughput

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 requesting the network resources, the protocol limitations etc.

Loss

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.

Delay / Jitter

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.

2.4.2 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 [38], [39]. One approach used to improve the multimedia streaming performances is the **end-to-end approach** [30] 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 29 is illustrating the end-to-end approach:

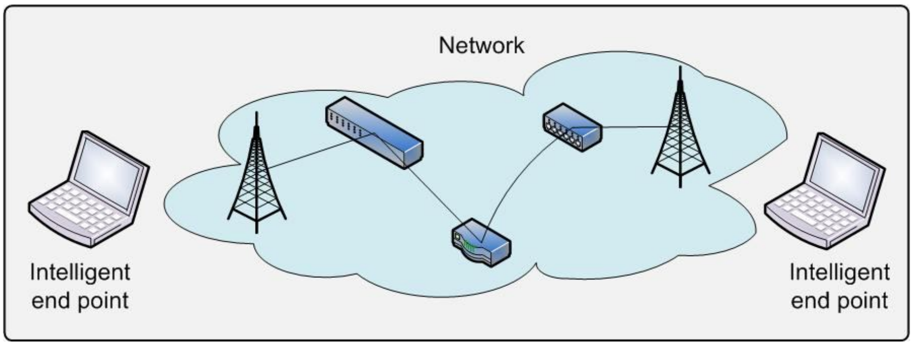


Figure 29 - End-to-end approach

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 [40] 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) [41] 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 [42]. 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) [40]. 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 [43] and it was proven that the performances obtained are lower than those of the TCP protocol. RAP performances can be improved if queuing routers [44] 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 [45], 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:

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

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 [46]. **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 [47] 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 [48]. 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 [49]. **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.

Video Transport Protocol (VTP) [50] 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 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) [9]. 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 [51] 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 30.

The simulation results are showing that compared to TFRCP, LDA+ and a nonadaptive 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 QoD_{score} 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 [52] 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.

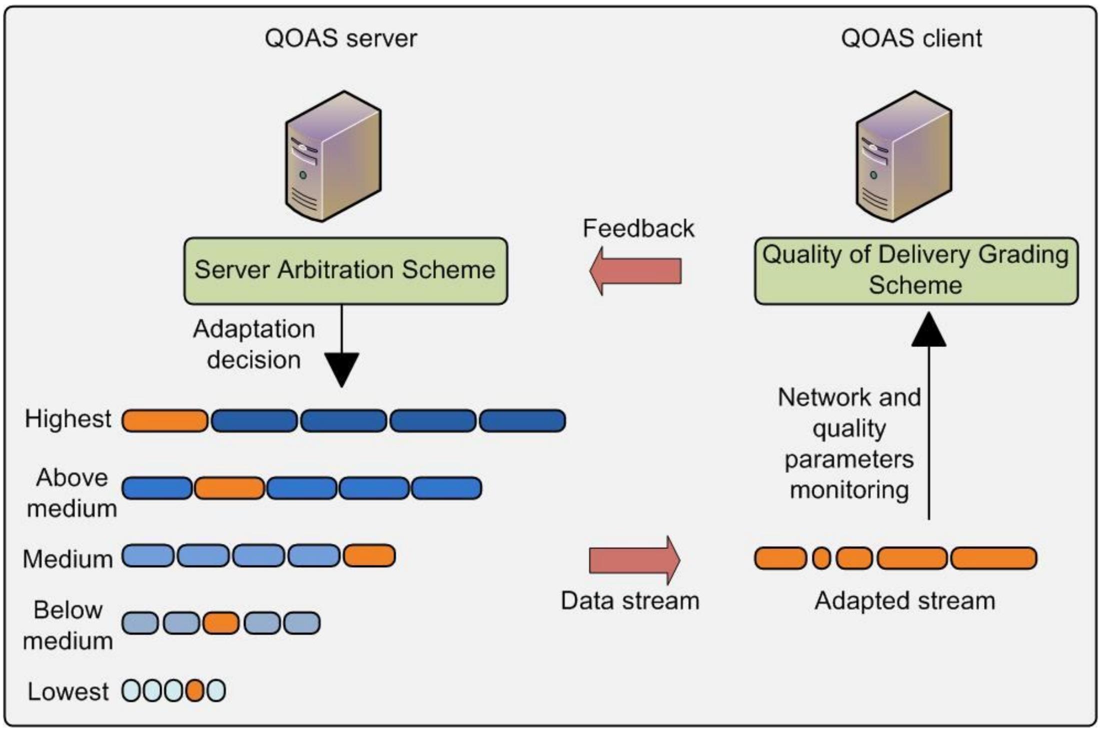


Figure 30 - QOAS adaptation principle

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 31 illustrates the network centric approach.

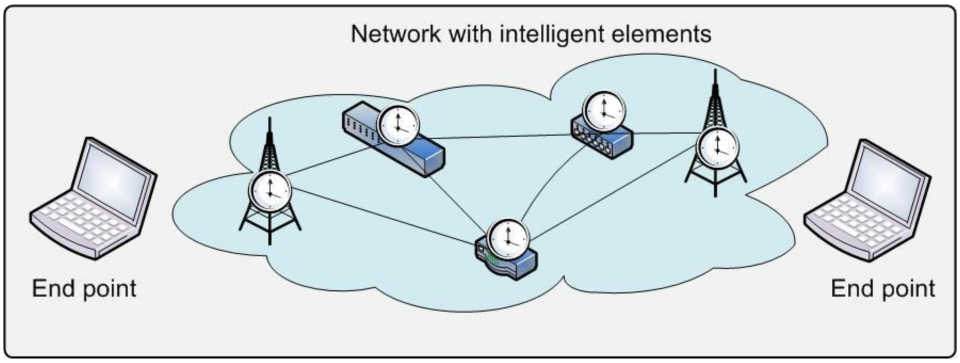


Figure 31 - Network-centric approach

One of the first architecture models used to achieve a certain QoS level was proposed by the IETF. The **Integrated Services** (IntServ) [53] model uses the Resource Reservation Protocol (RSVP) [54] for a per-flow signaling in order to provide the required QoS. For this model, routers in the network require major modifications to make possible the resource reservation for each data flow that needs QoS guarantees. The drawbacks of this model are the high complexity and the scalability problems that may appear when new elements are added to the network.

To overcome the issues introduced by the IntServ model, IETF proposed another model that uses a simple scalable mechanism to provide the required QoS guarantees for the data flows that are requiring a certain QoS level. The **Differentiated Services** (DiffServ) [55] working base is built from a group of routers that are set by an administrator to form the administrative domain that uses a predefined list of forwarding rules [56]. In order to signal that a data packet needs preferential treatment, the SERVICE TYPE field from the IP header is used.

2.4.3 Resource scheduling techniques 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 multimedia transmission is presented in [57] while a detailed description of available scheduling algorithms for LTE downlink can be found in [58].

The Distributed Weighted Fair Queue (DWFQ) [59] 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.

[60] 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 [61] as an extension to the Self-Clocked Fair Queuing [62]. 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 [63]. 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.

For LTE, reference [58] divides the scheduling algorithms into two categories, based on the type of traffic the scheduler was designed for: scheduling for elastic (non-real-time flows) [64] and scheduling for real-time flows [65]. LTE schedulers can also be classified based on their awareness parameter(s) into channel-aware schedulers [66], queue-aware schedulers and queue- and channel-aware schedulers [58]. 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.

3. DQOAS Design

This chapter describes in detail the proposed Dynamic Quality-Oriented Adaptive Scheme (DQOAS) multimedia streaming algorithm. The chapter begins with an introduction where the necessity of the proposed algorithm is outlined. It continues with the description of the architectural framework of the e-learning system where the algorithm was first introduced with good results, describing the main components of the adaptation mechanism. The last part of this chapter presents the DQOAS principle of functioning and its expected benefits.

3.1 Introduction

Before designing a competitive adaptation mechanism for congestion avoidance in wireless environments, capable to offer QoS guarantees and a satisfactory end-user perceived quality for the multimedia content being delivered, it is mandatory to consider both network and user-related problems. A first analysis of the most important network characteristics that affects the multimedia streaming process is done in chapter 2, where parameters like delay, throughput, packet loss rate or jitter were detailed, together with some existing solutions. The problem appears when we have to analyze the user-related parameters and the way they impact the streaming mechanism. The difficulty comes from the fact that there is no set of rules or parameters able to define the behavior or the requirements of an end-user or of a group of end-users; they rather have dynamic characteristics that are hard to quantify because they are continuously changing.

Generally, an adaptive behavior will allow the sender to manipulate its sending rate depending on the availability of network resources. One definition for adaptation is given in [67]:

> *"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 is it an adaptive process complete if only the network-related problems are addressed? Our opinion expressed in this thesis is that one adaptive process has to take into account also the end-user preferences and requirements, that can be translated into end-user satisfaction. DQOAS is a network and user-aware adaptation mechanism that was designed to improve the end-user quality of experience during a multimedia streaming session. 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, there are three options: server side adaptation, client side adaptation and in-network adaptation. Our solution uses the first option, server side adaptation. The server is adapting with the network conditions and user preferences by changing the transmission rate of the multimedia application according to the feedback information received from the client. This option was chosen because it is suitable for heterogeneous environments where the client can have a different access technique to the medium as compared to the sender and can experience a wide range of congestion scenarios. Also, compared with the other two options, the server side adaptation method offers more scalability.

The second question offers four solutions about how the adaptation should be performed:

- Adaptive Increase and Multiplicative Decrease (AIMD) [68], where the sender, based on the feedback messages, will either decrease the sending rate with a constant multiplication value or increase is with a constant additive value:

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

  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. The equation for AIPD is:

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

where f_i is the loss fraction.

o Adaptive Increase and Adaptive Decrease (AIAD), where the sending rate decreases and increases in a non-linear form. The way transmission rate is modified depends of 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.

o 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 section 2.4.2. This approach is not suitable for environments where the error rates are high and are considered to be media unfriendly.

For DQOAS, the selected approach is the AIAD because it gives the opportunity to consider both network and user-related parameters in the adaptation process.

To answer the third question, two solutions should be analyzed. First, the adaptation process can take place before any congestion occurs, adapting the transmission rate with the network conditions. The second solution is to start the adaptation process only after congestion is observed. The first approach is well suited for delay and packet loss sensitive applications but it has a high complexity in implementation. This is the reason that DQOAS uses the second method, based on constant probing the network channels and user parameters. In order to overcome the delayed responses to network congestion that can degrade the QoS, DQOAS uses a “fast decrease, slow increase” mechanism, ensuring a fast reaction during bad network conditions and allowing them to improve and stabilize before a quality increase.

3.2 Architecture

The proposed Dynamic Quality-Oriented Adaptation Scheme (DQOAS) is a user-oriented adaptation mechanism that dynamically adapts the multimedia content sent based on user preferences and network conditions. By employing this algorithm when an e-learning application is used, an increased number of simultaneous learners can be accommodated while maintaining the perceived quality levels above their individual "good" quality thresholds. This is done by dynamically modifying the adaptation policy during delivery, increasing or decreasing the granularity of the adaptation for some users, based on their preferences and actual network conditions.

DQOAS performs an optimal dynamic multimedia content management according to end-user profile, knowledge level, goals, preferences as well as network conditions in order to support higher end-user quality of experience during the learning process. The e-learning system used in this thesis is named Performance-Aware Multimedia-based Adaptive Hypermedia (PAMAH). This 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.

3.2.1 Architectural framework for PAMAH

Figure 32 presents the architectural framework for the Performance-Aware Multimedia-based Adaptive Hypermedia (PAMAH), which enhances 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 personalised educational content to e-learners via heterogeneous networks. Its goal is to optimise users' QoE and their learning outcomes by automaticaly adapting content and navigational support based on both user interests and knowledge levels and current network delivery conditions.

The delivery of multimedia content presents a significant challenge when it is combined with personalization. In isolation, multimedia content delivery may be thought of as attempting to deliver the highest quality information given 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 33).

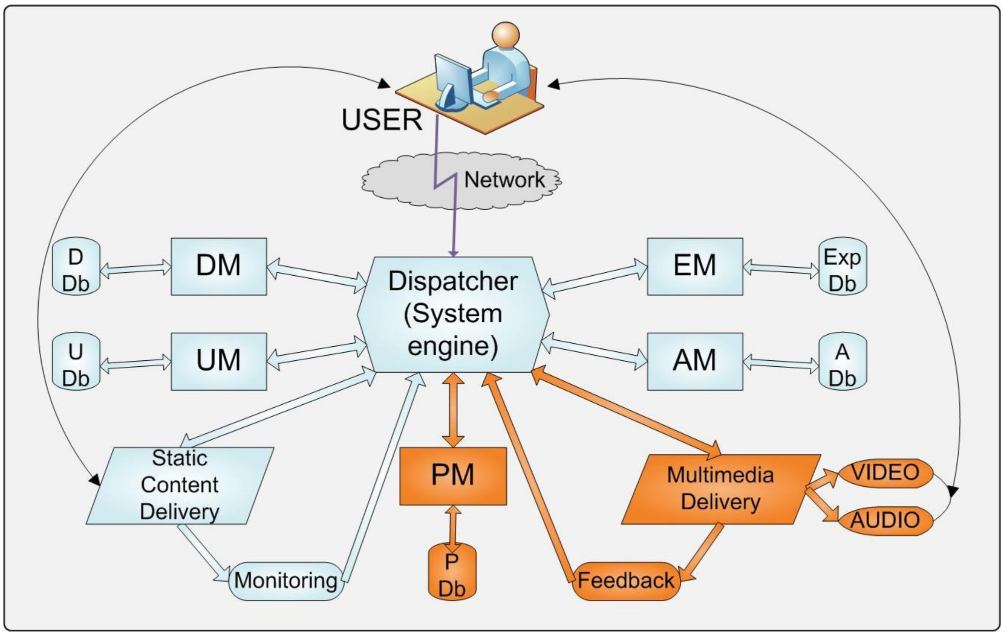


Figure 32 -PAMAH system – block-level architecture

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.

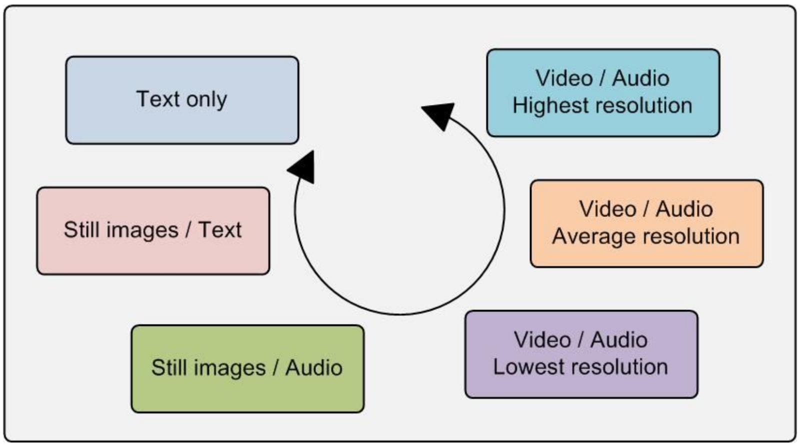


Figure 33 - Content adaptation in PAMAH

It is possible 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.

3.2.2 Models used in PAMAH architecture

The architectural framework for the PAMAH adaptive web-based system maintains five models: Domain Model (DM), User Model (UM), Experience Model (EM), Adaptation Model (AM) and Performance Model (PM). Every 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).

All these models are controlled 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 by the delivery mechanisms, both static content delivery and multimedia delivery. Except the parameters and rules offered to Multimedia Delivery module by the Dispatcher, there are other network-related parameters that are considered when rich media-content fragments are transmitted. 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.

3.2.3 Rich media content delivery blocks

From the block-level architecture of the PAMAH system presented in Figure 32, this thesis focuses on the highlighted parts, which are in charge with the rich-media content delivery. Multimedia Delivery block implements the DQOAS adaptation algorithm considering the input from the Dispatcher. The Feedback Module creates a report that summarizes the delivery conditions and the user feedback for the delivered stream. This report is sent 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. Taking into account the reports from the rest of the modules, the dispatcher builds a list of rules that will be sent to Multimedia Delivery block, as a new 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 to improve end-user quality of experience by adjusting dynamically its adaptation policy.

3.3 DQOAS Algorithm

3.3.1 Overview

DQOAS is developed to increase end-user quality of experience while also enabling higher number of simultaneously connected clients to communicate. In the system architecture presented in Figure 32, the Adaptation Model performs real-time bandwidth estimation for every user involved in the e-learning process. Based on this bandwidth estimation and also on the additional information from other models, the System Engine decides if a video stream is suited or not for a certain user. If the proper conditions for a streaming session are satisfied then a specific list of rules is built by the Dispatcher and forwarded to the N-Level Builder Module. Here, considering user requirements, the user-specific streaming parameters are computed and the session starts.

3.3.2 DQOAS principle

DQOAS enhances QOAS [9] by adding different user QoE expectation levels in the adaptation process. Consequently, DQOAS will perform a differentiate treatment of the users, adaptive process based not only on network conditions, but also on their requirements in terms of their QoE. Unlike QOAS, which uses static potential bitrate adaptive levels, DQOAS dynamically adapts them to suit the delivery process.

Figure 34 illustrates the major blocks of the DQOAS algorithm for improving user quality of experience during mobile and wireless e-learning, deployed on the PAMAH architectural structure described in section III.

On the client side there is a Decoding & Playing Module whose role is to decode and afterwards play the adapted video stream received from the server. Connected to this module is a dedicated module that performs estimation about the end-user perceived quality. The Feedback Module is monitoring delivery-interest parameters such as loss rate, delay and jitter and also assesses the quality of delivery, sending short reports to the server. There reports contain network-related information and user-specific parameters used on the server side by the Dispatcher and by the N-Level Builder Module. Based on this report the Dispatcher will update the new list of rules and parameters specific for every user and will also update the relevant system databases (A Db and U Db). The list of rules and parameters consists of user-related information (the QoE expectation levels for every user) and session-specific information (which users will be considered for a streaming session, video file to be selected for the current session, video metadata, etc).

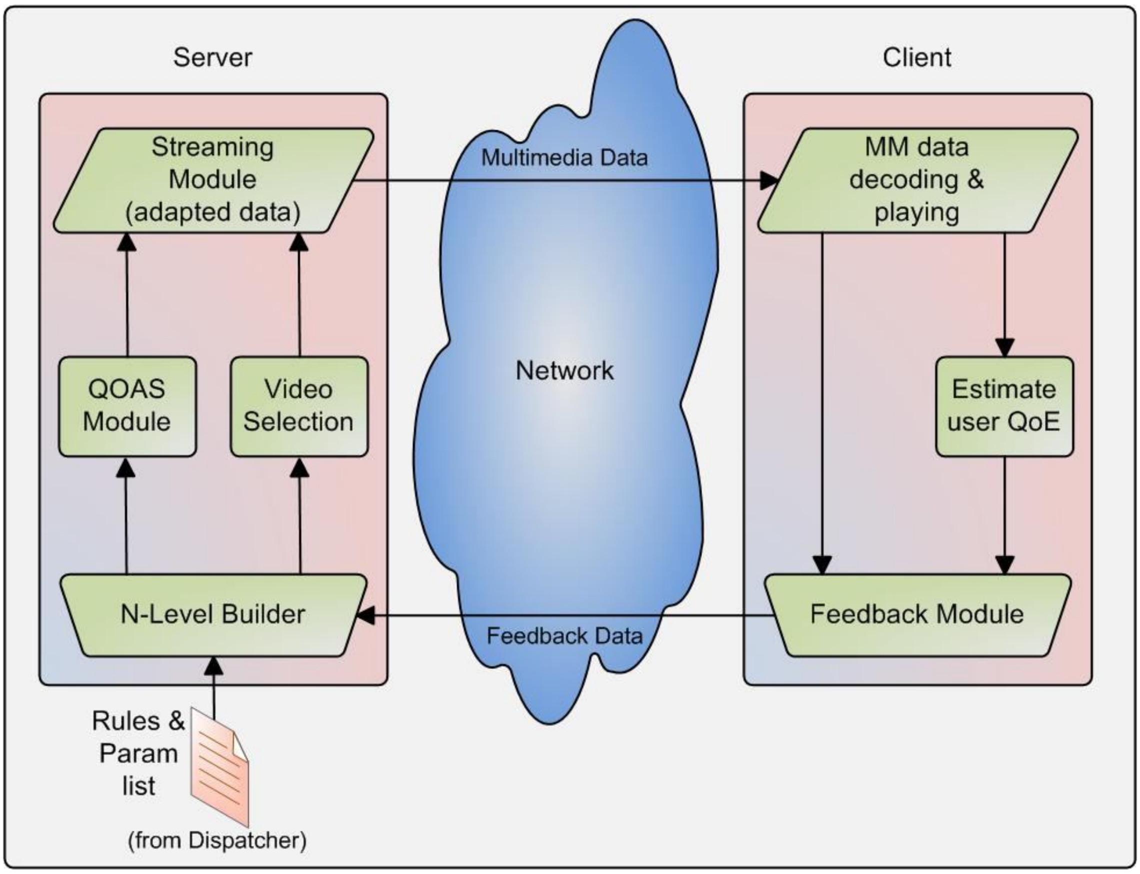


Figure 34 - DQOAS adaptation algorithm

On the server side, three important phases are described:

1. Initial Level Building

The initial level building is invoked when a new user in the system is requesting a multimedia stream. Because for this new user the Estimation Module has not established yet the QoE expectation level, the level building is done statically directly by the QOAS Module. This new user will appear in the list of rules and parameters with his QoE expectations level set to 0. When N-Level

Builder Module is analyzing the list and discovers the new user, it will forward a level request to QOAS Module for building the static levels just for this user.

For a new user ***Uk*** the QOAS algorithm builds M potential levels of quality: ***L1 L2 … Lj … LM***. ***L1*** represents a low bitrate while ***LM*** represents the maximum bitrate (the video bitrate).

During the session, the system estimates the QoE expectation level for this user and the quality levels will be updated dynamically. After QoE estimation, one of the inputs in the rules list corresponding to this user indicates a video-quality level which represents the acceptable quality expectation level associated to this user. Considering this level, DQOAS algorithm is building another set of levels, this time having new limits, other than the static levels: between the user QoE expected level and the maximum video level, ***LM***. Like this, a user oriented dynamic adaptation is performed. The procedure for this is presented below.

2. Updating Dynamically the Quality Levels

The procedure for dynamically update the levels for a user involved in a streaming session can be triggered by different causes: changes in the network parameters, users detaching/attaching to the network, updates of QoE expectation level for certain users, etc. The decision for dynamic update of levels is taken by the N-Level Builder based on the information received from the Dispatcher (the list of rules and parameters) and from the Feedback Module. The user QoE expected level found in the rules and parameters is unique for every user and represents the minimum video bitrate that will be transmitted for that user.

For user ***Uk***, after the QoE estimation is performed on the client side, the expectation level is decided to be ***Mk***. This level is situated in the interval defined by the minimum and maximum static levels, ***L1*** and ***LM***. Once ***Mk*** is known, the N-Level Builder Module proceeds and rebuilds the adaptation levels for user ***Uk***. Every time the user QoE expectation level is changing, the N-Level Builder will initiate the procedure for dynamically update the levels.

While a new user has a fixed number of levels, M, 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 number of users. By using the user-specific levels of quality the algorithm is able to control better the video quality adjustment. This in performed in order to increase end-user perceived quality since it was demonstrated that viewers tend to prefer a controlled reduction in the quality of the streamed multimedia content than random losses.

3. Adaptation Mechanism

Considering P users involved in the e-learning session: ***U1 U2 U3 … Uk … UP-1 UP***, for every user there is a QoE expectation level specifying the expected quality by that user: ***M1 M2 M3 … Mk … MP-1 MP***.

For any user ***Uk*** the algorithm builds the dynamic levels by dividing the amount of bandwidth between ***Mk*** and ***LM*** into intervals. The division depends on the network QoS parameters which are considered in real-time. After this step, the QOAS Module is tuned on these new adaptation levels of quality obtained for user ***Uk***. The Streaming Module is responsible for streaming the selected video file according to the user-specific levels. Video selection is performed by the Video Selection Module, based on the information presented in the list of rules and parameters.

Rate control adaptation process is applied by the N-Level Builder Module whenever media-rich content fragments transmitted needs to adapt their bit rate to match the network parameters and user requirements. Delivery of these fragments usually requires significant network resources and lasts over a longer period of time in which delivery conditions can vary. In this case, random losses have a greater impact on the end-user perceived quality than a controlled reduction in quality.

In conclusion DQOAS dynamically varies the quantity of information transmitted 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 any improvements in network conditions are detected, the N-Level Builder Model will increase step-by-step the bitrate, according to the user-specific levels, improving therefore user-perceived quality. On the other hand, if at a certain point the network conditions are degrading (increase in the background traffic, new e-learning users) then DQOAS will rebuild the dynamic levels for every user. The new levels will have a smaller granularity and as a consequence the maximum level considered might decrease.

3.3.3 Proposed QoS parameters mapping scheme for optimizing DQOAS in LTE networks

E-learning applications is usually generating more than one traffic type, so the data sent over the network to any user can be seen as a traffic mix composed mostly from video and web browsing traffic. Based on this fact and on the service classification in LTE by the SPI (Service Priority Information) field, there will be at least two queues corresponding to two different QoS classes, deserving the application. Considering service prioritization, the objects belonging to different queues have different probabilities of being scheduled, depending on the chosen scheduler model. The most common options for a scheduler are: Round Robin scheduler (RR), Proportional Fair scheduler (PF), and Maximum Throughput scheduler (MT).

Considering an application generating traffic belonging to two different service classes, this traffic is classified in different QoS queues and a RR scheduler is employed, we can write the equation describing the *i*-th user satisfaction, according to [69]:

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

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 $\beta = s_i^{max}/s_i$.

In our case, the maximum ratio of delayed and lost packets, ε, is different from one user to another, being a user-dependent parameter, not a service dependent parameter like it is in case of VoIP. Based on DQOAS description, every user has a minimum accepted quality level for the incoming video stream, representing the dynamically encoded bitrate of the stream, M_i. As this is the minimum accepted level that DQOAS can send for user *i*, ε as defined above has no significance because any lost or delayed packets will decrease the quality below M_i.

In order to overcome this problem, two solutions are possible. First solution is to add a guard to the minimum expected level, equal with the maximum agreed ratio of delayed and loss packets, ε: $M_i^{new} = M_i + \varepsilon$. Second solution implies $\varepsilon = 0$ only when DQOAS module is tuned on M_i. If the quality level is higher, decreasing this level with the maximum ratio of delayed and loss packets ε, keeps the video quality above the user satisfaction limit. The first proposed solution is easier to compute, because the changes are done inside the DQOAS module, based on the current network conditions and resource allocation scheme used.

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 QoS class (ρ=1). Considering that f_1 and f_2 represent the average packet transmission ratio of two flows generated by the same application, the satisfaction equation for user *i* reads:

$$\frac{(f_1 + \alpha f_2) \cdot T \cdot \left\lceil \frac{N}{n} \right\rceil \cdot \Delta}{(T + d^{max}) \cdot \beta} \leq \frac{1}{1 - \varepsilon} \qquad (M_i^{new} = M_i(1 + \varepsilon))$$

The case considered here, an e-learning application generates video streaming and web-browsing traffic. Following the assumptions, the second traffic type (with a lower priority) will have the same QoS class as video traffic – streaming class, as presented in Figure 35.

The advantage of being in the same QoS class is that the queue specific sorting algorithm will consider both flows with the same priority (the users' priority in the queue). In this way, the packets coming from the same application, even on different bearers, will have the same queuing delay, improving the QoE of the application as a whole.

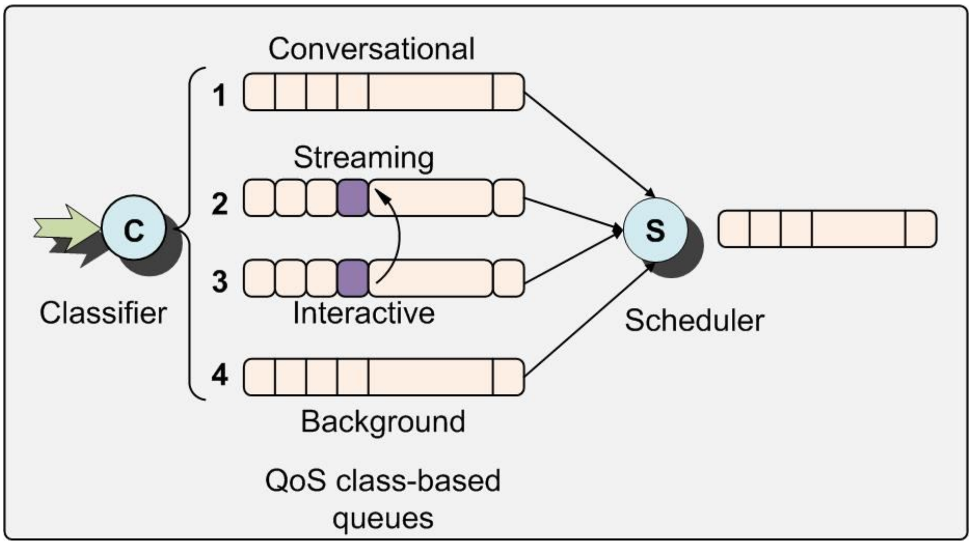


Figure 35 - Proposed scheduling framework for QoS class change of low priority e-learning flow

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 (web browsing traffic in our case).

4. Simulation and results

In this chapter, simulation results for this thesis are presented and discussed. The chapter starts with a brief description of the simulation environments used, including the simulation settings. Results of different simulations for every experiment performed in this study are compared with each other and discussed in separate sections.

4.1 Simulation Models

4.1.1 Network model for IEEE 802.11 wireless LANs and the experimental setup

The proposed algorithm was tested in IEEE 802.11 wireless environment using Network Simulator (NS) with NOAH (No Ad-Hoc) patch installed [70]. 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 Defence 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 tasks 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 36:

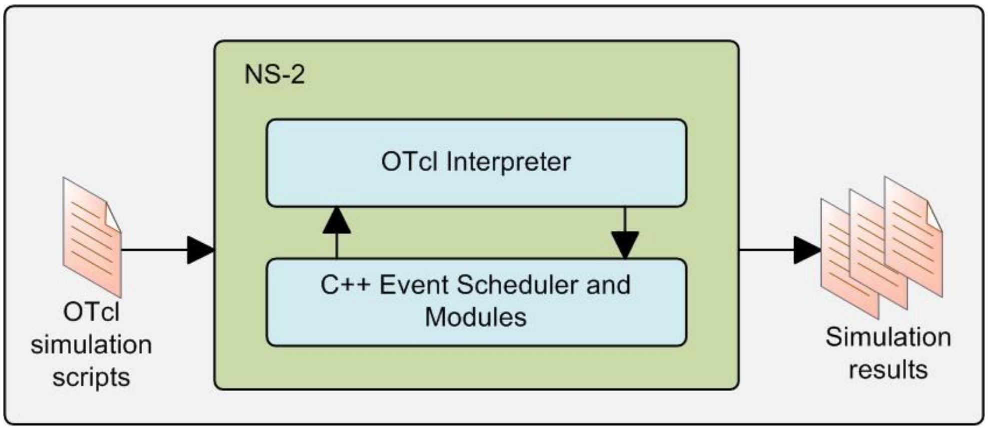


Figure 36 - NS-2 architecture

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.

The general network architecture used in all simulation scenarios is presented in Figure 37. 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 7:

Table 7 - 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

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.

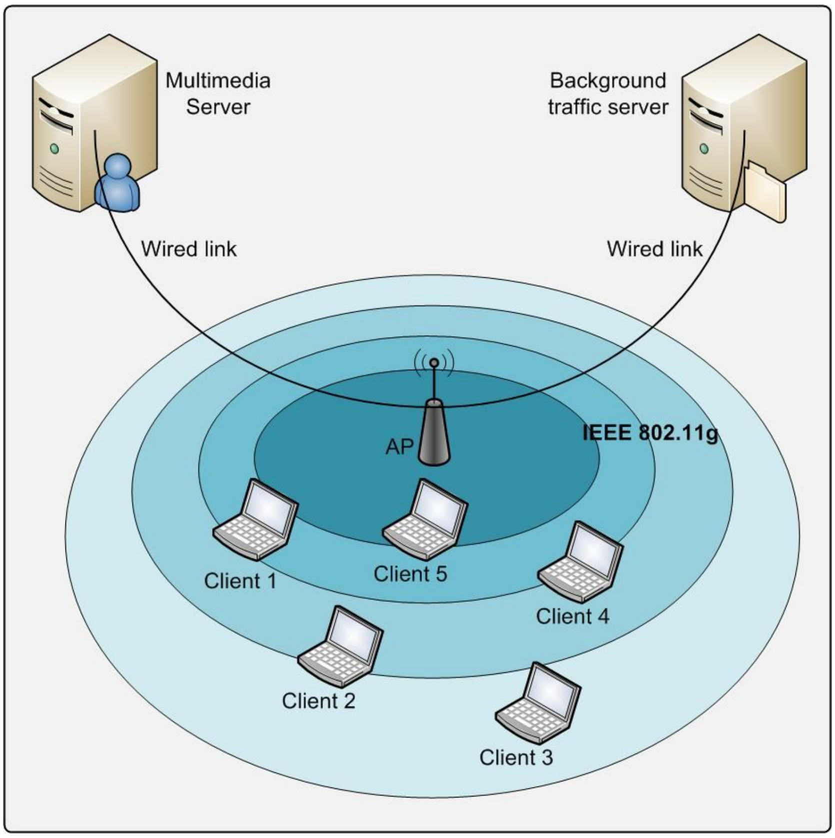


Figure 37 - General test architecture

The wired link between the AP and the Multimedia streaming 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 8.

Table 8 - 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

All simulations performed were considering the delivery of a stream in different scenarios, using a non-adaptive solution [52], the TFRC algorithm and the QOAS algorithm. The obtained results are compared with those obtained when the adaptive algorithm used was DQOAS.

4.1.2 Network model for 3GPP LTE network and the experimental setup

The proposed LTE QoS parameters mapping solution was tested using the LTE System Level Simulator [71], capable of simulating LTE SISO (Single Input Single Output) and MIMO (Multiple Input and Multiple Output) networks using TxD (Transmission Diversity) or OLSM (Open Loop Spatial Multiplexing) transmit 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 38 and like other system-level simulators, the core part consists of a link measurement model [72] and a link performance model [73]. 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 for easy adding of new functionalities and algorithms. The QoS parameters mapping scheme proposed in section 3.3.3 was implemented together with the DQOAS functioning principle in order to test the utility of an adaptation algorithm deployed over LTE networks. The simulation results obtained with different schedulers are compared with the results obtained when the original parameters mapping scheme is used and no adaptation algorithm is employed, leaving just the LTE QoS architecture to manage the multimedia flows.

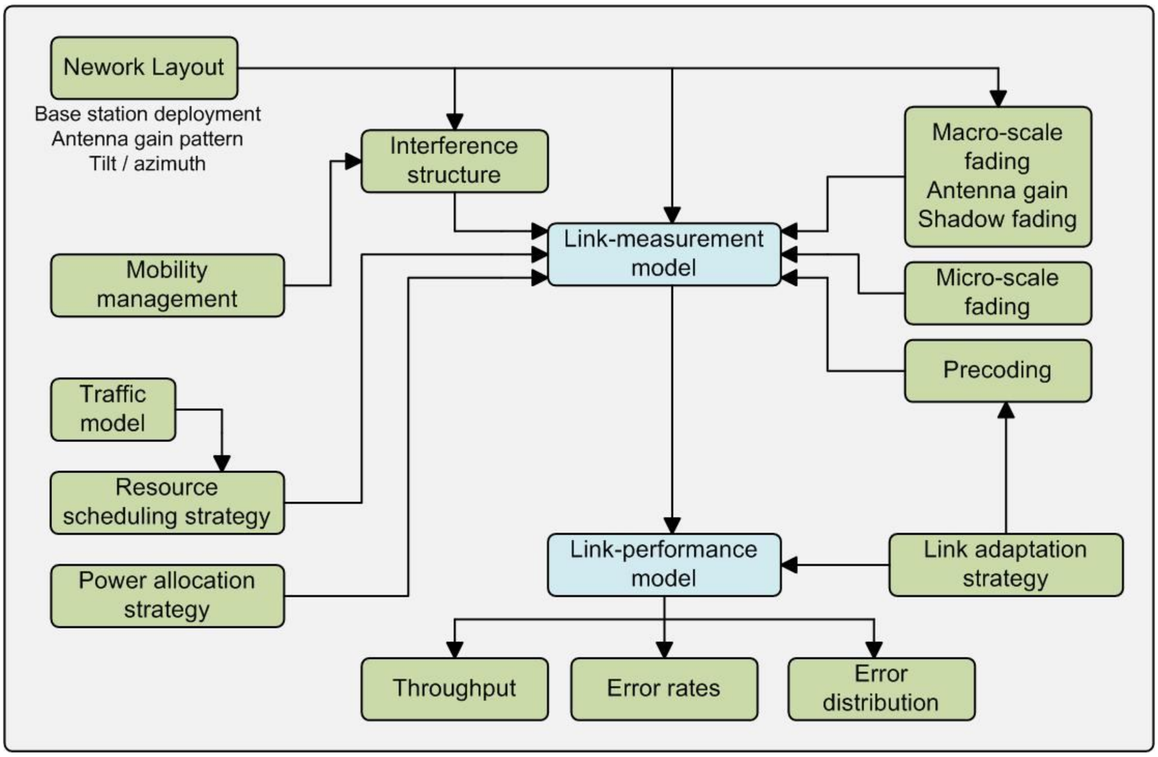


Figure 38 – LTE System Level Simulator architecture

The LTE network parameters used in the simulations are presented in the Table 9.

In each simulation scenario, two streams with different priorities are sent to each user in order to simulate a traffic mix. User *i* is considered to be satisfied with the received services if the quality level is kept above $M_i^{new} = M_i(1 + \varepsilon)$), where M_i represents the minimum accepted quality level for user *i*. For the ease of implementation, in the first step it is considered that all users have the same minimum accepted quality level, set at 0.5 Mbps.

Table 9 - 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 network map used to simulate the test scenarios is presented in Figure 39.It consists of 7 eNodeBs with 10 User Equipments (UEs) attached to each. The position of UE4 attached to eNodeB 1 is highlighted as it is one of the nodes used in the results discussions.

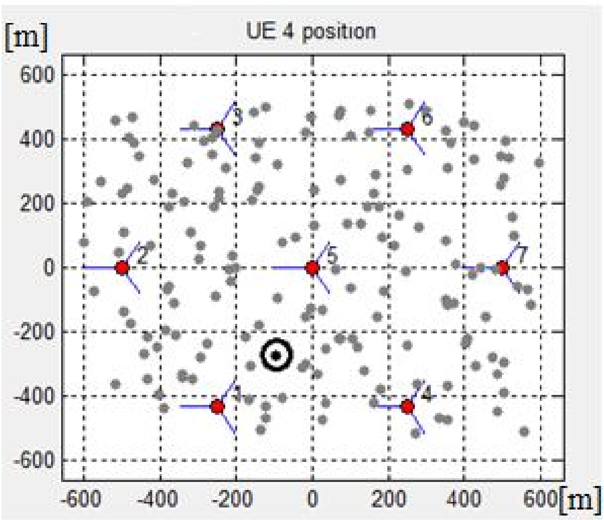


Figure 39 - LTE network map used in test scenarios

4.2 Experiments and Results

4.2.1 IEEE 802.11 wireless environment

DQOAS algorithm was tested in IEEE 802.11 environment by performing four different experiments, each lasting 60 seconds. In first two experiments, the five users are static and background traffic is considered only in the latest. Experiments three and four are employing user mobility, in both the absence and the presence of background traffic. The users are moving away from the AP with a speed of 0.5 m/s after the first 10 seconds of simulation. 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. For the non-adaptive method three different possibilities were analyzed, in which the transmission rate was set to the maximum rate (1.5 Mbps), to a medium rate (0.9 Mbps) and to the lowest rate (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.

Experiment 1: static users with no background traffic

In this experiment, all users are considered to have static positions inside the range of the AP. The positions are chosen so that the received signal strength is enough to satisfy their minimum QoE requirements. No background traffic is used in this simulation.

Figure 40 presents the multimedia throughput for the five users when QOAS adaptive algorithm is used while Figure 41 provides the same information obtained when DQOAS algorithm is employed for multimedia transmission. The non-adaptive method was used for all three transmission rates available (1.5, 0.9 and 0.3 Mbps). If the transmission rate is set to the maximum rate, only two users out of five can be served at the quality of experience expected by these users.

Table 10 - Results obtained with the non-adaptive solution

Transmission 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

If the quality is set to medium then the number of potential users increases, but still there will be an inefficient bandwidth use. Table 10 offers the results obtained using the non-adaptive solution.

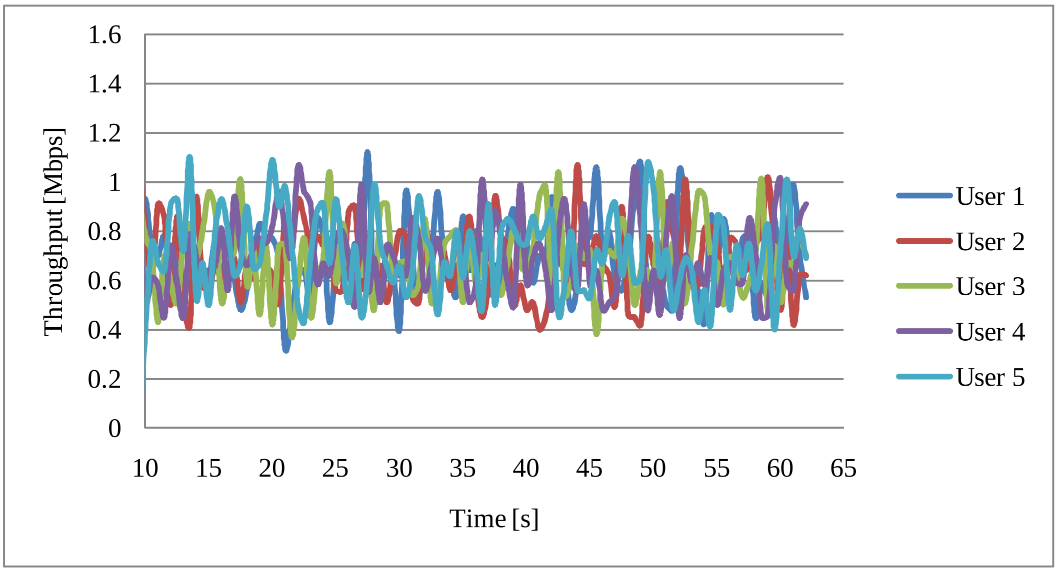


Figure 40 - Multimedia throughput when QOAS adaptive method is used

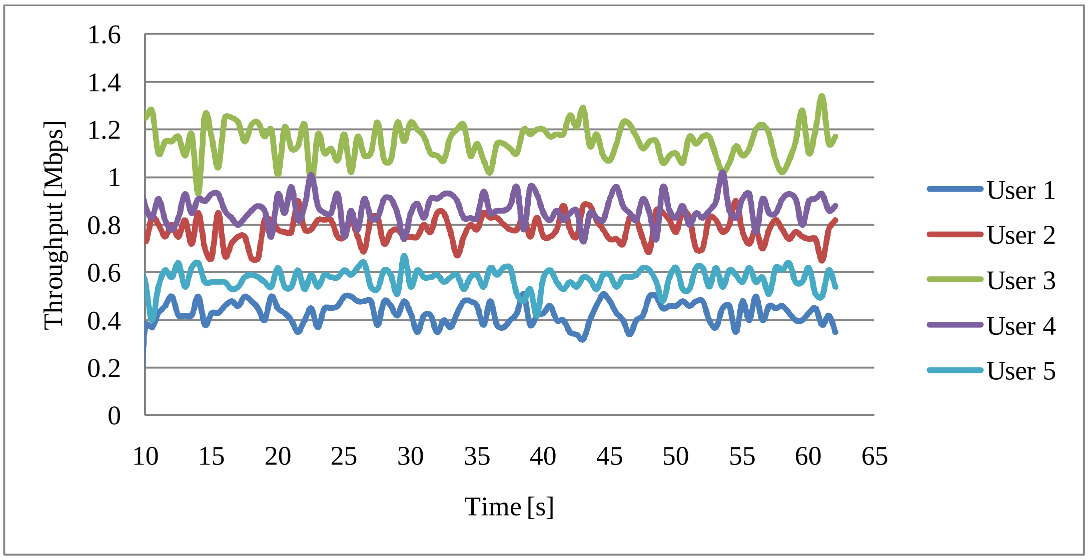


Figure 41 - Multimedia throughput when DQOAS algorithm is used

When TFRC is used, the throughput is comparable with the one obtained by QOAS algorithm, but the loss rate is bigger, reducing the perceived quality of the multimedia stream. Figure 42 presents the total throughput obtained by the adaptive algorithms during the simulation time:

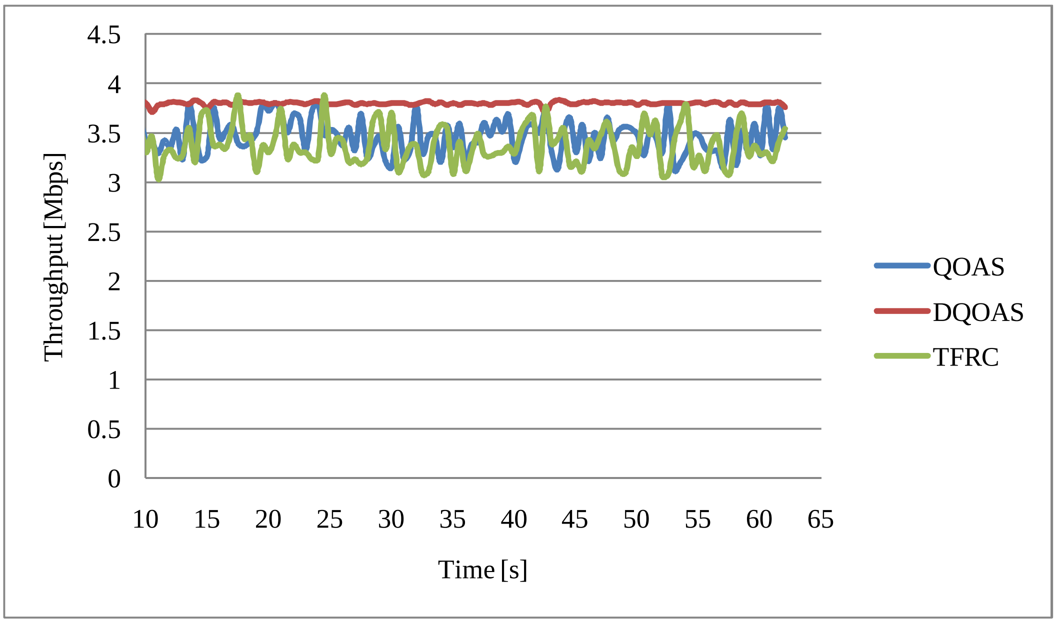


Figure 42 - Total throughput for the three adaptive solutions used

Table 11 concludes this experiment by presenting the average throughput for every uses when different adaptive algorithms are used. Compared to the minimum accepted quality level, it can be observed that DQOAS method is able to satisfy all five users.

Table 11 - Average throughput per user

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
Min. accepted level	0.300	0.600	1.0	0.800	0.450

The loss rate for the adaptive algorithms in the context of the first experiment is presented in Table 12.

Table 12 - Loss rate experienced when adaptive algorithms are used

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

Experiment 2: static users with background traffic

The conditions of the second experiment are similar with the one in the first experiment except the absence of the background traffic. For this experiment, one extra connection for every user is considered. The bandwidth of the background traffic connection varies between 0.10 Mbps and 0.11 Mbps. Figure

43 is presenting the multimedia throughput (not considering the background traffic) for every user when QOAS, TFRC and DQOAS algorithms are used, together with the minimum requirements. The variation of the throughput in time for all three algorithms is similar with the one obtained during the first experiment.

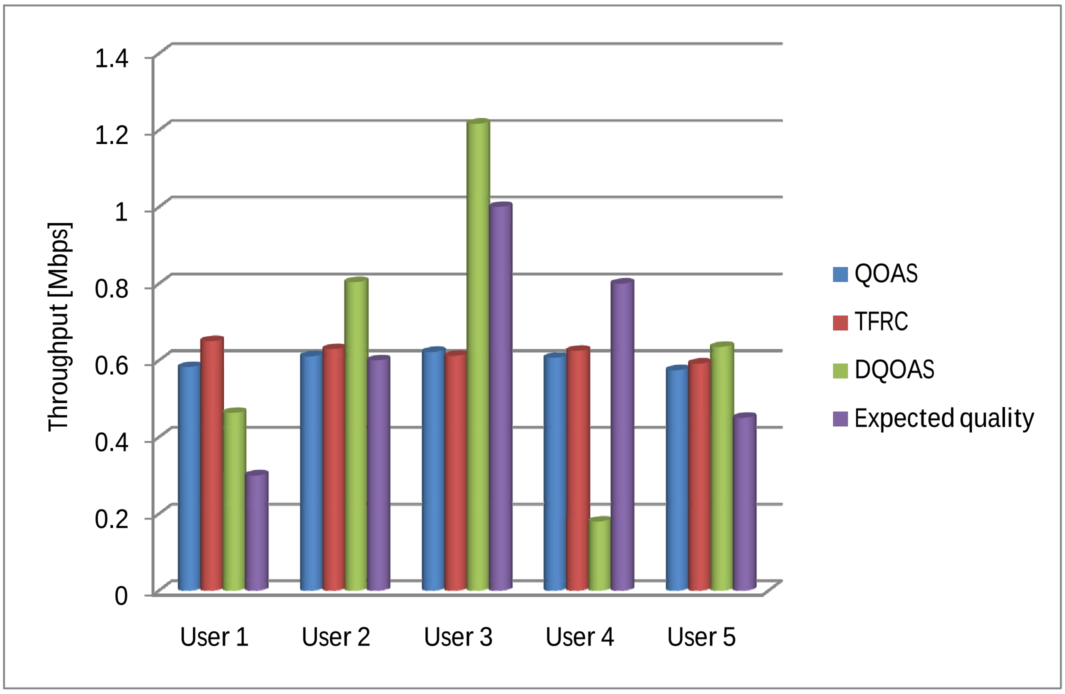


Figure 43 - Average user throughput when QOAS, TFRC and DQOAS algorithms are used

When DQOAS is used, the adaptation mechanism analyzes the user requirements and the network conditions and decides that user 4 is not receiving a satisfactory quality because of the presence of the background traffic. Therefore, it decides to terminate the session and free those resources so that they can be used by the other users. Figure 44 is illustrating the throughput variation for the five users.

Table 13 presents the total multimedia throughput obtained by the adaptive algorithms when background traffic in on. The total throughput for the background traffic used in this scenario is 0.54 Mbps.

Table 13 - Total multimedia throughput

Adaptive algorithm	Total multimedia throughput
QOAS	2.99
TRFC	3.10
DQOAS	3.30

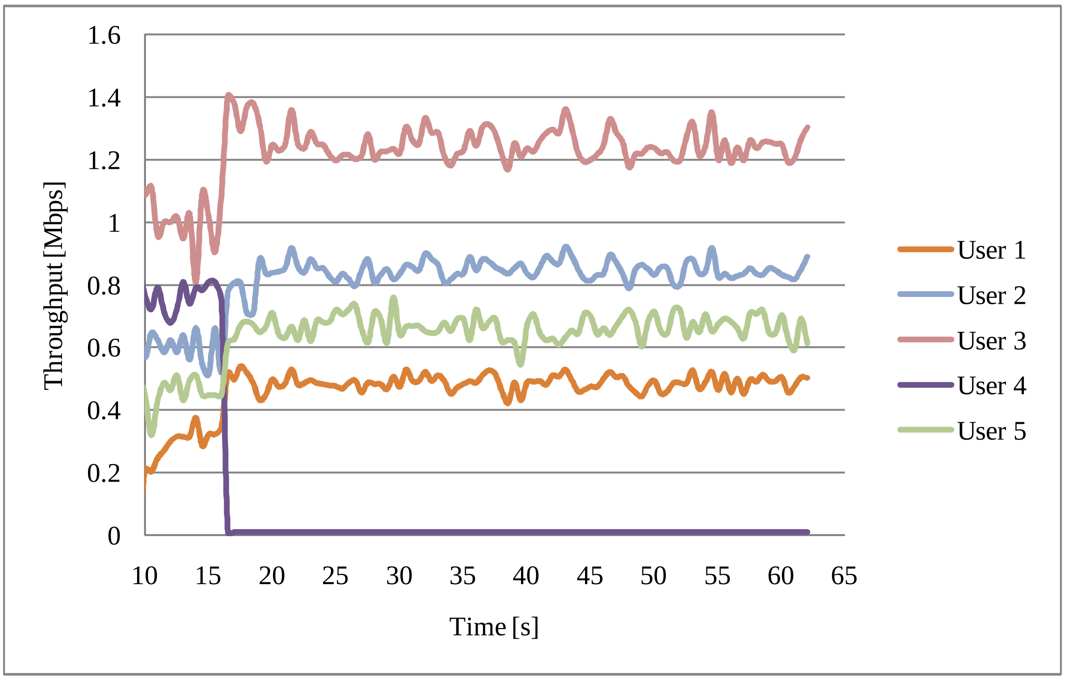


Figure 44 - Multimedia throughput when DQOAS algorithm is used

In case the non-adaptive solution is used, the results for highest and lowest transmission rates are the same with those from the first experiment. When the transmission rate is set to 0.9 Mbps, because of the background traffic the number of simultaneously satisfied users decreases from four, like it was in the case with no background traffic, to only three.

The loss rate for the adaptive algorithms 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	0.71
TFRC	3.14
DQOAS	0.86

Experiment 3: mobile users with no background traffic

The third experiment performed in order to test the performances of the proposed algorithm is having the same setup as the first experiment with the difference that node mobility is introduced. Here, all 5 users resides near the AP and during the first 10 seconds of simulation are starting to move towards the edge of the coverage area with a speed of 1.1 m/s. In the conditions of no background traffic, the multimedia throughput for the five users when TFRC and DQOAS are used is illustrated in Figure 45 and Figure 46 respectively, considering the hypothetical throughput of IEEE 802.11g standard reported to distance.

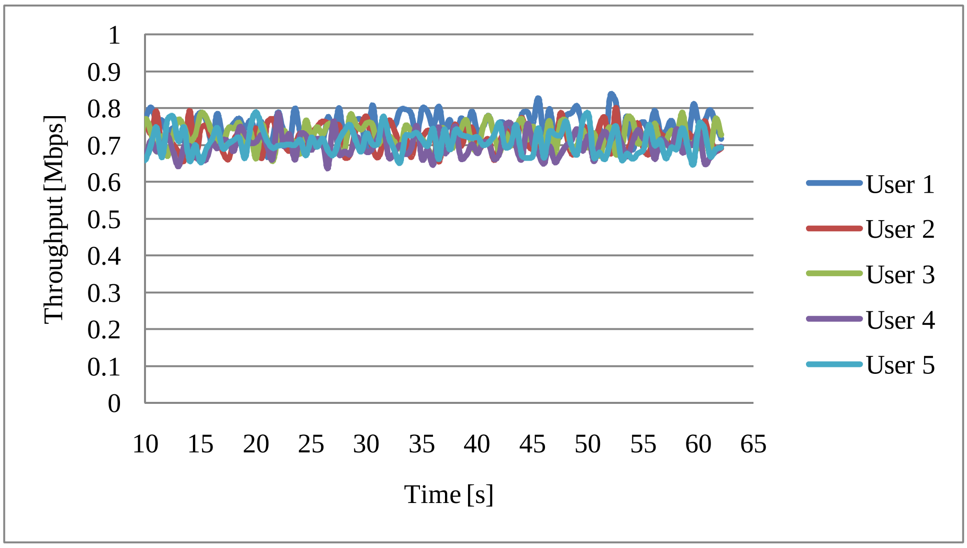


Figure 45 - Multimedia throughput when TFRC algorithm is used

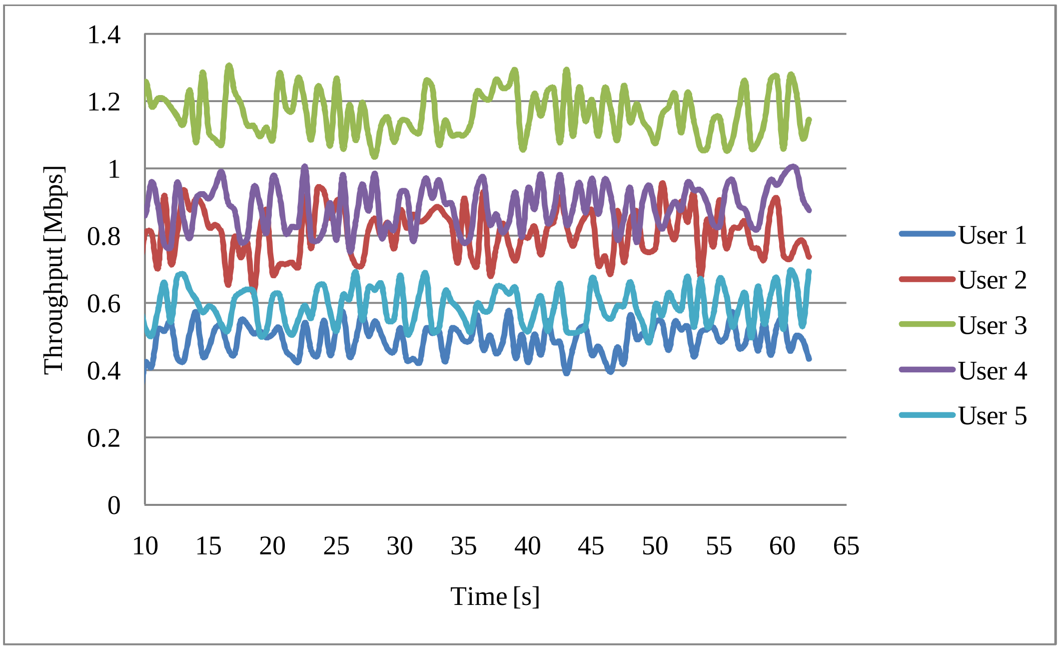


Figure 46 - Multimedia throughput when DQOAS algorithm is used

Table 15 presents the total throughput and loss obtained after the simulation scenario was run for all three adaptive algorithms.

Table 15 - Throughput and loss for all adaptive algorithms

Adaptive algorithm	Total 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. The user satisfaction for the adaptive methods is presented in Table 16. Like in the first experiment, 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.

Table 16 - Average throughput per user

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
Min. accepted level	0.300	0.600	1.0	0.800	0.450

Experiment 4: mobile users with background traffic

The last experiment is performed in the same simulation setup like the one used in simulation 3, but in conditions of background traffic with a total throughput of 0.54 Mbps. Each user has an established connection with the background traffic server with a throughput between 0.1 and 0.11 Mbps.

The presence of the background traffic is affecting the perceived quality of the multimedia stream and also the loss rate, as it can be observed in Table 17.

Table 17 - 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

Like in the second experiment, when used, DQOAS is trying to deliver the multimedia streams above the minimum satisfaction level for each user, but it decides to terminate one of the connections because of the limited resources. The multimedia throughput for every user is presented in Figure 47.

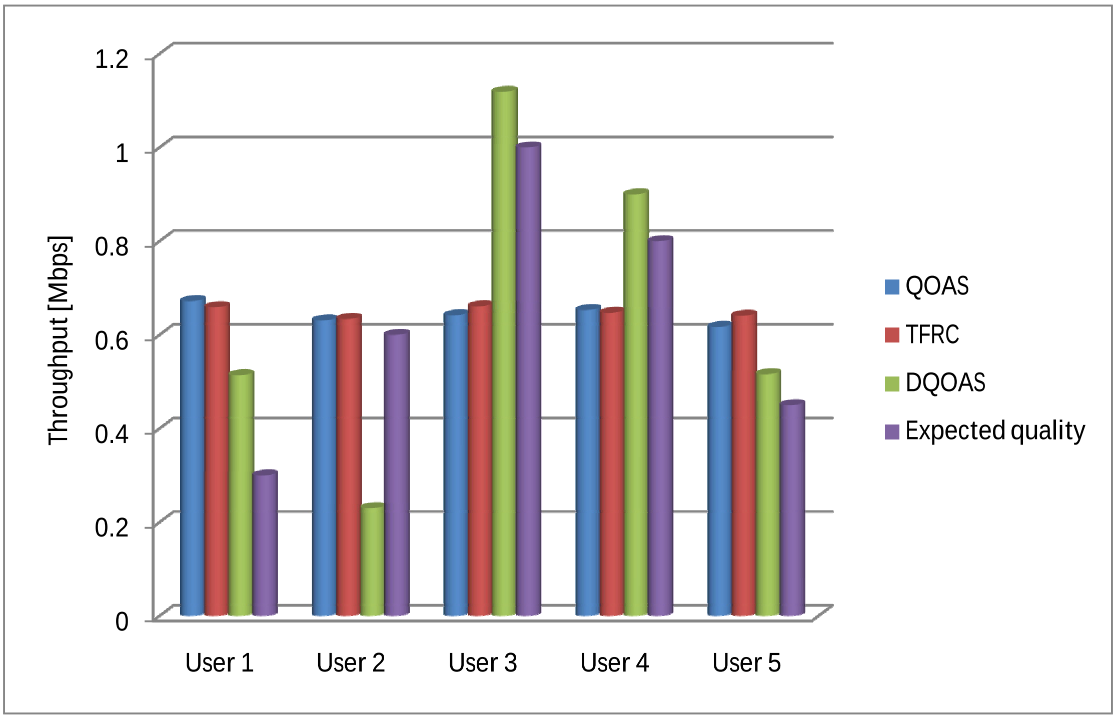


Figure 47 - Average user throughput when QOAS, TFRC and DQOAS algorithms are used

Used in other simulation environment, DQOAS decides to terminate the streaming process with user 2, not with user 4 like in scenario 2, when no mobility was present. The throughput diagram for users using DQOAS algorithm is presented in Figure 48.

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

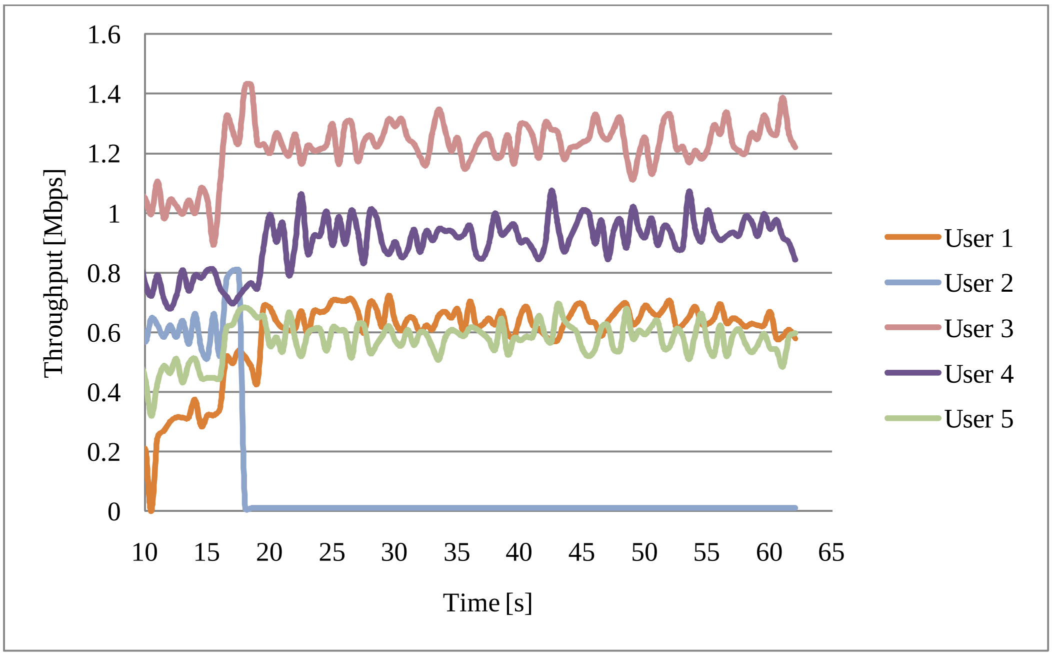


Figure 48 - Multimedia throughput when DQOAS algorithm is used

Discussions

The results shown in this section are indicating that DQOAS algorithm obtains very good results in scenarios with both static and mobile users. Being an extension of QOAS, the proposed algorithm is keeping the advantages of the first (very good loss rate, improved link utilization and dynamic adaptation procedure) while increasing the number of simultaneously satisfied users through its dynamic stream granularity adaptation. 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. Table 18 gives an overview of the four experiments in terms of link utilization and simultaneously satisfied users.

Table 18 - 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

Examples two and four are illustrating the case when DQOAS algorithm is not able to offer the minimum requirements to all users. Instead of reducing the resources used by other users in order to increase the transmission rate for the user with the poorest quality like TFRC and QOAS, the proposed method decides to terminate the streaming session 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. 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.

Another aspect that is important to be mentioned is related to the stream granularity in static and mobile scenarios. For the mobile case, the granularity is bigger when compared to the static case and the transmission levels are re-built more often because network conditions are changing faster. This approach improves the loss rate because the difference between the available streams data rates can cover the gap in available bandwidth introduced by the changes the radio channel when a user is mobile.

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).

4.2.2 3GPP LTE environment

DQOAS algorithm was tested in 3GPP LTE environment by performing three different experiments, each lasting 10 seconds. Each user receives two data streams, from which one has rich media content. The second stream simulates a web browsing connection that is being used simultaneously with rich media content streaming. One representative user from the 70 available was chosen in order to analyze the throughput and BLER. This is User 4, attached to eNodeB 1. Also, for every test scenario, three downlink schedulers are tested: Round Robin, Maximum Throughput and Proportional Fair.

First experiment uses the standard QoS parameters mapping scheme and the LTE QoS mechanism in order to deliver the two streams that have different priorities. In the second experiment, the LTE QoS mechanism is tested using the proposed mapping scheme, in which both data streams are considered to have the same priority with the same weight in the schedulers' queue. Third experiment is using DQOAS as the delivery algorithm. The minimum required level for the multimedia stream (first stream) was set to 0.500 Mbps for all users. The second stream, which is a TCP download, the minimum accepted level is set to 0.250 Mbps. All users are moving randomly through the network map with a speed of 5 km/h. The throughput and BLER average values are calculated over the last 40 second of each simulation.

Experiment 1: original mapping scheme + LTE QoS mechanism

In this experiment, the original 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 49 and 50. Each figure presents both the throughput and the BLER for each stream received by User 4.

Figure 49 - Throughput and BLER for User 4 when PF scheduler is used

The average throughput of stream 1 when PF scheduler is used is 1.447 Mbps, satisfying the minimum requirements for User 4, 0.500 Mbps. For the second stream, the average throughput calculated shows a value of 0.418 Mbps. Even if the average throughput is above the required satisfaction rate (0.250 Mbps), there is a period of 6 seconds, between second 39 and second 45, when the experienced throughput is below the requested quality. In terms of user perceived quality, this 6 seconds interval can determine the user to terminate the session for the second stream.

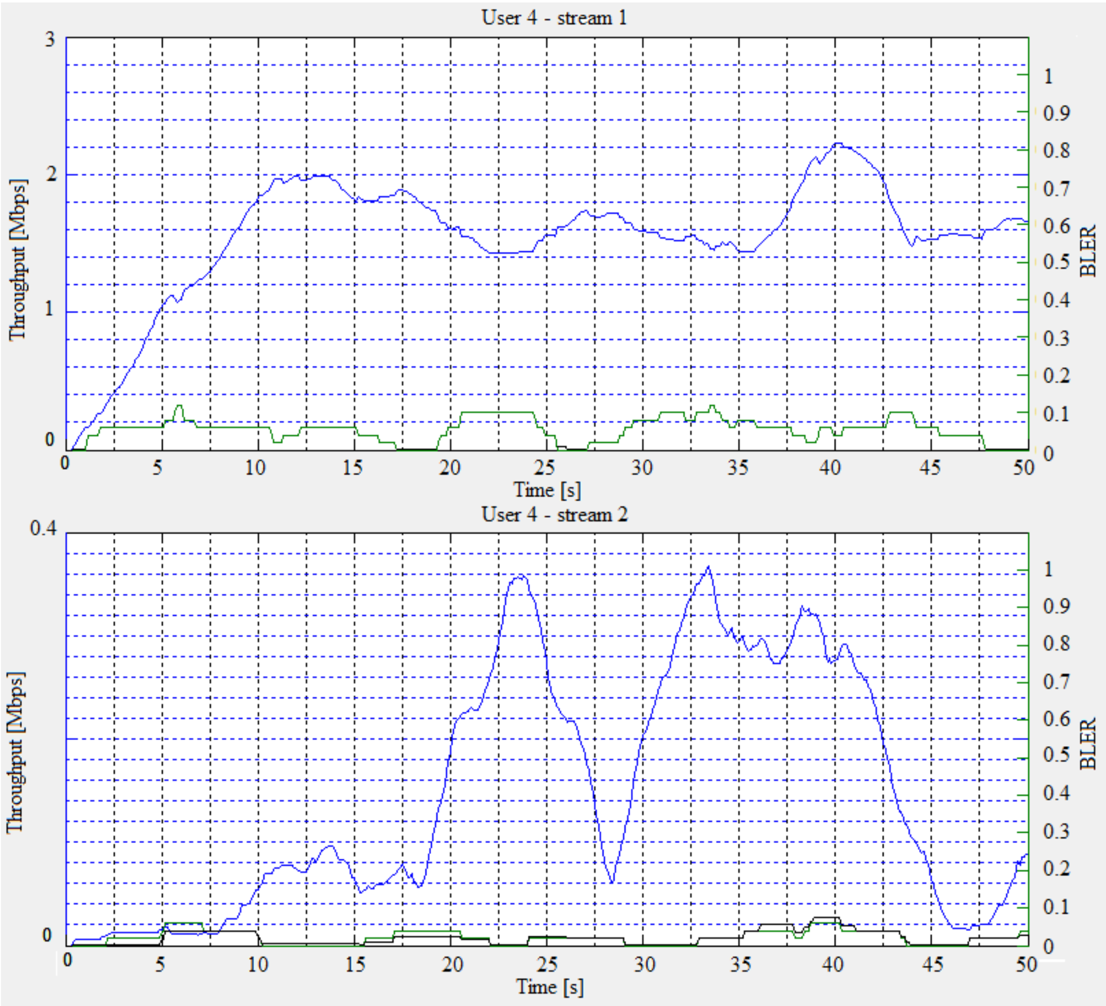


Figure 50 - Throughput and BLER for User 4 when MT scheduler is used

Using the Maximum Throughput Scheduler in the context of the first experiment, we obtain an average throughput value of 1.723 Mbps for the first stream and 0.221 Mbps for the second. In this case, the average throughput for the second stream experienced by the user is below the minimum required level, and as a consequence the overall user satisfaction will decrease. The low throughput value experienced for the second stream is a consequence of flow prioritization when used in combination with the MT scheduler.

The results when Round Robin scheduler is used – 1.161 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.

The BLER average values calculated during this experiment are presented in Table 19.

Table 19 - BLER values when different schedulers are used

	BLER in case of PF scheduler [%]	BLER in case of RR scheduler [%]	BLER in case of MT scheduler [%]
Stream 1	5.3	6.6	7.1
Stream 2	7.0	7.3	4.2

The results obtained during this experiment are suggesting that in the general LTE concept, the Proportional Fair scheduler performs better compared with Round Robin and Maximum Throughput.

Experiment 2: original mapping scheme + DQOAS algorithm

The second experiment performed using the 3GPP LTE technology is using the original flow prioritization mechanism in conjunction with the delivery algorithm proposed in this thesis, DQOAS. 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 QoS-based delivery mechanism. The User 4 throughput variation in time for the two streams is presented in Figure 51 for the PF scheduler and in Figure 52 for RR scheduler.

The average throughput value for the first stream obtained when the PF scheduler is used is 1.671 Mbps, which is greater than the one obtained in similar conditions in the context of the first experiment. The second stream average throughput value, 0.292 Mbps, is smaller than the one obtained when LTE QoS mechanism is used, but it never drops for more than 2 seconds under the minimum required level, which is 0.250 Mbps. This is an important aspect, because the chances User 4 will terminate the session due to a long period of unsatisfactory quality are reduced. The primary goal of DQOAS is to keep the transmission rate over the required level, and only after that, if the network conditions are allowing it, to increase the delivery rate.

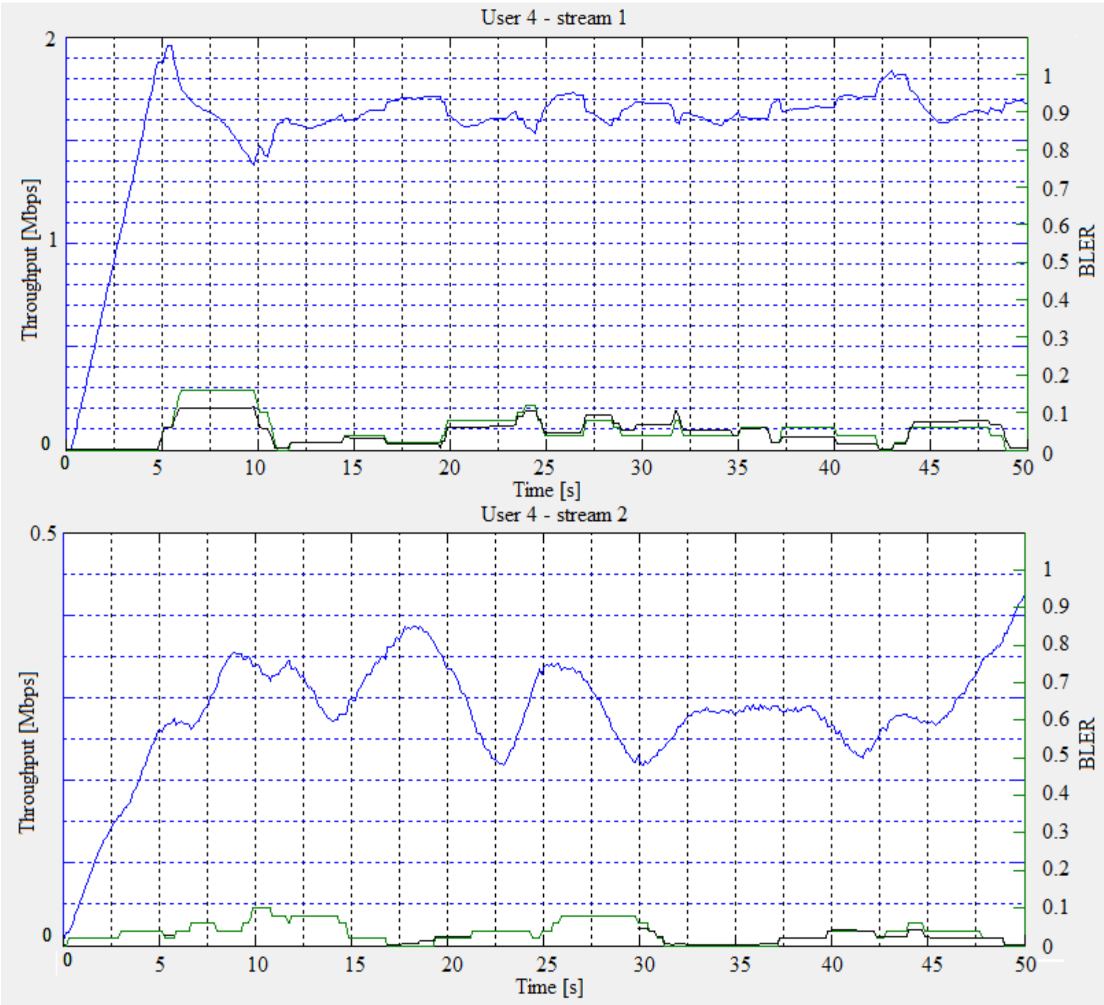


Figure 51 - Throughput and BLER for User 4 when PF scheduler is used

The average throughput simulation values when RR scheduler is used are 1.042 Mbps for the first stream and 0.338 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 bigger than the one obtained during the first experiment, but there is a period of almost 5 seconds when the throughput drops below the 0.250 Mbps limit. This quality drop is introduced by the characteristics of the scheduler and the radio channel, and during this interval the user can decide to terminate the session.

Figure 52 - Throughput and BLER for User 4 when RR scheduler is used

If the Maximum Throughput scheduler is used, the results are showing that DQOAS method performs better in terms of final user satisfaction. If during the first experiment, the second stream was not received in acceptable quality conditions, having an average throughput of just 0.221 Mbps, in the second setup the average value is 0.463 Mbps, without any drops below 0.300 Mbps. The second stream was also above the accepted quality level, averaging to a total of 1.392 Mbps during the last 40 seconds.

Considering also the BLER values presented in Table 20, we 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.

Table 20 - BLER values when different schedulers are used

	BLER in case of PF scheduler [%]	BLER in case of RR scheduler [%]	BLER in case of MT scheduler [%]
Stream 1	4.3	3.2	4.8
Stream 2	3.1	3.7	3.1

Experiment 3: proposed mapping scheme + DQOAS algorithm

This experiment proposes a new mapping scheme for LTE QoS parameters, modifying the flow priority settings for streams generated by the same application. The new approach, presented in detail in section 3.3.3 is designed to optimize the use of DQOAS over LTE network, increasing the total number of simultaneously per-application satisfied users and decreasing the BLER of individual streams. Like in the previous two experiments, three schedulers were used in order to perform an in-depth analyze of the proposed method, by comparing the results with those obtained in the above simulations.

For the first scheduler, PF, the simulations are showing an average throughput of 0.760 Mbps for the first stream and 0.430 Mbps for the second one, both big enough to satisfy User 4 quality requirements. First stream has a smaller average throughput compared with those obtained in the first two experiments because when the new mapping scheme is used, DQOAS can optimize the resource allocation process, by limiting the throughput upper value of stream 1 in order to release more resources for the nearby users. This is possible because during the adaptation process, DQOAS considers all users that are connected to one eNodeB. In the second experiment, DQOAS cannot manage optimal the two streams with different priorities because the scheduler is overriding it. 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 them, increasing the overall QoE of the application.

Figure 53 illustrates the results obtained when Proportional Fair scheduling method is used.

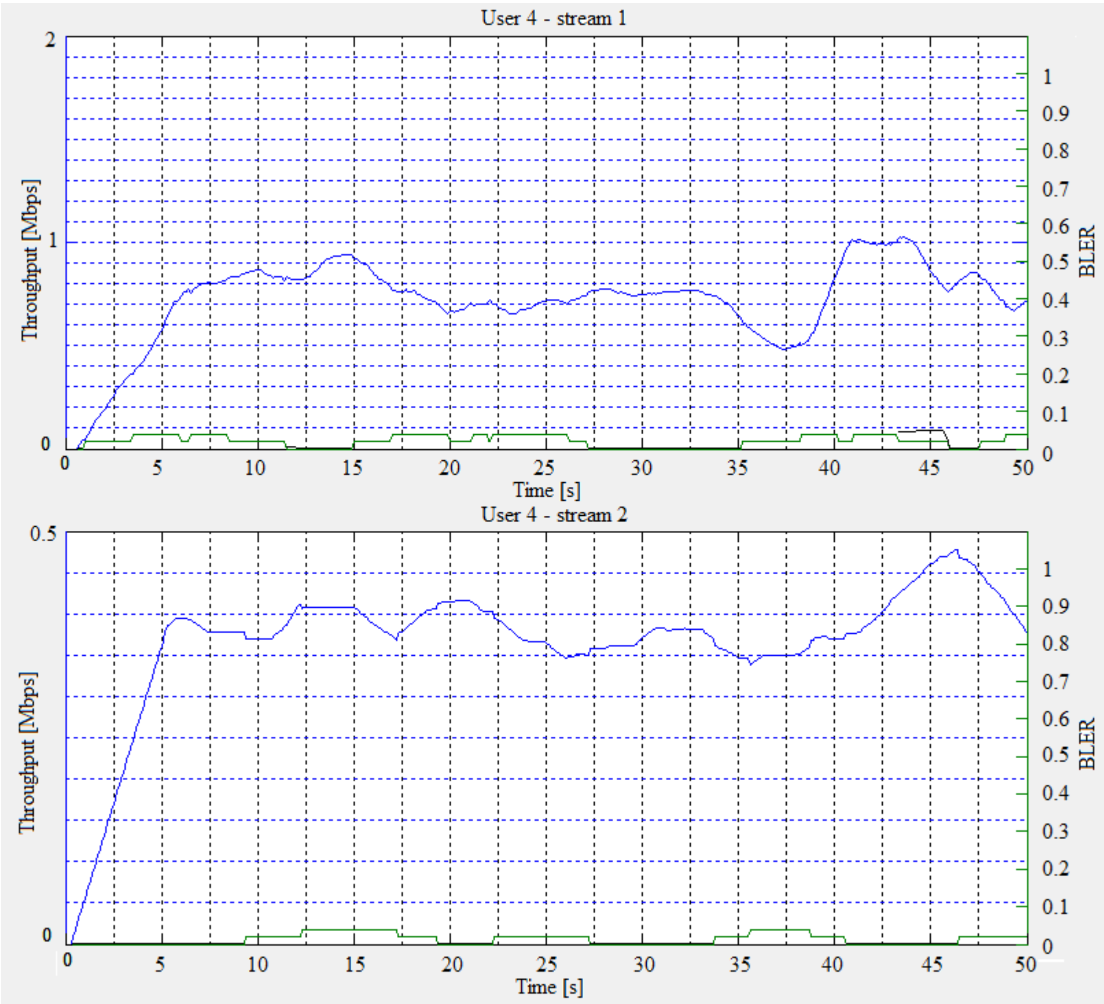


Figure 53 - Throughput and BLER for User 4 when PF scheduler is used

Figure 54 presents the simulation results obtained with RR scheduler. Throughput average values obtained in this case 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 first two simulations, 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, but the overall performance can be regarded as satisfactory. The values obtained when MT scheduler is used are 0.668 Mbps for stream 1 and 0.318 Mbps for stream 2. Figure 55 describes the throughput variation in the considered case.

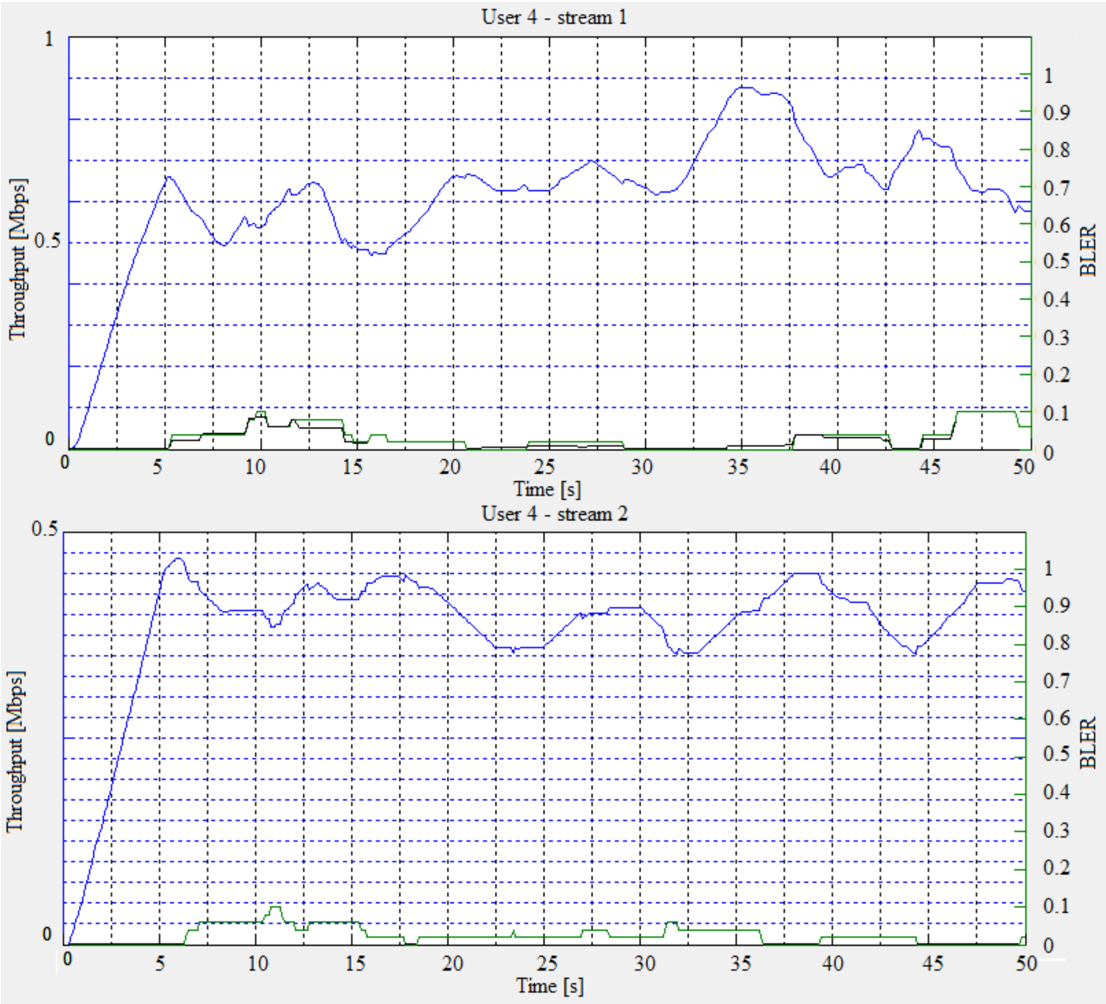


Figure 54 - Throughput and BLER for User 4 when RR scheduler is used

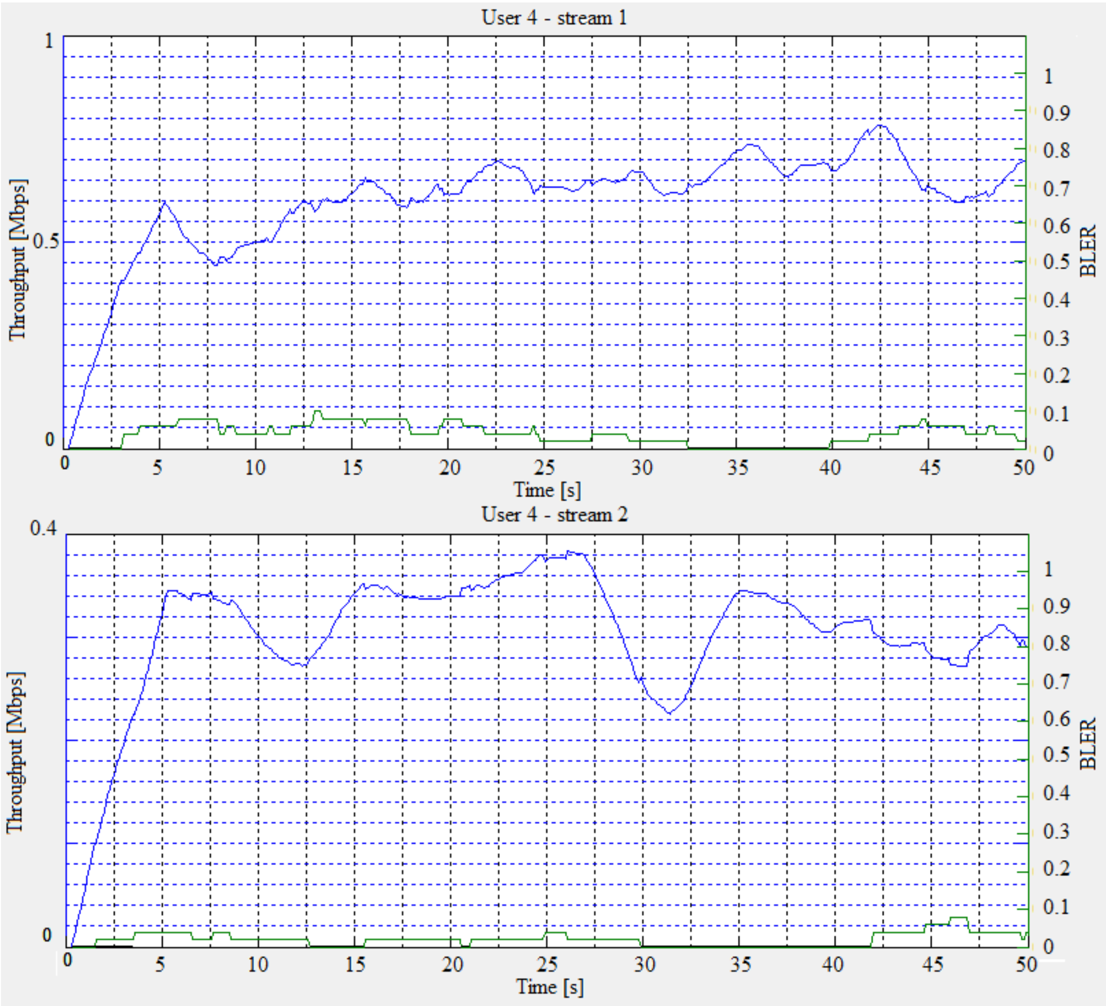
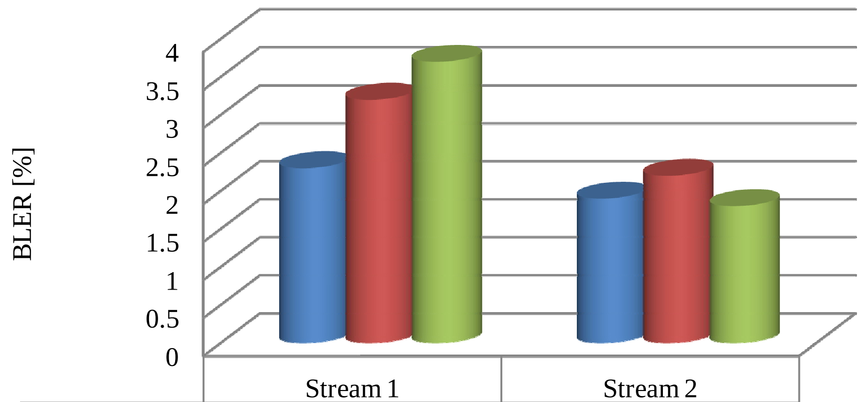


Figure 55 - Throughput and BLER for User 4 when MT scheduler is used

Figure 56 presents the BLER resulted for the two streams when all three schedulers are used. Like in the first experiment, considering the average throughput and BLER values for the two streams, the conclusion is that Proportional Fair scheduler performs better than Round Robin or Maximum Throughput. Compared with the first two experiments, the BLER values obtained in this experiment are lower, which means that the user perceived quality is improved.



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

Figure 56- BLER values when different schedulers are used

Discussions

The results presented in this section are suggesting that DQOAS algorithm can be used with good results in 3GPP LTE networks if the original service prioritization scheme is adapted in order to give DQOAS the opportunity to manage radio resources allocation in a more direct manner. This way, the scheduler will not override the algorithm's settings, letting DQOAS to improve the link efficiency and to increase the number of simultaneously satisfied users.

Table 21 gives an overview of the total satisfied users by both streams, considering all 7 eNodeBs with 10 users attached each.

Table 21 - The percentage of satisfied users per experiment

Experiment number	The percentage of satisfied users when the scheduler used is		
	PF	RR	MT
Experiment 1	48	46	23
Experiment 2	51	46	31
Experiment 3	60	50	30

Even if 60% satisfied users does not seems to be a good ratio, it is important to mention that this value was calculated considering all 70 users are using the same application simultaneously while moving randomly inside the network map. Like this, some of the 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 discriminating factor can be considered the distance from one user to the belonging eNodeB and for a more detailed analyze the satisfaction rate should be computed taking distance into account.

It can be observed that using DQOAS 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 and with 9% compared to DQOAS used with the original prioritization scheme.

Another advantage introduced by the proposed scheme is that the throughput fluctuations are small compared to those when other methods are used. This is because of the controlled difference between the transmission levels used by DQOAS. Big variations in throughput are unwanted during the multimedia delivery process because they rapidly decrease the perceived quality.

Looking at the BLER values obtained in all three experiments, the best results are experienced when the proposed solution is used. 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. It is in consequence important to have a reduced BLER, since it is a main characteristic for defining the systems' performance.

Taking into account these aspects, the advantages of DQOAS algorithm demonstrated in wireless LANs can be replicated also in 3GPP LTE networks if a new prioritization scheme is used, together with a competitive scheduling algorithm.

5. Conclusions and Future Works

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

1.1 Conclusion

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, 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 even a more difficult problem, especially when they are 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 we are to consider the adaptation process just between the delivery server and the device, than we can find many adaptation algorithms able to offer very good performances. But if we want to increase the end-used perceived quality, we have 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 see 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 with different strategies 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 controlling the radio resources allocation. The algorithms' decision to terminate a session when this is considered to be unsatisfying proves to be good, because it allows it 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 multiple streams. These results can be improved by using a new QoS parameters mapping scheme able to modify the streams' priority and thus allowing the algorithms decisions to not be overridden by the scheduler. Because the scheduler is on important component of the LTE QoS mechanism, some basic scheduling strategies were also analyzed and the results show that the Proportional Fair algorithm obtains better performances than Round Robin or Maximum Throughput.

1.2 Contributions

This thesis has the following contributions:

- It proposes a new user-oriented delivery algorithm able to improve the final users' QoE, the link utilization and the number of satisfied by dynamically building the stream granularity based on current network conditions and on user preferences; the benefits of the proposed algorithm are assessed using two different wireless technologies.

- A new QoS parameters mapping scheme for data flow prioritization in LTE was proposed in order to increase the performances obtained by DQOAS algorithm
- It shows that using the Proportional Fair scheduler for LTE technology gives better results, compared to Round Robin and Maximum Throughput schedulers.

Also, this thesis work was summarized into three scientific papers, from which two were published in international conferences and one was accepted for publication in a journal.

1.3 Future work

When studying mixed traffic scenarios in an all-IP system, there are a large number of parameters and possible scenarios that can be analyzed, while in this thesis, the simulations were performed using very specific traffic scenarios. Therefore, a natural extension of this work is to enlarge the testing grounds by using different traffic scenarios with an increased number of users. As this thesis was only concerned with the downlink aspect of LTE, there is still a need for a corresponding study for the uplink.

Another interesting development implies refining DQOAS as well as building a practical prototype for this system as subjective tests are needed to validate the preliminary results obtained by simulation. Studying different types of loss that can occur in a wired-cum-wireless network and fine tuning the adaptation to these is also attractive.

Furthermore, extensive testing using different propagation models available with the LTE System Level Simulator can be performed, together with extending the mapping scheme in order to increase the total number of satisfied users. This can be achieved by using different weights in the same queue based on the traffic type and by utilizing other schedulers developed by researchers in the field. Also, subjective assessments of user perceived quality for multimedia traffic can be performed.

Assessing the delivery performance of the algorithm can also be done using existing quality metrics, like PSNR or VQM. This will give a more precise indication about the user perceived quality and in consequence about the entire systems' performances.

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

Finally, the results presented in this thesis can be further verified by running more iterations of the simulation scenarios used.

Bibliography

[1] G.-M. Muntean C. H. Muntean, , C. Pahl, Ed.: IGI Global Publishing House, November, 2007, ch. End-User Quality of Experience Aware Personalized E-Learning.

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

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

[4] 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.

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

[6] 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.

[7] M. van der Schaar and H. Radha, "Adaptive motion-compensation finegranular-scalability (AMC-FGS) for wireless video," *IEEE Trans. Circuits Syst. Video Technol*, vol. 12, no. 6, pp. 360-371, June 2002.

[8] L. Murphy, and P. Perry N. Cranley, "Optimum adaptation trajectories for streamed multimedia," *ACM Multimedia Sys. J., New York: Springer-Verlag*, 2005.

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

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

September 2005.

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

[12] S. Mangold and L. Berlemann B. H. Walke,. The Atrium, Southern Gate, Chichester, England: John Wiley & Sons, Ltd., 2006, pp. 79-80, 84, 96, 103-105, 115–117.

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

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

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

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

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

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

[19] 3GPP, "3GPP TS 36.300 v8.7.0, Technical Specification Group Radio Access Network, Rel. 8," 3GPP LTE, Technical Report 2008.

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

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

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

[23] 3GPP, "3GPP TS 36.321, E-UTRA Medium Access Control (MAC) Protocol Specifications, Rel.8," 3GPP LTE, Technical Report 2008.

[24] E. Dahlman, *3G Evolution: HSPA and LTE for Mobile Broadband*, 2nd ed.: Academin Press, 2008.

[25] IETF, "Internet Protocol (ip)," IETF, Technical Report RFC 791 1981.

[26] S. Deering and R. Hinden, "Internet protocol version 6 (ipv6) specification," IETF, Technical Report RFC 2460 1998.

[27] J. Postel, "User datagram protocol," IETF, Technical Report RFC 768 1980.

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

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

[30] D. Reed, and D. Clark J. Saltzer, "End-to-end arguments in system design," *ACM Transactions on Computer Systems*, vol. 2, pp. 277-288, 1984.

[31] J. Qhang and M. E. Aydin J. Yang, "Optimization of WCDMA radio networks with consideration of link-level repformance factors," *Int. J. Mobile Networks Design and Innovation*, vol. 2, no. 1, pp. 26-32, 2007.

[32] H. Yu, M. Handley and D. Estrin R. Rejaie, "Multimedia proxy caching for quality adaptive streaming applications in the internet," *Proc. Nineteenth Annu. Joint Conf. IEEE Computer and Communications Societies 2000 (INFOCOM'00), Tel-Aviv, Israel*, pp. 980-989, March 2000.

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

[34] C. Perkins,. USA: Pearson Education, Inc., 2003, pp. 67-145.

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

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

[37] J. Nagle, "Congestion Control in IP/TCP Internetworks," IETF, TechnicalReport RFC 896 1984.

[38] J. Kurose, "Open issues and chalanges in providing quality of service guarantees in high speed networks," *ACM Computer Communicarion Review 23*, pp. 6-15, 1993.

[39] C. Elliot, "High quality multimedia conferencing through a long haul packet network," *Proc. of the first ACM International Conference on Multimedia*, pp. 91-98, 1993.

[40] M. Handley and D. Estrin R. Rejaie, "RAP: An end-to-end rate-based congestion control mechanism for realtime streams in the internet," *IEEE Computer and Communications Society (INFOCOM)*, vol. 3, pp. 1337-1345, March 1999.

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

[42] C. Casetti, M. Gerla, M. Y. Sanadidi and R. Wang S. Mascolo, "TCP westwood: Bandwidth estimation for enhanced transport over wireless," *Mobile Computing and Networking*, pp. 287-297, 2001.

[43] F. Menta and G. Schembra, "Effcient design of red routers for tcp/rap fairness optimizatio," *IEEE International Conference on Communications*, vol. 4, pp. 2186-2190, 2002.

[44] D. Clark, J. Crowcroft, B. Davie, S. Deering, D. Estrin, S. Floyd, V. Jacobson, G. Minshall, C. Partridge, L. Peterson, K. Ramakrishnan, S. Shenker, J. Wroclawski and L. Zhang B. Braden, "Recommendations on queue management and congestion avoidance in the internet," IETF, Technical Report RFC 2309 1998.

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

[46] P. C. Cosman and G. M. Voelker S. Cen, "End-to-end differentiation of congestion and wireless losses," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 703-717, 2003.

[47] M. Chen and A. Zakhor, "Multiple tfrc connections based rate control for wireless networks," *IEEE Transactions on Multimedia*, vol. 8, no. 5, pp. 1045-1061, October 2006.

[48] J. Crowcroft and P.Oechslin, "Differentiated end to end internet services using weighted fair sharing," *ACM computer Communications Review*, vol. 28, no. 3, July 1998.

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

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

[51] G.-M. Muntean, "Efficient delivery of multimedia streams over broadband networks using QOAS," *IEEE Transactions on Broadcasting*, vol. 52, no. 2, pp. 230-235, June 2006.

[52] G.-M. Muntean and N. Cranley, "Resource efficient quality-oriented wireless broadcasting of adaptive multimedia content," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 362-368, March 2007.

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

[54] L. Zhang, S. Berson, S. Herzog and S. Jamin R. Braden, "Resource reservation protocol (rsvp)," IETF, Technical Report RFC 2205 1997.

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

[56] A. Tanenbaum, *Computer Networks*, 4th ed.: Pearcon Education, Inc., 2003.

[57] H. Fattah and C. Leung, "An overview of scheduling algorithms in wireless multimedia networks," *IEEE Wireless Communications*, vol. 9, no. 5, pp. 76-83, October 2002.

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

[59] A. Banchs and X. Perez, "Distributed weighted fair queueing," *International Communications Conference*, vol. 5, pp. 3121-3127, April 2002.

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

[61] P. Bahl and S. Gupa N. Vaidya, "Distributed fair fcheduling in a wireless LAN," *Annual International Conference on Mobile Computing and Networking*, pp. 167-178, August 2000.

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

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

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

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

April 2008.

[66] G. Berardinelli et co., L. A. de Temino, "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.

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

[68] 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.

[69] 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.

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

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

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

[73] D. Astely, T. Salzer, S. Visuri, A. Alexiou, S. Karger and G.-A. Seraji K. Brueninghaus, "Link performance models for system level simulations of broadband radio access systems," September 2008.

Appendix

A Novel Adaptive Multimedia Delivery Algorithm for Increasing User Quality of Experience during Wireless and Mobile E-learning

Vasile Horia Muntean and Gabriel-Miro Muntean, *Member, IEEE*

***Abstract* — Multimedia content is distributed 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. The quality of experience influences in turn the quality of any e-learning process. This is as users in general and learners in particular are becoming increasingly quality-aware in their expectations. Therefore delivering a good quality video stream as a part of any e-learning process is very important. This paper proposes a new adaptive multimedia delivery algorithm which can be used in the context of e-learning. The Dynamic Quality Oriented Adaptation Scheme (D-QOAS) adapts the multimedia content sent based on both user preferences and network conditions, while adjusting dynamically its adaptation policy during delivery. Simulation results show that for different user profiles and various network conditions the improvement in end-user perceived quality is significant. Important benefits are obtained in terms of the total number of simultaneous users and in the link utilization, as well as in quality as measured by some video quality metrics.**

Index Terms — **dynamic content adaptation, e-learning, multimedia, rate control, wireless networks**

INTRODUCTION

E-LEARNING has become an important service offered over the Internet and despite the current economic downturn, is expected to increase in popularity [1]. In this context the global e-learning market which was worth $8 billion in 2006 is expected to grow to $13 billion by 2011 [2]. A recent study on e-learning in Europe [3] which highlighted future trends indicated that one of the most important is the provision of suitable services to people without broadband internet access, in rural or remote areas, potentially involving wireless connectivity.

Lately increased number of users is accessing learning content via wireless networks and by using

Manuscript received April 9, 2009. The support of Enterprise Ireland, Proof of Concept and Science Foundation Ireland Research Frontiers programmes is gratefully acknowledged.

A Novel Adaptive Multimedia Delivery Algorithm for Increasing User Quality of Experience during Wireless and Mobile E-learning

V. H. Muntean is with the Performance Engineering Laboratory, Dublin City University, Dublin 9, Ireland (e-mail: horia@eeng.dcu.ie).

G.-M. Muntean is with the Performance Engineering Laboratory, Dublin City University, Dublin 9, Ireland (e-mail: munteang@eeng.dcu.ie).

mobile devices. Most content is rich media-based and often puts significant pressure on the existing networks in order to support high quality of delivery. The problem of supporting high quality multimedia streaming is getting even more difficult when delivering multimedia over wireless LANs. This is as wireless technologies offer lower bandwidth and their service is highly affected by environmental factors, traffic load and number of clients, as well as their location and mobility patterns. In this context, offering a solution for improving user quality of experience regardless of users' speed and mobility pattern is highly important. Additionally during multimedia delivery in the context of e-learning, there are some extra parameters that should be considered in relation to the user profiles in order to support a user-oriented solution.

This paper introduces a novel solution for adaptive multimedia streaming over IEEE 802.11 Wireless LANs with focus on mobile e-learning. The proposed Dynamic Quality Oriented Adaptation Scheme (D-QOAS) is a user-oriented adaptation mechanism that dynamically adapts the multimedia content sent based on user preferences and network conditions. By employing this algorithm, an increased number of simultaneous learners can be accommodated while maintaining the perceived quality levels above their individual good quality thresholds. This is done by dynamically modifying the adaptation policy during delivery, increasing or decreasing the granularity of the adaptation for some users, based on their preferences and actual network conditions.

D-QOAS performs an optimal dynamic multimedia content management according to end-user profile, knowledge level, goals, preferences as well as network conditions in order to support higher end-user quality of experience during the learning process.

The paper is structured as follows. Section II presents some existing adaptive solutions and the progress done in the e-learning area, while section III describes the system architecture its main modules. Section IV presents the adaptation algorithm – D-QOAS and section V describes the testing environment and presents the results. Conclusions are drawn in section VI.

Adaptive Solutions for Multimedia Delivery

Much research work has been done in the area of adaptive streaming-based schemes. Existing adaptive solutions such as TCP-Friendly Rate Control Protocol (TFRC) [4] and the enhanced Loss-Delay Adaptation Algorithm (LDA+) [5] support certain level of multimedia quality in variable network delivery conditions, but they are mainly designed to offer good results when streaming multimedia over wired networks only. Using these algorithms over wireless networks often leads to results in medium/poor multimedia quality range as they do not include any end-user perceived quality assessment in the adaptation process.

One of the authors has already proposed the Quality Oriented Adaptation Scheme (QOAS), an adaptive streaming solution initially developed for broadband wired networks [6], but which also obtained very good results when it was used to deliver multimedia streams over wireless networks [7]. QOAS considers an estimation of end-user perceived quality as an active factor in the adaptation.

The area of network-oriented learning, known also as e-learning, is growing rapidly. Considering this interest and the diversity of the users, e-learning databases are becoming very rich in multimedia content. Some solutions were developed to adapt the content to user preferences and network conditions in the same time, but only static content adaptation was successful to date [8, 17]. Dynamic adaptation of multimedia content is the next step as web-based learning is using more video resources in order to provide higher user quality of experience levels.

In the past few years, a variety of solutions have been proposed for streaming scalable multimedia content over wireless networks [9] or wireless ad-hoc networks [10]. Among these, the adaptive algorithms that operate at the level of layers [10] or objects [11], fine-granular scalability schemes [12] and perception-based approaches [13] are most suitable for further research considering the results obtained. However none of these algorithms considers in conjunction user learning goals and interests as well as network delivery conditions.

System Architecture

Figure 1 presents the architectural framework for the Performance-Aware Multimedia-based Adaptive Hypermedia (PAMAH), which enhances 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 personalised educational content to e-learners via heterogeneous networks. Its goal is to optimise users' QoE and their learning outcomes by automaticaly adapting content and navigational support based on both user interests and knowledge levels and current network delivery conditions.

The block-level architecture of the proposed system is

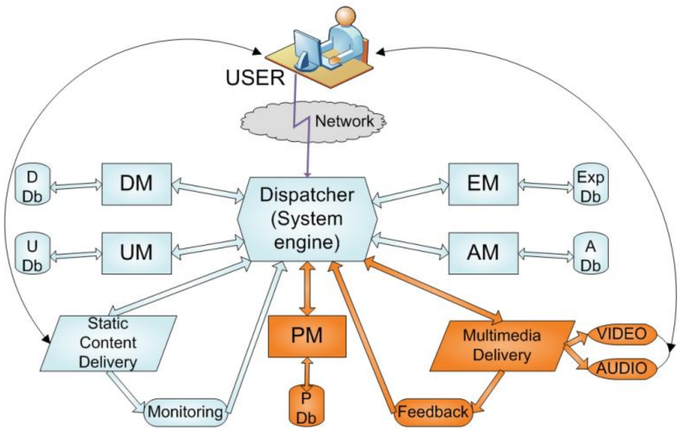


Figure 57 - Block-level system architecture

presented in Figure 1. This paper focuses on the highlighted parts, which are in charge with the rich-media content delivery.

This architectural framework for the PAMAH adaptive web-based system maintains five models: Domain Model (DM), User Model (UM), Experience Model (EM), Adaptation Model (AM) and Performance Model (PM). Every 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).

All these models are controlled 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 by the delivery mechanisms, both static content delivery and multimedia delivery. Except the parameters and rules offered to Multimedia Delivery module by the Dispatcher, there are other network-related parameters that are considered when rich media-content fragments are transmitted. 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.

In this paper the adaptive multimedia delivery is discussed in more details. Multimedia Delivery block implements the D-QOAS adaptation algorithm presented in the next section, considering the input from the Dispatcher. The Feedback Module creates a report that summarizes the delivery conditions and the user feedback for the delivered stream. This report is sent 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. Taking into account the reports from the rest of the modules, the dispatcher builds a list of rules that will be sent to Multimedia Delivery block, as a new input for the adaptation algorithm. As the e-learning process itself is highly dynamic, D-QOAS algorithm is designed to keep up with the continuously changing conditions and parameters involved in the e-learning process to improve end-user quality of experience by adjusting dynamically its adaptation policy.

D-QOAS Algorithm

Overview

D-QOAS is developed to increase end-user quality of experience while also enabling higher number of simultaneously connected clients to communicate. In the system architecture presented in Figure 1, the Adaptation Model performs real-time bandwidth estimation for every user involved in the e-learning process. Based on this bandwidth estimation and also on the additional information from other models, the System Engine decides if a video stream is suited or not for a certain user. If the proper conditions for a streaming session are satisfied then a specific list of rules is built by the Dispatcher and forwarded to the N-Level Builder Module. Here, considering user requirements, the user-specific streaming parameters are computed and the session starts.

D-QOAS Principle

D-QOAS enhances QOAS [8] by adding different user QoE expectation levels in the adaptation process. Consequently, D-QOAS will perform a differentiate treatment of the users, adaptive process based not only on network conditions, but also on their requirements in terms of their QoE. Unlike QOAS, which uses static potential bitrate adaptive levels, D-QOAS dynamically adapts them to suit the delivery process.

Figure 2 illustrates the major blocks of the D-QOAS algorithm for improving user quality of experience during mobile and wireless e-learning, deployed on the PAMAH architectural structure described in section III.

On the client side there is a Decoding & Playing Module whose role is to decode and afterwards play the adapted video stream received from the server. Connected to this module is a

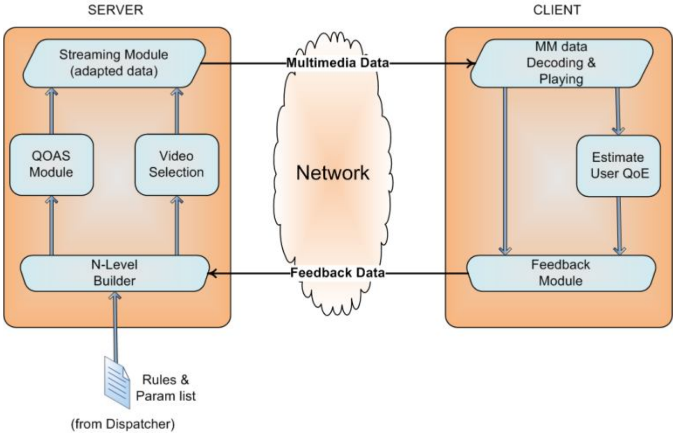


Figure 58 - Dynamic QOAS adaptation algorithm

dedicated module that performs estimation about the end-user perceived quality. The Feedback Module is monitoring delivery-interest parameters such as loss rate, delay and jitter and also assesses the quality of delivery, sending short reports to the server. There reports contain network-related information and user-specific parameters used on the server side by the Dispatcher (represented in Figure 1) and by the N-Level Builder Module. Based on this report the Dispatcher will update the new list of rules and parameters specific for every user and will also update the relevant system databases (A Db and U Db). The list of rules and parameters consists of user-related information (the QoE expectation levels for every user) and session-specific information (which users will be considered for a streaming session, video file to be selected for the current session, video metadata, etc).

On the server side, three important phases are described:

4. *Initial Level Building*

The initial level building is invoked when a new user in the system is requesting a multimedia stream. Because for this new user the Estimation Module has not established yet the QoE expectation level, the level building is done statically directly by the QOAS Module. This new user will appear in the list of rules and parameters with his QoE expectations level set to 0. When N-Level Builder Module is analyzing the list and discovers the new user, it will forward a level request to QOAS Module for building the static levels just for this user.

For a new user $\boldsymbol{U_k}$ the QOAS algorithm builds M potential levels of quality: $\boldsymbol{L_1}$ $\boldsymbol{L_2}$ ... $\boldsymbol{L_j}$ ... $\boldsymbol{L_M}$. $\boldsymbol{L_1}$ represents a low bitrate while $\boldsymbol{L_M}$ represents the maximum bitrate (the video bitrate).

During the session, the system estimates the QoE expectation level for this user and the quality levels will be updated dynamically. After QoE estimation, one of the inputs in the rules list corresponding to this user indicates a video-quality level which represents the acceptable quality expectation level associated to this user. Considering this level, D-QOAS algorithm is building another set of levels, this time having new limits, other than the static levels: between the user QoE expected level and the maximum video level, $\boldsymbol{L_M}$. Like this, a user oriented dynamic adaptation is performed. The procedure for this is presented below.

5. Updating Dynamically the Quality Levels

The procedure for dynamically update the levels for a user involved in a streaming session can be triggered by different causes: changes in the network parameters, users detaching/attaching to the network, updates of QoE expectation level for certain users, etc. The decision for dynamic update of levels is taken by the N-Level Builder based on the information received from the Dispatcher (the list of rules and parameters) and from the Feedback Module. The user QoE expected level found in the rules and parameters is unique for every user and represents the minimum video bitrate that will be transmitted for that user.

For user U_k, after the QoE estimation is performed on the client side, the expectation level is decided to be M_k. This level is situated in the interval defined by the minimum and maximum static levels, L_1 and L_M. Once M_k is known, 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 for dynamically update the levels.

While a new user has a fixed number of levels, M, 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 number of users. By using the user-specific levels of quality the algorithm is able to control better the video quality adjustment. This in performed in order to increase end-user perceived quality since it was demonstrated that viewers tend to prefer a controlled reduction in the quality of the streamed multimedia content than random losses [14].

6. Adaptation Mechanism

Considering P users involved in the e-learning session: ***U1 U2 U3 … Uk … UP-1 UP***, for every user there is a QoE expectation level specifying the expected quality by that user: ***M1 M2 M3 … Mk … MP-1 MP***.

For any user ***Uk*** the algorithm builds the dynamic levels by dividing the amount of bandwidth between ***Mk*** and ***LM*** into intervals. The division depends on the network QoS parameters which are considered in real-time. After this step, the QOAS Module is tuned on these new adaptation levels of quality obtained for user ***Uk***. The Streaming Module is responsible for streaming the selected video file according to the user-specific levels. Video selection is performed by the Video Selection Module, based on the information presented in the list of rules and parameters.

Rate control adaptation process is applied by the N-Level Builder Module whenever media-rich content fragments transmitted needs to adapt their bit rate to match the network parameters and user requirements. Delivery of these fragments usually requires significant network resources and lasts over a longer period of time in which delivery conditions can vary. In this case, random losses have a greater impact on the end-user perceived quality than a controlled reduction in quality.

In conclusion D-QOAS dynamically varies the quantity of information transmitted 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 any improvements in network conditions are detected, the N-Level Builder Model will increase step-by-step the bitrate, according to the user-specific levels, improving therefore user-perceived quality. On the other hand, if at a certain point the network conditions are degrading (increase in the background traffic, new e-learning users) then D-QOAS will rebuild the dynamic levels for every user. The new levels will have a smaller granularity and as a consequence the maximum level considered might decrease.

Test results

Simulation Models and Setup

The proposed algorithm is tested using Network Simulator with the NOAH (No Ad-Hoc) patch installed [15]. NOAH implements the direct wireless routing between base stations and mobile devices. The simulation models used are Dynamic QOAS, QOAS and a non-adaptive solution. Test scenarios will be deployed using the simulation setup, which involves a wired-cum-wireless network, presented in Figure 3.

Node 0 is considered to be the server, while nodes 3, 4, 5, 6 and 7 are the users. The server will stream multimedia traffic to five mobile users, who can attach/detach to the network at the same time or randomly. IEEE 802.11b WLAN covering nodes 2, 3, 4, 5, 6 and 7 and a wired LAN with a bottleneck link 0-1 are deployed. The bottleneck link was fixed to a value slightly greater than $\sum(\boldsymbol{Mk})$ (the sum of minimum QoE expectation levels for every user). In our case, considering the five users, the link bandwidth was set up at 3.8Mbps. The delay for this link is set to 2 ms.

For every user involved in the adaptation process some minimum quality-levels for the video stream were defined, as seen in Table 1.

The Dynamic QOAS algorithm implemented for testing uses specific adaptation rules for every user, considering user QoE expectation levels described in Table I and the network conditions. Depending on the number of users connected to the network and the actual network conditions these rules can change during the transmission of a multimedia stream so that the system can serve as many users as possible while still keeping the quality above their acceptance limit.

QOAS algorithm used here is deployed with five user QoE expectation levels, the lowest level being equal to the smallest quality level from the considered users (in our case the lowest level is 0.3 Mbps, the expected QoE level for U1). The highest level is the actual bitrate of the original video used (1.5 Mbps).

For the non-adaptive streaming three scenarios were considered: every client will try to obtain the maximum bitrate (1.5 Mbps), the average bitrate (0.9 Mbps) or the lowest bitrate (0.3 Mbps).

The multimedia clip used has an average/low motion level with a resolution of 840x480 and a frame rate of 23.97fps.

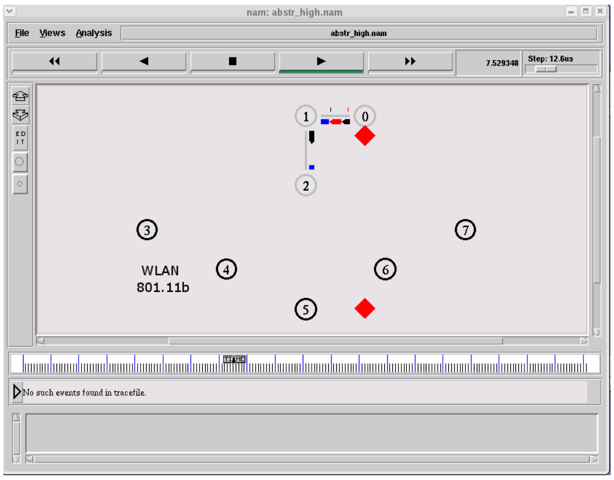


Figure 59 - Test topology - NS graphical interface

TABLE 22

USER-SPECIFIC THRESHOLDS FOR VIDEO QUALITY

User	*Minimum expected quality level*
U_1	0.3 Mbps
U_2	0.6 Mbps
U_3	1.0 Mbps
U_4	0.8 Mbps
U_5	0.45 Mbps

Simulation Scenarios and Performance Assessment

A number of simulation scenarios including multiple users and loaded network conditions determined by background traffic are run. Maximum number of users considered here is five. This number of clients was chosen because it enabled all considered streaming approaches to offer the clients a multimedia stream around half the original encoding movie bitrate in the current simulation conditions. Performance is assessed in terms of end-user perceived quality, average link utilisation and average loss for the three streaming methods. Using a quality assessment tool developed within the

Performance Engineering Laboratory at Dublin City University, Ireland, the estimated user perceived quality of the streaming was measured using VQM and PSNR metrics.

Results and Analysis

Figure 4 presents the throughput of the QOAS algorithm when 5 users are requesting a multimedia stream on different periods of time. It can be observed that even if the adaptation process improves quality of experience for all users involved in the streaming process compared to other solutions as also reported in [7] and [16], the algorithm is balanced and offers almost the same average stream quality for every user regardless of their characteristics. As some of the users involved in the E-learning process might have a higher QoE expectation level, their quality of experience will be affected negatively.

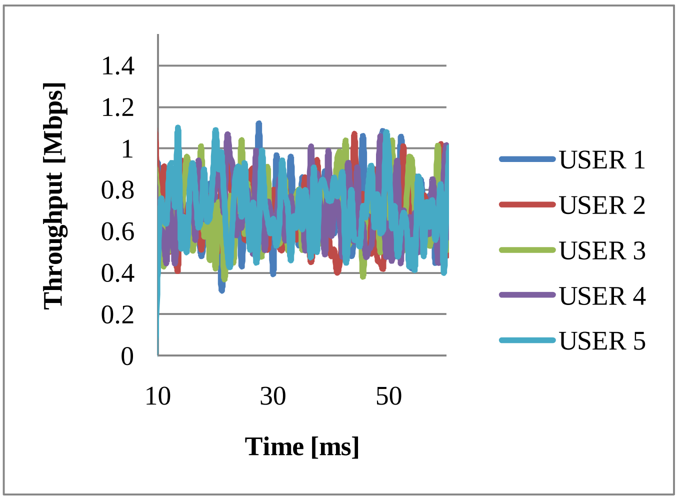


Figure 60 - Streaming session throughput with QOAS for 5 connected homogenous users

The same simulation conditions are used when Dynamic QOAS is deployed. The results are presented in Figure 5. It can be noticed that the user QoE expectation level represents a threshold for the video streaming throughput. This way the overall quality of experience for every user increases.

For the non-adaptive method (NoAd) we have considered the three scenarios specified and the results show that if the transmission rate is set to the maximum rate, only two users out of five can be served at the quality of experience expected by these users. If the quality is set to medium then the number of potential users increases, but still there will be an inefficient bandwidth use.

For the non-adaptive method (NoAd) we have considered the three scenarios specified and the results show that if the transmission rate is set to the maximum rate, only two users out of five can be served at the quality of experience expected by these users. If the quality is set to medium then the number of potential users increases, but still there will be an inefficient bandwidth use.

Table 2 presents the average throughput for every user receiving the multimedia stream considering all three simulation models. It can be observed that for NoAd method only 4 users can be served, as the loss rate is too high for the fifth. Also, the non-adaptive method provides an acceptable level of quality of experience for 3 users only in this setup. QOAS algorithm adapts the stream so that every user receives approximately the same amount of information. Being very efficient in terms of link utilization, QOAS provides the level of QoE expected to three out of five users only.

The proposed algorithm – D-QOAS – performs the dynamic adaptation according to user QoE expectation level and the current network conditions obtaining very good results. Using D-QOAS all five users are situated above their QoE expectation level, which means an important improvement in the overall quality of experience for every user.

In terms of link utilization, because of the dynamic adaptation performed, the proposed algorithm obtains very good results (over 96%). For the considered simulation testbed presented here, the link utilization obtained using D-QOAS was 99.8%.

TABLE 23

AVERAGE THROUGHPUT PER USER

	***U_1* [Mbps]**	***U_2* [Mbps]**	***U_3* [Mbps]**	***U_4* [Mbps]**	***U_5* [Mbps]**
DQOAS	0.431	0.786	1.176	0.879	0.573
QOAS	0.766	0.726	0.712	0.697	0.710
NoAd 1	0.900	0.900	0.900	0.900	0.0
NoAd 2	1.500	1.500	0.0	0.0	0.0
NoAd 3	0.300	0.300	0.300	0.300	0.300

The link utilization was calculated between second 9.5 and second 61, when all five users were connected to the streaming server.

Quality assessment was performed in terms of PSNR and VQM for all multimedia delivery methods discussed. The results confirm that D-QOAS offers a user quality above the acceptable user QoE expectation level and in the limits of the “fair” PSNR range, outperforming the other schemes considered. `

The results show that the Dynamic QOAS performance in terms of end-user perceived quality of experience is improved in comparison with QOAS and non-adaptive delivery.

Conclusions and further work

This paper proposes a new adaptive multimedia delivery algorithm which can be used in the context of e-learning. The Dynamic Quality Oriented Adaptation Scheme (D-QOAS) adapts the multimedia content sent based on both user preferences and network conditions, while adjusting dynamically its adaptation policy during delivery. Testing results show that compared with QOAS and a non-adaptive algorithm, the end user perceived quality has improved in terms of expected user quality of experience. Also there is an increase in the total number of simultaneous clients served as well as an increase in the link utilisation.

Further work implies refining D-QOAS as well as building a practical prototype for this system as subjective tests are needed to validate the preliminary results obtained by simulation. Studying different types of loss that can occur in a wired-cum-wireless network and fine tuning the adaptation to these is also envisaged.

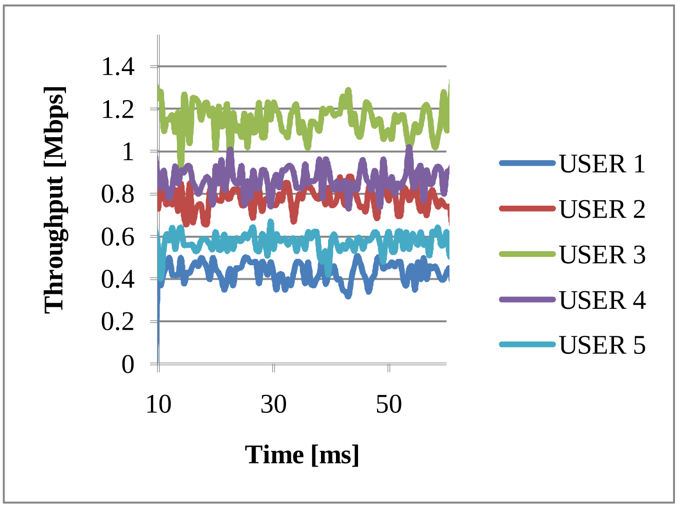


Figure 61 - Streaming session throughput with D-QOAS for 5 heterogeneous connected users

REFERENCES

[1] H. Liao, H. Lu, "The role of experience and innovation characteristics in the adoption and continued use of E-learning websites," in *Computers & Education*, vol. 51,Issue 4, 2008, pp. 1405–1416.

[2] G. Flood, "Make it specific," in *Human resources*, pp. 64–66, 2006.

[3] Enterprise Ireland, "Irish learning solutions," 2009 Available: http://www.enterprise-ireland.com

[4] J. Padhye, J. Kurose, D. Towsley and R. Koodli, "A model based TCP friendly rate control protocol," in *Proc. NOSSDAV*, 1999.

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

[6] G.-M. Muntean, P. Perry and L. Murphy, "Subjective Assessment of the Quality-Oriented Adaptive Scheme", in *IEEE Trans. on Broadcasting*, vol. 51, no. 3, Sept. 2005.

[7] G.-M. Muntean and N. Cranley, "Resource Efficient Quality-Oriented Wireless Broadcasting of Adaptive Multimedia Content," in *IEEE Trans.on Broadcasting*, vol. 53, no. 1, Mar. 2007.

[8] C. H. Muntean and G.-M. Muntean, "Open corpus architecture for personalised ubiquitous e-learning," in *ACM Personal and Ubiquitous Computing Journal*, 2007.

[9] 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, Apr. 2004.

[10] S. H. Shah, K. Chen, and K. Nahrstedt, “Dynamic bandwidth management for single-hop ad hoc wireless networks,” in *Proc. IEEE Int’l. Conf. Pervasive Comp. and Commun.*, Dallas-Fort Worth, USA, Mar. 2003.

[11] 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.

[12] M. van der Schaar and H. Radha, “Adaptive motion-compensation finegranular-scalability (AMC-FGS) for wireless video,” I*EEE Trans. Circuits Syst. Video Technol*, vol. 12, no. 6, pp. 360–371, Jun. 2002.

[13] N. Cranley, L. Murphy, and P. Perry, “Optimum adaptation trajectories for streamed multimedia,” in *ACM Multimedia Sys. J.*, New York: Springer-Verlag, 2005.

[14] G. Ghinea and J. Thomas, “QoS impact on user perception and understanding of multimedia video clips,” in *Proc. ACM Multimedia*, Bristol, United Kingdom, 1998, pp. 49–54.

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

[16] G.-M. Muntean, “Efficient Delivery of Multimedia Streams Over Broadband Networks using QOAS,” in *IEEE Trans.on Broadcasting*, vol. 52, no. 2, June 2006.

[17] C. H. Muntean, J. McManis, ”End-user Quality of Experience oriented adaptive e-learning system”, *Journal of Digital Information, Special Issue on Adaptive Hypermedia*, 7 (1), 2006, http://journals.tdl.org/jodi/issue/view/29.

QoS Parameters Mapping for the E-learning Traffic Mix in LTE Networks

V.H. Muntean*, M. Otesteanu* and G.-M. Muntean**
* "Politehnica" University Timisoara/Faculty of Electronics and Telecommunications, Timisoara, Romania
** Dublin City University/Department of Wireless Communications, Dublin, Ireland
*{horia.muntean, marius.otesteanu}@etc.upt.ro,**munteang@eeng.dcu.ie

***Abstract*—Next Generation Networks (NGN), like Worldwide Interoperability for Microwave Access (WiMAX) and Long Term Evolution (LTE) are expected to become the "anywhere and anytime" access networks to high speed wireless communications. This paper proposes a method of mapping the LTE QoS parameters in order to improve the quality of experience of the end-user when an e-learning application that uses Dynamic Quality Oriented Adaptive Scheme (DQOAS) and generates a traffic mix, is deployed over a Long Term Evolution network. Simulation results show that the proposed mapping method offers improved results compared to the normal mapping scheme. The best results are obtained in combination with the proportional fair scheduling algorithm, while round robin and maximum throughput schedulers' satisfaction rate does not go over 50% in our simulation setup.**

Index Terms **— content adaptation, QoS, E-learning, LTE, scheduling algorithms, wireless networks**

Introduction

During the last years, research efforts in the e-learning field were concentrated on developing of new delivery algorithms able to improve the quality of the learning process. This is a pressing problem, considering that an increasing number of users are accessing learning content via wireless networks and using mobile devices. E-learning process involves using multiple applications – web browsing, video and audio streaming, rich voice and ftp background traffic – that are generating a traffic mix. Managing the traffic mix flows is a difficult problem, especially when they are delivered over wireless networks, because wireless technologies are offering limited radio resources and the services in such networks are highly susceptible of being affected by environmental factors, traffic load, number of clients and their mobility pattern. The authors have proposed in [1] an adaptive multimedia delivery algorithm – DQOAS – designed to improve the end users' quality of experience (QoE) during the learning process, when this process is taking place in a IEEE 802.11 wireless LAN environment. The results 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. Taking into account the good results of DQOAS and the user-oriented approach, it is of further interest to analyze the behavior of this algorithm over other wireless networks, like NGNs. In this paper, LTE is the chosen technology as the next wide coverage wireless network.

LTE is an all-IP network standardised by 3rd Generation Partnership Project (3GPP) in Release 8 which uses new multiple access schemes on the air interface. Orthogonal Frequency Division Multiple Access (OFDMA) is used in the downlink and Single Carrier Frequency Division Multiple Access (SC-FDMA) is used in the uplink to fulfil all the ambitious requirements for data rate, spectrum efficiency,

This work was partially supported by the strategic grant POSDRU 2009 project ID 50783 of the Ministry of Labour, Family and Social Protection, Romania, co-financed by the European Social Fund – Investing in People and by Enterprise Ireland, Proof of Concept and Science Foundation Ireland Research Frontiers programs.

latency, and capacity. Another important technique used is Multiple-Input-Multiple-Output (MIMO) that involves using multiple transmitters and receivers to achieve higher bit rates and improved coverage [2].

The paper aims to offer a mapping alternative for LTE QoS parameters in case of an e-learning traffic mix, in order to obtain an optimal end-user QoE when DQOAS algorithm is used. The proposed solution enables DQOAS to update the quality levels (increase/decrease) of the selected multimedia stream based on user preferences, on instantaneous channel conditions and on the resource allocation scheme, with a minimum impact on the second service (web browsing traffic in our case).

The paper is structured as follows. Section II presents the studies done by researchers regarding LTE QoS parameters and scheduling in case of traffic mix connections for both Downlink and Uplink, while Section III describes DQOAS algorithm and the QoS concepts and architecture as defined in LTE technology. Section IV presents the proposed mapping scheme followed by test results in Section V. Conclusions are offered in Section VI.

Scheduling Algorithms for LTE Uplink and Downlink

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 [3] 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) [4] and scheduling for real-time flows [5]. LTE schedulers can also be classified based on their awareness parameter(s) into channel-aware schedulers [6], queue-aware schedulers and queue- and channel-aware schedulers [3].

Regarding the Uplink (UL) 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 [7] and it was proven that semi-persistent scheduling obtains better performances that group scheduling when no group interactions occurs for group scheduling. In [8], 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 [6] 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 (DL) 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 [3], 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 [9] showed that strict prioritization for session initiation protocol (SIP) packets over other packets – voice and data – can lead to better overall performances. References [10] and [11] 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.

LTE QoS Concept and DQOAS Adaptive Algorithm Description

LTE QoS Concept and Architecture Aspects

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 [12]. The use of SC-FDMA in the uplink reduces Peak-to-Average Power Ratio compared to OFDMA, increasing the battery life and the usage time on the User Equipments (UEs). In DL 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 [2].

In LTE, all network services available for users, are considered end-to-end, or from a Terminal Equipment (TE) to another TE. The provided services can be classified according to their QoS level, but finally is the user who can decide if the provided QoS for a certain service is satisfactory or not. Some of the most important general requirements for QoS attributes are stating that they must have an unambiguous meaning and the mapping should provide different levels of QoS by using UMTS specific control mechanisms. The technical specifications for 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 [13].

The 3GPP QoS concept is based on traffic differentiation and prioritization of data flows, using network-initiated bearers in conjunction with simple QoS profiles based on QoS Class Identifiers (QCIs). 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 [13]. As shown in the BS layered architecture depicted in Figure 1, the BS on a specific layer is offering its services to the bearer on the next level, using the services provided by the layer below.

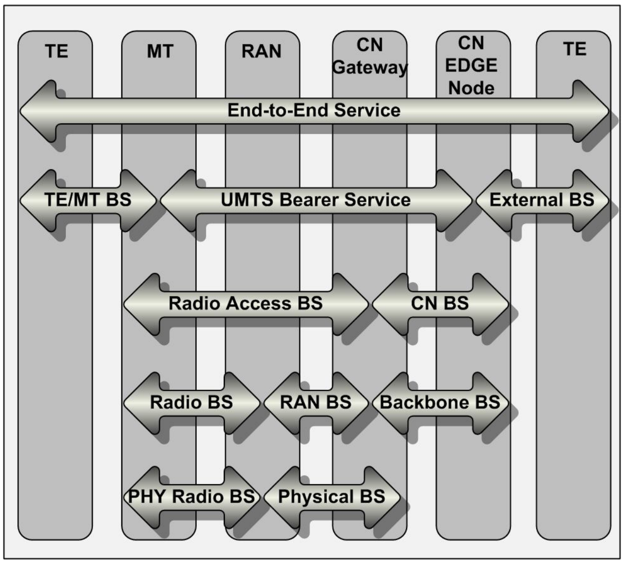


Figure 1 – UMTS QoS Architecture [13]

When two UEs are involved in a data communication, the data flow between them has to pass across different bearer services of the LTE 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 aspects of the radio interface transport and handles the part of the user flow that belongs to one subflow. Radio Access Network (RAN) bearer service takes part in the transport between RAN and CN, together with the Physical BS. Radio bearer service and RAN bearer service are composing the Radio Access bearer.

Unlike the complex QoS mechanisms defined in fixed networks, cellular networks use simple and robust mechanisms, able to offer a good QoS resolution, considering the fact that the air interface has different error characteristics. In LTE, the concept of traffic class was implemented, where a traffic class or a QoS class is defined considering the restrictions and the limitations of the radio interface. Based on the traffic sensitivity to packet delay, there are four classes defined as follows: conversational, streaming, interactive and background class. Conversational class is meant for traffic that has a high sensitivity to delay (e.g. VoIP), while background class deals with traffic that has a low sensitivity to delay (e.g. background download of files). As stated in [13], there is no strict one-to-one mapping between classes of services and the traffic classes defined above. For example, if a service is interactive by nature or if the user has strict requirements about delay, then that service can use the conversational traffic class for obtaining the desired QoS.

DQOAS Algorithm for E-learning Content Delivery

DQOAS algorithm was designed to perform an optimal dynamic multimedia content adaptation in wireless LAN environments, based on the end-users' profile and preferences and on the network conditions, in order to obtain high QoE during the learning process and to increase the total number of simultaneous connected users.

DQOAS extends the QOAS [14] algorithm by adding user QoE expectation levels as parameters in the adaptation process. Like this, the multimedia content will be delivered to users taking into account not only the network conditions but also their quality expectations. Figure 2 presents the block design of DQOAS algorithm. On the client side, the Feedback module monitors the network conditions and registering the delivery-interest parameters (loss rate, delay and jitter) and sends short reports to the N-level Builder, on the server side. Another input for the N-level Builder is the Rules&Param list, which contains user-related information (expected QoE level) and session-specific parameters.

Dynamic level building and adaptation is done on the server side in three steps. First step, Initial level building, is needed whenever a new user requests a multimedia stream. Because this user does not have a QoE expectation level, the levels will be set statically by the QOAS algorithm. As soon as the QoE level is estimated for this new user, DQOAS algorithm will dynamically build a new set of levels, based on the minimum quality accepted level. In the second step, dynamic update of the quality levels is performed. The procedure is triggered by a significant change in network conditions, users attaching/detaching to the network or new minimum quality levels for certain users. The minimum quality level represents the minimum video bitrate accepted by the user. The N-level Builder creates a number of M levels for a user, the lowest level being the minimum accepted quality. The maximum level can be as high as the original video bitrate, if network conditions are favorable. In the last step, the dynamic delivery takes place, by tuning the QOAS module on the new adaptation levels built for a user. This is a continuous process, as rate control adaptation is performed every time a media-rich fragment

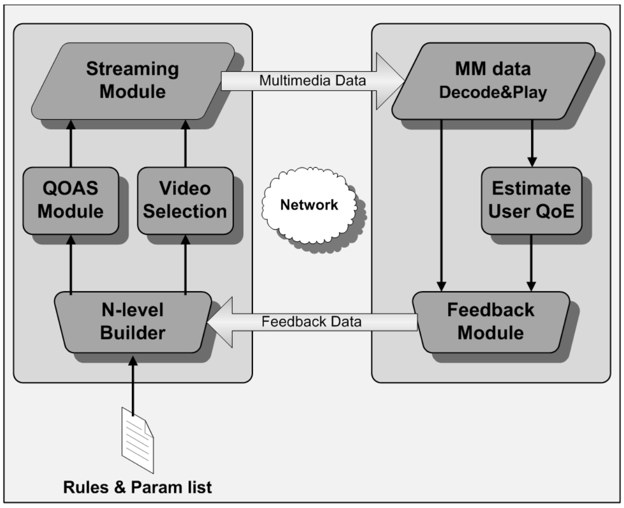


Figure 2 – DQOAS Adaptation Algorithm

needs to adapt the bitrate in order to match the network conditions and user requirements. As the e-learning process itself is highly dynamic, DQOAS algorithm is designed to keep up with the continuously changing conditions and parameters involved, improving end-user quality of experience by adjusting dynamically its adaptation policy [1].

Mapping of QoS parameters

E-learning applications are usually generating more than one traffic type, so the data to be sent over the network to a user can be seen as a traffic mix (e.g. video streaming and web browsing traffic). Based on this fact and on the service classification in LTE by the SPI (Service Priority Information) field, there will be at least two queues corresponding to two different QoS classes, deserving the application. Considering service prioritization, the objects belonging to different queues have different probabilities of being scheduled, depending on the chosen scheduler model. The most common options for a scheduler are:

- Round Robin scheduler (RR), where all users get equal time shares tor transmission, in an order based on their last scheduled time: $M_n = t_n$

- Proportional Fair scheduler (PF), where both the instantaneous channel conditions and the users' past average throughput are considered; it offers the same average 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

- Maximum Throughput scheduler (MT), based only on the instantaneous channel conditions, giving an advantage to users that have the best channel conditions at the given time: $M_n = d_n$

Considering an application generating two services classified in different QoS queues and a RR scheduler, we can write the following equation, describing the *i*-th user satisfaction condition, according to [11]:

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

where 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 $\beta = s_i^{max}/s_i$.

In our case, the maximum ratio of delayed and lost packets, ε, is different from one user to another, being a user-dependent parameter, not a service dependent parameter like it is in case of VoIP. Based on DQOAS description, every user has a minimum accepted quality level for the incoming video stream, representing the dynamically encoded bitrate of the stream, M_i. As this is the minimum accepted level that DQOAS can send for user i, ε as defined above has no significance because any lost or delayed packets will decrease the quality below M_i. To overcome this problem, two solutions are possible. First solution is to add a guard to the minimum expected level, equal with the maximum agreed ratio of delayed and loss packets, ε: $M_i^{new} = M_i + \varepsilon$. Second solution implies conditioning $\varepsilon = 0$ only when DQOAS module is tuned on M_i, because if the quality level is higher, then decreasing this level with the maximum ratio of delayed and loss packets, keeps the video quality above the user satisfaction limit. The first proposed solution is easier to compute, because the changes are done inside the DQOAS module, based on the current network conditions and resource allocation scheme used.

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 application QoE if both flows have the same QoS class (ρ=1). Considering that f_1 and f_2 represent the average packet transmission ratio of two flows generated by the same application, the satisfaction equation for user *i* reads:

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

In the case considered here, an e-learning application that generates video streaming and web-browsing traffic is used. Following the assumptions, the second traffic type (with a lower priority) will have the same QoS class as video streaming traffic – streaming class, as presented in Figure 3.

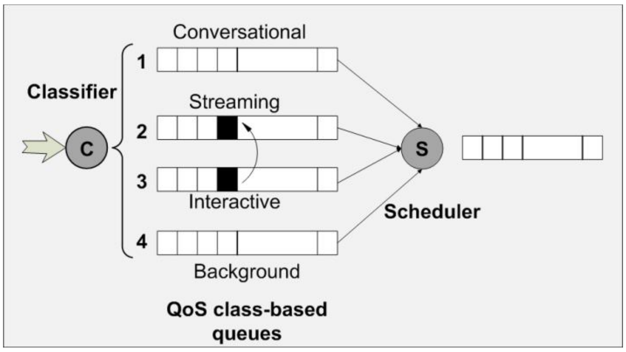


Figure 3 – Changing QoS Class for low priority E-learning traffic flow

The advantage of being in the same QoS class is that the queue specific sorting algorithm will consider both flows with the same priority (the users' priority in the queue). This way, the packets coming from the same application, even on different bearers, will have the same queuing delay, improving the QoE of the application as a whole. In these conditions, DQOAS can update the quality levels for the first multimedia stream based on user preferences, on instantaneous channel conditions and on the resource allocation scheme, with a minimum impact on the second service (web browsing traffic in our case).

Test Results

The proposed solution was tested using the LTE System Level Simulator [15], capable of simulating LTE SISO (Single Input Single Output) and MIMO networks using TxD (Transmission Diversity) or OLSM (Open Loop Spatial Multiplexing) transmit modes.

TABLE I.
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	1000 TTIs
Subcarrier averaging algorithm	EESM
UE speed	5 Km/h

PHY layer model is based on the post-equalization SINR, offering pre-calculated fading parameters and so reducing computational complexity at run-time. In the conducted tests three schedulers were used (round robin, proportional fair and maximum throughput), with two parallel streams for every UE.

A number of 7 eNodeBs with 10 UEs attached on each represents the LTE network used for testing. Table 1 presents the parameters used.

Figure 4 presents the LTE network used for simulation, highlighting the position of UE 4, attached to eNodeB 1. The two streams per user are considered to have the same priority, with the same weight in the schedulers' queue. The throughput of the two streams for UE 4 is presented in figures 5, 6 and 7 when proportional fair, round robin and maximum throughput schedulers are used. For TCP download, a user is considered satisfied if the experienced throughput is at least 300kbps [11], and for the video streaming the user i is satisfied if the stream quality level is kept above M_i^{new}. For user 4, it is observed that only the proportional fair scheduler is providing enough radio resources to be fully satisfied. If round robin scheduler is used, the user is at the satisfaction limit (throughput ~ 0.2 Mbps for both streams). If maximum throughput scheduler is used, user 4 is not able to initiate an e-learning session because of the poor quality conditions due to its relative position to eNodeB 1. Considering the other users, their behavior is presented in Table 2.

TABLE II.

USER SATISFACTION UNDER DIFFERENT SCHEDULING ALGORITHMS

Scheduler used	*Satisfied users*	*Unsatisfied users*	*Maximum Throughput*
Proportional Fair	60%	40%	1.5 Mbps
Round Robin	50% (20% of them are at the limit)	50%	1.8 Mbps
Maximum Throughput	30%	70%	5.5 Mbps

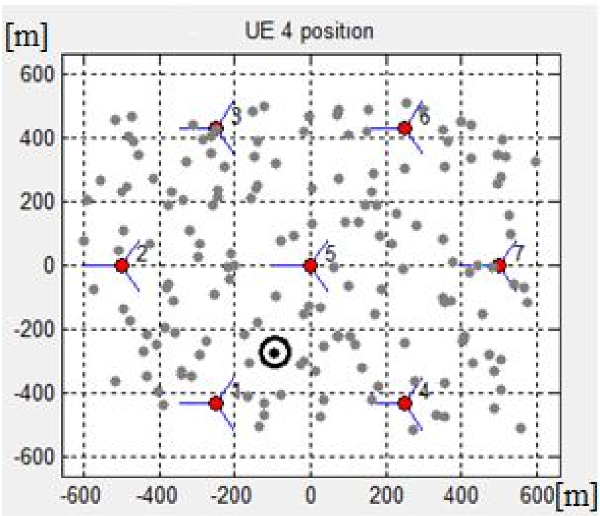


Figure 4 - LTE network map; 7 eNBs with 10 UEs attached to each

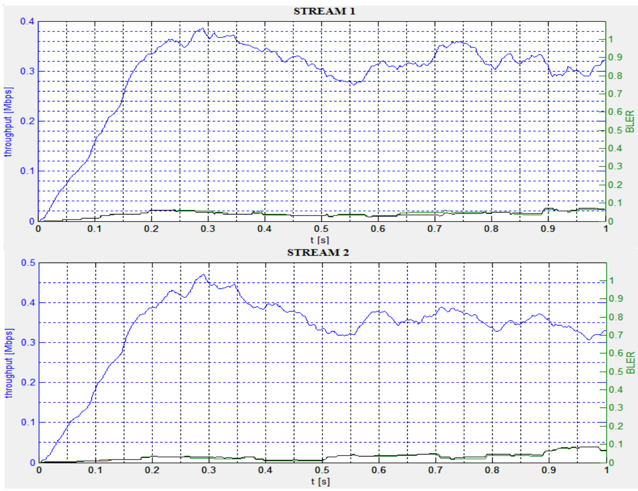


Figure 5 - Throughput and BLER for UE 4 using Proportional Fair scheduler

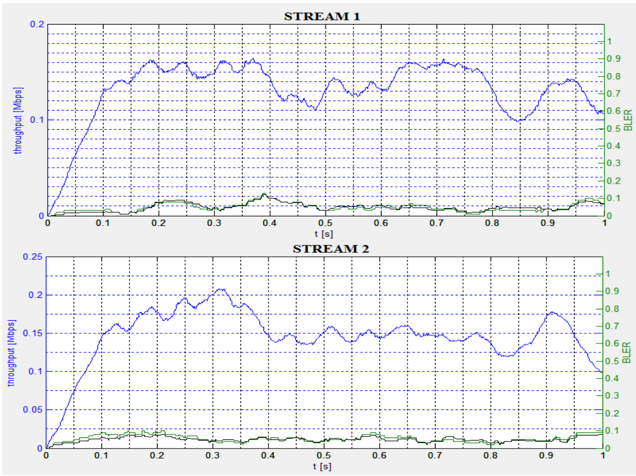


Figure 6 - Throughput and BLER for UE 4 using Round Robin scheduler

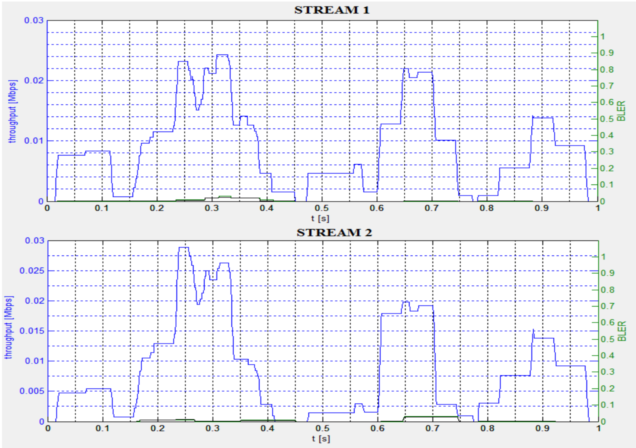


Figure 7 - Throughput and BLER using Maximum Throughput scheduler

It is noticed that in case of a maximum throughput scheduler, the number of users that can be served using the proposed algorithm decreases drastically. Still, the three satisfied users are experiencing very high data rates in comparison with the maximum data rates obtained when PF or RR schedulers were used.

Conclusions and Further Work

This paper proposes a new method of mapping the QoS parameters in order to improve the quality of experience of the end-user when an e-learning application generating a traffic mix is deployed over a Long Term Evolution network. Two different streams coming from the same application were considered for each user in the network, while three scheduling algorithms were applied. The results show that proportional fair algorithm in combination with the proposed mapping scheme offers improved performances compared to the normal mapping scheme used in LTE. If maximum throughput algorithm is used, only 30% of the users are satisfied with the offered quality, but these users are experiencing a very high throughput, as it was expected.

Further work implies extensive testing using different propagation models available with the simulator and also extending the mapping scheme in order to increase the total number of satisfied users. This can be achieved by using different weights in the same queue based on the traffic type and by utilizing other schedulers developed by researchers in the field.

References

[1] V. H. Muntean, G.-M. Muntean, "A Novel Adaptive Multimedia Delivery Algorithm for Increasing User Quality of Experience during Wireless and Mobile Learning" in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, Bilbao, Spain, May 2009.

[2] B. Furht and S. Ahson, Long Term Evolution – 3GPP LTE Radio and Cellular Technology. CRC Press, 2009.

[3] B. Sadiq, R. Madan and A. Sampath, "Downlink Scheduling for Multiclass Traffic in LTE" in *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 510617, 18 pages, July 2009.

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

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

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

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

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

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

[10] J. Puttonen, N. Kolehmainen et co., "Mixed Traffic Packet Scheduling in UTRAN Long Term Evolution Downlink" in *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2008)*, Cannes, France, September 2008.

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

[12] K. Bogineni, R. Ludwig et co. "LTE Part I: Core Network – Guest Editorial" in *IEEE Communications Magazine*, vol. 47, No. 2, pp. 40-43, February 2009.

[13] 3GPP TS 23.107 v9.0.0, "Technical Specification Group Services and System Aspects; Quality of Service (QOS) concept and architecture," Rel. 9, December 2009.

[14] G.-M. Muntean, "Efficient Delivery of Multimedia Streams Over Broadband Networks using QOAS" in *IEEE Trans. on Broadcasting*, vol.52, No.2, pp. 230-235, June 2006.

[15] J. C. Ikuno, M. Wrulich and M. Rupp, "System level simulator of LTE networks" in *Proceedings of IEEE 71st Vehicular Technology Conference*, Taipei, Taiwan, May 2010. Available online at http://www.nt.tuwien.ac.at/about-us/staff/josep-colom-ikuno/lte-system-level-simulator/