UNIVERSITÀ DI PADOVA FACOLTÀ DI INGEGNERIA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

SCUOLA DI DOTTORATO IN INGEGNERIA DELL'INFORMAZIONE

INDIRIZZO IN SCIENZA E TECNOLOGIA DELL'INFORMAZIONE

XXVI Ciclo

Advanced Resource Management Techniques for Next Generation Wireless Networks

Dottorando

MARCO MEZZAVILLA

Supervisore:

Chiar.mo Prof. Michele Zorzi

Direttore della Scuola:

Chiar.mo Prof. Matteo Bertocco

Coordinatore di Indirizzo:

Chiar.mo Prof. Carlo Ferrari

Anno Accademico 2013/2014

To my special family, and my amazing girl.

Contents

List of Tables

List of Figures

Abstract

The increasing penetration of mobile devices in everyday life is posing a broad range of research challenges to meet such a massive data demand. Mobile users seek connectivity "anywhere, at anytime". In addition, killer applications with multimedia contents, like video transmissions, require larger amounts of resources to cope with tight quality constraints. Spectrum scarcity and interference issues represent the key aspects of next generation wireless networks. Consequently, designing proper resource management solutions is critical. To this aim, we first propose a model to better assess the performance of Orthogonal Frequency-Division Multiple Access (OFDMA)-based simulated cellular networks. A link abstraction of the downlink data transmission can provide an accurate performance metric at a low computational cost. Our model combines Mutual Information-based multi-carrier compression metrics with Link-Level performance profiles, thus expressing the dependency of the transmitted data Block Error Rate (BLER) on the SINR values and on the modulation and coding scheme (MCS) being assigned. In addition, we aim at evaluating the impact of Jumboframes transmission in LTE networks, which are packets breaking the 1500-byte legacy value. A comparative evaluation is performed based on diverse network configuration criteria, thus highlighting specific limitations. In particular, we observed rapid buffer saturation under certain circumstances, due to the transmission of oversized packets with scarce radio resources. A novel cross-layer approach is proposed to prevent saturation, and thus tune the transmitted packet size with the instantaneous channel conditions, fed back through standard CQI-based procedures.

Recent advances in wireless networking introduce the concept of resource sharing as one promising way to enhance the performance of radio communications. As the wireless spectrum is a scarce resource, and its usage is often found to be inefficient, it may be meaningful to design solutions where multiple operators join their efforts, so that wireless access takes

place on shared, rather than proprietary to a single operator, frequency bands. In spite of the conceptual simplicity of this idea, the resulting mathematical analysis may be very complex, since it involves analytical representation of multiple wireless channels. Thus, we propose an evaluative tool for spectrum sharing techniques in OFDMA-based wireless networks, where multiple sharing policies can be easily integrated and, consequently, evaluated.

On the other hand, relatively to contention-based broadband wireless access, we target an important issue in mobile ad hoc networks: the intrinsic inefficiency of the standard transmission control protocol (TCP), which presents degraded performance mainly due to mechanisms such as congestion control and avoidance. In fact, TCP was originally designed for wired networks, where packet losses indicate congestion. Conversely, channels in wireless networks might vary rapidly, thus most loss events are due to channel errors or link layer contention. We aim at designing a light-weight cross-layer framework which, differently from many other works in the literature, is based on the cognitive network paradigm. It includes an observation phase, i.e., a training set in which the network parameters are collected; a learning phase, in which the information to be used is extracted from the data; a planning phase, in which we define the strategies to trigger; an acting phase, which corresponds to dynamically applying such strategies during network simulations.

The next generation mobile infrastructure frontier relies on the concept of heterogeneous networks. However, the existence of multiple types of access nodes poses new challenges such as more stringent interference constraints due to node densification and self-deployed access. Here, we propose methods that aim at extending femto cells coverage range by enabling idle User Equipments (UE) to serve as relays. This way, UEs otherwise connected to macro cells can be offloaded to femto cells through UE relays. A joint resource allocation and user association scheme based on the solutions of a convex optimization problem is proposed. Another challenging issue to be addressed in such scenarios is admission control, which is in charge of ensuring that, when a new resource reservation is accepted, previously connected users continue having their QoS guarantees honored. Thus, we consider different approaches to compute the aggregate projected capacity in OFDMA-based networks, and propose the E-Diophantine solution, whose mathematical foundation is provided along with the performance improvements to be expected, both in accuracy and computational terms.

Sommario

L'esplosiva penetrazione di dispositivi mobili nella vita di tutti i giorni pone molteplici sfide nel campo della ricerca nelle comunicazioni 'senza fili', al fine di sostenere la crescente mole di dati generata dagli utenti cellulari, i quali richiedono connettività "in ogni momento, in ogni dove". Inoltre, le applicazioni più richieste, dotate di contenuti multimediali quali le trasmissioni video, richiedono l'utilizzo di molte risorse per sostenere stringenti vincoli di qualità. La limitatezza dello spettro, congiuntamente ai problemi legati all'interferenza, rappresentano i fattori chiave delle reti cellulari di nuova generazione. Conseguentemente, il design di tecniche per la gestione delle risorse risulta essere un aspetto particolarmente critico. A questo fine, proponiamo in primo luogo un modello per valutare le prestazioni simulate delle reti cellulari basate sulla tecnologia Orthogonal Frequency-Division Multiple Access (OFDMA). Un modello di astrazione del canale associato alla trasmissione in downlink di dati fornisce un'accurata metrica valutativa a basso costo computazionale. Il nostro modello combina metriche di compressione multi-portante basate sull'Informazione Mutua con profili prestazionali generati a livello di canale, esprimendo così la dipendenza del rate d'errore associato al blocco di dati trasmesso con i valori di SINR, e l'indice di codifica e modulazione (MCS) assegnato dall'allocatore di risorse. Inoltre, ci proponiamo di valutare l'impatto della trasmissione di Jumboframes in reti LTE, ovvero pacchetti la cui dimensione supera il massimo valore tradizionale di 1500 Bytes. Una valutazione comparativa viene eseguita relativamente a varie configurazioni di rete, in modo da mettere in luce specifiche limitazioni. In particolare, abbiamo potuto osservare una rapida saturazione del buffer di trasmissione legato alla trasmissione di maxi pacchetti attraverso link di bassa qualità. Abbiamo dunque proposto un'architettura di rete 'cross-layer' che ci permetta di prevenire tale esubero di risorse disponibili; si tratta di un approccio che rende possibile la regolazione della dimensione dei pacchetti a seconda della capacità istantanea del canale,

nota attraverso procedure standard basate sulla conoscenza di predefinite sequenze pilota alle quali vengono associate valori di qualità (CQI).

Nella ricerca applicata alle reti wireless è stato recentemente introdotto il concetto di condivisione delle risorse, visto come promettente approccio attraverso cui migliorare le prestazioni delle comunicazioni radio. Lo spettro radio è limitato, e il suo utilizzo risulta spesso inefficiente. Per questi motivi appare significativo proporre soluzioni nelle quali diversi operatori uniscano le proprie forze al fine di fornire accesso wireless a bande condivise piuttosto che proprietarie. Diversamente dalla semplicità concettuale di tale idea, l'analisi matematica che ne deriva può essere molto complessa. Per questo motivo proponiamo uno strumento atto a valutare le prestazioni delle tecniche di condivisone dello spettro nelle reti cellulari basate sulla tecnologia OFDMA, al cui interno è dunque possibile integrare, testare e valutare ogni politica di condivisione.

D'altra parte, relativamente all'accesso a banda larga basato sulla contesa al mezzo, ci soffermiamo su un'importante problematica all'interno delle reti mobili ad hoc WiFi, ovvero l'intrinseca inefficienza del protocollo di trasporto universalmente riconosciuto come standard, il TCP. Quest'ultimo presenta ridotte prestazioni, principalmente legate alle politiche di controllo della congestione. Infatti il TCP è stato originariamente pensato per le reti cablate, dove le perdite di pacchetti indicano una congestione. Diversamente, gli eventi di perdita nelle reti wireless possono essere legati alle variazioni del canale radio, o alla contesa sul collegamento. Intendiamo dunque definire un'architettura 'cross-layer' sufficientemente snella e dinamica, basata sul paradigma delle reti cognitive. Questo framework include una fase di osservazione, i.e., un 'training set' all'interno del quale vengono collezionati svariati parametri di rete; una fase di apprendimento, in cui viene estratta l'informazione da utilizzare per il miglioramento delle prestazioni di rete; una fase di pianificazione, in cui vengono definite le strategie da utilizzare con le informazioni 'imparate'; infine, una fase di azione che rappresenta l'esecuzione 'online' di tali strategie all'interno della rete.

La più recente frontiera per le infrastrutture di rete di prossima generazione si sviluppa intorno al concetto di reti eterogenee. Tuttavia, la presenza di una moltitudine di dispositivi diversi fra loro, in quanto a tecnologia e tecniche di accesso al mezzo, pone nuove sfide. Fra tutte, l'incremento dell'interferenza legato alla densificazione dei nodi e alla pianificazione autonoma. Proponiamo dunque un approccio atto a supportare il reindiriz-

zamento del carico di rete dalle macro celle alle femto celle, attraverso una cooperazione fornita dagli utenti mobili in modalità 'idle', che operano a tutti gli effetti come nodi relay. In questo modo aumentiamo la probabilità che un utente connesso alla macro cella possa alternativamente connettersi ad una femto cella (procedura nota come offload). Abbiamo così definito un modello di ottimizzazione congiunto per l'allocazione delle risorse e la determinazione del collegamento stazione radio base - utente. Un ulteriore tema particolarmente importante riguarda il controllo per l'accettazione di nuovi utenti nel sistema. Tale modello deve garantire il mantenimento dei margini di qualità (QoS) associati ai nodi precedentemente connessi alla rete. A questo fine consideriamo diversi approcci per il calcolo della proiezione di capacità allocata in reti wireless basate sulla tecnologia OFDMA. Infine proponiamo la soluzione 'E-Diophantine' basata sulla teoria diofantina, di cui forniamo le basi matematiche, e mostriamo l'incremento delle prestazioni che ne risulta.

List of Acronyms

3GPP	Third Generation Partnership Project
ABR	Available Bit Rate
ACK	Acknowledgment
AM	Acknowledged Mode
AMC	Adaptive Modulation and Coding
AP	Access Point
API	Application Programming Interface
ARQ	Automatic Repeat reQuest
AS	Access Stratum
AWGN	Additive White Gaussian Noise Vector
BE	Best Effort
BER	Bit Error Rate
BLER	BLock Error Rate
BN	Bayesian Network
CA	Carrier Aggregation
CB	Code Block
CBN	Conditional Bayesian Network

CDF Cumulative Distribution Function

CDMA Code-Division Multiple Access

CDN Content Delivery Network

CESM Capacity ESINR Mapping

CPU Control Process Unit

CQI Channel Quality Indicator

CRC Cyclic Redundancy Check

CSI Channel State Information

CSMA Carrier Sense Multiple Access

CTS Clear-To-Send

DAG Directed Acyclic Graph

DCI Data Control Indicator

DMM Distributed Mobility Management

DSA-REQ Dynamic Service Addition REQuest

E2E End-to-End

ECDF Empirical Cumulative Distribution Function

ECR Effective Coding Rate

EESM Exponential ESINR Mapping

ELFN Explicit Link Failure Notification

EPC Evolved Packet Core

EPS Evolved Packet System

ERT-VR Extended Real Time Variable Rate

ESINR	Effective SINR
EUTRAN	Evolved Universal Telecommunications Radio Access Network
FEC	Forward Error Correction
FI	Framing Information
FP7	Seventh Framework Programme
FTP	File Transfer Protocol
GBR	Guaranteed Bit Rate
GW	GateWay
HARQ	Hybrid Automatic Repeat reQuest
HC	Hill Climbing
IE	Information Element
IMS	IP Multimedia Subsystem
IP	Internet Protocol
IR	Incremental Redundancy
LCM	Least Common Multiple
LESM	Logarithmic ESINR Mapping
LENA	LTE-EPC Network Simulator
LI	Length Indicator
LL	Link Level
LLR	Log Likelihood Ratio
LSM	Link-to-System Mapping
LTE	Long Term Evolution

LTE-A	Long Term Evolution-Advanced
MAC	Medium Access Control
MANET	Mobile Ad hoc NETwork
MCS	Modulation and Coding Scheme
MEDIEVAL	MultiMEDia transport for mobIlE Video AppLications
MI	Mutual Information
MIESM	Mutual Information Effective SINR Mapping
MIMO	Multiple-Input-Multiple-Output
MICS	Media Independent Event Service
MIES	Media Independent Command Service
MIIS	Media Independent Information Service
MIH	Media Independent Handover
MIHF	Media Independent Handover Function
MN	Mobile Node
MPEG	Moving Picture Experts Group
MQAM	M-ary Quadrature Amplitude Modulation
MRTR	Minimum Reserved Traffic Rate
MSTR	Maximum Sustainable Traffic Rate
MTU	Maximum Transmit Unit
NAS	Non Access Stratum
NL	Network Level
ODTONE	Open Dot Twenty ONE

OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OLSR	Optimized Link State Routing
OOP	Object Oriented Programming
PDCCH	Physical Downlink Control Channe
PDN	Packet Data Network
PDU	Packet Data Unit
PHY	Physical
PoA	Point of Attachment
PoS	Point of Service
PP	Process Priority
QCI	QoS Class Identifier
QoE	Quality of Experience
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RB	Resource Block
RLC	Radio Link Control
RRC	Radio Resource Management
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RTS	Request-To-Send
RTO	Retransmission TimeOut

RT-VR Real Time Variable Rate

SAP Service Access Point

SDU Segment Data Unit

SE Spectral Efficiency

SER Symbol Error Rate

SINR Signal to Interference and Noise Ratio

SISO Single-Input-Single-Output

SL System Level

SN Sequence Number

SNR Signal to Noise Ratio

SVC Scalable Video Coding

TBLER Transport BLock Error Rate

TC Topology Control

TCP Transmission Control Protocol

TCPD Tabular Conditional Probability Distribution

TDM Time Division Multiplexing

TSPEC Traffic Specification

TTI Transmission Time Interval

UE User Equipment

UDP User Datagram Protocol

UGS Unsolicited Grant Service

UM Unacknowledged Mode

UMTS	Universal Mobile Telecommunications System
UTRAN	Universal Telecommunications Radio Access Network
VAD	Voice with Activity Detection
VoIP	Voice over IP
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network

Chapter 1

Introduction

The increased usage of mobile phones that connect to a plethora of on-line services and multimedia content is putting mobile operators under pressure. In addition, the different nature of the kind of traffic traversing the wireless medium, shared by thousands of concurrent users, further creates complex scenarios. For these reasons, mobile operators have started to research novel approaches, which aim to tackle these performance requirements under increasingly complex scenarios. A main course of action by standardization bodies, such as the Third Generation Partnership Project (3GPP), has led to the enhancement of the radio access of the cellular networks, leading to packet loss reduction and increased throughput [1]. In this Thesis, we propose a range of advanced resource management solutions to further enhance the radio access, which mainly include 1) cross-layer functionalities, 2) cognitive exploitation of the network dynamics, and 3) mathematical optimization of Medium Access Control (MAC) algorithms. In addition, the increasing number of cellular users connected on-line has motivated the interest in capitalizing on available non-3GPP accesses such as Wireless Local Area Networks (WLAN). Towards that end, different offloading mechanisms, where the mobile operator is able to provide discovery, routing and flow information about both 3GPP and non-3GPP accesses, are being defined and employed [2]. As such, these considerations further pave the way for the definition and experimentation of novel mechanisms, aiming to optimize the user experience of mobile users looking for rich content. The purpose of this introductory chapter is to briefly describe the key technologies and paradigms studied throughout this Thesis, and summarize the main contributions.

1.1 Key Topics

Most of our contributions relate to the 3GPP Long Term Evolution (**LTE**) [3], which is today's most advanced cellular network technology, and is expected to be massively deployed in the upcoming years. It is a very versatile system, featuring many possible configurations able to achieve different performance tradeoffs. Like the Worldwide Interoperability for Microwave Access (**WiMAX**) standard, its non-3GPP counterpart, this technology is based on the OFDMA digital modulation scheme, which results in a coordinated radio access.

With spectral efficiency of point-to-point links approaching its theoretical limits, node densification is a potential solution to meet this demand. While densification of macrocells would result in unacceptably high costs, deployment of small cells (low power nodes) is instead perceived as a cost-effective option [4]. Therefore, **heterogeneous networks** are foreseen as the next generation of cellular network infrastructure [5]. However, the coexistence of multiple types of access nodes, such as macro, pico, relay and femto base stations, brings new challenges because of complex interference conditions from node densification and, in some cases, self-deployed access.

On the other hand, the massive proliferation of mobile devices, along with the ever-increasing data demand, has been pushing interest in the use of multi-hop ad hoc networks for more than a decade. The last-mile multiple hops crossed by each connection before reaching their respective destinations can be seen as a "wireless appendix", which is attached to the worldwide (wired) web. These types of networks are named mobile ad hoc networks (**MANET**), and are characterized by a self-configuring infrastructureless architecture, which can handle communications in a highly dynamic network topology. It can provide connectivity also in areas with no network infrastructure, and can be a valuable and inexpensive solution for many networking problems. For these reasons, a broad set of research contributions has been carried out in this area.

The evolution of cognitive techniques for self-adaptive communication systems can be crucial for next generation wireless technologies, due to both the complexity and high dynamics that characterize these incoming network paradigms. In this Thesis, we propose to exploit i) **spectrum sharing** techniques for the analysis of an inter-operator cooperative approach, and ii) **bayesian inference** for designing a cross-layer framework to optimize the TCP throughput in MANETs.

Further communication issues defining a key role for the optimization of next generation wireless networks broadly relate to MAC strategies. In particular, we focus i) on resource allocation and user association through a **nonlinear optimization** model, and ii) on admission control through a capacity estimator based on the **Diophantine theory**, which deals with indeterminate polynomial equations.

Finally, in the perspective of deploying and managing future wireless networks, operators and equipment vendors are strongly interested in identifying the configurations and solutions that can achieve the best possible performance in a variety of scenarios according to their needs. Ideally, this identification would be carried out experimentally by analyzing measurements obtained from an actually deployed network. However, doing so by either testbed experiments or field trials is a very time consuming and expensive process. Thus, **network simulation** is an appealing alternative to perform some pre-tuning of the selected algorithms and protocols before they are deployed. In this respect, there is a tradeoff between choosing a very accurate simulation model, which typically has a high computational complexity and only allows for the simulation of a few network elements, and a more simplified model, which can scale to larger scenarios. This justifies the number of extensions that we have applied to the network simulator ns-3 [6], which is a worldwide open source simulator used by both academia and industry.

1.2 Organization of the Thesis

This Thesis addresses a number of issues related to radio resource management in next generation wireless networks. We propose i) cost-effective solutions based on alternative technologies and architectures, and ii) advanced techniques to enhance state-of-the-art scenarios. In both cases, we introduce compliant frameworks and novel algorithms.

In Chapter 2, we provide an overview of the proposed architecture to better support mobile video over wireless. In particular, in Section 2.1, the main reference scenario consists of an operator supporting connectivity through heterogeneous access technologies. Our goal is to describe the architectural solutions envisioned to provide enhanced video delivery in the last (wireless) hop, mainly focusing on the existing access techniques. In Section 2.2, in order to achieve more reliable results, we present a link abstraction model for the simulation of downlink data transmission in LTE networks. The purpose of this model is to

provide an accurate link performance metric at a low computational cost by relying solely on the knowledge of the SINR and of the modulation and coding scheme. To this aim, the model combines Mutual Information-based multi-carrier compression metrics with Link-Level performance curves matching, to obtain lookup tables that express the dependency of the Block Error Rate on the SINR values and on the modulation and coding scheme being used. In addition, we propose a 3GPP-compliant Channel Quality Indicator evaluation procedure, based on the proposed Link Abstraction Model, to be used as part of the LTE Adaptive Modulation and Coding mechanisms. Finally, in Section 2.3, we aim at evaluating the impact that packets breaking the 1500-byte legacy value, called Jumboframes, have in LTE networks, by exploiting and extending the network stack in the ns-3 simulator. We first provide an overview about the key features of LTE starting from a physical layer perspective, to logical functions like the adaptive modulation and coding scheme, together with a detailed description of the Radio Link Control (RLC) segmentation capabilities. A comparative evaluation is performed based on diverse network configuration criteria, such as user position, density and mobility. We aim at assessing the benefits and caveats that derive from Jumboframes usage in LTE networks. Moreover, a novel cross-layer approach is proposed to mitigate the effect of rapid buffer saturation, due to the transmission of oversized packets with scarce radio resources. To conclude, we test our framework through the analysis of realistic video traces.

In Chapter 3, we describe the implementation of our proposed cognitive techniques for the optimization of next generation wireless networks. In particular, in Section 3.1, we introduce solutions where multiple operators join their efforts, so that wireless access of their terminals takes place on shared, rather than proprietary to a single operator, frequency bands. In spite of the conceptual simplicity of this idea, the resulting mathematical analysis may be very complex, since it involves analytical representation of multiple wireless channels. In this sense, simulation studies may be extremely useful to grasp a correct performance characterization of wireless networks with shared resources. In this spirit, this section introduces and evaluates an original extension of the well known ns-3 network simulator which focuses on multiple operators of the most up-to-date cellular scenarios, i.e., the Long Term Evolution of UMTS employing OFDMA multiplexing. Spectrum sharing is represented through a proper software architecture where several sharing policies can be framed. A detailed sim-

ulation campaign is run to validate the proposed architecture, both in terms of modeling realism and computational performance. On the other hand, in Section 3.2, we target an important problem in mobile ad hoc networks, namely, the intrinsic inefficiency of the standard transmission control protocol (TCP), which has not been designed to work in these types of networks. After an initial training phase, we predict the mobility status of the network through a probabilistic approach, and we propose a series of ad hoc strategies to counteract the TCP inefficiency based on this prediction. Via simulation, we show the performance improvements in various wireless scenarios, in terms of increased average throughput and decreased length of the outage intervals. The significant performance improvements shown here will be verified in a future work by implementing our approach in a real testbed.

Finally, in Chapter 4, we introduce the mathematical models proposed to enhance the radio access of heterogeneous networks. More specifically, in Section 4.1, we propose a novel approach that aims at extending femtocell coverage by enabling nearby idle User Equipment (UEs) in the proximity of femtocells to serve as relays, called UE-Relays. The users that would nominally connect to macrocells are offloaded to femtocells via the coverage extension provided by the UE-Relays. The problem of determining the optimal subset of idle UEs to be enabled as UE-Relays and user association to these relays is non-trivial because it depends on the projected capacity of the relay access and backhaul. We formulate the user association as an optimization problem that maximizes the network-wide log utility of the user rates. Our simulation results corroborate our expectations: overlaying a femtocell-tier with UE-Relays will increase the user throughput and the network throughput. In addition, we compare our approach with a MAX-SINR based user association, and evaluate the performance profiles at varying scenario settings. The proposed rate-based scheme outperforms the SINR-based cell selection, thus validating the idea that, in heterogeneous cellular networks, user association should be performed on a load-basis rather than relying merely on link quality indicators. In addition, in Section 4.2, we aim at estimating the increase in peak allocated capacity when admitting a new resource reservation in the system. Thus, we consider different approaches and, based on their limitations, propose our solution, which is based on the Diophantine theory. The properties of the designed algorithms are derived through a mathematical analysis and their accuracy and computational load characteristics evaluated in a generic scenario.

Chapter 2

Mobile Video Over Wireless

During the last few years, a joint evolution of demand and offer for rich contents over wireless has taken place. The diffusion of new mobile devices which combine multimedia features and Internet connectivity, yet realized with different access techniques, has reached high penetration rates worldwide. Nowadays, wireless networks include smart phones, netbooks, or other small and portable devices all equipped with multimedia encoding and decoding capabilities. On the other hand, multimedia content exchange is starting to dominate the Internet traffic, with applications for watching on-line videos, exchanging user-generated video content, or video-chatting that impose constraints in terms of latency and data transmission reliability. However, the current Internet and its wireless extensions do not glue these two sides of the evolution well together. Since they were designed having in mind a different paradigm for content exchange, demand and offer of multimedia content are still mismatched in terms of provided Quality of Service (QoS) and Quality of Experience (QoE) for the end users. More in general, the explosion of multimedia content demand actually poses several challenges in terms of efficient *delivery* of this content throughout the network to the end user. In this spirit, our main goal is to tailor future architecture for the wireless Internet in order to support enhanced multimedia support, as described in [7]. Wireless network access and the exchange of multimedia flows over the Internet are becoming more and more pervasive in the everyday life. However, simple technological advances in terms of improved network capacity cannot satisfy the increasing demand of such services, since a paradigm shift from the current Internet architecture is required. We tackle this issue by addressing novel architectural frameworks and viable strategies to efficiently

deliver video services in a wireless Internet context, as part of our contribution to the EU FP7 MEDIEVAL [8]. In particular, we focus on the identification of useful techniques for the considered access technologies (WLAN and LTE) and the general definition of architectural schemes to efficiently support video flows.

This chapter is organized as follows. In Section 2.1, we provide an overview of the proposed architecture for video delivery optimization. In Section 2.2, we describe our link abstraction model for an improved evaluation of the simulated LTE networks. In Section 2.3, we evaluate the impact of transmitting Jumboframes in LTE networks, and propose a cross-layer framework to dynamically tune the size of the transmitted packets based on the radio conditions. Finally, we summarize our contributions in Section 2.4.

2.1 Architecture Overview

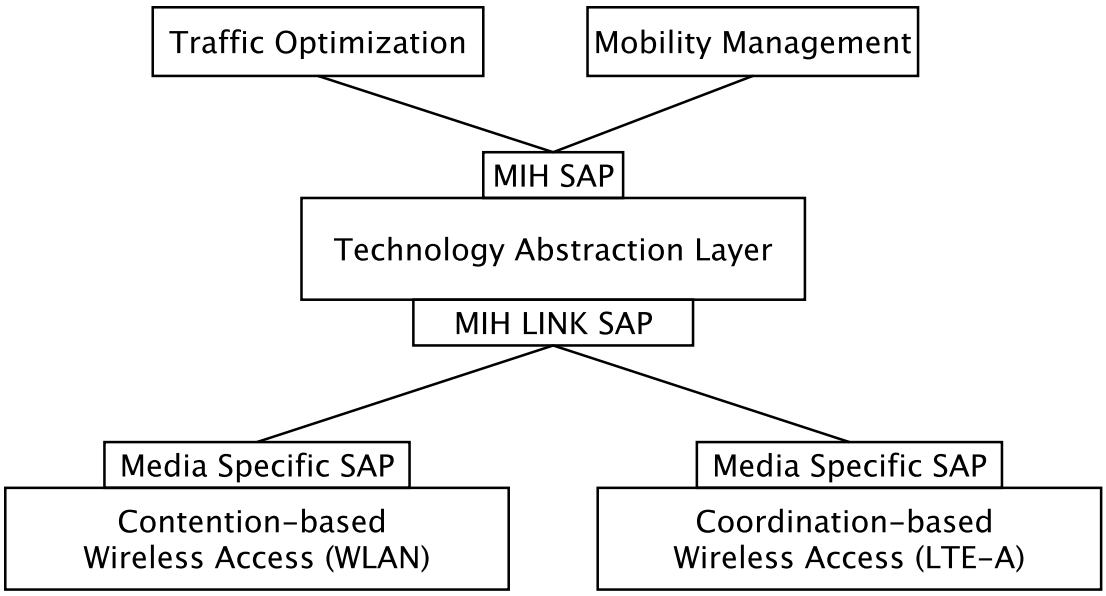


Figure 2.1. *Wireless Access Architecture and Cross-layer Design*

Several works focus on cross-layer optimization techniques adopted to improve multimedia transmission over wireless channels: in [9] an approach is presented that combines the flexibility and programmability of application layer adaptations with low delay and bandwidth efficiency of link layer techniques, while [10] shows the significant improve-

ments that can be obtained by deploying a joint application-layer adaptive packetization and prioritized scheduling along with a MAC-layer retransmission strategy. Our architecture encompasses the cross-layer optimization through enabling mechanisms for the different layers to communicate among them. In this sense, cross-layer signaling is implemented between the lower layers and the video application and services, as well as with mobility services. An example of this cross-layering can be seen as follows: information gathered into a Media Specific SAP (Service Access Point) such as video services characterization (that is, bit rate guarantees, priority, tolerated delay/jitter/packet dropping), together with link quality indications perceived by all the LTE subscribers, are used to generate a utility function that is associated to each user, and will be used to optimally allocate the LTE wireless resources. Another example of cross-layer interaction as listed previously is packet marking. The video services or transport optimization components insert some specific marking in the Internet Protocol (IP) header of each video packet, based on video coding and avoiding deep packet inspection at the wireless access level. According to some pre-defined signaled agreement and on the wireless access type, each packet is processed individually, which results in forwarding the packet on the wireless link, prioritizing it or even dropping it if the available bandwidth on the wireless link is not sufficient.

The rest of this Section is organized as follows. In Section 2.1.1, we discuss a set of optimization techniques for a coordination-based wireless access. In Section 2.1.2, we provide a description of the abstraction layer along with a novel transmission technique represented by the usage of Jumboframes to increase system throughput. In Section 2.1.3, we discuss our proposed extended functionalities of the standard IEEE 802.21 [11].

2.1.1 Enhanced Coordination-based Wireless Access

A first task to be performed at the LTE radio access is to gather channel measurements performed at lower layers and inform the related decision entities. In fact, knowing the load of the cell at the network level, or the quality of the links perceived by the mobile terminal, improves the decision-making algorithms in the upper layers entities. Then, a novel medium access strategy will be applied, taking into account the characteristics of video flows and identifying techniques to interact with upper layers, to design access techniques, which are aware of the video content delivery in a network-wide perspective. In partic-

ular, allocation strategies, proposed at the MAC level, exploit the cross-layer techniques to allocate as many users as the LTE wireless channel can admit, trying not to affect the perceived QoE by the end users. Moreover, equipment capabilities announced at network attachment must be exploited to improve the bandwidth occupation of the video signal flowing in the last wireless hop of the network. Then, based on the information received through the abstract interface, the control plane instructs the user plane entities to select and transmit towards the air interface only those video frames which are adapted to the receiving device capabilities. This operation is accomplished before video frames enter in the wireless access technology protocols, then such information is received when the User Equipment (UE) connects to the network, and can be correlated with some pre-defined coding levels of the video signal to determine the optimal quality level corresponding to the equipment. The frame level of each video packet is marked by the application or service layer in the IP header and is provided to the transmitting entities, which then prioritize or even drop the video packets, thus reducing the bandwidth occupation in the downstream nodes, especially in the last wireless hop. The statistics related to this filtering are reported at the wireless network entry node, e.g., the base station.

Ns-3 network simulator [6] is the simulation tool used for validating the proposed optimization approaches for LTE networks. A basic description of the protocol stack entities can be found in [12], while the whole framework can be downloaded and executed in [13]. We propose an improvement of the LTE module, as reported in Section 2.2.

2.1.2 Abstracted Mechanisms

To achieve efficient video transport in heterogeneous networks, high transparency and seamless intercommunication within the system components is needed, able to operate in both types of wireless technologies. This is accomplished by a set of functions defined in the abstraction layer, together with some ad hoc features designed to further enhance the video flows transfer over the air. In the following, we present a key concept: the abstraction layer, and its mutual interactions with the transport and mobility layers.

Abstraction Layer: The abstraction layer is at the heart of this video delivery optimization process, as it provides the means for the interaction between the radio access network and upper layers in order to accomplish cross layer functionalities and underlying tech-

nologies transparency. This twofold goal is realized by using abstract interfaces defined to establish a functional connection with both upper and lower layers: video transport and mobility management on the one hand, wireless access on the other hand. These interfaces are designed in such a way that they provide a common functionality regardless of the specific underlying technology. In the following we present the different interfaces and functionalities provided by the abstraction layer.

Transport Optimization: This abstract interface provides the upper layers with information and configuration mechanisms regarding the video characteristics available at the lower wireless technologies in an abstract way, translating generic video parameters into video specific extensions defined by the underlying heterogeneous technologies. The objective of this cross-layer interface is to enable an efficient and optimized video transport by providing information about the availability of the wireless accesses and the dynamic variations of the radio channel, and informing about specific Jumboframe functionalities such as aggregating non-related traffic (or traffic from different flows), as well as providing indications necessary for the transport of jumbo frames, such as frame size and queue delay. Then, it reports information about the Mobile Nodes (MNs) capabilities such as the size of the screen, the list of available network interfaces, active links and flows, or the counting of session receivers in the case of multicast. Moreover, it receives information to enhance the configuration of the wireless access: flow requirements (for QoS and multicast mechanism), flow identification, marking criteria, configuration parameters for multicast and jumbo frames procedures. The use of this information at the transport layer allows the operator or video service provider to implement smart optimizations such as: *i)* design evolved cross-layer algorithms and mechanisms between the video services and the network layer and to dynamical optimization of video services with suitable network support, *ii)* develop new procedures and algorithms to better adjust the video streaming taking into account the network dependencies or *iii)* improve cross-layer communication in order to have Forward Error Correction (FEC) mechanisms in a joint coordination between the wireless access module and the service layer controller. Then, we achieve an end-to-end interface through the transport optimization layer and, at the same time, a dynamic cross-layer interface targeting video improvements and content delivery. This will provide support to video coding: FEC/Automatic Repeat reQuest (ARQ) [14], Available Bit Rate (ABR), scheduling, transcoding according to device

capability and QoS availability. Other important features to be addressed are the Process Priority (PP) settings (marked packet for prioritization and selection), the indication of Jumboframes availability, the quality of the link information provisioning, and the support of signal split across multiple channels.

Mobility Management: This functionality interacts with the mobility management mechanisms by reporting abstracted parameters from the wireless accesses and forwarding downwards the received commands relative to interface or flow mobility. The objective of this cross-layer interface is to extend the media-independent signalling functionalities provided by IEEE 802.21 and support extra mobility functionalities and video aware services to optimize the user experience while performing handover. Hence, in order to provide a better mobile user experience within a network, the Media Independent Handover Function (MIHF) will support flow mobility (identification of flows, flow granularity at the primitive level) and new mobility protocols which work as MIH Users and that are not currently contemplated in the standard, such as Distributed Mobility Management (DMM) approaches. One of the most interesting functionalities offered by IEEE 802.21 is its ability to provide dynamic and static information, through the Media Independent Event and Information Services. In order to enable mobility decisions, information about link quality and QoS levels must be provided to the mobility management entities, along with static or semi-static information regarding the video specific features implemented in the network, such as signalling the availability of surrounding access networks, Point of Attachment (PoA) link layer parameters, PoA capabilities (IP multicast, mobility scheme support, Jumboframes support), upper layers capabilities (video services supported, transport optimization mechanisms available), and availability of multicast support in the network.

The above interfaces represent the exchange of information between the higher layers and the media abstraction provided by the abstraction layer. In order to be able to actually work with the real technology, the abstraction layer talks with the underlying technologies through media dependent abstract interfaces, which are in charge of gathering the technology measurements to inform the upper layers transparently and dynamically about the current radio conditions. Although the role of the abstraction layer may appear as a simple translator of media independent events/commands/information, this layer is designed as an intelligent module able to provide also extra-functionality apart from translating between

abstracted and media specific languages. The novel functionality included in the abstraction layer is called the Abstract QoS Mapper. The objective is to provide a comprehensive mechanism for the higher layers to be able to map their requirements in terms of QoS or even QoE to a set of abstract parameters that the lower layers can understand. In this way, the higher layers will provide through the video transport interface, the flow requirements in terms of QoS and multicast. By using this information, the flow requirements are mapped to a certain traffic class or general protocol parameters that will be translated so that the specific technology can be configured accordingly.

Finally, in the following Section, we introduce the concept of Jumboframes, which is another generic technology that is being investigated for video delivery optimization, and will be better discussed in Section 2.3.

Jumboframes: Jumboframes, or jumbograms, are packets with a size larger than 1500 bytes, where 1500 bytes is the normal Ethernet size being used since its creation (around the 80s). The major benefit of using a larger frame size is that, when compared with a lower one, the same amount of information can be sent in fewer packets (i.e., less fragmentation occurs). Each frame, both when sending or receiving, requires CPU processing and has header overhead associated, thus fewer frames will require less CPU processing and generate less overhead, as well as increased throughput. The designed mechanisms will consider a flexible approach and therefore will aim to have a negligible impact in the current or upcoming network developments. Also, the deployment of Jumboframes in this manner requires some important issues to be tackled. One of these issues is the mobility experienced by terminals which can seriously impact and alter conditions such as packet loss, which highly impact the usage of larger frames. Also, considering the number of users associated to the same AP, the usage of greater Maximum Transmission Unit (MTU) sizes can cause fairness issues in multi-user environments. As such, the development of new resource allocation through, for example, channel hopping, is required. Otherwise, jumbo ability can be emulated at higher layers of the protocol stack, impacting only queues at the lower layers. The usage of Jumboframes considers not only the last hop towards mobile users (which is wireless), but also the full path from mobile users to the content provider. This means that different media will be crossed at different points of the network, which need to be aware of Jumboframes. However, the wireless part, due to its inherent operation constraints and changing condi-

Reported to RRC and upper layers	Executed but not reported	To be reported
RSRP	UE, eNodeB: PHY	eNodeB: MAC/RLC
RSRQ	eNodeB: MAC/RLC	UE Capabilities

Table 2.1. *LTE Measurement Parameters*

tions of the wireless medium, provides the most challenging part for the development of performance-increase mechanisms. As such, the usage of Jumboframes is being addressed in the wireless environment, i.e., the last hop, between MN and the PoA. In order to simplify things, evaluation will contemplate the PoA acting as a Point of Service (PoS), in order not to mix wireless Jumboframes with wired Jumboframes.

2.1.3 IEEE 802.21

The IEEE 802.21 Media Independent Handover (MIH) [15] is a standard that aims to facilitate and optimize handover procedures between different access technologies. It adds an abstraction layer (in the form of the Media Independent Handover Function (MIHF)) that abstracts the different link technologies to high-level entities, here deemed MIH-Users. This abstraction is achieved through the provision of a set of services: the Media Independent Event Service (MIES), which allows MIH-Users to receive events about link conditions, the Media Independent Command Service (MICS) which enables MIH-Users to exercise control over the links and the Media Independent Information Service (MIIS) which provides network information that can be used for optimal handover candidate selection. These services can be accessed remotely between MIH-enabled entities via the MIH Protocol, enabling the network and Mobile Terminals to exchange media independent information with which network management and handover control can be optimized.

Here, the role of IEEE 802.21 is to provide media independent interfaces to the decision entities in the architecture, towards control of the network access links, while facilitating handover procedures. The events, command and information defined in the protocol provide the common interaction fabric with which the different aspects (video services, wireless technologies, mobility and traffic optimization) can interact, in order to provide an enhanced

RRM	MAC/PHY	RRM → MAC/PHY
RLC Mode	Spectrum Info	E2E Bearer Specs
HARQ	Transmission Mode	Buffer Size
Buffer Size	Transceiver Specs	

Table 2.2. *LTE Configuration Parameters*

mobile video experience in wireless environments.

The IEEE 802.21 standard defines reference models for IEEE (i.e., 802.3, 802.11 and 802.16), 3GPP and 3GPP2 technologies. For these technologies it provides a direct mapping for the link events, commands and parameters that are available for MIH-Users to interact with links. However, later developments of cellular technologies, such as 3GPP LTE, provide significant differences which were not considered by 802.21. As such, we are developing a mapping between 802.21 parameters and LTE parameters [16] with the objective to replace the existing mapping between 802.21 and UMTS parameters. For a detailed explanation, see [15]. As an example, in Tables 2.1 and 2.2 a list of measurement and configuration parameters is provided in order to define the services to be handled in the LTE Media Specific SAP to perform an efficient protocol extension; in fact, dynamic channel information such as Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ) are reported to higher layers, whereas UE physical capabilities and system resource usage attributes are exchanged in the lower layers for radio access purposes. Then, such information might be integrated into standardized MIES to extend the protocol capabilities, in the same way as parameters listed in Table 2.2 can be used to configure the air interface, or Access Stratum, based on information received from the upper layers (MICS).

2.2 Link Abstraction Model

Our goal is to provide an accurate and, at the same time, computationally lightweight Link Abstraction Model (LAM) for the simulated LTE networks. This model shall allow the accurate prediction of transport block errors at the MAC layer taking into account channel fluctuations, multi-user interference as well as physical layer configurations (bandwidth assignment, modulation, coding, etc.). Thus, we integrate this model into the ns-3 Network-

Level (NL) simulator of LTE, which accounts for multiple UEs and eNodeBs, and allows the simulation of a complete end-to-end LTE system, including architectural components. To this end, one might of course come up with a detailed implementation of the eUTRAN procedures [17] and especially of the LTE PHY (e.g., modeling its operations at the symbol level). However, this would lead to a very complex and computationally demanding model, which would not scale up to the medium and large network sizes typically considered for NL simulation. A more suitable approach, which is the one that we take here, is instead that of performing some offline pre-processing based on Link-Level (LL) simulations, so as to derive a simplified model of the influence of channel and system parameters on the PHY performance, represented by the Transport Block error rate. This pre-encoded mapping allows to retain a good amount of the accuracy of LL simulation when modeling phenomena such as multi-user interference, OFDMA bandwidth allocation and random channel realizations, while not retaining their complexity, thereby allowing for better scalability. The main contribution of this work, which is discussed in the following Sections, is the design and implementation of a lightweight link abstraction model for the downlink transmission of LTE systems. In addition, we show how this model can be exploited to design an algorithm for reporting channel quality indicator (CQI) feedback according to the 3GPP guidelines in order to test the online selection of the Modulation and Coding Scheme (MCS) for each user, subject to given BLock Error Rate (BLER) requirements.

2.2.1 Related Work

Evaluating the error distribution in OFDMA-based wireless systems is very challenging for a number of reasons. First of all, OFDM transmissions are typically used in scenarios affected by frequency selective fading, meaning that subcarriers may perceive very different channel gains. Besides, OFDMA further increases the system complexity as subcarriers are assigned to different users, whose signals are typically generated with different transmission powers and MCSs. This makes the task of predicting the error distribution per user rather complex, in terms of both collecting a reduced subset of parameters to describe performance trends, and generating a flexible error model in order to cover all possible scenarios. To address this problem, Link-to-System Mapping (LSM) has been previously proposed for use with generic multicarrier systems [18–20]. The first practical application of the LSM

approach to modeling a real wireless technology was the evaluation of the Worldwide Interoperability for Microwave Access (WiMAX) radio technology by the IEEE 802.16 task force; in this context, several LSM techniques were applied and evaluated [21]. On this matter, two extensions of the well known network simulator 2 (ns-2) [22] called WINSE [23] and WiDe [24] had these solutions integrated; however, their code is not publicly available.

More recently, LSM has been extensively investigated for application to the LTE technology. Many papers have leveraged on LSM as part of simulation models aimed at the evaluation of interference management and allocation schemes [25, 26]. However, only a few of them [27, 28] made the simulation tool publicly available. [27] refers to a set of Matlab simulators that aim at providing a comprehensive framework for the simulation of link and MAC layer performance. The design choices of Matlab and the focus on lower layer aspects do not give to this tool the possibility of evaluating complex network scenarios, featuring mobility and traffic constraints. Some of these assumptions have been relaxed in [28], where c++ was adopted as the programming language and some networking functionalities were included. However, this simulator is designed primarily to evaluate MAC-level performance and does not properly model the Evolved Packet Core (EPC), in charge of handling, among other aspects, bearers, their Quality of Service (QoS), and mobility.

Recently, a new module called LTE-EPC Network Simulator (LENA) [29] has been developed for LTE as an extension of the ns-3 simulator [6]. LENA already includes EPC functionalities [30] and is designed in a product oriented fashion (i.e., it implements the Scheduling APIs defined by the Small Cell Forum [31], formerly known as Femto Forum). It is to be noted that LENA has all the advantages of a large open source project, including the support of a lively community for debugging, validation and maintenance.

2.2.2 Simulator Overview

The simulation platform used to test and validate our contributions is ns-3, an open source discrete-event network simulator for Internet-based systems, available online at [6]. Our work extends the LTE module currently under development within the project LENA [29], which comprises architectural LTE-compliant features that refer to both the Evolved Packet Core (EPC), presented in [30], and the Evolved Universal Telecommunications Radio

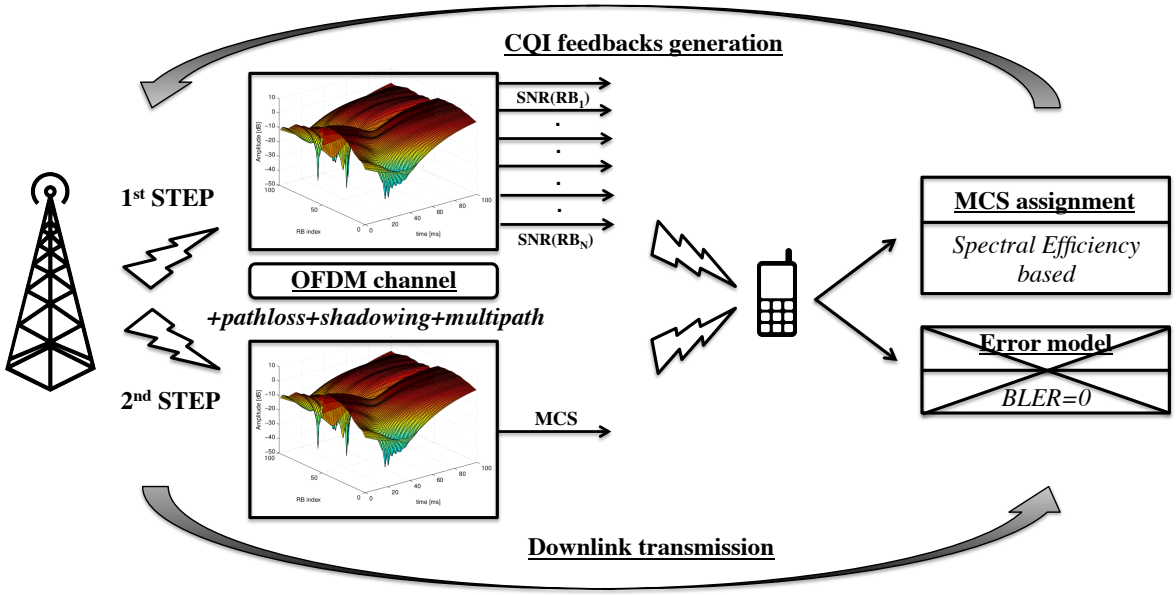


Figure 2.2. *ns-3 transmission diagram*

Access Network (EUTRAN), detailed in [32], based on the first LTE framework configuration for ns-3 [12]. The overall framework is extensively documented in [33].

In Fig. 2.2 is depicted a simple flow diagram that represents the LTE transmission procedure in ns-3. Most of this components will be described in the next paragraphs, thus we conceptually focus now on the importance of the error model. It is to be noted that the actual ns-3 LTE implementation lacks of a prediction model, which means that once the UE receives a packet because it is in the eNodeB transmission range, and got a resource in its allocation scheme, it will always decode the incoming data correctly (BLER = 0). Our goal is to design a *smart* coin to be tossed when the users receive a packet, in order to model the reception side by taking into account also decoding capabilities described in the link level performance curves. In the following, we provide a brief description of the key features of the ns-3 module for LTE networks.

Spectrum Model: The spectrum framework adopted in ns-3 was firstly proposed in [34]. The frequency model designed by Baldo et al represented a fundamental building block for the realization of the LTE radio spectrum described in [35]. *fc* denotes the LTE absolute radio frequency channel number, which identifies the carrier frequency on a 100 KHz raster, whereas *B* is the transmission bandwidth configuration expressed in number of Resource Blocks (RB), as shown in Table 2.3. The LTE frame is composed by 10 subframes, for a total

Bandwidth (MHz)	Number of RBs
1.4	6
3	15
5	25
10	50
15	75
20	100

Table 2.3. *Bandwidth configuration*

Scenario	UE speed (kmph)
Pedestrian	0, 3
Vehicular	30, 60
Urban	0, 3, 30, 60

Table 2.4. *3GPP propagation scenarios*

duration of 10 milliseconds. As shown in Fig. 2.3, each subframe can be seen as a *time* vs *frequency* grid.

Channel Model: Each LTE device can be assigned with all the channel propagation loss models defined by the *ns-community*. As to fading phenomena, the propagation defined in Annex B.2 of [36] have been used to generate channel traces for three different scenarios, as shown in Table 2.4.

Interference Model: The PHY model is based on the well-known Gaussian interference channel, according to which interfering signals' powers, in linear unit, are summed up together to determine the overall interference power. The resulting SINR expression for a given subcarrier is given by

$$SINR = \frac{|h_0|^2 P_{t,0}}{\sum_{i=1}^{N_I} |h_i|^2 P_{t,i} + \sigma_0^2}, \tag{2.1}$$

where N_I is the number of interferers, σ_0^2 is the variance of the thermal noise, whereas $|h_i|^2$ and $P_{t,i}$ represent the channel gain and the transmission power, respectively.

CQI feedbacks: Prior to transmit, each eNodeB broadcasts a signaling pilot sequence. All the UEs within its coverage area decode it, and generate a list of Channel Quality Indicators (CQI) on a per sub-channel basis, in order to provide the eNodeB with an overall quality information snapshot. CQI feedbacks, whose guidelines are described in [17], represent an indication of the data rate which can be supported by the channel for allocation and scheduling purposes, as shown in Table 2.5. In LENA, the generation of CQI feedbacks is done accordingly to what specified in [31].

Scheduling & Resource Allocation: The scheduler is in charge of generating specific structures called Data Control Indication (DCI), which are then transmitted by the eNodeB

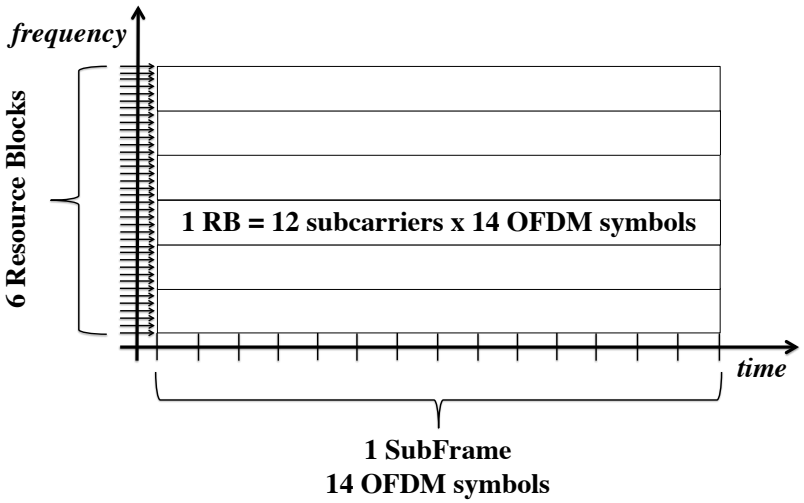


Figure 2.3. *LTE subframe structure (BW = 1.4 MHz)*

through the Physical Downlink Control Channel (PDCCH) to the connected UEs, in order to provide them with a resource allocation map on a per subframe basis. This control messages contain information such as the MCS to be used, the MAC Transport Block (TB) size, and the allocation bitmap which identifies which RBs will carry the data transmitted by the eNodeB to each user.

2.2.3 Proposed Error Model

As previously mentioned, we aim at introducing a prediction error scheme to be used at the reception side. In Fig. 2.4 is depicted the contribution applied to the actual implementation, which is twofold: on the one hand, is the MCS assignment definition based on our error model, while on the other hand is the proposed prediction error framework.

For an accurate evaluation of the user's performance, besides the MCS assigned by the LTE scheduler, it is important to track the residual errors, i.e., after link layer processing, that are due to channel phenomena such as fading, multiple-access interference, etc. However, as discussed above, a comprehensive simulation of link layer procedures would entail a high computational complexity, which is undesirable for multi-user scenarios. With BLER we refer to the residual error rate after all PHY layer procedures, i.e., affecting the code blocks at the output of the turbo decoder at the receiver side. The physical layer model of the LTE simulator returns SINR values $SINR_1, SINR_2, \ldots, SINR_N$ for each Resource

MCS	Modulation	Spectral Efficiency	ECR
1	QPSK	0.15	0.08
2	QPSK	0.19	0.1
3	QPSK	0.23	0.11
4	QPSK	0.31	0.15
5	QPSK	0.38	0.19
6	QPSK	0.49	0.24
7	QPSK	0.6	0.3
8	QPSK	0.74	0.37
9	QPSK	0.88	0.44
10	QPSK	1.03	0.51
11	16QAM	1.18	0.3
12	16QAM	1.33	0.33
13	16QAM	1.48	0.37
14	16QAM	1.7	0.42
15	16QAM	1.91	0.48
16	16QAM	2.16	0.54
17	16QAM	2.41	0.6
18	64QAM	2.57	0.43
19	64QAM	2.73	0.45
20	64QAM	3.03	0.5
21	64QAM	3.32	0.55
22	64QAM	3.61	0.6
23	64QAM	3.9	0.65
24	64QAM	4.21	0.7
25	64QAM	4.52	0.75
26	64QAM	4.82	0.8
27	64QAM	5.12	0.85
28	64QAM	5.33	0.89
29	64QAM	5.55	0.92

Table 2.5. *LTE MCS*

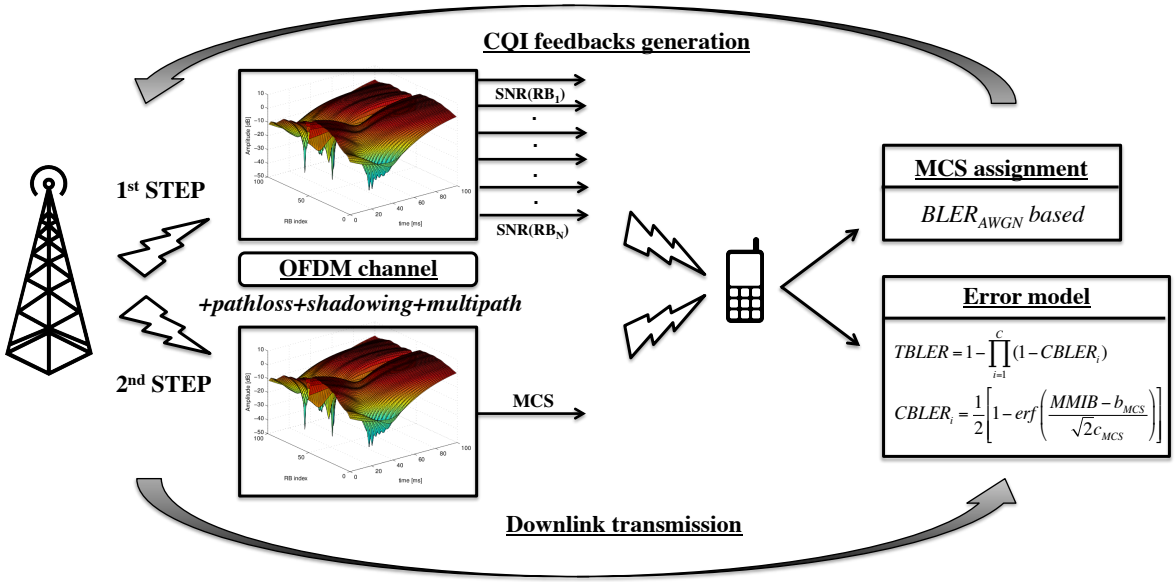


Figure 2.4. *Improved ns-3 transmission diagram*

Block (RB) $n \in \{1, 2, \ldots N\}$ for all users, calculated using an AWGN model and a Gaussian interference model. We recall that a RB corresponds to the allocation quantum in LTE, and is composed of 12 sub-carriers (15 kHz each) and 14 OFDM symbols, transmitted over a Time Transmission Interval (TTI) of 1 ms. Following the Mutual Information Effective SINR Mapping (MIESM) method [37], for the transmission of each block we pick the instantaneous SINR vector $(SINR_1, SINR_2, \ldots, SINR_N)$ and map it onto a Mean Mutual Information per coded Bit (MMIB) metric. The obtained MMIB is a time-varying compressed representation of the channel quality as perceived by any given user at any given time. In addition, we store offline calculated curves returning the BLER as a function of the SINR for each valid (MCS, CB_{size}) pair, where MCS is a modulation and coding scheme and CB_{size} represents the code block size. These curves have been obtained with the Vienna link level simulator [38, 39].

Finally, this offline calculated SINR to BLER mapping is utilized, together with the instantaneous MMIB information, to obtain the BLER traces for each user. This procedure is explained in greater detail in the following.

2.2.3.1 Effective SINR Mapping Model

The link level simulations executed to build our abstraction model assume a frequency flat channel response at any given SINR (the so-called AWGN channel). Let us consider a

given LTE user and let $SINR_n$ be the instantaneous SINR value associated with RB n, where $n = 1, 2, \ldots, N$ and N is the number of RBs allotted to this user. Given this, let us assume that the simulator returns an instantaneous SINR sample for each RB, which means a vector $(SINR_1, SINR_2, \ldots, SINR_N)$. In order to obtain a lightweight and effective mapping from this vector to a single BLER metric we consider the *effective SINR mapping* (ESM) method, see [40]. Briefly, the instantaneous SINR vector is mapped onto a single scalar value as follows [20]:

$$eSINR = \alpha_1 I^{-1}\left(\frac{1}{N}\sum_{n=1}^{N} I\left(\frac{SINR_n}{\alpha_2}\right)\right), \tag{2.2}$$

where $I(\cdot)$ represents the *information measure* function, $I^{-1}(\cdot)$ is its inverse, whereas α_1 and α_2 are two scaling parameters that are tuned as a function of the selected MCS. These parameters will be defined in the next Section, along with function $I(\cdot)$. As shown in Table 2.6, several approaches have been proposed in the literature.

Effective SINR Mapping	**Information Measure**
Capacity (CESM) [41]	$I(x) = \log_2(1 + x)$
Logarithmic (LESM) [42]	$I(x) = \log_{10}(x)$
Exponential (EESM) [43]	$I(x) = e^{-x}$
Mutual Information (MIESM) [44]	$I(x) = MI(x)$

Table 2.6. *Effective SINR extraction techniques*

The results provided in [37] demonstrate that the MIESM method outperforms all the other mapping approaches in terms of approximation accuracy for the BLER curves. Thus, we adopted the Mutual Information (MI) metric for our implementation.

2.2.3.2 MIB Mapping

The MIB is defined in [40] as the mutual information between the bit input belonging to a specific constellation (MCS), and the corresponding log-likelihood ratio (LLR) output at the receiver.

As reported in [40], it can be approximated through the following function:[1]

$$J(t) = \begin{cases} 0, & t < 0.001 \\ a_1 t^3 + b_1 t^2 + c_1 t, & 0.001 \leq t < 1.6363 \\ 1 - e^{(a_2 t^3 + b_2 t^2 + c_2 t + d_2)}, & 1.6363 \leq t \leq 50 \\ 1, & t > 50 \end{cases}, \tag{2.3}$$

where the parameters have been obtained through numerical fitting and are reported in the following Table 2.7.

$a_1 = -0.04210661$	$a_2 = 0.00181492$
$b_1 = 0.209252$	$b_2 = -0.142675$
$c_1 = -0.00640081$	$c_2 = -0.0822054$
–	$d_2 = 0.0549608$

Table 2.7. *J-function approximation parameters*

Specifically, it has been demonstrated [40] that the MIB of any modulation m can be approximated as a mixture of $J(\cdot)$ functions as follows:

$$I_m(x) \simeq \sum_{k=1}^{K} \alpha_k J(\beta_k \sqrt{x}) \tag{2.4}$$

where $\sum_{k=1}^{K} \alpha_k = 1$ for some $K \geq 1$ and the argument x is the SINR associated with the transmission channel under study. Numerical fittings have been carried out (see again [40]) to obtain K, α_k and β_k for QPSK, 16-QAM and 64-QAM, as reported in the following Table 2.8.

[1]Note that, compared to [40], we have truncated the function to avoid that it takes values outside the [0,1] interval.

Modulation m	**MIB function** $I_m(x)$
QPSK	$J(2\sqrt{x})$
16-QAM	$\frac{1}{2}J(0.8\sqrt{x}) + \frac{1}{4}J(2.17\sqrt{x}) + \frac{1}{4}J(0.965\sqrt{x})$
64-QAM	$\frac{1}{3}J(1.47\sqrt{x}) + \frac{1}{3}J(0.529\sqrt{x}) + \frac{1}{3}J(0.366\sqrt{x})$

Table 2.8. *Numerical approximations for MIB mapping*

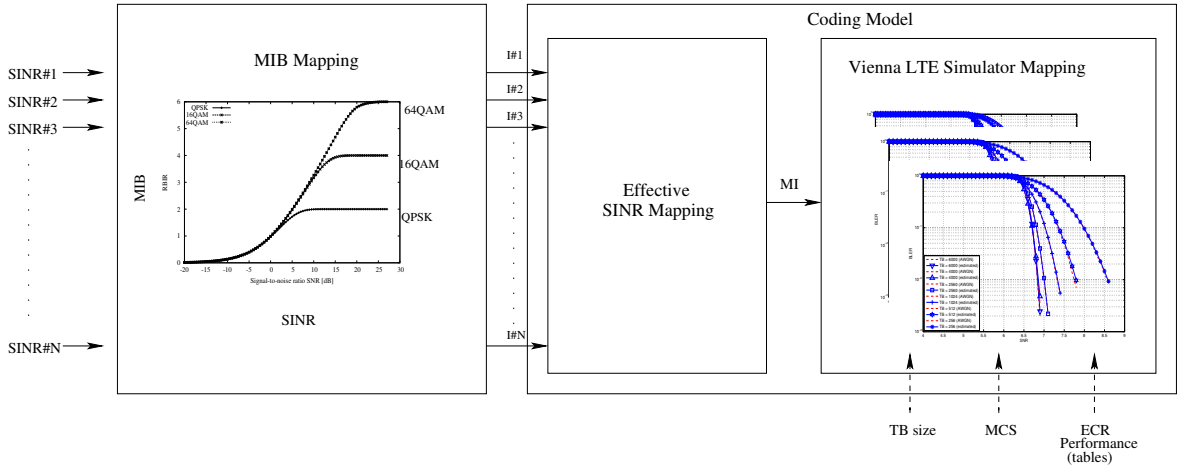


Figure 2.5. *MIESM computational procedure diagram*

As above, with $SINR_n$ we mean the instantaneous SINR value associated with RB n, where $n = 1, 2, \ldots, N$ and N is the number of RBs allotted to the user. According to the above discussion, the function $I_m(x)$ can be used to map $SINR_n$ onto the corresponding mutual information domain, where m is the adopted modulation scheme. Note that the argument x corresponds to $SINR_n$ and in LTE, for each sub-frame, the same modulation is picked for all RBs. Given all that, the Mean Mutual Information per coded Bit (MMIB) can be obtained as follows:

$$MMIB = \frac{1}{N} \sum_{n=1}^{N} I_m(SINR_n)\,, \tag{2.5}$$

where N is the number of RBs assigned to a specific user and m is the modulation that this user is exploiting. To sum up, the model starts by evaluating the mutual information value for each RB from the corresponding SINR samples. Subsequently, the MMIB is computed by averaging (effective SINR mapping) the corresponding mutual information values as per Eq. (2.5). The implemented scheme is depicted in Figure 2.5: the model starts by evaluating the MI value for each RB, represented by the SINR samples, then is computed the equivalent MI per TB basis by averaging the corresponding MI values.

2.2.3.3 BLER Prediction

The data at the MAC layer of the LTE protocol stack (right above the LTE PHY) is arranged in Transport Blocks (TB), whose size depends on the specific configuration of the underlying PHY. TBs are split into a number of CBs which are independently encoded by

the turbo encoder at the PHY layer. Each CB is then encoded and transmitted over the channel exploiting the N RBs allotted to the user. Here we show how to efficiently compute the Transport BLock Error Rate (TBLER) based on the results presented in the previous Section.

For the moment, let us focus on the i-th CB of a given TB. As mentioned in Section 2.2.3, link-level simulations (whose results were obtained using the Vienna LL simulator) have been used to obtain the PHY layer performance in terms of BLER vs SINR over AWGN channels, accounting for the configuration of the PHY layer turbo encoder in terms of Code Block (CB) length and selected MCS. The 3GPP standard has been considered to assess the correct CB sizes in the simulations, according to [17]. As an example, the dotted lines in Fig. 2.6 show the BLER as a function of SINR for MCS 1. These curves have been calculated offline considering the LTE PHY layer procedures implemented in the LTE Downlink LL Vienna Simulator [38], as described in [45]. As can be seen from these plots, the CB size highly impacts the actual BLER performance for a given MCS. The selected CB i is transmitted over the channel using the N RBs that are assigned to the user. At the receiver side, a reference SINR value[2] is made available by the PHY layer model of the ns-3 simulator for each of these RBs, returning the SINR vector $(SINR_1, SINR_2, \ldots, SINR_N)$, as discussed in Section 2.2.3. From here, we obtain the MMIB metric using Eq. (2.5), which corresponds to an equivalent SINR for the transmission of CB i over the allotted RBs. As a last step to obtain the residual error rate of CB i, we need to map its MMIB onto the corresponding BLER, which is referred to here as $CBLER_i$. This is done according to the following procedure.

In order to reduce the computational burden at simulation time as much as possible, an approximation based on the Gaussian cumulative model has been adopted. According to this, the estimated BLER curves as a function of MMIB are parameterized as follows:

$$CBLER_i(x) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{x - b_{S,M}}{\sqrt{2}c_{S,M}}\right)\right], \tag{2.6}$$

where $b_{S,M}$ and $c_{S,M}$ are the mean and the standard deviation of the Gaussian cumulative distribution, respectively, and x is the MMIB associated with CB i. S is the code block size and M is the MCS, which dictates the actual transmission rate. What we did at this point, was to find suitable pairs $(b_{S,M}, c_{S,M})$ for each MCS and code size. We did so through numerical fitting so that the curves from Eq. (2.6) would match those obtained from the Vienna LL simulator. The result of this procedure is shown in Fig. 2.6, where the solid

[2]We assume no frequency selectivity among the 12 sub-carriers composing the resource block.

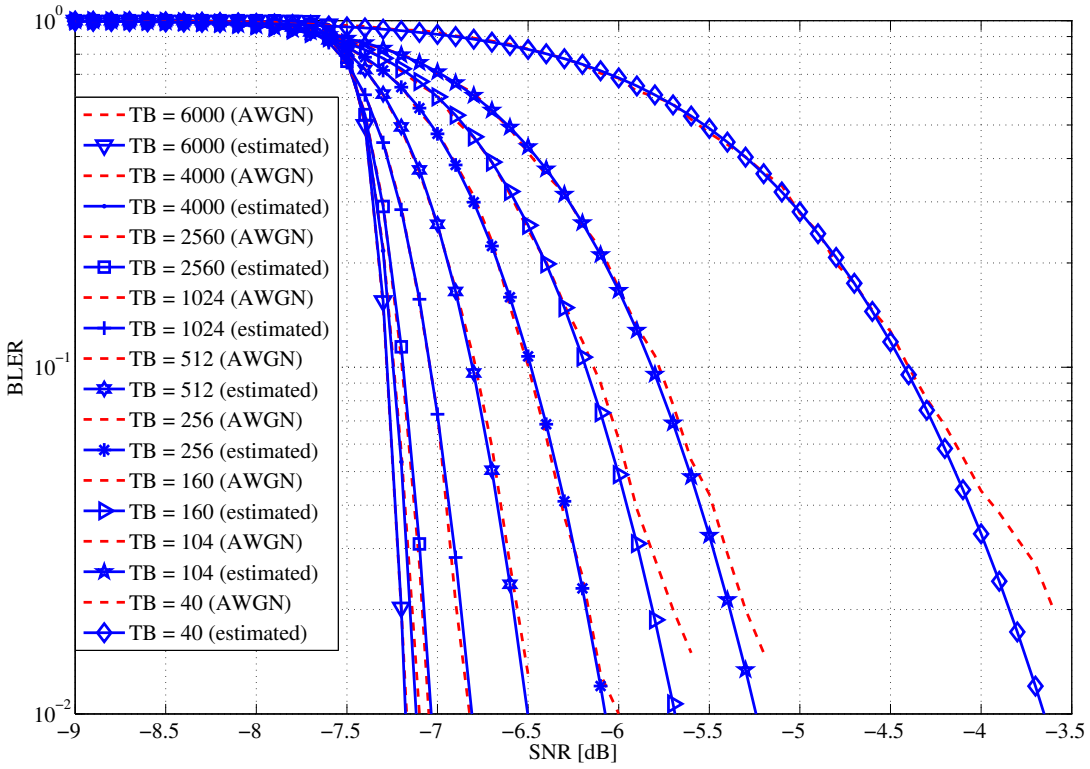


Figure 2.6. *BLER vs SNR for MCS 1*

curves represent the result of Eq. (2.6) where we have used the best fitting $(b_{S,M}, c_{S,M})$ pair for each MCS and code size. As can be seen from these plots, the approximated BLER from Eq. (2.6) (solid lines) closely match the BLER obtained through the numerical simulation of LTE PHY procedures (dotted lines). The overall Transport BLock Error Rate (TBLER) is thus found as:

$$TBLER = 1 - \prod_{i=1}^{C} (1 - CBLER_i)\,, \tag{2.7}$$

where C is the number of CBs contained in the TB.

Lookup tables: To limit the computational complexity and the memory space taken by the proposed link abstraction model, we only considered a subset of CB sizes, i.e., $\mathcal{S} = \{40, 104, 160, 256, 512, 1024, 2560, 4000, 6000\}$ bits. This choice is aligned with the typical performance of turbo codes, where large CB sizes do not strongly affect BLER performance. However, we note that for CB sizes smaller than 1000 bits, the BLER performance might significantly differ as we vary the block size (up to nearly 3 dB). Therefore, we accounted

for an unbalanced quantization of CB sizes in order to get more accuracy in the critical zone (small code blocks). This is particularly evident from Fig. 2.6 that shows a similar BLER profile for large CB sizes (e.g., 2500, 4000 and 6000 bits), whereas the performance gap increases as the CB size gets smaller. Thus, $(b_{S,M}, c_{S,M})$ parameters have been tabulated for all valid combinations of MCS and block sizes in set $\mathcal{S}$. We remark that high MCS values with high order modulations and efficient coding rate schemes, such as 64-QAM with an Effective Coding Rate (ECR) of 0.92 (i.e., MCS 29), allow for a minimum CB size of 2560 bits. The latter is much larger than the minimum size at small MCS values, e.g., MCS 1, where the minimum size is 40 bits, see Fig. 2.6. This reflects the fact that turbo coding offers better performance as the code block size increases; thus, for high order modulations such as MCS 29, small code block lengths are inefficient as the resulting BLER performance is unacceptable.

2.2.4 Link Adaptation Improvement

In this Section, we propose an improved MCS assignment scheme supported by a new CQI evaluation scheme based on 3GPP guidelines. Note that this novel algorithm for CQI evaluation could not be tested in the previous ns-3 distribution as it is based on residual error estimates. Link adaptation plays a fundamental role in modern wireless communications systems, which need to face issues such as strong interference from multiple users and their mobility, which makes the wireless channel frequency selective. These facts are coped with by LTE adaptive modulation and coding algorithms. Focusing on the downlink scenario, AMC has the role of tracking the perceived SINR and sending back to the base station (eNodeB) a so called CQI report. Hence, periodically, the UE reports to the eNodeB a single CQI value for all the RBs (the so called *wideband* CQI). This information is a "compressed" representation of the quality experienced by the UE in a specific sub-frame and is used at the base station side for the selection of the MCS. This process is continuously executed so as to adapt to channel and network dynamics.

Our proposed MCS assignment scheme relies on an SINR to CQI mapping approach based on the link error abstraction model presented in the previous Section. As a competing approach we consider the algorithm that is currently implemented in the LENA ns-3 simulator, which is inspired by the *spectral efficiency* concept, see also [46].

Spectral efficiency-based approach: consider the generic RB n, and let $SINR_n$ be the corresponding SINR value, in linear units. We obtain the spectral efficiency η_n of RB n using the following equations:

$$\Gamma = -\frac{\ln(5\,\mathrm{BER})}{1.5}\,, \tag{2.8}$$

$$\eta_n = \log_2\left(1 + \frac{SINR_n}{\Gamma}\right), \tag{2.9}$$

where BER is the Bit Error Rate and Γ is the so called *SNR gap*, as it models the discrepancy between practical implementations and information-theoretic results.

Upon the calculation of η_n, which lies in the continuous interval $[0.15, 5.55]$, the procedure described in [47] is used to derive the corresponding CQI, which is a quantized version of η_n.

Error model-based approach: this model relies on the exploitation of our link abstraction model. Thanks to this approach, we can dynamically select the MCS that better complies with a given target transport block error rate for the connection, referred to as $TBLER_{\mathrm{th}}$. In the following, we describe our improved CQI evaluation procedure by abstracting away from the actual implementation details, i.e., on the actual representation of CQI values (at the receiver, e.g., number of CQI levels, etc.) and the subsequent mapping of these CQIs onto a suitable MCS (which is done at the eNodeB).

As detailed in Algorithm 1, our procedure works as follows. Periodically, each UE computes its received power spectrum profile, i.e., an SINR sample is acquired for each possible RB.[3] For any user $i = 1, 2, \ldots, N_{UE}$, it starts from MCS 29, which corresponds to the most aggressive transmission scheme, and evaluates the TBLER performance considering a transport block composed of all possible LTE RBs. The transport block error rate for user i, $TBLER_i$, is estimated through Eq. (2.6), taking as input the vector of SINRs for the selected user, and the MCS that we are currently evaluating. If $TBLER_i$ is larger than or equal to the target BLER defined by 3GPP (i.e., 0.1) [17], we keep on searching for a better (less aggressive) MCS; otherwise, the procedure stops. Thus, the corresponding CQI is obtained in order to satisfy the spectral efficiency constraints defined by the standard [17].

[3] In this case, all RBs allowed by the selected LTE channel bandwidth are accounted for.

Algorithm 1 MCS assignment performed by each UE, every TTI. Prior to this computation, the users decode the pilot sequence sent by the eNodeB in order to get all the RB SINR samples that are needed to retrieve the predicted error rate.

Require: UE Target BLER ($BLER_{t,UE}$), $SINR_{UE}$

```
for i = 1 → N_UE do
    MCS ← 0
    while MCS < 29 do
        BLER ← GetTbError(SINR_i, MCS + 1)
        if BLER > BLER_{t,i} then
            break
        else
            MCS++
        end if
    end while
    MCS_i = MCS
end for
```

2.2.5 Simulation Results

In the following we provide some technical results for different LTE scenarios. Our main goals are: 1) to illustrate the usability of the proposed link abstraction model, and 2) to prove the efficiency of the link adaptation improvement proposed in Section 2.2.4. We would like to note that the validation of the error model proposed is simplified for the sake of readability; a more in depth validation can be found in the LENA documentation [29]. In Table 2.9 we report the considered system parameters. First of all we would like to present some quantitative results on the computational complexity of the model. The simulator presented in this Section takes 23 seconds for simulating a 10 seconds scenario with one UE transmitting continuously at full bandwidth to an eNB, while the same scenario takes 38 minutes when simulated with the LTE Vienna Link Layer simulator.[4]

[4]The reference hardware platform is an Intel Core2 Quad CPU Q8400 2.66GHz.

Parameter	Value
Frequency	2.1 GHz
Channel Bandwidth	5 MHz
Propagation Model	Friis free-space
TX power	30 dBm

Parameter	Value
Number of RBs	25
$RB_{bandwidth}$	180 kHz
$RB_{subcarriers}$	12
$RB_{OFDMsymbols}$	14

Table 2.9. *PHY Configuration*

1) Error model: We consider a downlink transmission from an eNodeB to a single static UE. For the wireless channel, we account for a Friis free-space propagation model, but note that the conclusions that we draw here are general and apply to more sophisticated models as well. The UE is placed 2150 meters away from the eNodeB and, according to the considered propagation loss model, it experiences an SINR of 15.1 dB for all its RBs. We first evaluate the BLER performance resulting from the selection of a "safe" transmission

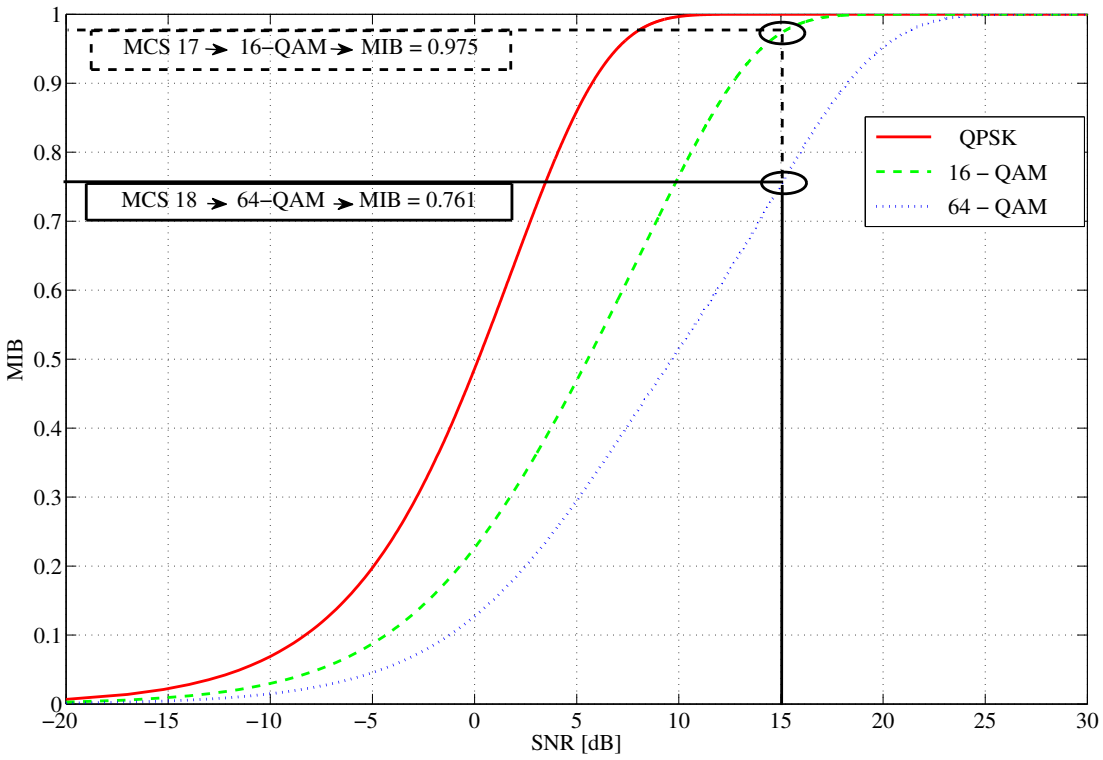


Figure 2.7. *MI extraction*

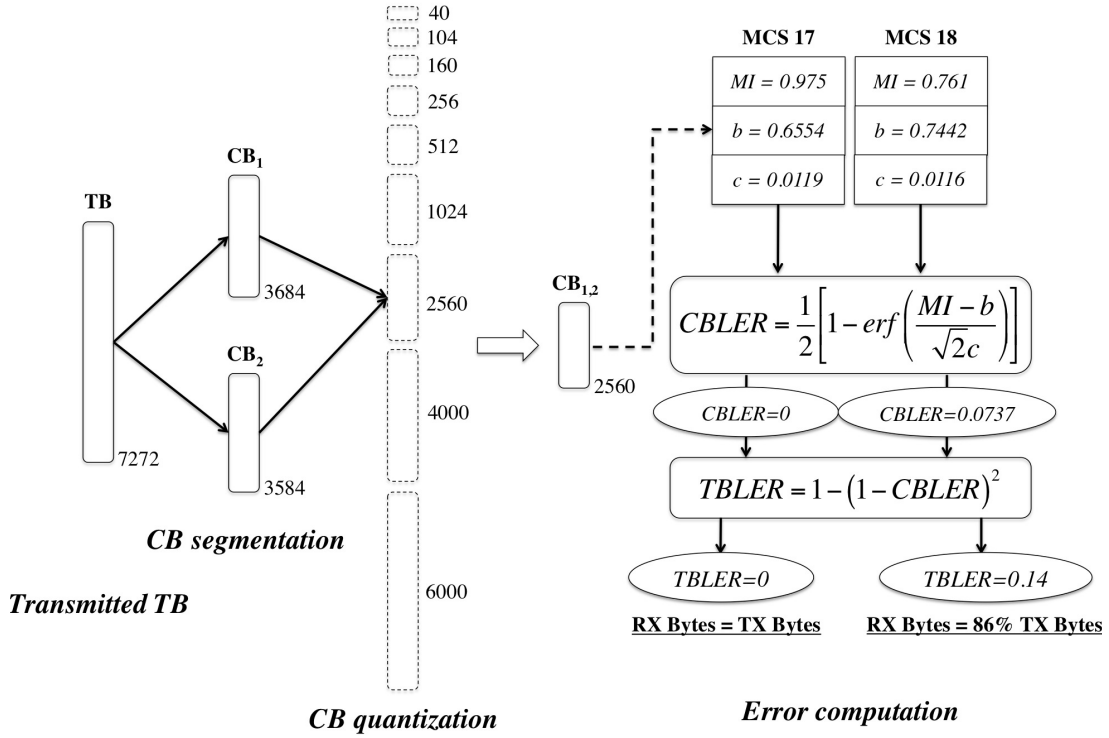


Figure 2.8. *TBLER computation*

scheme, MCS 17, which corresponds to a 16-QAM. As we show, for this MCS the estimated error rate (through Eq. (2.6)) is below the standard TBLER threshold of $TBLER_{\text{th}} = 0.1$. First of all, we extract the *mutual information* value associated with the experienced SNR, as shown in Fig. 2.7 (where we plot the approximation functions of Table 2.8). According to [17], from the selected MCS and the maximum number of assignable RBs, the TB size is 7272 bits (including the header). Following [17], the TB is split into two code blocks, CB_1 and CB_2, of size 3684 and 3584 bits, respectively. As shown in Fig. 2.8, these code sizes are mapped onto the closest CB size in set $\mathcal{S}$. In fact, our ns-3 simulator fitting parameters are only stored for a subset of all possible CB sizes. Thus, the resulting CB size that will be used for the prediction of the CBLER performance is 2560 bits. Now, using the $b_{S,M}$ and $c_{S,M}$ parameters associated with the latter code block size and the previously extracted MI with Eqs. (2.6) and (2.7), we obtain an estimated transport block error rate of $TBLER = 0$.

Next, we try to allocate a more aggressive modulation and coding scheme, MCS 18, for which the modulation order amounts to 6 bits per OFDM symbol (64-QAM). Thus, we repeat the procedure illustrated in the previous paragraph, obtaining the mutual information MI, and the fitting parameters $b_{S,M}$ and $c_{S,M}$. These quantities, together with Eqs. (2.6) and

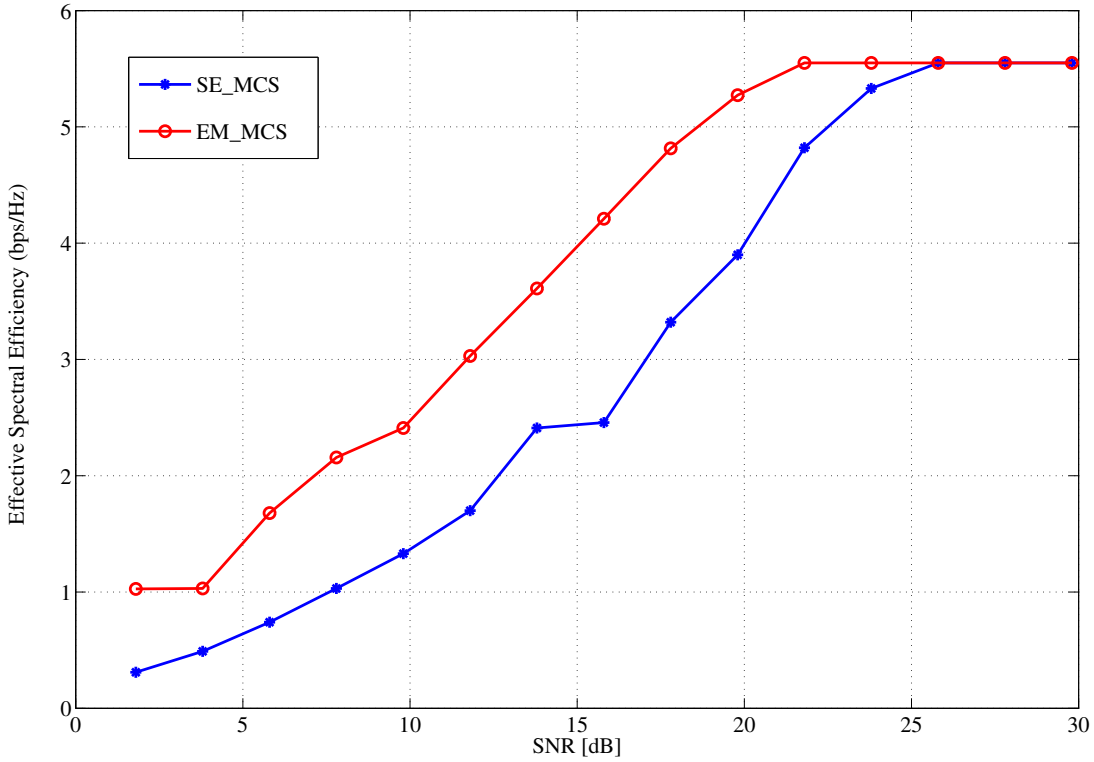


Figure 2.9. *MCS assignment comparison*

(2.7), return $TBLER = 0.14$, which means that MCS 18 is not compatible with the considered error requirements.

2) Link adaptation: We now consider a scenario with a single UE, and we vary its distance from the eNodeB. This leads to SNR values ranging from about 2 to 30 dB. Also, we consider the standard target transport block error rate of $TBLER = 0.1$. In Fig. 2.9, we show the effective spectral efficiency as a function of the SNR for the spectral efficiency-based (SE_MCS) and the error-based (EM_MCS) MCS selection schemes. The effective spectral efficiency metric reflects the actual bits per second per unit of frequency that are successfully transmitted from the eNodeB to the UE, by also accounting for the residual error after PHY layer processing. As can be seen from Fig. 2.9, EM_MCS outperforms the current approach, at all SNR levels. This indicates that SE_MCS tends to be too conservative, even though a more aggressive technique can be used while still adhering to the target error requirements.

2.2.6 HARQ Implementation

The proposed LTE-HARQ scheme is based on a soft combining full incremental redundancy (IR), also called IR Type II, meaning that each retransmission carries only new redundant information. With no repetition of coded bits, the performance of the decoder at each stage is that corresponding to a binary code with the modified equivalent code rate and code size as illustrated in Figure 2.10.

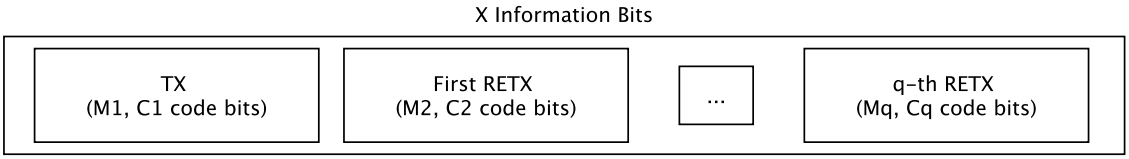


Figure 2.10. *MI-based parameter update after each transmission*

Our proposed PHY error model has been extended for considering IR HARQ according to [48]; the required input parameters for AWGN mapping function are given below

$$R_{eff} = \frac{X}{\sum_{i=1}^{q} C_i} \tag{2.10}$$

$$L_{eff} = \sum_{i=1}^{q} C_i \tag{2.11}$$

$$MI_{eff} = \frac{\sum_{i=1}^{q} C_i M_i}{\sum_{i=1}^{q} C_i} \tag{2.12}$$

where R_{eff}, L_{eff}, and MI_{eff} are the effective code rate, block size and mutual information after q retransmissions, respectively. X is the number of original information bits, C_i is the number of coded bits, M_i is the mutual information per HARQ block received at each q retransmission. Thus, in order to be able to return the error probability through our framework, we first compute the corresponding R_{eff} to pick the proper AWGN performance curve (based on the same transmitted MCS and resulting CB size); then, we compute the MI_{eff}, we numerically derive the corresponding effective SINR, and finally obtain the error probability by matching this value with the selected AWGN curve.

2.3 Cross Layer Framework for Variable Packet Size Allocation

A typical assessment factor of the performance of an access network is its throughput capability. As of today, wired fiber optic links are one of the fastest networking technologies available. In fact, the throughputs made available by these environments [49] have motivated the increment of the frame size parameter, the Maximum Transmission Unit (MTU), which before was associated to a 1500 bytes legacy value from early Ethernet deployments. Taking advantage of the inherently better performance and reliability of the fiber optic links, the increase of the MTU parameter reduces the number of packets sent by enforcing less fragmentation, while at the same time reducing overhead and CPU processing since fewer packet headers need to be analyzed by the network stack.

The Section is organized as follows. Section 2.3.2 provides a description on the LTE system model regarding frame and fragmentation mechanisms related to the paradigm of Jumboframes. In Section 2.3.3, our Jumboframe-enabled LTE framework is analyzed in different scenarios. In particular, Section 2.3.3.1 presents a study on the overhead and the buffer status for the LTE architecture without mobility. Mobile users are considered in Section 2.3.3.2 with a solution to the problem of buffer saturation. The validation concludes in Section 2.3.3.3 with a study of real-time video transmission.

2.3.1 Related Work

The usage of these larger packets, deemed Jumboframes, has been amply studied in the past, and different analyses on the performance impact in TCP [50] have been presented. Likewise, such evaluations have also targeted the shortcomings of not just the MTU increase itself (e.g., different MTU values on the path [51], burst drop, delay jitter and application sensitivity [52]), but also its impact on other mechanisms relying on the legacy MTU value (e.g., IP packets Total Length fields and their CRC limit [53]). However, wired links have not been the only technologies evolving in this direction. For example, [54] has explored the impact of Jumboframes in WLANs with a real testbed, showing benefits in terms of throughput in different scenarios. In particular, the authors highlighted the fact that the wireless medium dynamics and shared access of its contention-based access mechanism not only exacerbate the previously mentioned shortcomings of large frame usage, but also

create new issues on medium fairness usage and end-user experience. As such, resilience mechanisms, such as partial frame size selection, packet recovery and rate adaptation, need to be employed in order to circumvent these issues.

This contribution aims to design and evaluate the usage of enhanced packetization mechanisms for deployment in LTE networks. It is well-known that coordinated-based wireless access technologies (e.g., LTE and LTE Advanced) are strictly coupled with adaptive resilience and medium fairness mechanisms [55]. Therefore, this work provides a feasibility study evaluation of the opportunistic advantage of using frames of larger size in LTE networks, contributing with a concrete assessment of employing Jumboframes in different scenarios. Unlike in [54], here the study is more focused on the description of the key aspects of LTE networks (e.g., RLC header) that have a direct impact on Jumboframe implementation, leading to a complete understanding of the addressed problem. Due to the particular complexity of the considered cellular architecture, we study the impact of Jumboframe transmissions using ns-3, a network simulator that permits to evaluate both the the wireless access and the Core Network, and to perform an overall system evaluation without recurring to a real testbed, as done in [54].

2.3.2 System Model

The simulation platform used to evaluate Jumboframes feasibility is ns-3, an open source discrete-event network simulator for Internet-based systems, available online at [6]. Our work exploits the module currently under development within the LENA project [29], which comprises the LTE core network, EPC, presented in [30], and the radio access, eUTRAN, detailed in [32] and based on the first LTE simulation framework for ns-3 [12]. The simulator is extensively documented in [33].

RLC segmentation/concatenation: As shown in Fig. 2.11, depending on the channel conditions and the adopted scheduling strategy, a user might be allotted a number of resources that enable the transmission of multiple Segment Data Units (SDU) (i.e., *concatenation*); conversely, the SDU has to be split (i.e., *segmentation*). Generally speaking, the overhead plays a fundamental role for the evaluation of Jumboframes, so we focus on the LTE-specific guidelines defined in [56]. The RLC mode selected for our simulations is Unacknowledged (UM).[5]

[5]Please note that in LTE there are two more RLC modes: Transparent Mode (TM), and Acknowledged Mode (AM). Please refer to [56] for further details.

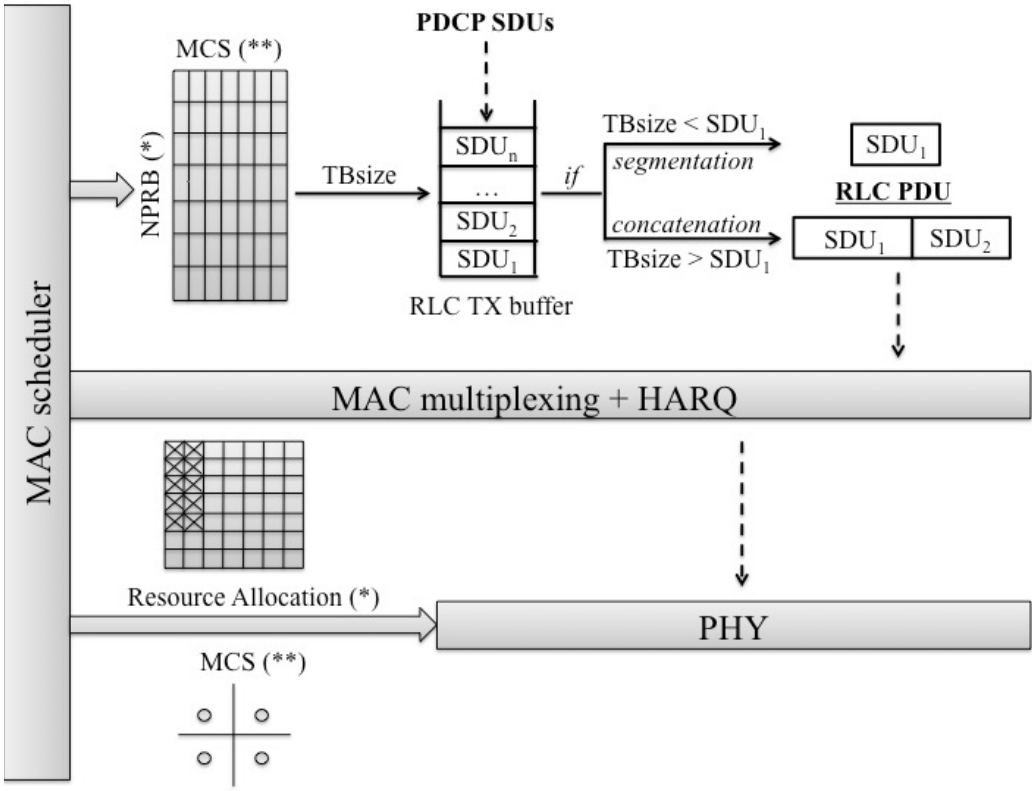


Figure 2.11. *Overview of the Radio Access mechanisms involved for Jumboframe transmissions in LTE networks.*

As shown in Fig. 2.12, the RLC header of a Packet Data Unit (PDU) containing more than one SDU is more complex. The following additional fields are introduced to describe the packet concatenation structure.

- FI (Framing Information): the first bit indicates whether the first byte of the Data field is the first byte of an RLC SDU, whereas the second bit indicates whether the last byte of the Data field corresponds to the last byte of an RLC SDU.

- E (Extension): when this field is set to 1, it means that a new RLC SDU will be included in the current PDU;

- SN (Sequence Number): indicates the sequence number of the current PDU;

- LI_i (Length Indicator): represents the Data field size of the i–th SDU, in Bytes.

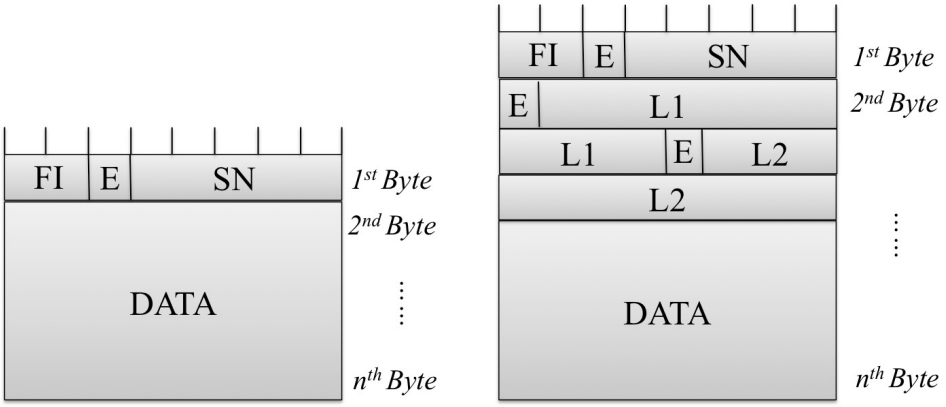


Figure 2.12. *RLC PDU example:* 1 *single SDU (left),* 2 *concatenated SDUs (right).*

2.3.3 Jumboframes Evaluation Framework

Jumboframes in ns-3 are enabled by appropriately setting to 10 KB the MTU $i)$ in the application server (*udp-client*[6]), $ii)$ in the internet (*point-to-point-helper*), and $iii)$ in the LTE core network (*epc-helper*). In the application server, Jumboframes are generated with packets of 9 KB, whereas the traditional frames have 1 KB long packets. We take into account the delivery of three files with different sizes: 4.5, 45 and 450 MB.

2.3.3.1 Feasibility Study

In order to provide an adequate insight about the impact of Jumboframes in LTE networks, we configure different simulations characterized by diverse numbers of users and their locations, in a pedestrian scenario with no mobility. The variability offered by these parameters permits to highlight the interplay between the RLC segmentation/concatenation mechanism, the size of RLC PDU and the number of RBs assigned by the MAC scheduler to each user.

User-location: The considered scenario is composed of 3 users located at different distances from the eNB: 200, 1000, and 2000 m away, respectively. Each user perceives highly different channel conditions, resulting in a diverse bandwidth utilization (i.e., different MCS value). The downlink flow is designed so that, in each sub-frame, the transmission queues are always full (i.e., backlogged). As shown in Fig. 2.13, UE1, that corresponds to the user

[6]The UDP client represents a UDP packet generator, where it is possible to set the generation rate and the packet size.

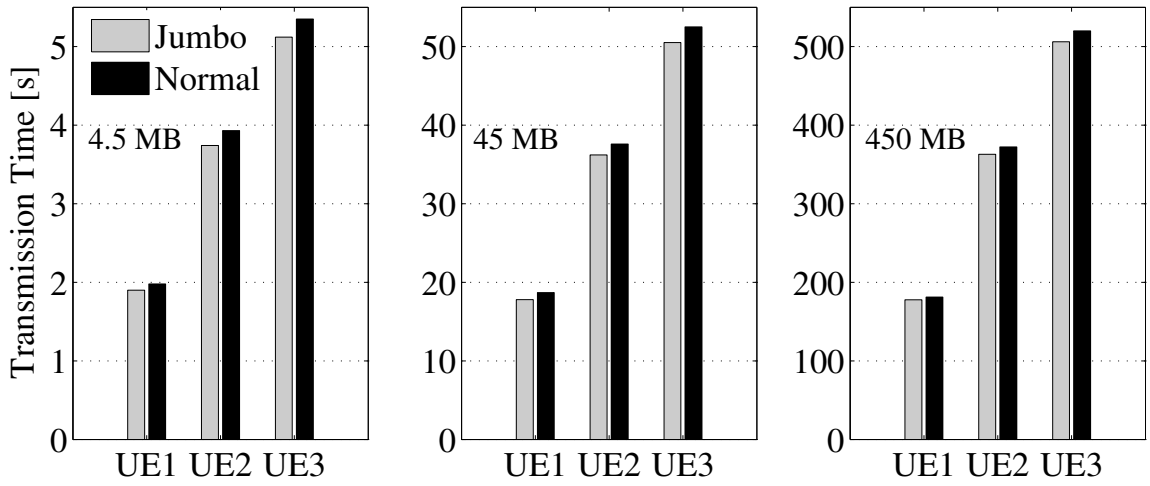


Figure 2.13. *Transmission time of three files with different sizes ($\{4.5, 45, 450\}$ MB) for three users. Gray and black lines represent Jumbo and normal frames, respectively.*

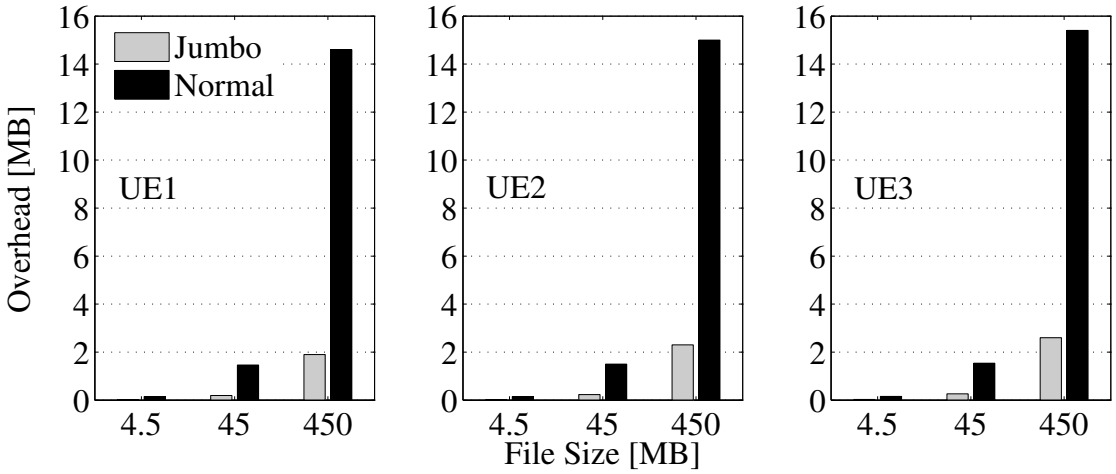


Figure 2.14. *Total overhead sent during the transmission of the 450 MB file for Jumbo and normal frames.*

in the proximity of the eNB, is the first one that completes its download, thanks to a better exploitation of the assigned resources. The usage of Jumboframes slightly outperforms the traditional approach in terms of transmission time reduction (3%). It is interesting to note the large overhead savings of Jumbo with respect to normal frames usage (87%) shown in Fig. 2.14, where the total amount of overhead (i.e., UDP, IP and RLC) is depicted for the transmission of a 450 MB file. This is due to the reduced amount of RLC concatenation performed by the eNB to adapt the incoming Jumboframe to the reserved region assigned by the MAC scheduler to each user. In fact, in addition to the higher layer headers such

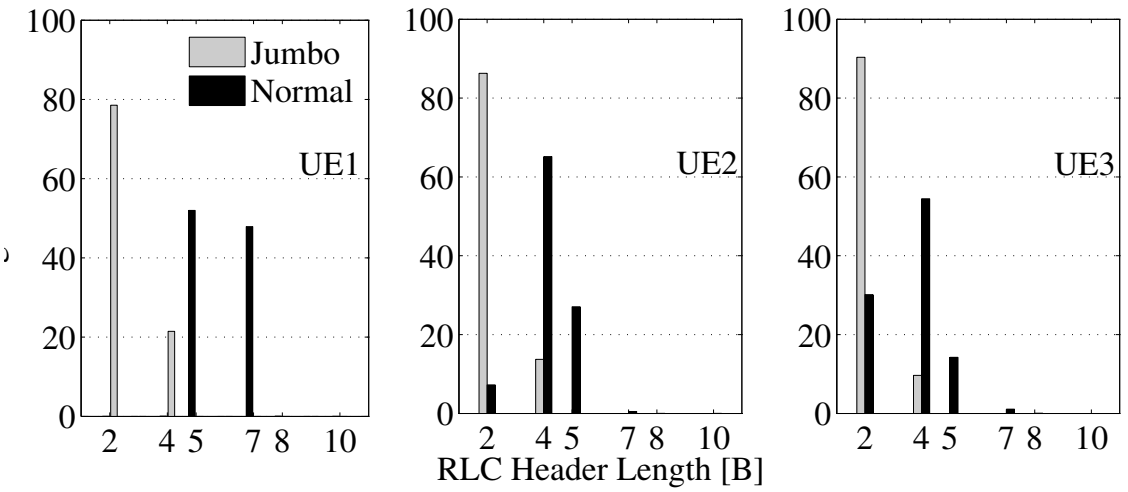


Figure 2.15. *Comparison of RLC Header length expressed in Bytes between Jumbo (gray bars) and normal (black bars) frames for the* 450 *MB file delivery.*

as UDP and IP, each transmitted packet presents an RLC header whose size depends on the fragmentation/concatenation performed at the RLC layer (see Fig. 2.11): the more RLC operations accomplished the bigger the RLC header. Thus, the lower amount of overhead required by Jumboframes permits to reduce the processing complexity with respect to the normal frames case, as illustrated in Fig. 2.15. Here, the depicted distributions collect the different RLC header lengths used by the eNB to transmit all the RLC PDUs for each user in the case of the 450 MB file delivery. For example, in case of Jumboframes (gray bars), UE1 (the user with the best channel conditions) transmits PDUs with 2 B RLC header size for slightly 80% of the total transmissions, and 4 B otherwise. On the other hand, by using normal frames, the eNB performs multiple concatenations to produce the RLC PDU suitable for the number of RBs assigned by the MAC scheduler,[7] thus increasing the overall transmission processing complexity. This effect is slightly reduced for the other users, due to the lower number of bits that can be transmitted over the same number of RBs, avoiding complex concatenations between different SDUs.

User-density: Here we vary the density of the users in the cell, aiming to evaluate the performance of Jumboframes when the available resources (i.e., RBs) allotted to each user decrease. In particular, we place all the users at the same distance with respect to the eNB

[7] In fact, the amount of information that UE1 can receive is higher than the size of a single packet with 1500 B.

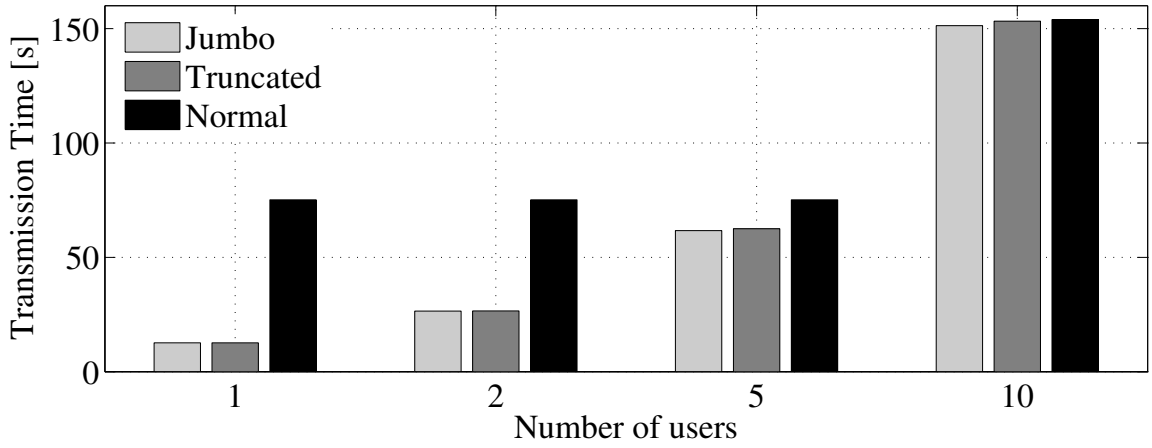


Figure 2.16. *Transmission time for Jumbo,* truncated *Jumbo and normal frames as a function of the number of users served.*

and we vary their number from 1 to 10. The packet generation time is equal to the LTE Transmission Time Interval (TTI), i.e., is 1 ms. Thus, a new packet is added to the transmission buffer at each subframe. The download file size is 112.5 MB. In order to see the impact of Jumboframes in a more realistic scenario, the source is no longer backlogged. Recalling that regular frames and Jumboframe contain 1 KB and 9 KB of data, respectively, here we introduce a third type of frame, named *truncated* Jumboframe. This special version of Jumboframe is made of 9 KB as before, but undergoes fragmentation according to the traditional 1500 MTU limit in the EPC. This new classification permits to show the potential of Jumboframes when the source admits an extended frame size, but the network cannot sustain a higher MTU limit (1500 B). It can be noted from Fig. 2.16 that the total transmission of normal frames up to 5 users takes the same amount of time; this is due to the fact that each user can transmit the generated PDU in less than a TTI, thus underutilizing the available bandwidth. On the other hand, concerning the comparison between Jumbo and *truncated* Jumboframes, we can observe that the transmission time is the same with up to 2 users, while it slightly differs when the number of users increases. With a few users, *truncated* Jumboframes are transmitted in the same amount of time required for the transmission of Jumboframes; in fact, the resources assigned by the MAC scheduler allow to transmit, in each subframe, the additional overhead generated when concatenating more SDUs.

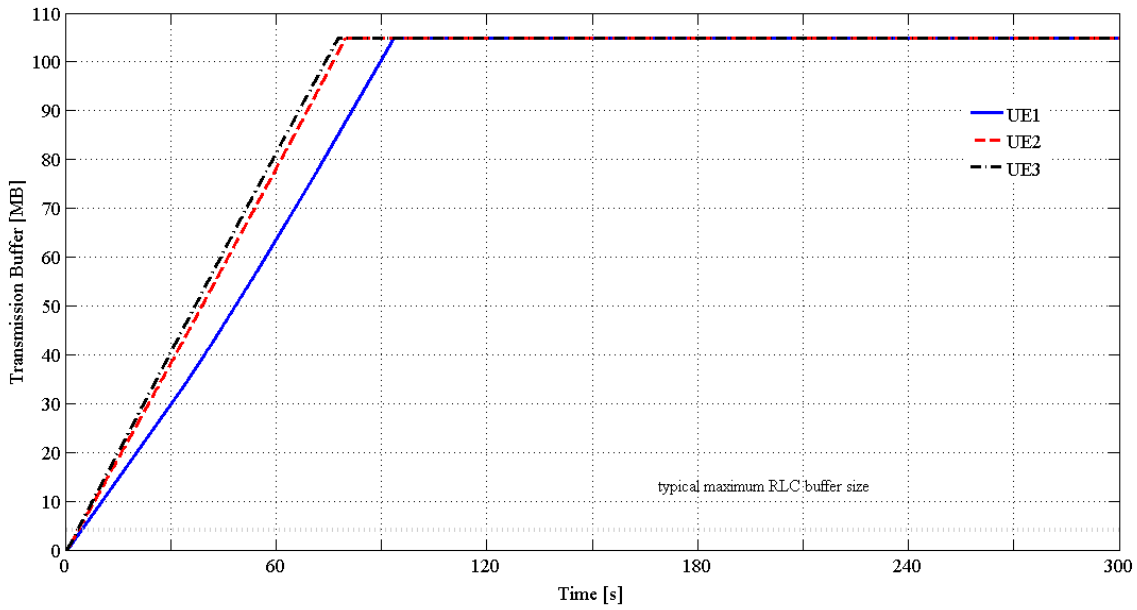


Figure 2.17. *Status of the RLC transmission buffer for Jumboframes transmissions. The typical size of the buffer is depicted for comparison.*

2.3.3.2 Buffer Saturation

We noted above that transmitting bigger frames directly from the source can result in a transmission time and overhead reduction. However, focusing the attention on the status of the RLC transmission buffer as in Fig. 2.17, it is possible to note the main drawback of Jumboframe transmission in LTE networks: due to the higher amount of information to be sent in the same amount of time, there is a concrete risk of saturation. This figure is obtained by moving the users in opposite directions in order to simulate a continuous change of channel quality, based on a vehicular propagation scenario (30 kmph). In particular, user number 1 (UE1) leaves its initial position (200 m from the eNB), and approaches the cell-edge (2 km) at a constant speed, and then goes back to its original position; UE2 moves from an intermediate position (1 km) to the cell edge, and then back to the eNB (200 m); UE3 follows the opposite path of UE1.

To overcome this problem, we propose a dynamic technique that permits to mitigate the congestion by creating a cross-layer interaction between the radio access (i.e., eNB) and the core network (EPC), as depicted in Fig. 2.18. Thanks to this approach, we can tune the packet size before entering the access network by relying on the channel feedbacks and the

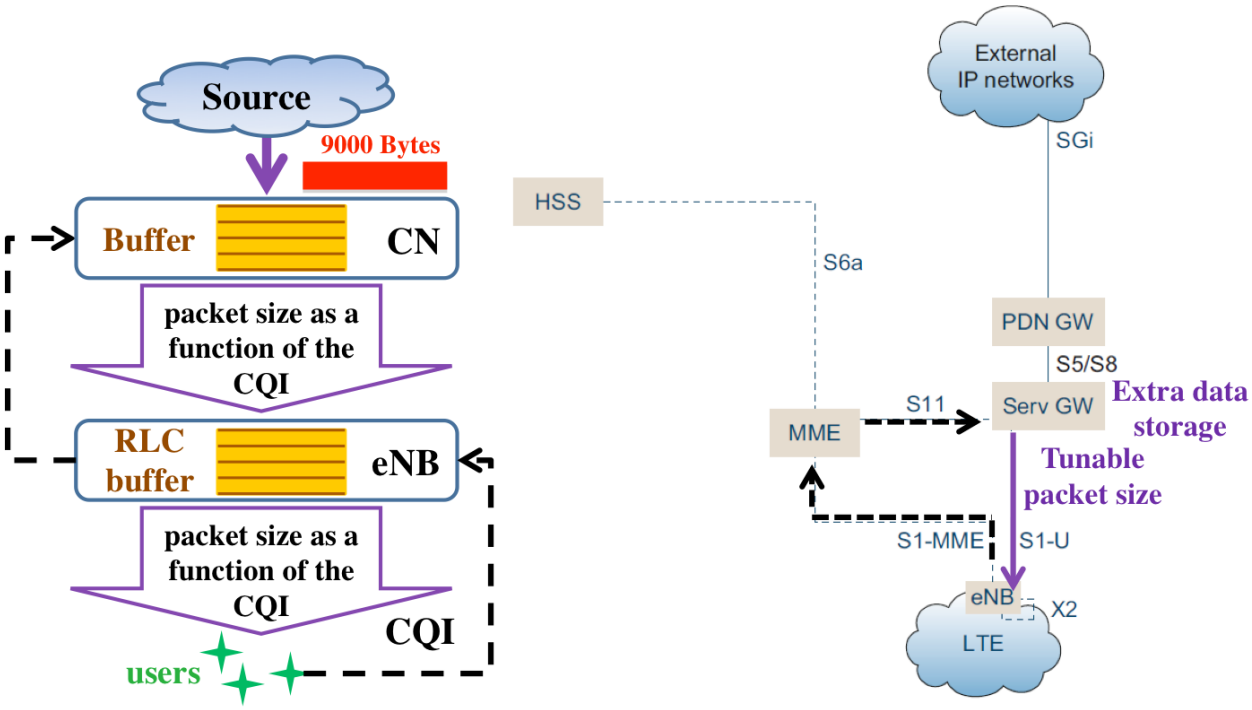


Figure 2.18. *EPC architecture: ns-3 implementation (left), 3GPP view (right).*

scheduling decision made at the eNB, thereby avoiding saturation at the RLC layer. It is to be noted that this technique has been implemented in ns-3 through *c++* programming strategies for evaluation purposes, thus it is not part of a 3GPP compliant structure that should rely on S1 control signaling. A small delay is necessarily introduced during the communication between the eNB and the Service Gateway; if we consider a scenario without mobility, this delay does not affect the transmission buffer because the *tuned* frame size fits with the instantaneous TB size assigned by the scheduler (except for minor multipath channel oscillations). On the other hand, if we introduce mobility, the transmission buffer varies with the user speed as well as where the user is going. In fact, if the user moves closer to the eNB, the delayed channel feedback is more conservative with respect to the actual position; otherwise, if the user is going away from the eNB, the delayed TB size indication overestimates the effective allocatable packet dimension, thus imposing fragmentation, and a consequent increase of the RLC buffer. This behavior is confirmed by Fig. 2.19, where the RLC transmission buffer of each user is illustrated. The transmission buffer of UE1 increases stepwise (because of quantized CQI levels) and, at around 40 s, it starts decreasing: this is because the user has reached the cell-edge, and returns back to its initial position. On the

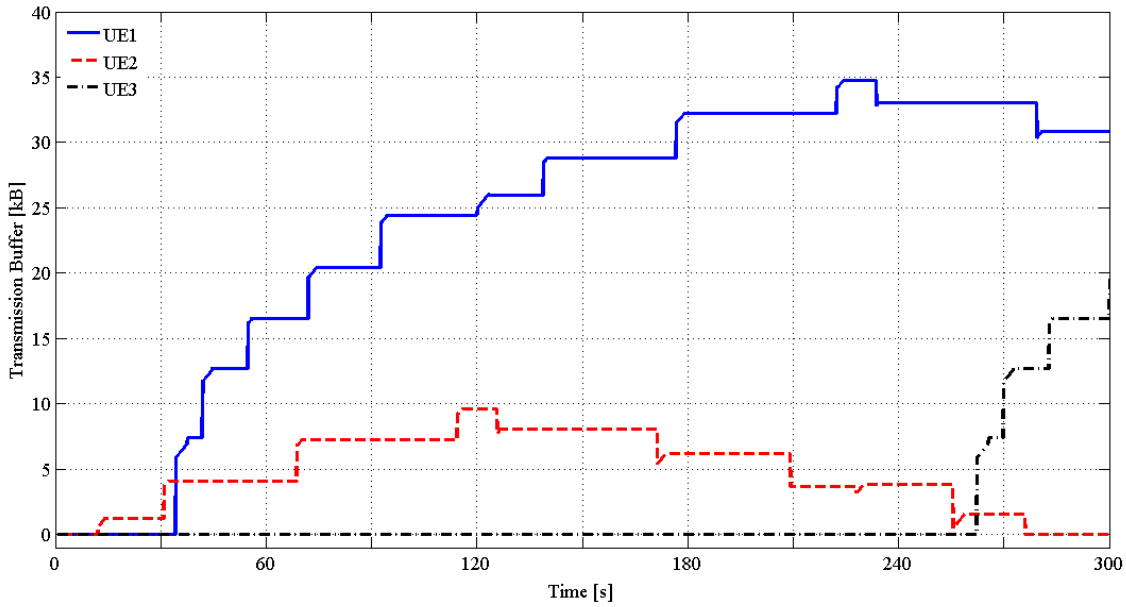


Figure 2.19. *Status of the RLC transmission buffer for* tuned *frame transmissions during the file delivery.*

contrary, UE3 initially shows a stable transmission buffer (it starts from the cell-edge and approaches the eNB), then it goes back to its initial position, showing an increase of the RLC buffer. Similarly, UE2 stores extra information at the beginning of the simulation and then reduces the stored data (from the cell-edge the user approaches the eNB), confirming our explanation.

2.3.3.3 Video Transmission Evaluation

In this Section we present the evaluation of realistic video traces transmitted through Jumbo and normal frames, in order to show the effectiveness of our approach also with real applications, where the size of each frame is variable. To do so, we selected a subset of video packets (16 MB), taken from *Jurassic Park I (H.263 - 256 kbit/s target bit rate)*, downloaded from [57].

The results shown in Figs. 2.20 and 2.21 confirm the behavior described in the previous Sections. For a low number of users, if compared to the Jumboframe transmission, the additional overhead is transmitted in the same amount of time because of the available resources. Concerning the overhead, we can observe a smaller gain as compared to our example shown in Fig. 2.14; this is because the size of these time-varying packets is, on average, 4.3 MB.

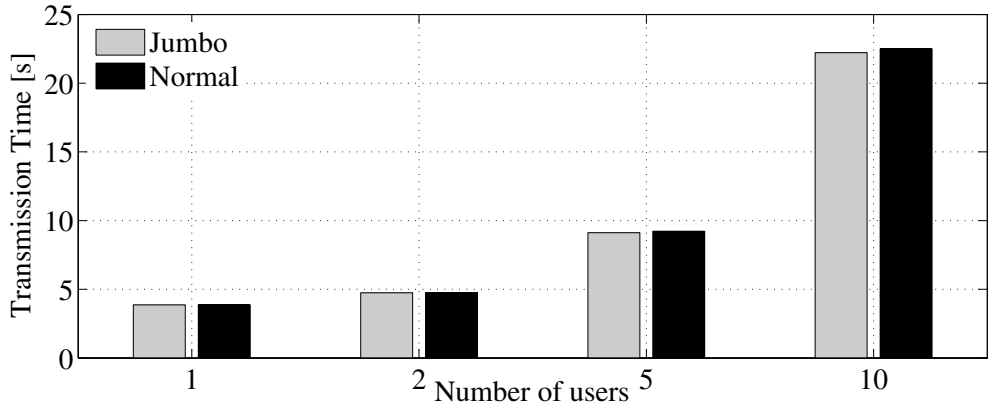


Figure 2.20. *Transmission time of a part of* Jurassic Park *movie for Jumbo and normal frames as a function of the number of users served.*

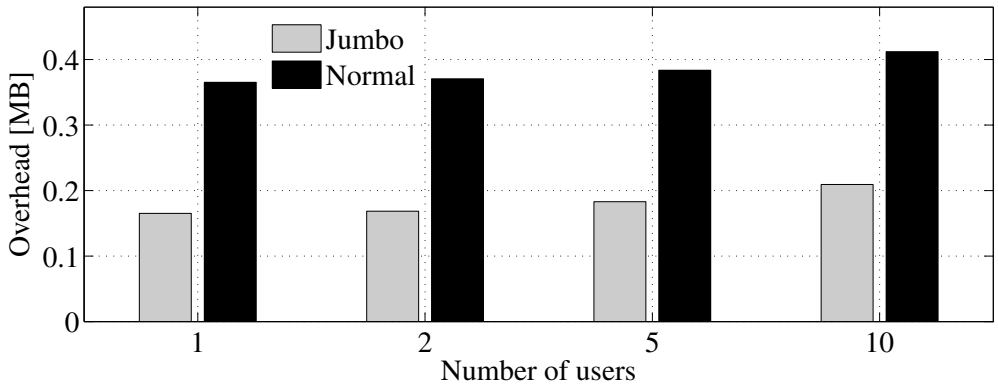


Figure 2.21. *Total overhead sent during the transmission of a part of* Jurassic Park *movie for Jumbo and normal frames as a function of the number of users served.*

2.4 Summary

In Section 2.1, we have provided an overview of the wireless access components in a novel network architecture, which is designed to better support video delivery in next generation networks. The radio access aims at optimizing the last hop for video flows, with target technologies classified in contention-based, such as WLAN, and coordination-based, such as LTE. Several cross-layer functions and enhancements have been proposed within this architecture, with particular focus on LTE. To achieve an efficient communication with heterogeneous access, an abstraction layer has been introduced, which hides the specifics of each technology to the upper layers by introducing generic parameters and primitives. This

layer is based on an extension of the MIH services offered in the IEEE 802.21 standard.

In Section 2.2, we have introduced a link abstraction model for the simulation of Transport Block errors in the downlink of an LTE system. Our objective is to provide a lightweight but still accurate procedure for the computation of the residual errors, after PHY layer processing, without having to go through the detailed simulation of LTE PHY procedures. Toward this end, we combined Mutual Information-based multi-carrier compression metrics with Link-Level performance curves matching. This allowed us to obtain pre-calculated lookup tables, which can be used in an online fashion to track residual bit errors after physical layer modulation and coding procedures. In addition, we have proposed a CQI evaluation procedure which can be used as part of the LTE Adaptive Modulation and Coding scheme, showing its superiority in terms of achievable spectral efficiency with respect to current ns-3 solutions. Finally, based on the proposed error model, we have introduced a lightweight approach to analyze retransmission errors in LTE-HARQ.

In Section 2.3, we have presented a simulation framework for the evaluation of the impact of Jumboframes in LTE networks. A comparative observation based on diverse network configuration criteria highlighted some benefits and caveats related to this approach. On the one hand, under certain channel and scheduling conditions, we observed a slight transmission time reduction, and a great overhead saving, mainly due to the reduced amount of higher layer headers. In addition, Jumboframes permit to further reduce the size of the RLC headers by requiring less segmentation/concatenation mechanisms. On the other hand, scarce resources lead to rapid buffer saturation. In order to mitigate this effect, a cross-layer approach has been proposed, allowing the radio access (i.e., eNB) and core network (EPC) layers to interact and tune the packet size dynamically. We also tested our evaluation framework through the analysis of realistic video traces. There are many other issues to consider relatively to the impact of Jumboframes in wireless networks, such as error control techniques, ad-hoc allocation schemes, LTE-A Carrier Aggregation exploitation, and so forth. This work represents a first contribution for the Jumboframes feasibility in LTE networks, paving the way for further research leveraging future mobile accesses with these enhancing mechanisms.

Chapter 3

Cognitive Exploitation of Radio Dynamics

Next generation wireless networks can be further enhanced by properly exploiting cognitive solutions. On the one hand, the concept of resource sharing can be seen as a promising way to optimize the performance of radio communications. In particular, the wireless spectrum is scarce, and its usage is often found to be inefficient. For this reason, spectrum sharing has been gaining increased attention in the last few years. On the other hand, one of the main issues affecting mobile ad hoc networks is related to the fast changing topology, which creates significant problems to the standard transport layer protocol. Indeed, the transmission control protocol (TCP), which is the standard for reliable connection-oriented communications, was designed for wired networks. Thus, a dynamical cross-layer self-adaptive framework based on cognitive techniques can provide large gains.

This Chapter is organized as follows. In Section 3.1, we describe our spectrum sharing framework, along with a detailed simulation campaign which is run to validate the proposed architecture, in terms of both modeling realism and computational performance. In Section 3.2, we design a model to predict the mobility status of a network through a probabilistic approach, and we propose a series of ad hoc strategies to counteract TCP's inefficiency based on this prediction. Finally, we summarize our contributions in Section 3.3.

3.1 Spectrum Sharing in Multi-Operator LTE Networks

Although game theory started as a mathematical formulation of problems mostly belonging to economic and political systems, its application to wireless networking is becoming common practice [58]. Indeed, game theory is well suited to study situations where a scarce resource is contended for by multiple agents (players), as well as situations where these players have contrasting objectives, mostly because they are selfish, i.e., interested in their own good only. Incidentally, this scenario describes the vast majority of wireless access problems, even though it can be noted that spectrum scarcity is more due to bad frequency planning than to a real lack of available frequencies [59]. In any event, it is the selfishness of the players, i.e., the network operators, that makes the radio access inefficient. This gives a strong motivation for replacing the classic scenario where network users are driven by egoism with another where they cooperate [60].

A related early attempt at using game theory within wireless scenarios involves *cognitive networks*. In such a model, as defined by [61], two kinds of users co-exist, i.e., primary and secondary. The former are licensed users which access the frequency bands they are entitled to; the latter dispose of the leftovers, i.e., the subchannels which are unused by their legitimate owners, (the primary users) for a given amount of time. The secondary users can act only after the primary users have made their decision. To some extent, cooperation is present in the sense that the primary users are aware of the secondary, but they let them be; after all, they are not threatening as they only exploit unwanted resources. This situation of cognitive networks is also reminiscent of some game theoretic investigations describing duopoly situations with an incumbent enterprise and an outsider one. Such a scenario is known as a *Stackelberg game*, and there have been applications of it to wireless networks [62]. Actually, an even more general form of collaboration can be thought of, without classifying players into primary or secondary, but rather considering an egalitarian approach, where similarly-minded players desire to use a common resource, or share a portion of their properties with the others. In the wireless network context, this would mean that network operators join their licensed frequencies for common wireless access. As hinted by several studies, such an idea may be beneficial for all the involved players if a collaborative access to the wireless resource is achieved [63–65]. Possible ways to quantify a gain can be in a larger number of users served, a wider network coverage, a higher network

throughput.

However, the main challenge for analyzing this problem is in the adoption of realistic models for the physical layer. In principle, it can be easily argued that certain physical characteristics of the wireless channel, for example multi-user and frequency diversity, make it appealing to share its access, rather than competing for it [66]. Yet, an exact characterization of the wireless channel for several players is mathematically difficult. For this reason, the considered scenarios are often limited to small networks with few transmitter-receiver pairs, most of the times just two, i.e., a total of four nodes. We believe that a realistic performance evaluation of larger networks is key to get a clear understanding of the usefulness of the sharing concept in wireless scenarios.

In the scientific community it is quite common to resort to network simulation instruments to assess the performance of large networks which are not easy to tackle exactly. For example, the well known network simulator-3 (ns-3) [6] is currently considered as one of the most advanced and modular software tools to perform this task. The ns-3 simulator is entirely open source and comprises the entire protocol stack, from the physical layer up to the application. Although the focus of network simulation is often on the intermediate layers (data link, network, transport), ns-3 is extremely modular and therefore admits integration with detailed models of the physical layer, especially with the most up-to-date wireless technologies. A bottom-up representation of the protocol stack can be particularly appealing for the analysis of spectrum sharing, that involves both lower and higher layers. These reasons motivate our choice to employ an existing implementation [12] within ns-3 of the Long Term Evolution (LTE) of the Universal Mobile Telecommunications System (UMTS) [1].

The main goal here is to introduce a novel software extension of this ns-3 version to characterize spectrum sharing scenarios where cooperation is established among multiple operators, each with a considerable number of nodes. To realize this enhancement, original software structures are introduced. First of all, a class describing a virtual *frequency market* has been inserted in the simulator structure. This class implements the functionalities of a virtual arbitrator, and does not represent a physical entity of the network, but rather it determines the sharing policy of the frequencies belonging to the common pool. In other words, its role is to abstract the set of rules agreed by the operators when determining the shared portion of the spectrum. Moreover, two main sharing meta-policies can be utilized,

namely *orthogonal* and *non-orthogonal* sharing. In the former case, the frequencies of the shared pool are used by one and only one operator, although not necessarily the one that detains the legal property of the access on that frequency. In the latter, also simultaneous access on the same frequency is possible. In both cases, the arbitrator structure is required to represent in an abstract manner the details of the sharing policy, such as priority rules among the operators in case of conflicting assignments. It is worth noting that the definition of efficient sharing policies is out of the scope of this work. For the sake of simplicity, we focus on orthogonal sharing, which is immediate to describe and does not require to detail any power control policy for shared frequencies. However, as the code developed is entirely modular, an extension to non-orthogonal sharing would be immediate. Moreover, only competitive sharing was modeled, leaving the issue of identifying efficient and collaborative sharing mechanisms for a future analysis, possibly with more advanced game theoretic instruments.

Finally, besides introducing the details of the software extensions implemented within ns-3, this contribution also provides the results of an extensive simulation campaign meant to assess the effectiveness of the resulting simulator as a benchmark for testing spectrum sharing algorithms. A simple sharing algorithm is employed, and the evaluation of the modified version of the simulator in terms of computational requirements is given as well. The results confirm the ability of such a software instrument to give realistic assessments of the usefulness of spectrum sharing, and at the same time motivate further efforts with game theoretic approaches to implement efficient sharing algorithms where collaborative sharing is sought.

The rest of this Section is organized as follows. In Section 3.1.1 we review related works on simulation platforms for spectrum sharing analysis. In Section 3.1.2 we describe the system model, detailing the theoretical rationale behind the spectrum sharing characterization, while in Section 3.1.6 we discuss the modifications applied to the software architecture. In Section 3.1.9 we outline the simulation scenario and in Section 3.1.10 we present numerical results to validate our proposed contribution.

3.1.1 Related Work

Simulation platforms are a very common reference point to test protocols and assess the network performance, in particular for all those scenarios where the mathematical analysis becomes complex or cannot produce a solution in closed form.

In the literature, most of the work dealing with complex network systems include a performance evaluation part which leans on a simulator. This can be either a single-purpose simulator, specific to the scenario under investigation, or an adapted version of a general-purpose simulator. We believe that the latter alternative better fulfills scientific generality and reproducibility of the results, and enables future developments of individual findings. However, the software instruments chosen by the scientific community to this end are quite heterogeneous, from extremely general software platforms like SIMLIB [67] or MATLAB [68], which are properly customized to the particular context under evaluation, to more application specific tools which refer to particular systems, such as OMNET++ [69].

One of the most used tool in the research community is the Network Simulator ns [6], whose latest version is ns-3. It is an open source, free software managed by an active community of developers. The whole Internet suite protocol stack is implemented together with the most important protocols at the transport, network, and datalink layers. Therefore, many different network scenarios can be created and simulated. One of the last implemented modules realizes LTE cellular networks [12]. The introduced framework enables the creation of Base Stations (called eNodeBs, or eNBs) and user terminals (called UEs) which can communicate with the eNBs. Most of the functionalities of the physical channel and medium access have been implemented. We aim at extending this basic framework by introducing the multi-cell scenario and allowing eNBs to share part of their frequencies in the downlink direction. This is particularly interesting when the eNBs are managed by different cellular network operators.

Although the problems of interference channels and spectrum sharing have been addressed in several papers, e.g., [70, 71], the scenario of inter-cell spectrum sharing has been considered in a small number of them so far, and even fewer papers have focused on multi-operator networks. However, since in current network deployments the coexistence of multiple operators in adjacent regions is quite common, it makes sense to investigate the efficiency of the spectrum division policies adopted in common practice. The interest in this

area has increased during the last years and has been involving not only researchers, but also telecommunication companies and regulatory bodies.

A first simple concept of spectrum sharing has been introduced and analyzed in [72]. Base Stations try to face their incoming requests first by using their initial spectrum, and then by exploiting frequencies not used by the others. Two algorithms for resource allocation are presented and evaluated, but the presence of a centralized network is assumed, together with a coordinating unit that manages the whole network. In [73, 74] the authors introduce the concept of resource sharing in broadband cellular networks and show its impact on achievable capacity and packet delay. In this case, the resources shared among the different cells are the time slots (time division multiple access is employed), and operators use their allocated slots to transfer data to their mobiles. While in [73] sharing is seen only as a "last resort" solution, in [74] a new way of implementing radio networks is explored where mobiles are always connected to the best base station, regardless of whether it belongs to their home operator or not. This point is quite far from the implementation that we present in this work, where the resource shared is the band and mobile terminals are always connected to their home operators.

Another paper where the inter-operator spectrum sharing context is taken into consideration is [75], where a game theoretic analysis is given for a cognitive context where operators are classified into primary and secondary. This is slightly different from the system modeled in our simulator, where eNBs are not supposed to have sensing capabilities and such a hierarchy is not present.

3.1.2 System Model

The focus of this work is spectrum sharing in OFDMA networks, particularly concerning the LTE standard [1]. As mentioned in the introduction, it is important to clarify the orthogonality of the access scheme in the pool of common frequencies: "orthogonality" means "impossibility of simultaneous usage by more than one operator." Considering that a non-orthogonal approach would require a longer discussion about the convergence of the contention for shared frequencies, and the introduction of power control mechanisms for the users (i.e., the eNBs), we will focus on the orthogonal sharing case. Note that, in any event, this choice is made only for the sake of simplicity and, thanks to the modularity of the pro-

posed framework, it can be easily extended to work with non-orthogonal sharing. Therefore, from this point on, we will assume that eNBs share orthogonally the pool of common frequencies, that is each frequency resource can be assigned to at most a single operator within an allocation time slot, in our case corresponding to the LTE subframe duration (1 ms).

Therefore, we focus on the definition of a modular framework developed to test different solutions and efficiently evaluate the performance in terms of throughput and execution time. The resulting software can be used to validate several sharing policies, possibly derived within a game theoretic analysis. In this work, we will show sample results for orthogonal competitive sharing. However, given the modular nature of the simulator, more complex game theoretic approaches can be framed, even resorting to dynamic games, Stackelberg games and so forth [58, 62].

The proposed framework can be divided into three parts. First of all, the provision of spectrum management parameters, i.e., physical details such as center frequencies, channel bandwidth, and sharing percentages (see Section 3.1.3). After that, local scheduling and resource allocation algorithms must be executed in each eNB in order to generate an allocation map, the downlink serving scheme, as detailed in Section 3.1.4. Finally, a virtual market is in charge of collecting this information and derive the serving schemes that must be adopted by each eNB, according to the chosen contention solving policy, as will be illustrated in Section 3.1.5.

3.1.3 Spectrum Management

Each eNB is associated with a set of physical parameters. Then, it has to be defined a set of frequencies on which we plan to let them interoperate. The policy behind such a cooperation agreement is out of the scope of this work, as it is more related to the economic agreement between the operators and their business models. However, along with different allocation and coordination techniques, it represents an interesting research topic and, thanks to this contribution, various approaches can be quantitatively evaluated. Figure 3.1 shows the scheme adopted to define the system sharing capabilities. According to the selected bandwidth percentage to be shared, the eNBs will allow partial access to UEs belonging to other domains.

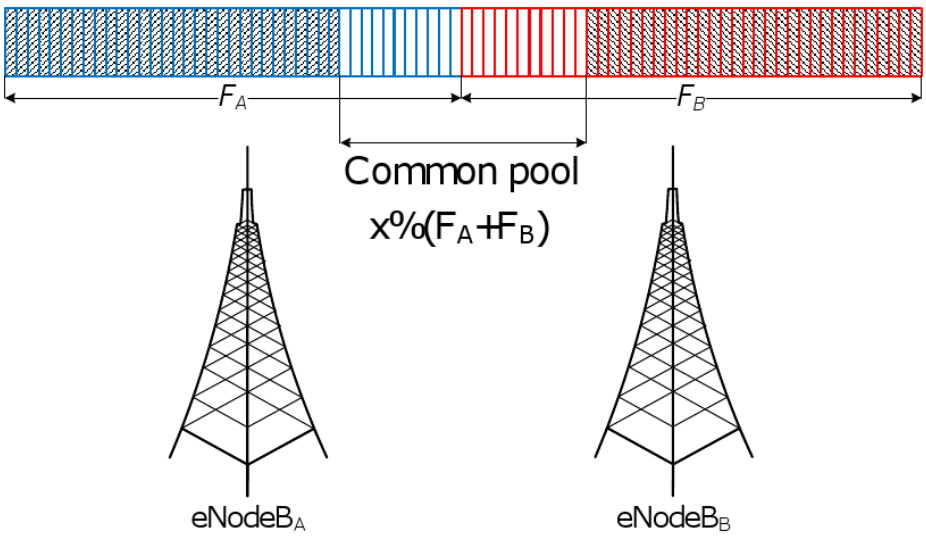


Figure 3.1. *Spectrum sharing*

3.1.4 Intra-Cell Allocation

The cell capabilities are fully characterized when the physical components have been defined. Then, we provide a joint scheduling and resource allocation algorithm to properly design a downlink transmission scheme. However, our goal is to integrate the proposed spectrum sharing framework for LTE systems into a simulation tool, ns-3. Conversely, the definition and the analysis of efficient game theoretic schemes which can be fed to this simulator are not directly investigated here, but are left for future work, possibly within a game theoretic context.

Here, two basic algorithms have been implemented and compared: on one the hand, *max throughput* represents an allocation scheme for which the resources are allocated to the UEs better perceiving the channel, without taking into account fairness among users. On the other hand a fair approach, denominated *fairness*, is proposed where the available system resources are distributed among the users in a Round Robin way, thus reducing the overall throughput, but increasing the average level of satisfaction of each UE. Figure 3.2 depicts a sample scenario, where 10 UEs and 10 resources, hereinafter referred as resource blocks (RBs), are considered. By selecting the first approach, *max throughput*, all the available resources are allocated to the UEs with the best channel quality indicator (CQI), as discussed

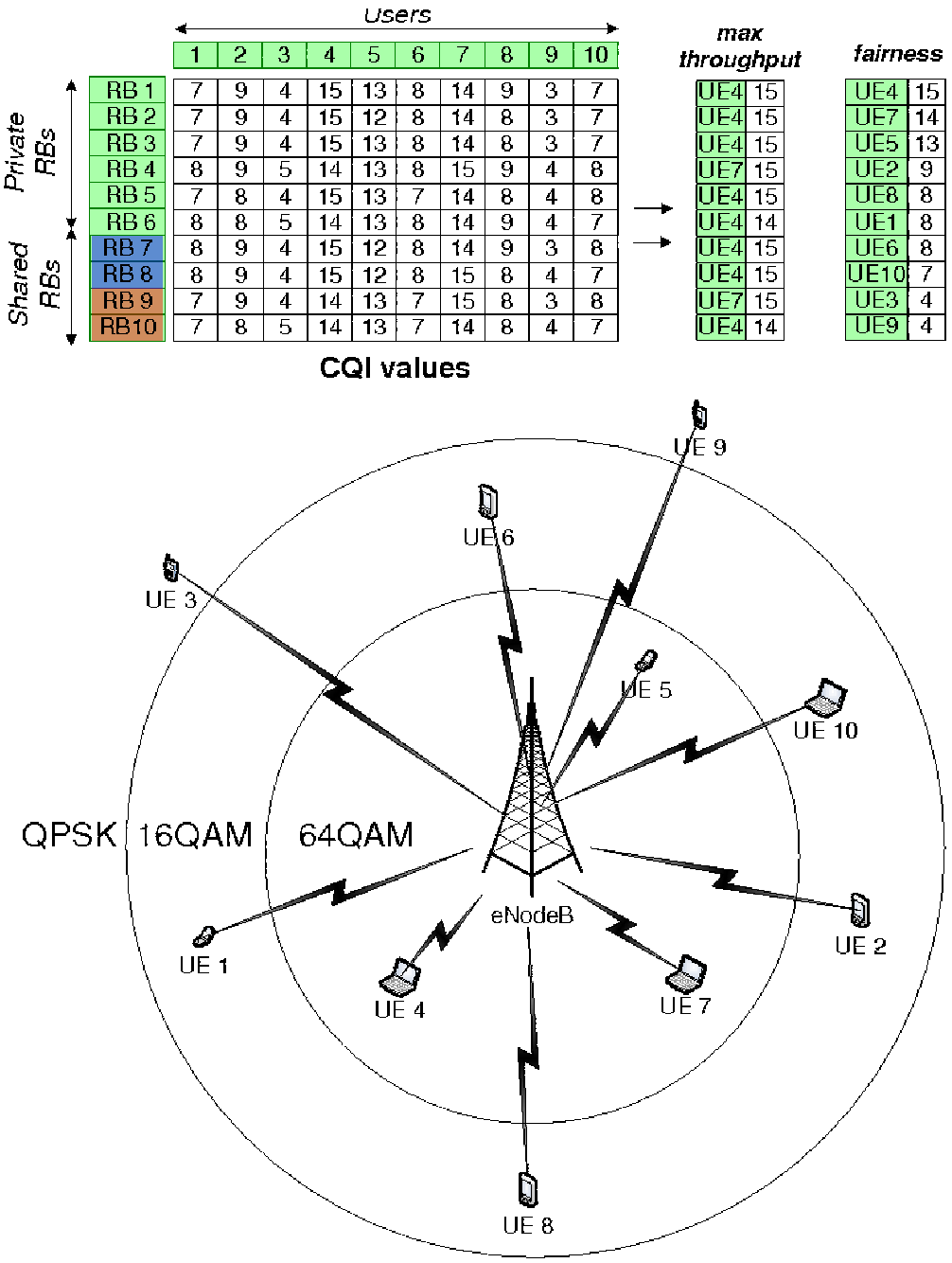


Figure 3.2. *Intra-cell allocation*

later in Section 3.1.10. Thus, by exploiting multiuser diversity, the system throughput can be very high. However, UEs with lower CQIs will never be served. Therefore, an additional technique has been introduced, i.e., the *fairness* mechanism. As visible in the figure, this approach will provide service to all the registered UEs. As in the previous case, each RB is

allocated to the best UE, but each user is allocated at least a certain amount of resources in order to prevent starvation. In particular, the distribution of the RBs happens in a Round Robin way with the pooling starting from the UEs with the best CQI and moving to those in a worse condition. During the first allocation round, each UE is given

$$TH_{min} = \left\lfloor \frac{N_{RB}}{N_{UE}} \right\rfloor , \tag{3.1}$$

where N_{RB} represents the total number of RBs, and N_{UE} is the number of registered UEs requesting admittance in the system. Once that this minimum threshold has been guaranteed to all the users, then all the remaining RBs are distributed again by adopting a Round Robin policy and assigning 1 RB per UE starting first from those with better channel conditions. In the proposed example, the threshold in Equation (3.1) is equal to 1, so all the UEs will be allocated a single RB.

3.1.5 Inter-Cell Multi-Operator Coordination

The sharing contention policy is implemented in a separate module, hereinafter called *virtual market*. The relevant class (we refer to an Object-Oriented Programming, or OOP, paradigm) implements an arbitration rule which defines how the operators bargain the access to the common portion of the licensed spectra. Any complex strategy can be implemented within this class, possibly involving further extensions. In particular, this may be the place where to implement, in an entirely modular manner, some procedures inspired by game theoretic principles. As shown in Figure 3.3, each eNB, after generating its own allocation map, sends it to the *virtual market* that gathers all the cells' allocation information and rearranges the allocation map according to the sharing policy. For the sake of simplicity, we propose immediate implementations of scheduling and resource allocation algorithms, as well as a simple procedure to handle the contentions among operators. Each eNB is assigned a *priority* value per frequency sub-channel, defined as

$$PR_{eNB_j, RB_{pool,i}} = \begin{cases} p, & RB_{pool,i} \in F_{eNB_j} \\ 1-p, & \text{otherwise} \end{cases} , \tag{3.2}$$

where $j \in \{1, ..., m\}$ represents the eNB identifier, m is the total number of eNBs involved in the sharing process, $p \in [0, 1]$ is the priority level given to the eNB, $F_{eNB_j} = \left\{RB_{j,1}, ..., RB_{j,n_j}\right\}$, n_j is the total number of RBs available at eNB$_j$, and $RB_{pool,i} \in F_{eNB_j} \cup ... \cup F_{eNB_m}$. In other

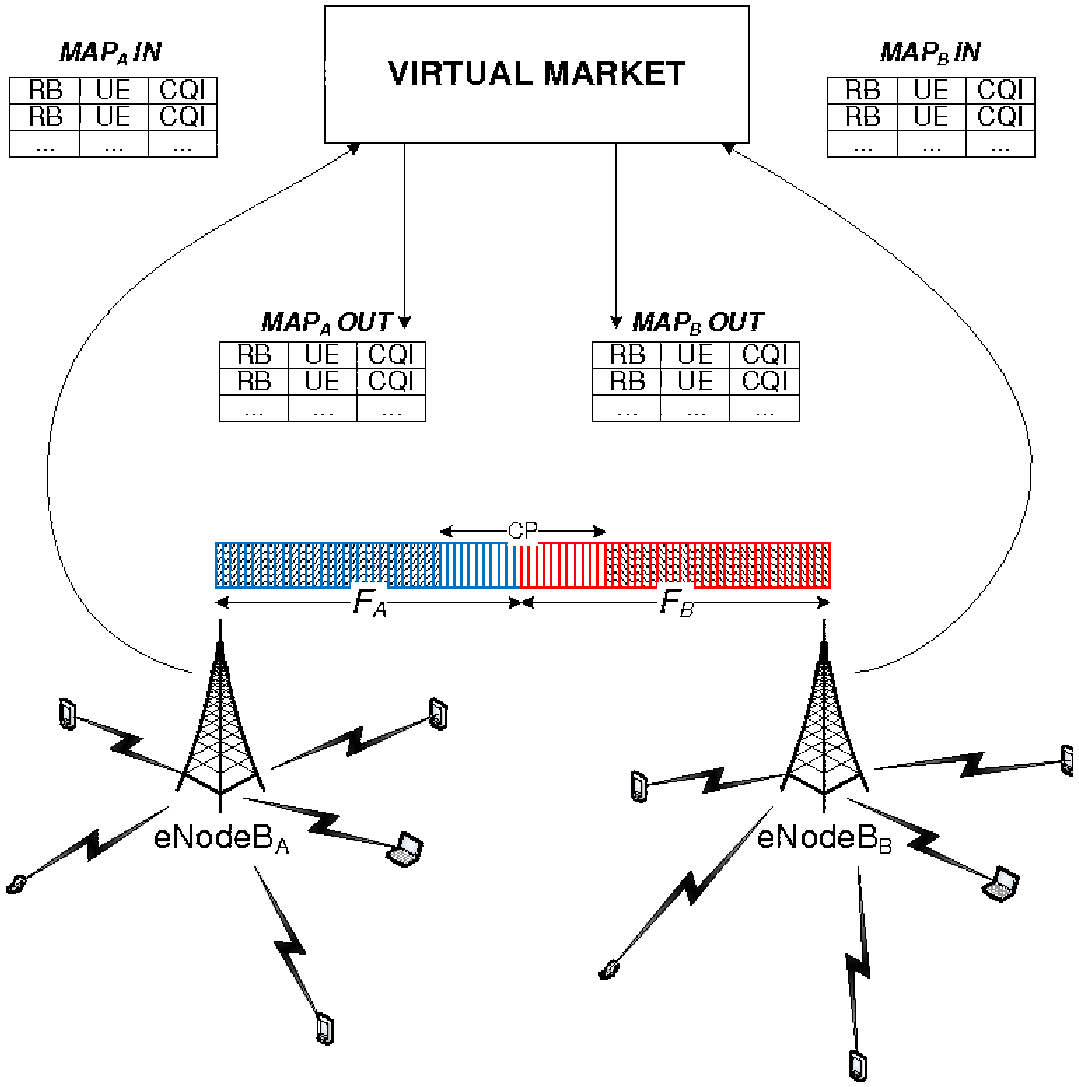


Figure 3.3. *Inter-cell coordination*

words, shared resources are assigned based on these priority levels; obviously, the UEs associated with eNB$_j$ will always have higher priority than all other competing users. In our work, the proposed approach is even simpler: we assume $p = 1$ and $m = 2$, so an eNB will assign to its UEs the shared resources belonging to the *competitor* eNB, referred to as eNB$_c$, only if these are not allocated to UEs belonging to eNB$_c$. Thus, when multiple players request the same resource, only the one with the highest priority will get it. The others end up with no assignment, which is in general inefficient.

We stress that this general strategy is not given as an optimal allocation, which ought

to be derived from a (game) theoretic perspective. Rather, such sub-optimal policies serve to show the effectiveness of our software implementation. Moreover, it can be thought of as a characterization of the inefficient Nash equilibria in the games with *competitive* sharing, while the goal of spectrum sharing should rather be a *collaborative* assignment of frequencies. Thus, our reference allocation policy correctly reflects that, if the whole common pool is shared competitively, in the long run only inefficient and unfair allocations will be achieved. However, we also remark that more efficient solutions derived through game theory, either available in the literature or originally developed, can be tested and validated within the modular framework proposed in this Section, so as to determine the choice that better suits the operator needs.

3.1.6 NS-3 LTE Extension

Here we provide the technical details about the extension of the ns-3 original code that we implemented. The reference implementation of LTE to which we applied our modifications is the one presented in [12] and included in ns-3 starting from the release ns-3.9. All the changes and/or additions done in the code aim at enabling the multi-cell multi-operator LTE scenarios and defining a flexible architecture for inter-operator downlink spectrum sharing to test sharing policies. In this way, we have prepared a framework that can be used as is or extended again to simulate a broader range of situations. This is made possible by the extreme modularity of ns-3. It is also worth mentioning that our extension is entirely backward compatible with the previous releases of ns-3. The code is publicly available [76].

3.1.7 Multi-Cell Multi-Operator Scenario

The first step for the definition of a multi-cell scenario is the allocation of several objects of the class *LteHelper*, one for each cell. Such an object contains all the information needed to create and manage a cell, including a reference to the eNB and all its UEs.

An important modification that was required with respect to [12] regards the management of the time by of each eNB. The PHY layer of eNBs and UEs is implemented in the classes *EnbLtePhy* and *UeLtePhy* respectively, where the operations of signal transmission and reception are managed. Both these classes are derived from the base class *LtePhy*, which contains all the properties and methods common to both types of nodes. Among the others,

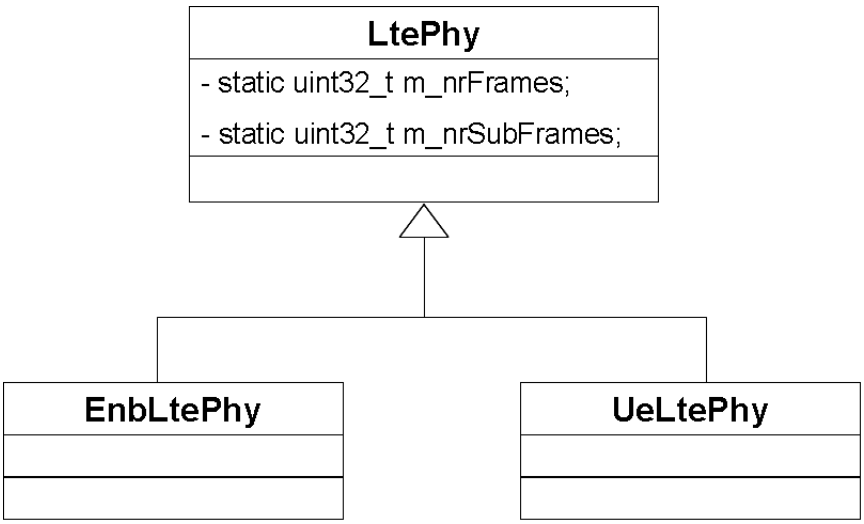


Figure 3.4. *UML class diagram for the PHY layer in ns-3 LTE*

this class has in its private fields two *static* counters, one for the frame index ($m\_nrFrames$) and another for the subframe index ($m\_nrSubFrames$) within the current frame (see Figure 3.4). They are incremented every time a new frame/subframe is started, a functionality that is implemented by the *EnbLtePhy* class, methods *StartFrame* and *StartSubFrame*, since it is up to the eNB to decide when to start the new frame/subframe. In a certain sense, they represent the timeline for the base station (and for the related cell as well). In a multi-cell scenario there are many eNBs, each with its own *EnbLtePhy*, but the values of the counters are common since they are declared as static. This means that if n eNBs increment the counter, then this will have a value n times greater than what it should have. Two possible solutions are available to solve this problem: either only an eNB increments those counters (a kind of master eNB) or each one of them has its own counters and increments them independently. In our implementation we have chosen the latter, thus each eNB has its private view of the time index. In our case, they are all synchronized, hence they start each (sub)frame at the same time, but this choice does not prevent further more realistic variations where the eNBs are non-synchronized.

3.1.8 Downlink Spectrum Sharing

The implementation of the inter-cell downlink spectrum sharing involved several modifications with the aim to first introduce the spectrum sharing data structures and then develop the support to the conflict resolution mechanisms.

We made eNBs aware of the additional sub-channels they can use for downlink resource allocation by modifying the implementation of the *EnbLtePhy* and *UeLtePhy* classes. Originally, each one of them was assigned a vector of sub-channels which represents the available resources they can use. In our implementation we have associated to each node an extended vector containing not only the sub-channels originally assigned to it, but also those that the other eNBs are willing to share together with the sub-channel priority access information (i.e., a value indicated the level of priority of that node on each sub-channel). This vector of frequencies is the one actually used for all the allocation, transmission and reception operations. In particular, the resource allocation is performed by the class *PacketScheduler*. To test some new scheduling and allocation policies it is sufficient to extend it and override the method *DoRunPacketScheduler*, which is the routine called at the beginning of each subframe when a new set of packets must be selected for transmission to the UEs.

Once that the support for sharing the frequencies among the eNBs (belonging to different operators) has been implemented, thus putting them in the position to choose among a broader range of resources, it is necessary to define a software architecture for the communication among the base stations and the synchronization on the access to the common pool. Each eNB determines its allocation map independently, according to an internal scheduling and resource allocation policy. Then, we decided to manage the inter-cell communication (i.e., map exchange) and trading (i.e., conflict resolution) by introducing a virtual entity. In a real system, this phase requires that the eNBs communicate (e.g., through a backhaul) and agree on a final allocation map to which all of them must adhere. In our implementation, this communication is abstracted by such virtual entity, which is an object defined as an instance of the class *VirtualMarket*. At the beginning of each subframe, it receives the resource allocation maps proposed by all the eNBs (competitors) and decides the final map according to some policy. Developers can implement whatever policy they need, by just modifying that class or extending it and overriding the method that deals with contention resolution, i.e., *GetAllocationMap*. The *VirtualMarket* has a collection of eNB entities, which can communicate with it through its public interface. In Figure 3.5 an example of such a communication is shown by means of an UML sequence diagram, which is also able to catch the temporal dimension of the activity. The particular communication scheme shown is the one described in the previous Section for conflict resolution, i.e., the priority mechanism. An iteration is

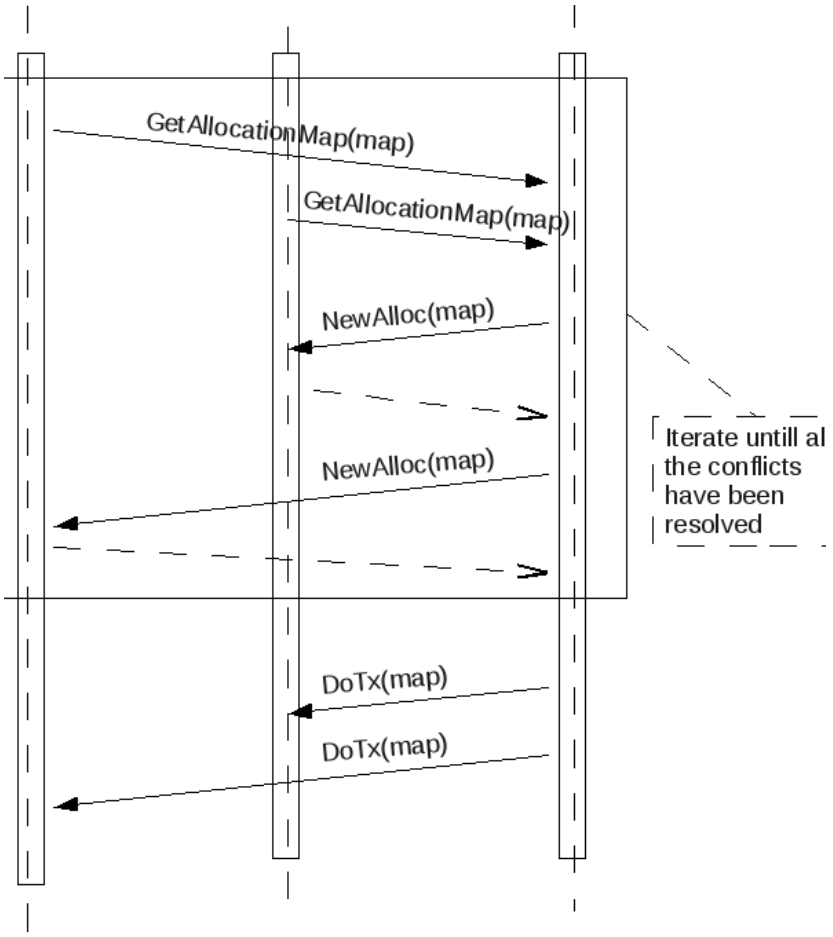


Figure 3.5. *Sequence diagram for allocation conflict resolution*

indicated since every time a competitor cannot use a sub-channel for some UEs (i.e., it loses the contention), it is invited to reschedule those UEs on other free resources (if any). Of course, this is just an example given to describe how the architecture works.

3.1.9 Simulation Scenario

Here we describe the simulation scenarios that we considered to validate the software architecture presented above. With that aim, we tested the sample sharing algorithms discussed in the previous Sections. However, the focus is on the performance and usability of the extended simulation platform instead of the algorithms themselves. These are not expected to be optimal, they are to be considered as possible examples that anybody can replace. The results obtained cannot be compared with those coming out from a test bed in

Parameter	Value
Center Frequencies	2.115 GHz (eNB_0), 2.125 GHz (eNB_1)
eNB Downlink Channel Bandwidth	10 MHz
Subcarrier Bandwidth	15 kHz
Doppler frequency	60 Hz
$RB_{bandwidth}$	180 kHz
$RB_{subcarriers}$	12
$RB_{OFDMsymbols}$	14
eNodeB TX power per sub-channel	26.98 dBm
Noise figure (F)	2.5
Noise spectral density (N_0)	-174 dBm/Hz
Macroscopic pathloss	$128.1 + (37.6 \cdot \log_{10}(R))$ dB
Shadow fading	log-normal ($\mu = 0, \sigma = 8$ dB)
Multipath	Jakes model with 6 to 12 scatterers
Wall penetration loss	10 dB
Simulated interval	10 s
Frame duration	10 ms
TTI	1 ms

Table 3.1. *Main system parameters*

a real environment. Nonetheless, they are meaningful and match the expectations.

The scenario consists of two eNBs (i.e., eNB_A and eNB_B) positioned in the same field, both with a coverage of 1500 m. A certain (fixed) number of UEs, characterized by low mobility, is registered to each eNB and generates traffic. The UEs are distributed within the coverage area of the corresponding base station according to a two-dimensional uniform distribution. For the sake of completeness, we considered two types of situation: (i) both cells have the same number of UEs and the same traffic load (*symmetric cell load*), and (ii) the two cells have a different number of UEs and thus a different total traffic load (*asymmetric cell*

load). In the latter case we supposed that the eNB_A has 40 UEs to serve while the eNB_B has a variable number, and each one of them is allocated a maximum number of 2 RBs according to a *max throughput* criterion.

As previously mentioned, each user perceives a different quality of the channel according to its position and other radio propagation related factors. According to the corresponding level of SINR, each UE calculates the CQI for each RB and sends it back to the eNB through a control channel. This is information is used for the selection of an adequate Modulation and Coding Scheme (MCS). As reported in Table 2.5, LTE technology provides multiple options [47]. Hence, each CQI determines the number of bits that can be transmitted in a single RB:

$$b_{RB} = RB_{subcarriers} \cdot RB_{OFDMsymbols} \cdot ECR_{CQI} \cdot b_{mod}, \tag{3.3}$$

where $RB_{subcarriers}$ and $RB_{OFDMsymbols}$, that represent the number of sub-carriers per RB and the number of OFDM symbols per RB respectively, are provided in Table 3.1, together with the main system parameters, while b_{mod} is the number of bits per symbol, determined by the type of modulation adopted:

$$b_{mod} = \begin{cases} 2, & \text{QPSK} \\ 4, & \text{16QAM} \\ 6, & \text{64QAM} \end{cases}. \tag{3.4}$$

The simulation campaign is executed to investigate the reliability of the proposed framework, in terms of cell sum capacity, aggregate throughput, and execution time. In fact, as will be extensively detailed in Section 3.1.10, the system performance behavior follows the trend that we expected: on the one hand, increasing the number of UEs in the system corresponds to a throughput increase, while on the other hand increasing the sharing percentage induces a smooth decrease of the system throughput, according to the simple conflict resolution approach implemented. More specifically, the performance metrics taken into consideration are:

—**Cell Sum Capacity**, which represents the sum of the Shannon capacity reached in a cell on each sub-channel. It is given by

$$C = \sum_{i=1}^{N_{UE}} \sum_{j=1}^{N_{RB}} \left(B \cdot \log_2(1 + SINR_{ij} \cdot \delta_{ij})\right), \quad \delta_{ij} = \begin{cases} 1, & \text{UE}_i \text{ allocated to RB}_j \\ 0, & \text{otherwise} \end{cases} \tag{3.5}$$

where N_{RB} is the total number of RBs that can be exploited in the downlink of the cell (i.e., including those shared by the other eNBs), and $SINR_{ij}$ is the $SINR$ at UE$_i$ on RB$_j$.

—**Cell Sum Throughput**, which represents the aggregation of the data rates delivered to all UEs, and is computed as

$$T = \frac{\sum_{i=1}^{N_{RB}} b_{RB_i}}{TTI}, \tag{3.6}$$

where b_{RB_i} represents the resource block size referred to the ith RB, and N_{RB} is the total number of RBs available in the system.

—**Execution time**, which represents the time required for the execution of a simulation run. We expect an increasing behavior in the number of UEs and in the sharing percentage because of the higher computational complexity needed to perform a greater number of operations. The duration of a single run is of 10 ms. According to the Monte Carlo simulation method, 1000 runs have been executed for each parameters combination in order to have a good characterization of the channel behavior, with each UE replaced at each repetition in order to simulate the mobility. The reference machine is a computation server with 48 Pentium CPUs, 64 GB RAM and running GNU/Linux Ubuntu 11.04 as the operating system. It must be noted that, even though the number of available processors is considerable, the ns-3 software is inherently non parallel and thus all the runs were always executed on a single processor as if it were a single CPU machine. The only advantage of having more CPUs derived from the possibility to execute several simulations in parallel, one for each different combination of the input parameters (i.e., number of UEs and sharing percentage).

3.1.10 Numerical Results

Figures 3.6–3.7 show the performance in terms of sum capacity and throughput achieved by each cell for both *max throughput* and *fairness* intra-cell allocation algorithms for a different number of cell users. In this case we are considering the symmetric cell load, thus both cells have the same number of UEs and are statistically equivalent. For this reason only the results for one of them are reported. As expected, the actual throughput value is significantly below the cell sum capacity, as defined in Equation 3.5, which represents the upper bound on the data rate achievable with such a channel condition. The actual amount of data transmitted depends on the ECR and is upper bounded by the cell sum capacity. However, the behavior of both sum capacity and throughput as functions of the sharing percentage for

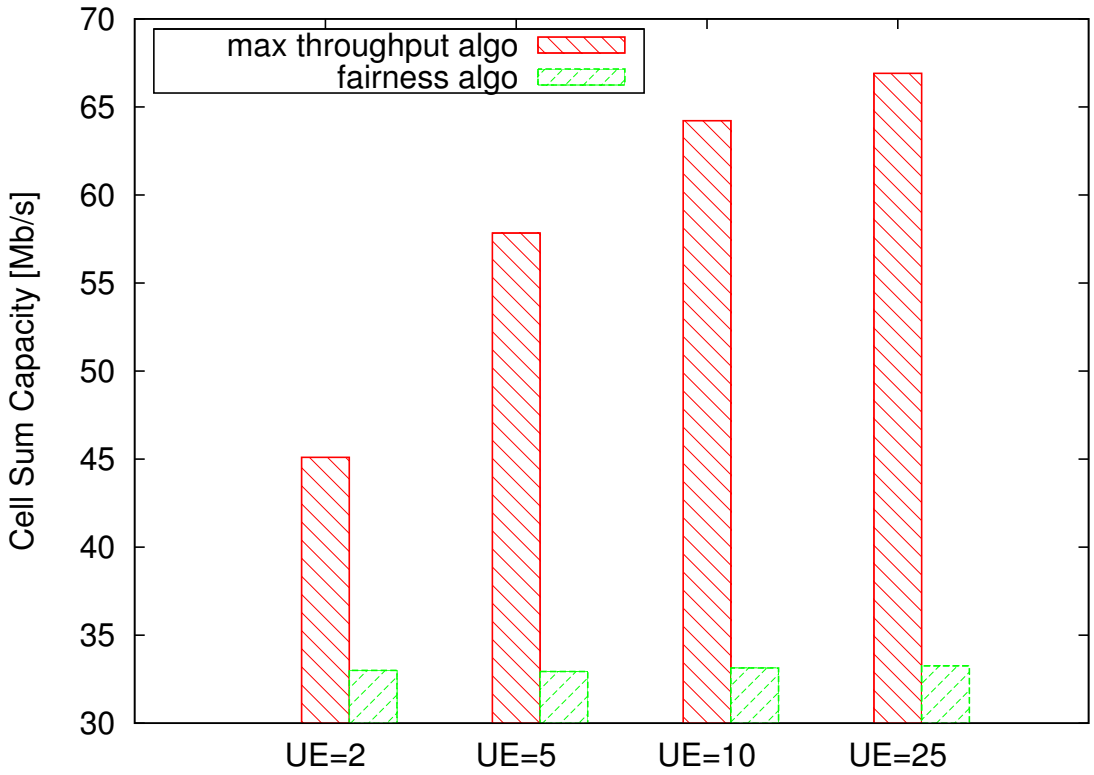


Figure 3.6. *Comparison of the Cell Sum Capacity for the* max throughput *and the* fairness *allocation algorithms, with a sharing percentage of 100%*

different numbers of users is qualitatively similar, meaning that they differ only by a scaling factor due to the use of real coding and modulation schemes.

In both figures the trade-off between the *max-throughput* and the *fairness* allocation algorithms is clearly shown. The former always makes the system reach a greater performance because the application of a fair scheduling policy requires the allocation of RBs also to the UEs with lower CQI. This is true for all the values of UE.

Another important effect that can be noted from Figures 3.6–3.7 is the increment of both performance indices with the number of UEs. As expected, this is due to the multiuser diversity effect: the greater the number of UEs, the higher probability that for each subchannel there is at least one of them with a good CQI. Of course, this might lead to some (short term) unfairness in favor of the users with a good channel quality. On the contrary, if the fairness constraint must be taken into consideration, then the effect of the multiuser diversity is significantly reduced. That is the reason for which in both figures, the increment

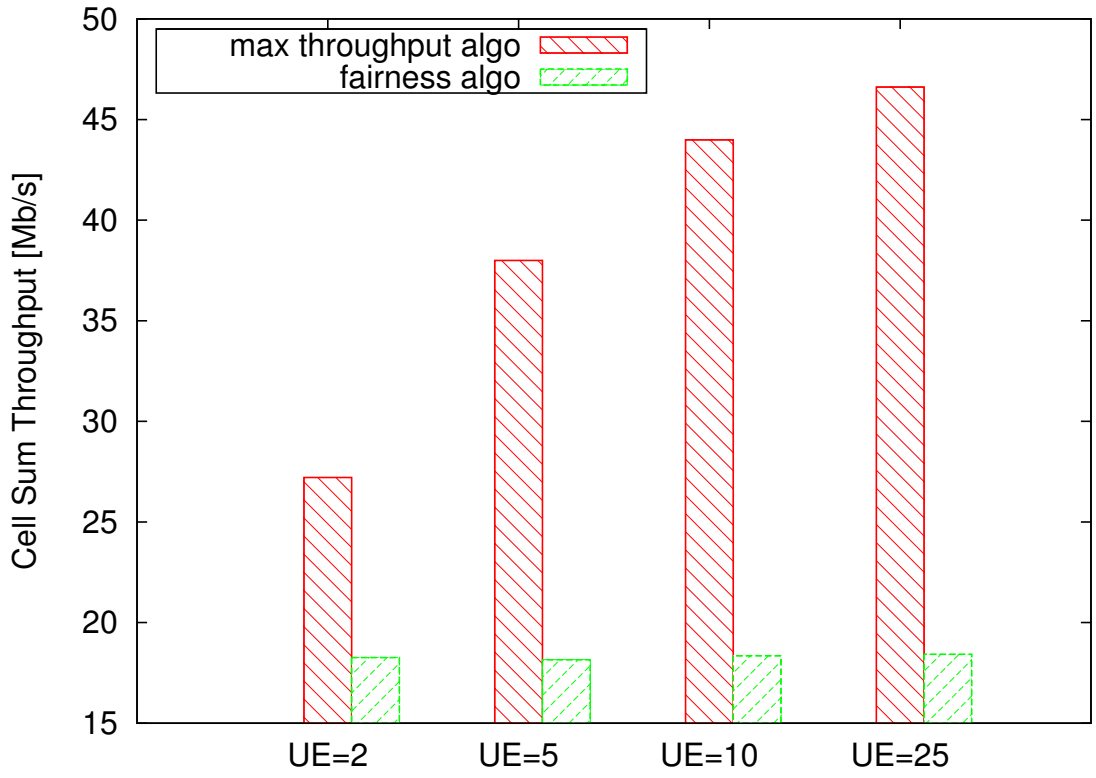


Figure 3.7. *Comparison of the Cell Sum Throughput for the* max throughput *and the* fairness *allocation algorithms, with a sharing percentage of 100%*

of the performance indices for the *fairness* approach is almost negligible. For a possible discussion of this trade-off in a game-theoretic perspective see [77,78]. Moreover, the marginal increment of efficiency decreases when a certain user density has been reached in the cell. When more users are in the system, then for almost all the sub-channels there is a user with good CQI. Thus, a saturation effect appears.

To sum up, the results validate the reliability of our model. Thanks to the modularity introduced, the contention technique can be adapted to different needs, and in particular to pursue a cooperative sharing, where system capacity and throughput increase when the spectrum sharing percentage becomes higher.

In Figures 3.8–3.9 the sum capacity for both cells is shown in the asymmetric load scenario. In this case, since the total amount of traffic is different, the two cells are no longer statistically equivalent. The two figures show the variation of the performance index when several values of sharing percentage (parameter α) are considered. In such a scenario the

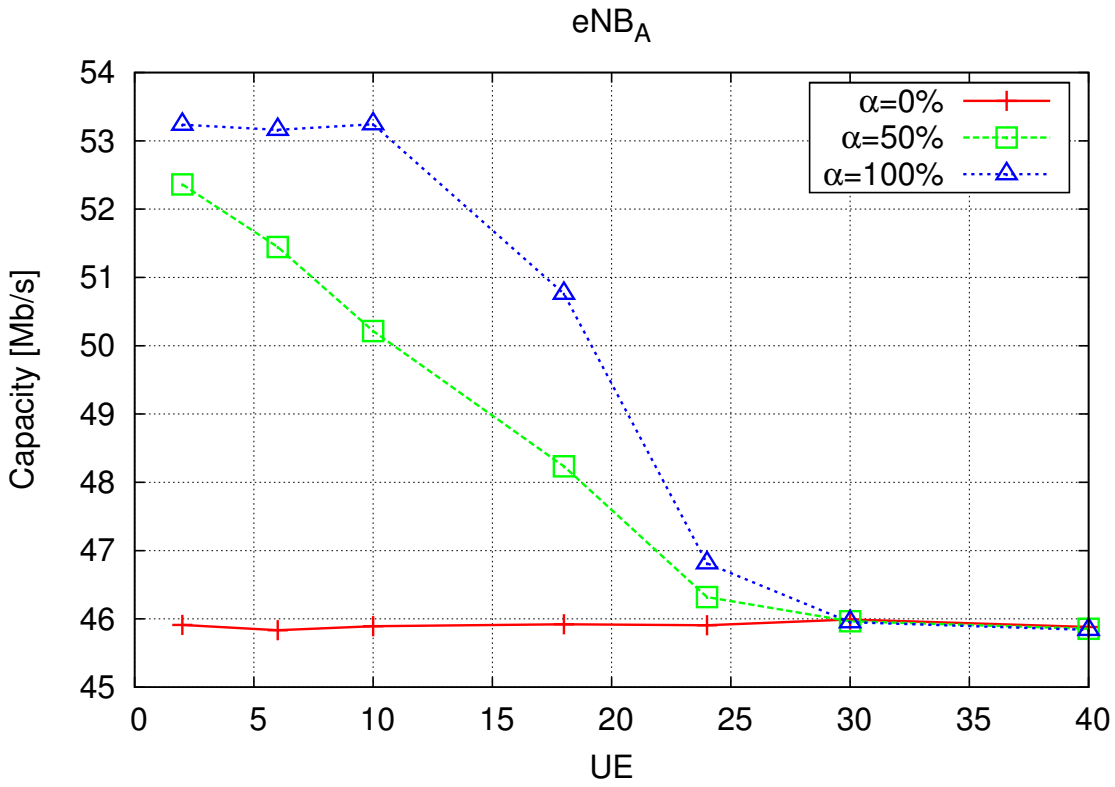


Figure 3.8. *Cell Sum Capacity for eNB_A versus the number of UEs in cell B in the asymmetric load scenario*

spectrum sharing gain can be better appreciated since the eNB overloaded can opportunistically exploit the RBs not used by the other. Of course, when $\alpha = 0\%$ the total capacity achieved in the first cell does not depend on the number of UEs in the second, since it can never use any of the spare resources, thus resulting in a remarkable waste of spectrum efficiency. This means that the eNB_A cannot serve all its 40 UEs, which would require the access to 80 RBs while only 50 RBs are available to it. On the other hand, when the sharing percentage increases the first eNB is entitled to use some of the resources of the second one if this does not need them. This implies an average increment in the total capacity of eNB_A with α. Of course, also eNB_B is entitled to use some of the sub-channels in eNB_A's original pool but, since this one is in saturation, it is very unlikely to find some spare resources and thus it will end up in using mainly its portion of the spectrum. Therefore, the sum capacity in cell B increases in the number of UEs because more users are served but it does not vary significantly in α. It must be noted that the amount of this increment decreases at a

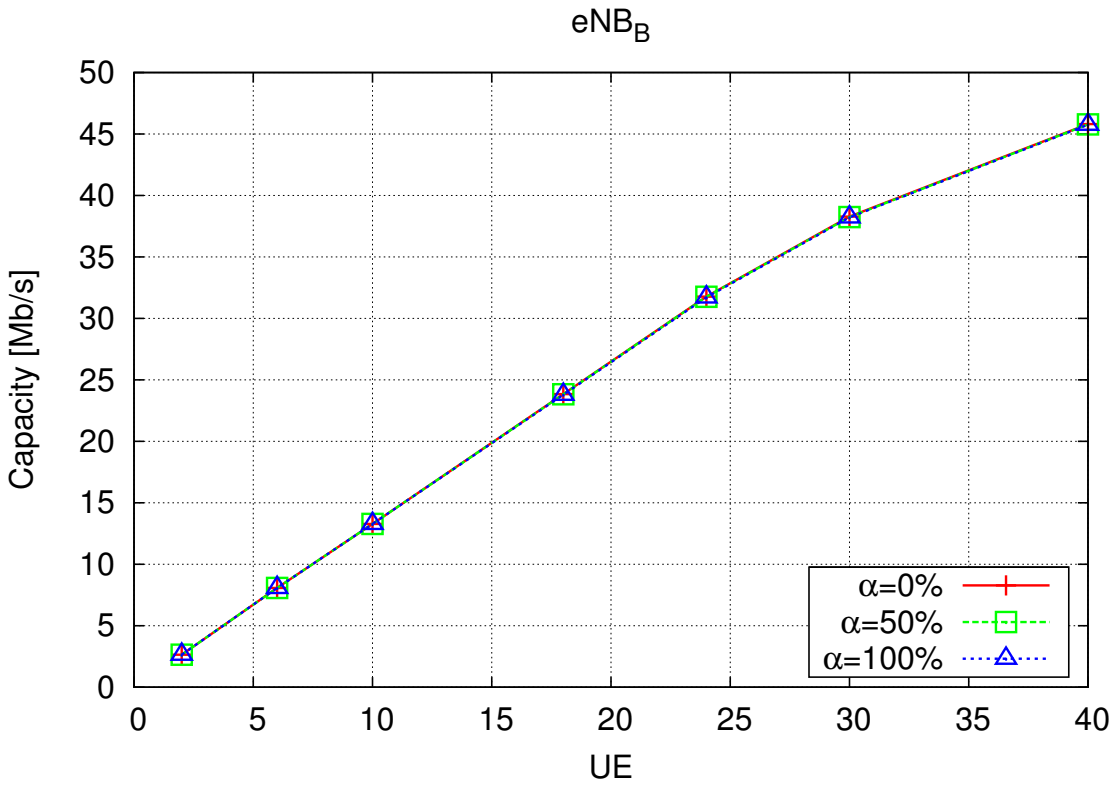


Figure 3.9. *Cell Sum Capacity for eNB_B versus the number of UEs in cell B in the asymmetric load scenario*

certain point, i.e., after $UE = 25$. Indeed, while below such threshold all the users can be served, beyond that value it is not possible to serve all of them (consider that the other cell is in saturation, so no spare frequencies can be found) and the only degree of freedom that eNB_B can exploit regards the scheduling of an UE instead of another one for the multiuser diversity. Regarding cell A, the total capacity for $\alpha = 50\%, 100\%$ decreases with the number of UEs in cell B since the greater the load in that cell, the greater the number of RBs needed and thus the lower the number of spare resources that can be accessed by eNB_A (consider that a priority scheduling policy is adopted, so if eNB_B needs one of its sub-channels it will get it disregarding eNB_A's requests). It is useful to remark that the aim of the work here discussed is the validation of the software architecture, and not the identification of the optimal sharing policy. A joint gain might be achieved by introducing some coordination between the base stations, according to what stated by the cooperative game theory [79].

Finally, in Figure 3.10 the execution time resulting from a wide range of simulations is

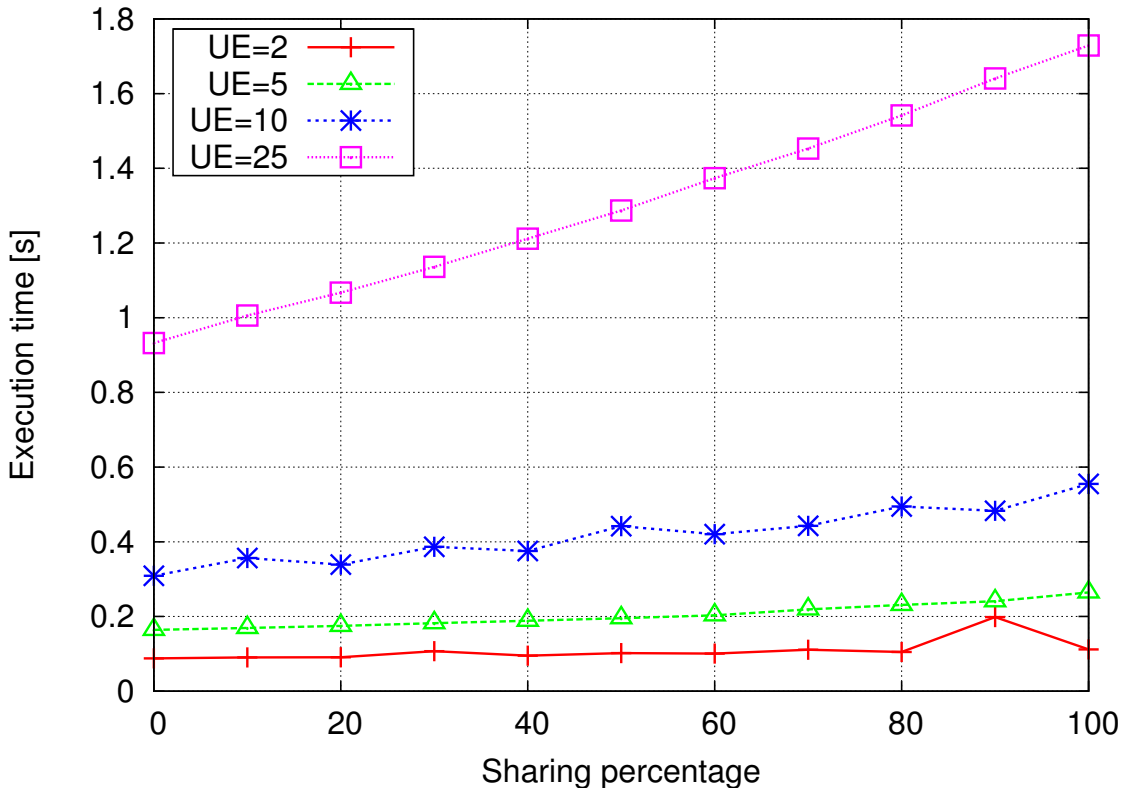


Figure 3.10. *Execution time*

shown. It refers to the symmetric load scenario with a *max throughput* allocation algorithm. There is an obvious increment of the time required to run the simulations with respect to the increase of the number of UEs and spectrum sharing percentage. The simulation of more UEs requires more memory and computational resources to store and manage all those objects and thus a larger execution time. On the other hand, a greater number of shared resources implies more contention and thus more iterations of the conflict resolution algorithm. Execution times also increase for greater sharing percentages since the (intra-cell) resource allocator has a greater number of degrees of freedom. Moreover, we remark that the tracing option was enabled in order to log the performance indices and calculate statistics. Disk accesses are quite time consuming and can slow down the execution by more than 10 times the normal duration. However, in spite of all these points, the computational complexity scales almost linearly with the number of users and the sharing percentage, and can thus be considered acceptable for realistic and detailed simulation campaigns.

3.2 Cognitive Mobility Prediction in Wireless Multi–Hop Ad Hoc Networks

In a MANET, the TCP throughput shows degraded performance profiles. This is due to the congestion control and avoidance mechanisms which were modeled for wired networks, and thus assuming i) static nodes, i.e., a stable topology, and ii) packet losses caused by buffer overflow, i.e., network congestion. Conversely, in a MANET there may be mobile nodes and rapidly changing channels. Therefore, most packet losses are due to channel errors or link layer contention, as pointed out in [80]. Furthermore, also the network topology can rapidly change due to mobility, which can cause a temporary link breakage and seriously affect the TCP performance, since this protocol was not designed to keep into consideration this kind of events.

In this work, we aim at designing a light-weight cross-layer framework to counteract the aforementioned TCP limitations, and to propose a valid solution to some specific problems of the TCP and routing protocols that we observed in highly mobile scenarios. Our approach, differently from many other works in the literature, is based on the cognitive network paradigm [81]. It includes an observation phase, i.e., a training set in which the network parameters are collected; a learning phase, in which the information to be used for network improvement is extracted from the data; a planning phase, in which the strategy, in terms of protocol modifications exploiting the learned information, is decided; and an acting phase, which corresponds to running such strategies in the network and observing the communication performance improvement. The general workflow behind our contribution is structured as follows.

1) We observe the overall TCP throughput degradation by means of simulations performed with *ns3* [6], an open-source discrete-event network simulator for Internet systems. We consider both a static scenario and a mobile scenario, and we vary the channel characteristics to simulate different realistic scenarios. We identify a set of critical network states, and in this phase we also collect the values of some network parameters as a function of time, which are stored in a training dataset.

2) We exploit the training dataset to learn the probabilistic relationships among the communication parameters, and we organize this probabilistic information in a Bayesian net-

work (BN). The BN is designed in order to provide real-time information on the mobility status of the network.

3) We define a set of actions to be adaptively taken in order to address the problem of each critical network state, once the network state has been inferred by means of the BN.

4) Finally, we design a cross-layer framework that allows to dynamically take actions at the TCP and IP levels, i.e., to apply the corresponding strategy defined in 3). We also perform a simulation campaign to show the performance improvements in terms of increased average throughput and decreased length of the outage intervals, i.e., the time intervals in which the communication is frozen due to topology or network problems.

The rest of the Section is organized as follows. In Section 4.1.1 we briefly overview the state-of-the-art dealing with the degradation of TCP performance over MANETs and the proposed solutions. In Section 4.1.2 we detail the network scenario and the main communication problems we observed. In Section 4.1.3 we describe our system model, including the probabilistic graphical model approach for learning and the strategies to address the main communication problems we identified. Then, in Section 4.1.6 we validate our framework through a simulation campaign and we show the performance improvements.

3.2.1 Related Work

In wireless networks, TCP suffers from poor performance because of packet losses and corruptions due to the wireless channel [82]. In [83, 84], a comprehensive overview of the main limitations of TCP over MANETs is provided, and the performance of different TCP techniques is evaluated via simulation. We report here a few examples addressing these issues from different perspectives. An adaptive congestion control mechanism based on link layer measurements and performed by each node along the path is proposed in [85]. In [86], a dynamic slow start threshold mechanism, as a function of the number of outstanding packets, is designed. In [87], the maximum congestion window is adapted as a function of the channel bandwidth and the packets' delay profiles; in [88], instead, a comparative analysis of several end-to-end, link-layer or split-connection techniques to improve the performance of TCP over lossy wireless hops is provided. Alternatively, in [89] some reliable transport protocols, optimized to better support MANETs, are detailed. Moreover, other contributions deal with the TCP degradation due to node mobility [90–92], where unnec-

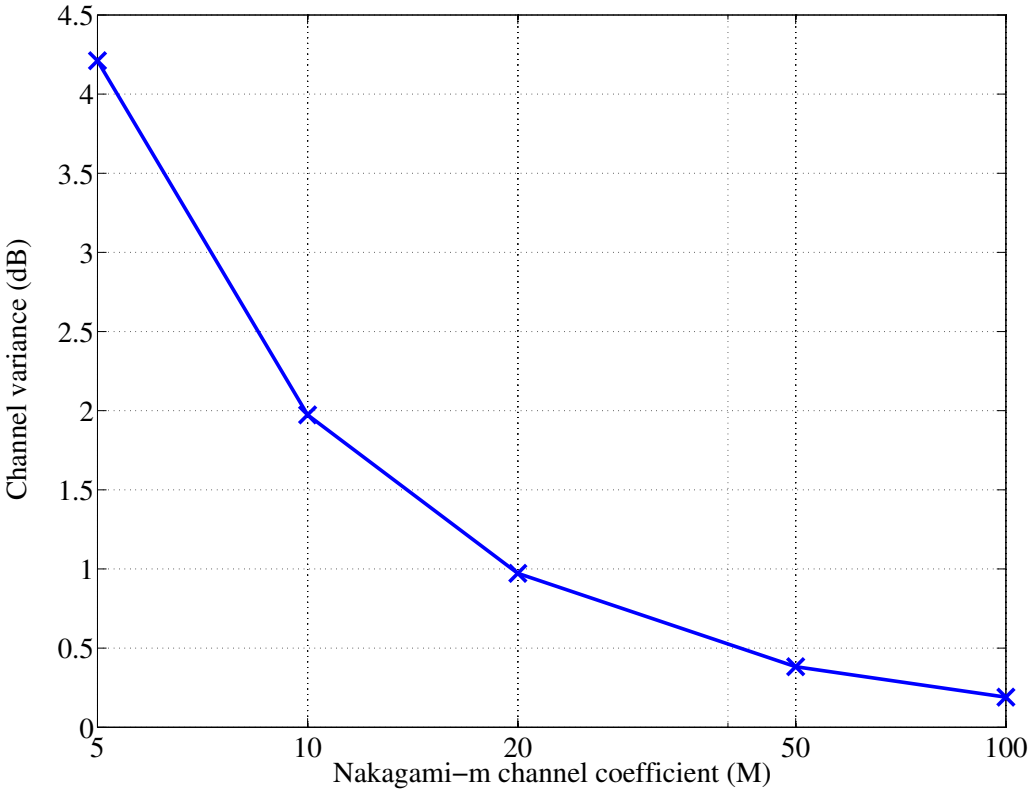
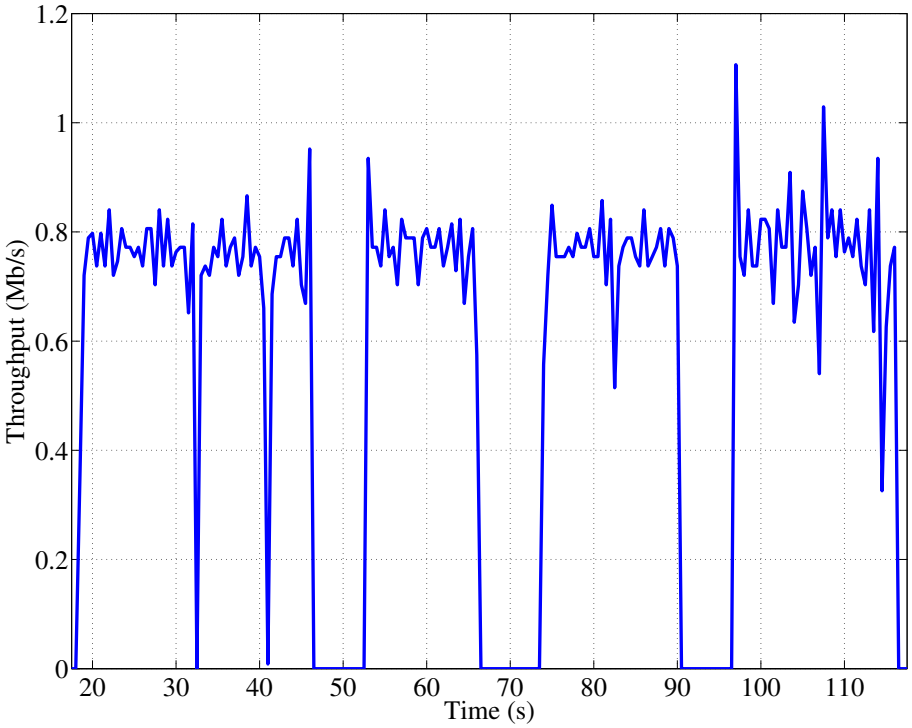


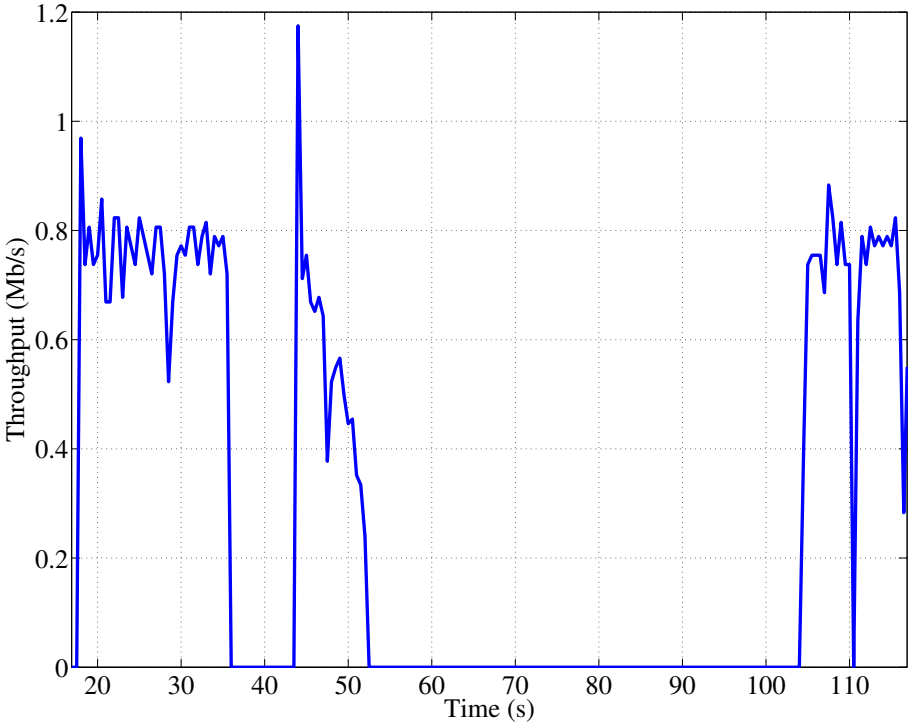
Figure 3.11. *Received power variance as a function of the Nakagami-m fading channel coefficient M.*

essary retransmissions are triggered because of route failures or route changes. To address this problem, modifications of the routing protocol are proposed in [93,94], in order to better support mobility during the topology discovery phase. Finally, there are many experimental analyses of TCP performance in a MANET with mobility, e.g., see [95–98].

In our previous work, we have applied the cognitive network approach [81] to a various set of networking problems. We adopted a learning phase similar to the one used in this paper, which makes use of Bayesian networks (BN), in order to infer the presence of congestion in a multi–hop static network [99], to learn the information needed by a game theoretical inter–network node sharing strategy [100], and for a call admission control protocol in LTE [101].



(a)



(b)

Figure 3.12. *Throughput with a Nakagami-m fading channel ($M = 10$) (a) in a static scenario, and (b) in the presence of mobility.*

3.2.2 Scenario Overview

As a first step, in order to more easily assess the effects of mobility, we consider a scenario with six nodes, connected through a chain topology with five wireless hops, where the inter-node distance is 100 m. [1] We simulate an FTP file transfer, where the data is sent via TCP New Reno from node 0 to node 5.

During the simulation time there are a series of mobility events, in which two adjacent nodes exchange their positions, by moving in opposite directions at a constant speed of 2 m/s. The nodes are disconnected if their distance becomes approximately larger than 130 m. During these events, the connections among the nodes break and the network topology needs to be reconstructed once all the nodes are connected again. We also noticed that the effects of these mobility events on the communications among the nodes change as a function of the channel variability, thus we consider in our simulations different Nakagami-m fading channel models, in which the variance of the received power[2] decreases at increasing values of the parameter M, as depicted in Fig. 3.11.

We make use of the optimized link state routing (OLSR) network protocol [102], which is the most popular open source proactive routing protocol for MANETs. It builds up a route for data transmission thanks to the dissemination of two types of periodic control messages. HELLO messages are broadcasted by each node to find all the one hop and two hop neighbor nodes. Then, topology control (TC) messages are broadcasted by each node with the list of its neighbor nodes [96].

In a scenario with wireless links and mobile nodes, as noted in [82], and as we have observed through our first simulation campaign, there is room for improvement at both the transport and the network layers in order to adapt to the network dynamics. We consider as a performance metric the TCP throughput $t^{(k)}$, which is defined as the number of bits acknowledged by the sender during a time interval k and divided by the length of the time interval (equal to 0.1 seconds). In particular, we seek a solution to the following network problems.

Problem 1. In a scenario with or without mobility, the measured TCP throughput $t^{(k)}$ can go to zero for a certain time interval, as shown in Fig. 3.12-(a) in the case of a static scenario.

[1] We plan to evaluate random deployments of the nodes in a real testbed, as part of our future work.

[2] Evaluated on 500 samples, extracted from a 50 s long simulation

Parameter	Value
WiFi technology	802.11a
Spectrum	OFDM PHY (5 GHz band)
Channel model	Nakagami-M
Transmission Power	17 dBm
Inter-node distance	100 m
Speed of the mobile nodes	2 m/s
Transmission range	∼130 m
Routing Protocol	OLSR
Transport Protocol	TCP New Reno

Table 3.2. *simulation Scenario Settings*

We observed in our simulations that these transmission holes are due to route failures, which occur when the TC messages of the OLSR protocol are dropped due to failures in the wireless transmission. This results in a topology breakage, which blocks the data transmission for a few seconds.

Problem 2. In the presence of mobility, there can be a route failure when a node falls out of the connection range of its neighbors; e.g., in Fig. 3.12-(b), two nodes exchange their positions in the time interval between 40 and 90 seconds. In this example, the time needed to restore the data transmission is significantly longer than the time spent to restore the topology.

Problem 3. In both static and mobile scenarios there is another problem due to the nature of the TCP protocol, which was not designed for a wireless multi-hop network. TCP regulates the retransmission mechanism assuming that the unacknowledged packet has been dropped because of congestion. Thus, at each retransmission, the retransmission timeout (RTO) timer is doubled to prevent any further congestion. Nevertheless, packet losses in wireless networks are dominated by link failures. Therefore, increasing the RTO value at each retransmission may not be a suitable solution, and can turn out to be highly inefficient, since data transmissions might be prevented despite a good channel and a stable topology.

We design a light-weight flexible approach that aims at dynamically detecting whether the network is in a static or in a mobile scenario, and at taking specific actions to mitigate

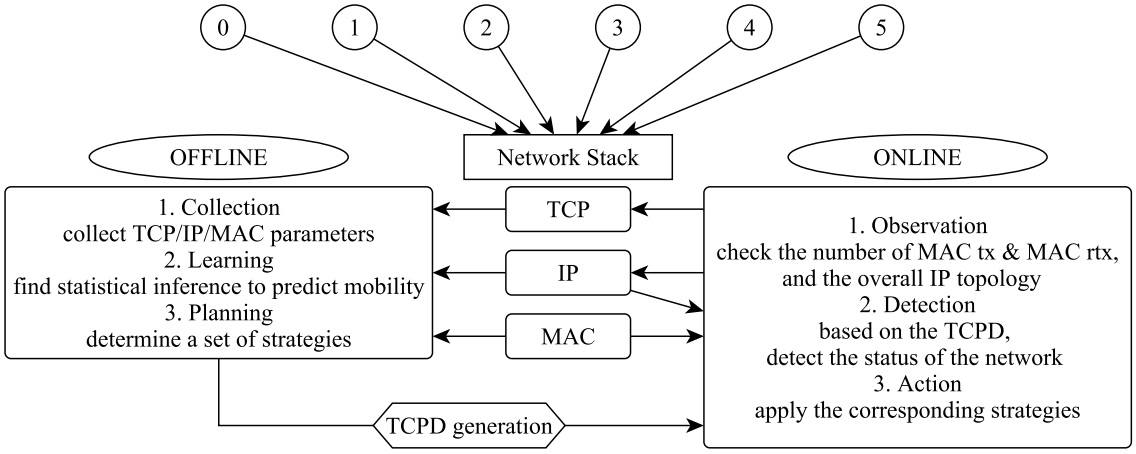


Figure 3.13. *The two modes of the cognitive framework.*

TCP degradation based on such prediction. More in detail, the core of our proposed framework relies on learning network parameters' statistical dependencies for an accurate prediction of mobility. To collect an observable data set, we perform a number of simulations characterized by the parameters reported in Table 3.2. Then, we propose a set of strategies to adaptively counteract the main TCP limitations, by detecting mobility, and by appropriately modifying some key aspects of the transport and network layer protocols.

3.2.3 System Model

Our cognitive framework has been designed to address the problems highlighted in Section 4.1.2 by means of a probabilistic approach, which can infer a mobility event, and ad hoc solutions at the transport and network layers, which exploit such knowledge. The framework, depicted in Fig. 3.13, is divided into two modes, an offline and an online (real-time) mode.

The offline mode involves an initial analysis of the data, which can be collected during a training period in a real network. It is composed of three points: 1) the TCP, IP and MAC parameters are observed at each node of the network during a training period; 2) the probabilistic relationships among these parameters are learned through a probabilistic graphical model approach, which allows to infer the presence of a mobility event as a function of the observation of certain network parameters; and 3) a set of strategies to address specific network problems as a function of the presence or absence of a mobility event is defined. The

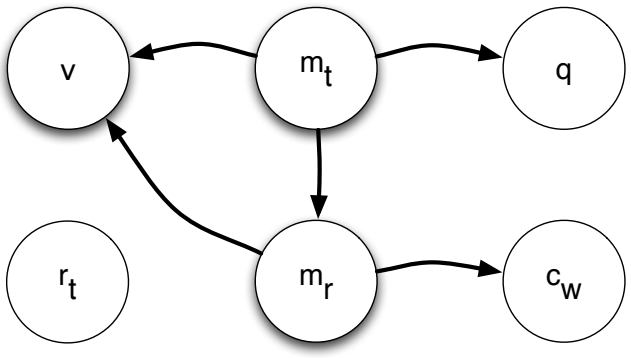


Figure 3.14. *The best fitting DAG $\mathcal{D}$, which describes qualitatively the probabilistic relationships among the network parameters considered.*

probabilistic graphical approach is detailed in Section 3.2.4, while the network strategies are outlined in Section 3.2.5.

The second part of the cognitive framework involves a real-time (online mode) action on the network. It is composed of three points: 1) a subset S of the network parameters is observed; 2) the presence of a mobility event is inferred in real time, as a function of the observation of the network parameters in S, using the probabilities learned in the offline mode; and 3) the corresponding strategies are applied to the network protocol, as a function of the inferred presence of mobility.

3.2.4 Probabilistic Graphical Model Approach

Our probabilistic graphical model approach to infer the presence of mobility is divided into two phases. During the first phase, we study the conditional independence relationships among the set of network parameters available. We represent such probabilistic relationships in a Bayesian Network (BN), which is a probabilistic graphical model [103].

The value of each network parameter is collected at each node, with the exception of the source and the destination nodes, and at every time sample k, which corresponds to a time interval of 0.1 seconds. The collected values for each network parameter are represented as independent samples of a random variable. The set of network parameters includes:

1. c_w is the maximum number of slots the node will wait before transmitting a packet at the MAC level, according to a random back off interval between zero and c_w;
2. q is the number of packets in the transmission queue;
3. m_t is the number of original MAC packets transmitted in the sampling interval;
4. m_r is the number of MAC retransmissions;
5. r_t is the number of missing entries in the IP table; a value larger than 0 indicates that the topology is corrupted.

The parameter we want to infer is v, which indicates the absence ($v = 0$) or the presence ($v = 1$) of mobility.

The structure of the probabilistic relationships among these variables is represented by a directed acyclic graph (DAG). A DAG is a graphical representation of the conditional dependencies among the variables, that defines the structure of the joint probability among these variables. We use a structure learning algorithm [103] to select the DAG that best represents the probabilistic relationships among the variables, using the samples in the training dataset. Unfortunately, the number of DAGs grows super-exponentially with the number of variables, so we need to exploit a local search algorithm, the hill climbing (HC) random search [104], which is not optimal, but provides a good approximation of the best fitting DAG. Furthermore, in order to choose the DAG that best fits the data, we use the Bayesian information criterion (BIC) scoring function [105]. We assign a score to each DAG as a function of how well it fits the data in the training dataset, and penalizing it based on the number of edges of the DAG, thus favoring simpler DAG structures. The best fitting DAG is denoted by $\mathcal{D}$ and it is shown in Fig. 3.14.

In the second phase of our approach, we select from $\mathcal{D}$ the set of nodes S which separate the parameter to infer (v) from the rest of the graph according to the d-separation rule [103]. According to this rule, the observation of the parameters in S is sufficient to make the variable v independent from the other variables in $\mathcal{D}$, or in other words, to make the inference of v depend only on the observation of the variables in S. Given the structure of $\mathcal{D}$, we obtain $S = \{m_t, m_r\}$.

We can now build a simplified probabilistic model, i.e., a conditional Bayesian network (CBN) in order to study how by observing the variables in S it is possible to infer the value

		m_r				
		0-6	7-12	13-19	20-25	>26
m_t	0-34	0.44	0.25	0.87	1	1
	35-68	0.11	0.14	0.27	0.8	1
	69-103	0.08	0.1	0.14	0.45	1

Table 3.3. *Tabular conditional probability distribution in a Nakagami-m fading channel, with* $M = 50$.

of v. This is a simplified model, in which there is an arrow from each variable in S pointing to v. A CBN does not contain the information on the joint probability distribution among all the variables, i.e., $P(c_w, q, m_t, m_r, r_t, v)$, but only on the conditional probability $P(v|m_t, m_r)$, which is simpler to learn and can be approximated more accurately from the observation of a finite dataset.

The parameters of the CBN are learned from the data with a maximum likelihood approach and can be summarized in a tabular conditional probability distribution (TCPD), which is a probability matrix that indicates the probability of $v = 1$, for each value of m_t and m_r. In particular, the values of m_t and m_r are quantized to 5 and 3 levels, respectively, with a uniform quantization. We provide a numerical example of the TCPD in Tab. **??**, where the columns represent the quantized values of m_r, while the rows represents the quantized values of m_t.

The information is exploited by our online framework. At each time sample k, we observe m_t and m_r for each link, we gather the corresponding probability value in the TCPD, and we infer the status of the network (static network or presence of a mobility event). According to the inferred status of the network, we can apply the corresponding network strategy, which is detailed in the next Section.

3.2.5 Strategies Definition

Here we describe the two strategies which can be adopted as a function of the scenario inferred by the probabilistic graphical model approach.

Strategy 1. If the probabilistic model recognizes a static scenario, we increase the holding time of the topology from the default value of 6 s to 100 s, in order to make sure that the

topology does not rely on discovery messages, since we expect the scenario to be *static*. In this way, we aim at reducing the probability of a route failure due to Problem 1.

Strategy 2. In the presence of mobility, we increase the HELLO and TC generation rate by a factor of 10, from the default values of 0.5 and 0.2 messages per second, respectively, to 5 HELLO messages and 2 TC messages per second. Thanks to these modifications, once the physical connections are re-established, the OLSR protocol can recover the network topology more quickly, reducing the long interval with zero TCP throughput observed in Fig. 3.12-(b).

Furthermore, we also adopt an ad hoc solution to Problem 3. Both in the case of route failures for a static network and in the presence of mobility, at each packet loss we do not increase the RTO until the overall topology is restored.[3] In this way, we make sure that the retransmission is performed as soon as the complete topology is re-established.

A possible drawback of our approach is that, since we modify the TCP protocol to promptly react in a mobile wireless multi-hop scenario, it may not behave properly when a congestion really occurs. Dealing with the occurrence of a congestion is out of the scope of this contribution. Anyway, it is possible to design a cognitive approach which can predict the occurrence of a congestion in the network, as in [99]. If a congestion is detected, the standard TCP retransmission mechanisms should be applied.

3.2.6 Performance Evaluation

Here we evaluate the performance of our model to predict the mobility events in a simulated multi–hop wireless mobile scenario. Then, we show the performance improvements achieved by adopting our set of strategies in this scenario.

3.2.7 Prediction Analysis

We evaluate the prediction accuracy of our probabilistic graphical network approach on 5 scenarios, which differ for the wireless channel adopted. In particular, we use a Nagakami-m fading channel model with a parameter $M \in \{5, 10, 20, 50, 100\}$. Our goal is to discriminate between two network conditions, i.e., a static network in which the topology is stable,

[3] This ad hoc solution follows the rationale behind the explicit link failure notification (ELFN) technique for ad hoc networks in [82].

	prediction accuracy	false positives
$M = 5$	17/24	1
$M = 10$	20/24	5
$M = 20$	24/24	5
$M = 50$	24/24	2
$M = 100$	24/24	2

Table 3.4. *Prediction analysis at varying Nakagami-m fading channels*

and a mobile network, where a mobility event occurs. For each choice of the parameter M we run a training simulation of length 2000 s. We observe the value of each network parameter for every time sample k, which corresponds to a time interval of length 0.1 s. The collected data becomes the input for the BN structure learning algorithm and then for inferring the probabilistic parameters of the CBN, as described in Sec 3.2.4. Then, we obtain the TCPD needed to predict the state of the network.

In order to discriminate between the two network conditions, we need to set up a probability threshold, which is used to make the decision after the observation of the parameters in S, and as a function of the corresponding TCPD. We have selected a threshold which represents the best tradeoff between the number of false positives, i.e., the prediction of a mobility event in the case in which the scenario is static, and the number of false negatives, i.e., the prediction of a static network when a mobility event occurs.

We evaluate the accuracy of this network status prediction by running 6 simulations (of 500 s each) for every Nakagami-m fading channel coefficient M, with 4 disruptive mobility events in each simulation. In Tab. 3.4 we show the results in terms of prediction accuracy, i.e., the fraction of mobility events detected, and in terms of false positives. We notice that for $M = 5$, which corresponds to an outdoor scenario in a residential area, the prediction is still good, but less accurate than in the case in which M is larger, which corresponds to a scenario with more stable radio propagation conditions.

3.2.8 Performance Improvements

Our approach is compared to the standard TCP with OLSR protocol stack in different wireless scenarios. The performance is evaluated in terms of the average TCP throughput,

which is defined as

$$\bar{t} = \frac{1}{K} \sum_{k=1}^{K} t^{(k)} \ , \tag{3.7}$$

where $t^{(k)}$ is the instantaneous throughput and K is the total observation time. We evaluate also the outage probability p_o, which is defined as the fraction of time in which $t^{(k)}$ is below τ, the throughput threshold [4], i.e.,

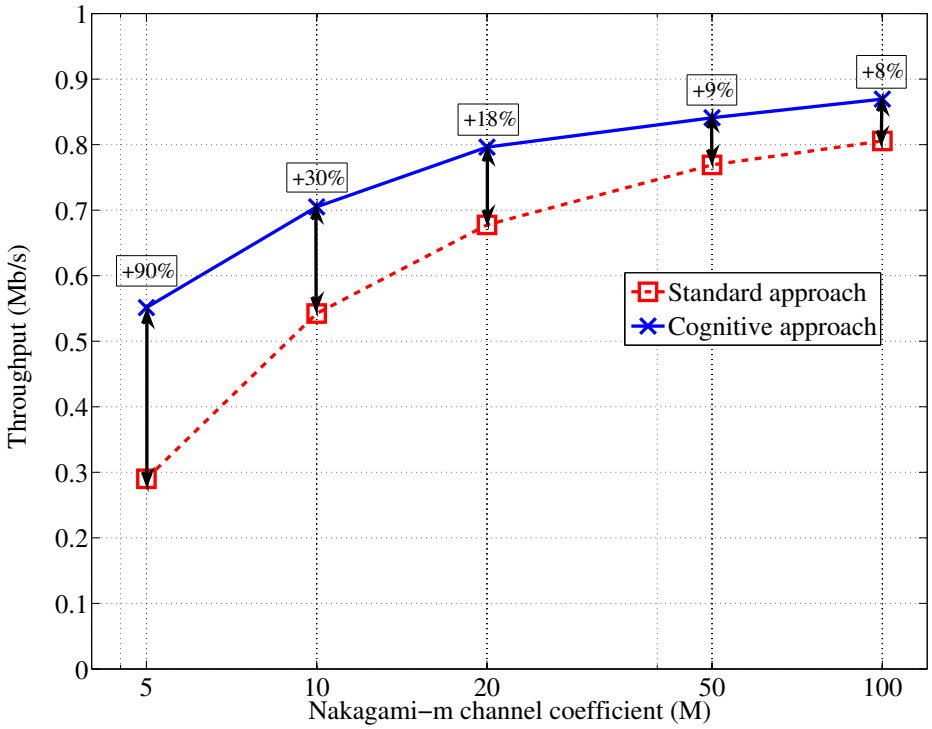
$$p_o = \frac{1}{K} \sum_{k=1}^{K} \chi^{(k)} \ , \ \text{where} \ \chi^{(k)} = \begin{cases} 1, & \text{if } t^{(k)} < \tau \ , \\ 0, & \text{otherwise} \ . \end{cases} \tag{3.8}$$

In Fig. 3.15-(a) we compare the throughput obtained in a static scenario when adopting our proposed framework as opposed to standard procedures. We show also the percentage of throughput improvement obtained with our approach. Similarly, in Fig. 3.15-(b) we show the same throughput comparison in the case of a scenario with mobility events. From these figures, we obtain three important insights. i) The throughput increases as M increases, as expected, due to a more stable channel. ii) Our approach provides increasing gains at lower values of M. iii) By comparing the two scenarios, we observe that our model introduces better performance compared to the standard approach when mobility is introduced, and the topology varies as a function of time. Thus, our approach shows a significant performance improvement in both a static scenario and in the presence of mobility events.

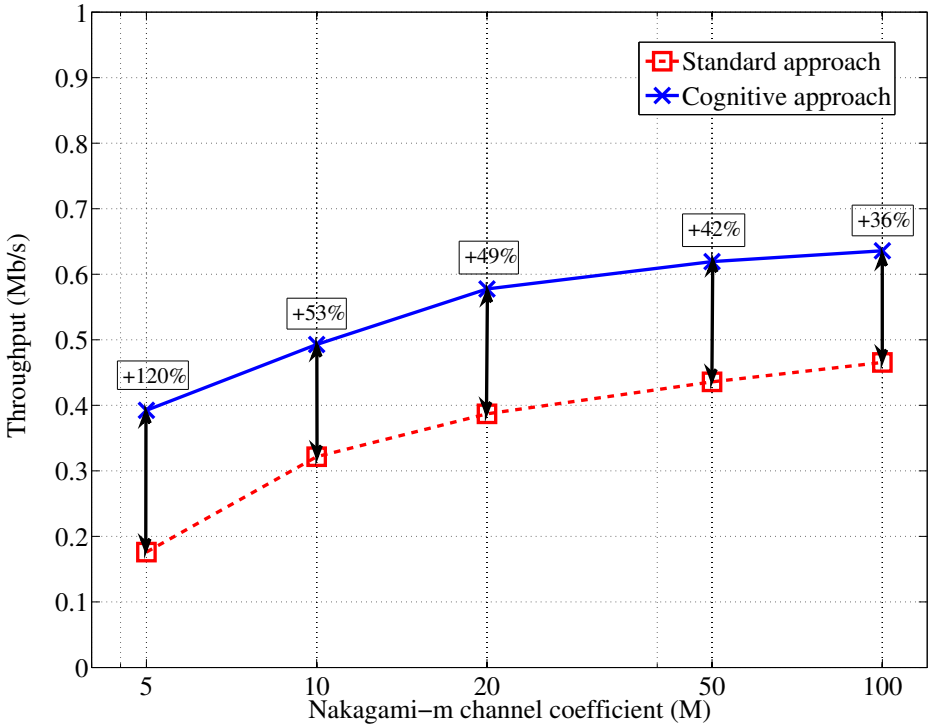
Before describing the performance improvement in terms of the reduction of the outage probability, we study the complementary cumulative distribution function (CCDF) of the duration of the outage intervals, which is the complementary of the cumulative distribution function. In Fig. 3.16-(a) we show the CCDF in a static scenario for the standard protocols and for our approach, while in Fig. 3.16-(b) we show the CCDF for a scenario with mobility events. We show that our approach can decrease the average duration of the outage intervals, which may be an important requirement to meet the requested quality of service (QoS) for some specific applications, in both civilian and military scenarios.

Finally, in Fig. 3.17 we show the reduction in the outage probability for the static scenario, in Fig. 3.17-(a), and in the presence of mobility events, in Fig. 3.17-(b). It can be noted that, in both cases, this probability is significantly reduced, thus corroborating the validity of our proposed model.

[4] In this work, the TCP throughput threshold is set to $\tau = 1$ KB/s.

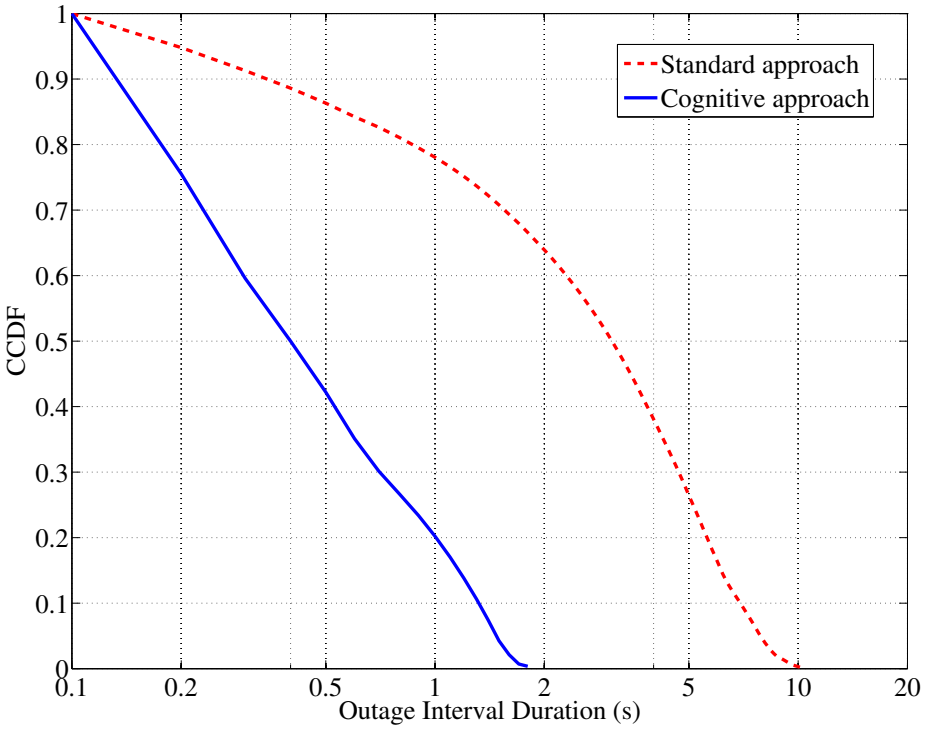


(a)

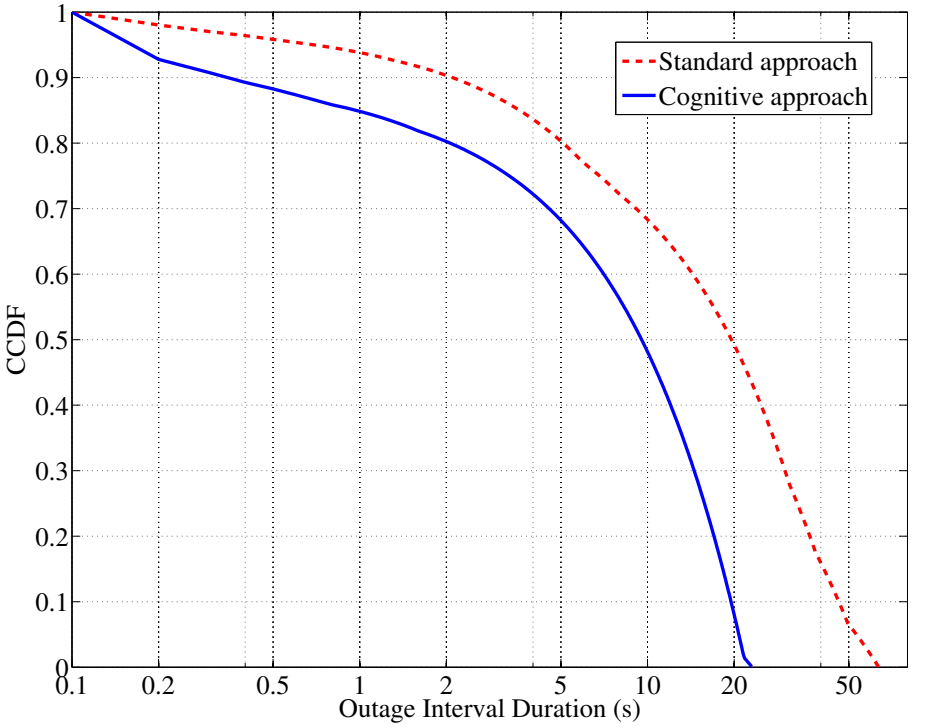


(b)

Figure 3.15. *Average throughput for different Nakagami-m fading channel coefficients (M) (a) in a static scenario, and (b) in the presence of mobility.*

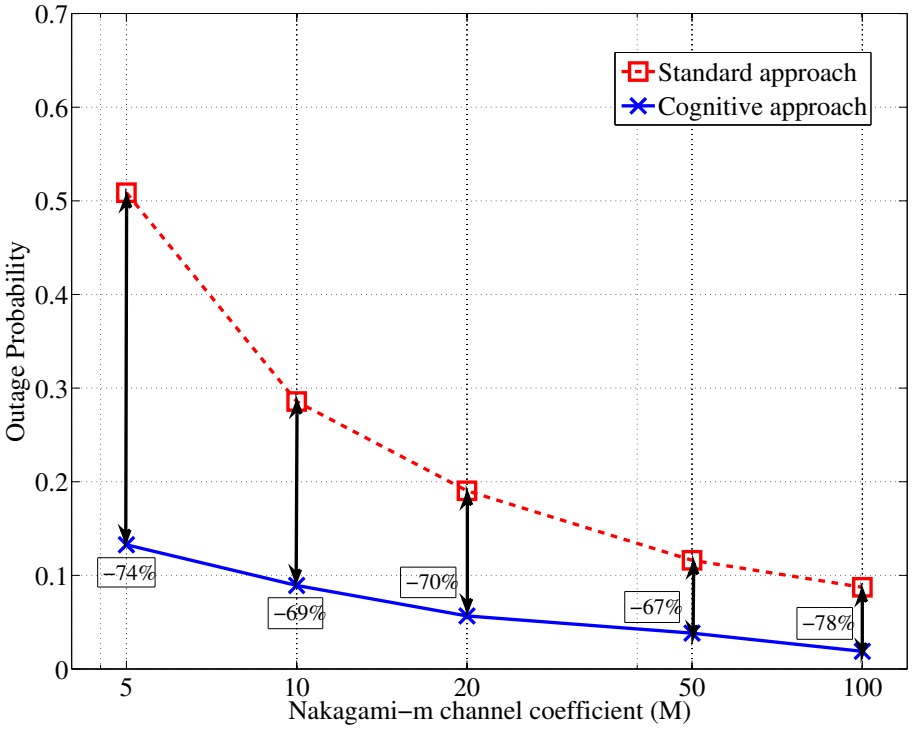


(a)

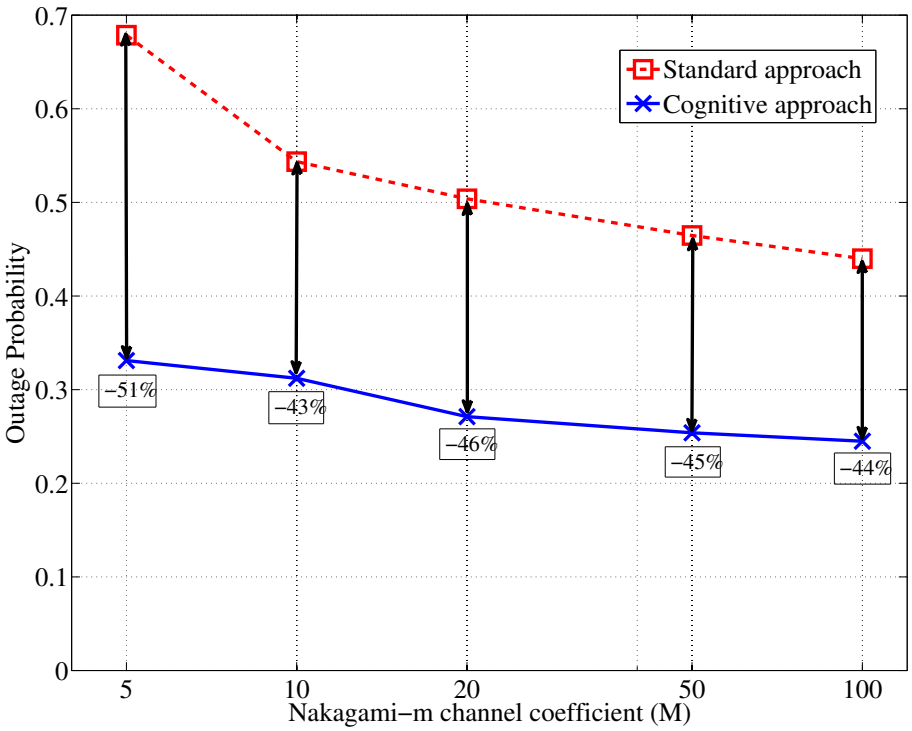


(b)

Figure 3.16. *CCDF of the length of the outage intervals for Nakagami-m fading channel ($M = 5$) (a) in a static scenario, and (b) in the presence of mobility.*



(a)



(b)

Figure 3.17. *Outage probability for different Nakagami-m fading channel coefficients (M) (a) in a static scenario, and (b) in the presence of mobility.*

3.3 Summary

In Section 3.1, the main contribution is represented by the design and implementation of a framework for multi-operator spectrum sharing mechanisms within an LTE implementation of the well-known network simulator ns-3. The aim is to provide the scientific community with an effective and flexible simulation tool that can be easily used, and eventually extended, for the investigation of such a challenging research field. Besides an in-depth description, the resulting software has been thoroughly tested to evaluate its correctness and reliability in achieving spectrum sharing functionalities. Two different algorithms for intra-cell allocation have been implemented in order to show the flexibility of the architecture and its importance for performance comparisons. Of course, the focus of this phase was on the simulator itself and not on the algorithms, whose performance are not expected to be optimal. However, the results have been satisfactory under all aspects, showing that our proposed extension can serve as a practical tool to evaluate resource sharing mechanisms in next generation wireless networks. The code has been released and is publicly available [76].

In Section 3.2, we have adopted the cognitive network paradigm to address the intrinsic inefficiency of standard TCP in mobile ad hoc networks. With our probabilistic approach, we have been able to identify in real time the presence of mobility events, and we have estimated also the prediction accuracy. Through a simulation campaign, we have shown that our approach can significantly outperform the standard TCP with OLSR protocol both in a static and in a mobile scenario, in terms of increased average throughput and decreased outage probability.

Chapter 4

Mathematical Models for Enhanced MAC Strategies

With the advent of new data services and improved device capabilities, cellular data traffic demand is growing rapidly; data demand is expected to double annually over the next five years [5]. The so-called 1000x data challenge is mainly addressed through the definition of heterogeneous networks. In such increasingly complex scenarios, MAC strategies play a key role. Thus, we aim at providing solid contributions that better fit with the resulting access requirements. In particular, we focus i) on resource allocation and user association through a nonlinear optimization model, and ii) on admission control through a capacity estimator based on the Diophantine theory, which deals with indeterminate polynomial equations.

This Chapter is organized as follows. In Section 4.1, we propose a novel approach that aims at extending femtocell coverage by enabling nearby idle UEs in the proximity of femtocells to serve as relays, called UE-Relays. In Section 4.2, because wireless networks providing QoS guarantees need to estimate the increase in peak allocated capacity when admitting a new resource reservation in the system, we consider different approaches to estimate the aggregated capacity and, based on their limitations, propose the E- Diophantine solution. Finally, we summarize our contributions in Section 4.3.

4.1 Joint User Association and Resource Allocation in UE-Relay Assisted HetNets

We consider heterogeneous networks in which open-access indoor femtos are deployed co-channel in a traditional macrocell deployment, where femtos and macros share the same frequency bands. Due to high transmit power differential between them, femto coverage is very limited. For this reason, we introduce UE-Relays: idle UEs that can serve as relays through femto backhaul. We foresee this to be a cost-effective solution to extend femto coverage, which enables greater offload from the macro cells. These UE-Relays may be opportunistically activated (turned on to serve as cells) for users in its vicinity, thereby creating a dynamic relay network. Thus, (1) we introduce the concept of opportunistic UE-Relays to extend the nominal coverage of femtocells. (2) We propose a log utility maximization formulation for optimal relay activation and association in a UE-relay assisted heterogeneous cellular networks. And finally (3) using different nonlinear optimization techniques we show that load-aware cell selection in heterogeneous scenarios is more efficient than classical SINR maximization approaches, as shown in [106].

The Section is organized as follows. In Section 4.1.2, we provide a brief scenario overview. In Section 4.1.3, we introduce our log utility maximization formulation along with our MAX-SINR based association algorithm. In Section 4.1.6, we provide a set of numerical results obtained by jointly varying user association and resource allocation schemes for different scenarios.

4.1.1 Related Work

[107] studies the problem of user association in heterogeneous networks, which serves as a starting point of our work. The resource allocation problem has been investigated in [108] and [109], which makes use of time-sharing relaxation to transform combinatorial optimization problems into convex optimization problems for OFDMA systems. We extend this work to encompass femtocells and UE-Relays. The non-triviality for our problem arises from the fact that we do not know, apriori, the subset of UE-Relays that will be activated, and hence the interference condition is unknown to setup the optimization problem as posed in [107–109]. Note that the modeling and analysis approach based on Poisson Point Process

in [110] cannot be extended to our deployment scenario, where the UE-Relays are clustered around the femtocell. Other related work includes the following.

Femtohaul: In [111], it is proposed using femtocells with relays to increase macrocell backhaul bandwidth. In such scenarios, the relays establish a wireless backhaul connected to the macrocell.

Femtorelay: In [112, 113], the authors propose an access point operating both as a femto and as a relay, thus providing a dual-backhaul to each connected user. In [114], the authors propose an accurate outage analysis based on the same concept. A similar approach, covering the uplink side of the problem, is proposed and analyzed in [115].

UE-Relays for SINR increase: The authors in [116] investigate a cooperative usage of the UE-Relays; the main goal is to increase the SINR, thus allowing the femto to transmit at lower power and, consequently, generating less interference on the macro UEs. In [117], the authors propose to cooperatively use the UE-Relays to increase the uplink SINR, in order to better face the cases of macro users too close to the femtocell, which severely interfere with the femto users.

UE-Relays for offloading: In [118], similarly to what we propose, users connected to the femtocells are used to extend femtocell coverage, thus promoting the offload from the macros to the femtocells through UE-Relays. Nonetheless, the main assumptions show some limitations; i) the offload is performed only if the macrocell is fully loaded, i.e., has not enough resources to serve additional users. In addition, ii) the outage probability is computed solely based on the SNR, thus not considering the interference, which is critical in heterogenous networks. On the other hand, the authors in [119] investigate the same scenario from a different perspective. The network model is based on PMIPv6 to support the UE relay paradigm. Finally, the authors in [120] propose a network model where some of the mobile users which are close to the BS can be used as mobile relays for the users which are far away. Cell edge users connect to a fixed set of WLAN APs. Contrary to our model, where UE-Relays connect to femtocells, here the backhaul is given by the wireless link with the macrocell, thus representing a traditional solution that aims at extending the coverage area of the macro base stations.

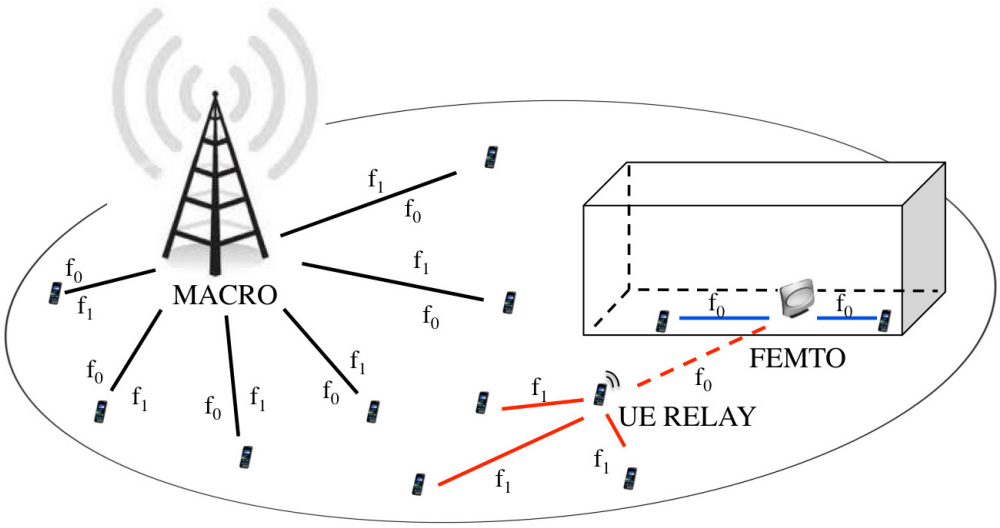


Figure 4.1. *UE-Relay assisted heterogenous scenario.*

4.1.2 Scenario Overview

Our goal is to extend the coverage of femtocells in order to better offload macrocell traffic towards the small cells tiers, which is critical in heterogeneous networks. We introduce the concept of UE-Relays, idle UEs that are connected to femto small cells. We assume that these idle UEs are willing to relay traffic for other users based on some incentives, which is outside of the scope of this contribution. In this work, we aim at studying and evaluating the performance improvements, if any, that result from adopting the proposed proportional rate-based cell selection and resource allocation strategies. The complexity of our centralized model could be controlled in a distributed manner as part of our future works, along with simplifications and lightweight online policies implementation.

As shown in Fig. 4.1, we focus on a specific carrier configuration, which might be interpreted as a carrier aggregation (CA) scheme for LTE-Advanced. Nonetheless, it has to be noted that, for the sake of simplicity, in this Section we do not consider the resource block (RB) granularity of the LTE OFDMA frame. However, it might be easily introduced in the optimization model at the price of an increased complexity. Macrocells transmit on two carriers, f_0 and f_1, whereas femtocells operate only on f_0. The UE-Relays communicate with their connected users (the *access* links) on carrier f_1, and with the femtocells (the *backhaul* links) on carrier f_0. Potential relays can wirelessly connect to a femtocell in the proximity.

Therefore, these devices may extend femtocells' coverage, as depicted in Fig. 4.1, or increase indoor users' throughput by exploiting the access carrier otherwise connected to the macrocell.

4.1.3 System Model

In this Section, we focus on the downlink cell association and resource allocation for UE-Relay assisted heterogeneous cellular networks. As pointed out in [107], unlike what is usually done in traditional cellular networks, the key metric for cell selection is the service rate rather than the SINR, promoting a user association based on load-aware policies.

4.1.4 Log Utility Maximization

In the following, we introduce the proposed optimization problem. The objective is the association vector x, which explicitly carries the resulting user association scheme, and implicitly defines a proportional resource allocation. c_{ij} represents the Shannon point-to-point link capacity

$$c_{ij} = \log_2 (1 + SINR_{ij}) \tag{4.1}$$

where i is a user connected to a base station j. We denote the sets of users, macrocells (one per carrier), femtocells, and relays as U, M, F, and R, respectively. The union of all the base stations classes is $B = M \cup F \cup R$. The log utility for each cell depends on the base station class because of the different carrier configurations. The utility maximization problem is posed as follows.

$$\begin{aligned} \max_{x} \quad & f_{UM}(x) + f_{UF}(x) + f_{UR}(x) \\ \text{s.t.} \quad & \sum_{b \in B} x_{ub} = 1, \quad \forall u \in U \cup R, \\ & 0 \leq x_{ub} \leq 1, \quad \forall u \in U \cup R, \quad \forall b \in B \end{aligned} \tag{4.2}$$

where, the utility functions per base station class is

$$f_{UM}(x) = \sum_{u \in U} \sum_{m \in M} x_{um} \log \left(\frac{c_{um}}{\sum_{i \in U} x_{im}} \right) \tag{4.3}$$

$$f_{UF}(x) = \sum_{u \in U} \sum_{f \in F} x_{uf} \log \left(\frac{c_{uf}}{\sum_{r \in R} x_{rf} \sum_{k \in U} x_{kr} + \sum_{i \in U} x_{if}} \right) \tag{4.4}$$

$$f_{UR}(x) = \sum_{u \in U} \sum_{r \in R} x_{ur} \log \left(\frac{\min(b_r(x), c_{ur})}{\sum_{j \in U} x_{jr}} \right) \tag{4.5}$$

$$b_r(x) = \sum_{f \in F} x_{rf} \frac{c_{rf} \sum_{k \in U} x_{kr}}{\sum_{t \in R} x_{tf} \sum_{i \in U} x_{it} + \sum_{l \in U} x_{lf}} \tag{4.6}$$

We next provide a detailed description of each class utility function.

Eq. (4.3): The logarithmic resource allocation scheme for users connected to macrocells, as per the results presented in [107], provides an equally partitioned rate among all the UEs connected to the same device.

Eq. (4.4): As to the rate achieved by users connected to femtocells, we divide it by the number of other UEs connected to the same femtocell, plus all the users that connect to the relays backhauled by the same femtocell.

Eq. (4.5): On the other hand, concerning the rate achieved by users connected to UE-Relays, we divide the bottleneck rate by the number of UEs connected to the same relay. Here, the bottleneck is defined as the minimum among the backhaul and the access rates.

Eq. (4.6): This equation describes the backhaul rate, i.e., the rate achieved by a relay that connects to a femtocell. It is obtained by dividing the achievable rate by the number of users connected to the relays which share the same femtocell, plus the number of users directly connected to the same femto, as depicted in Fig. 4.2

It can be noted that we have performed a relaxation of the optimization variables, that should be binary values, in order to avoid the combinatorial nature of the problem, and thus apply traditional nonlinear programming strategies. In other words, we have allowed each UE to simultaneously connect to more than a serving station on the same carrier. As shown in the next Section, we observe a negligible gap if comparing the resulting performance with the one obtained by rounding the association to a single serving base station.

4.1.4.1 Convexity Study

We have rewritten the rate functions describing UE-to-femto and UE-to-relay connections, and obtained, respectively, the following forms to make the convexity evaluation easier:

$$x \cdot \log(x + y), \quad \frac{x}{x + y} \tag{4.7}$$

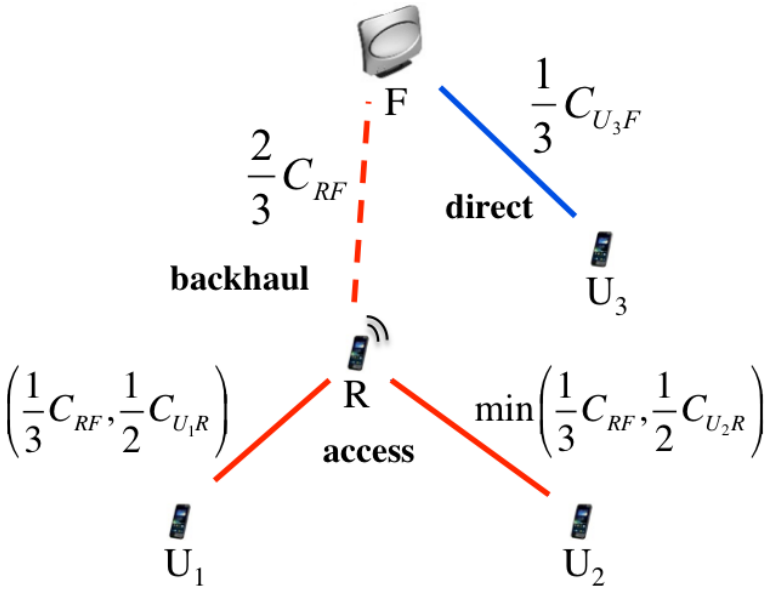


Figure 4.2. *Rate-based proportional user association.*

Because the hessian matrix of both functions is neither positive nor negative semidefinite, the maximization function is nonconvex. Therefore, we cannot state anything about the solutions found, whether they represent global or local optima. For this reason, we compare the one-shot solution, which is obtained by applying the interior point method, with a number of solutions found with global search optimization algorithms, as detailed in the following.

4.1.4.2 Global Search Optimization

In Fig. 4.3, we show a range of solutions found in several scenarios, based on different optimization techniques:

- **Interior Point**: The interior point method leads to an optimal solution by traversing the interior of the feasible region, through multiple iterations which turn the gradient of a logarithmic barrier function to zero [121].
- **Genetic Algorithm**: Genetic algorithms operate by mimicking the principles of biological evolution; repeated modifications of groups of individual points are modeled based on gene combinations in biological reproduction and, because of its random nature, increase the probability to find a global solution [122].

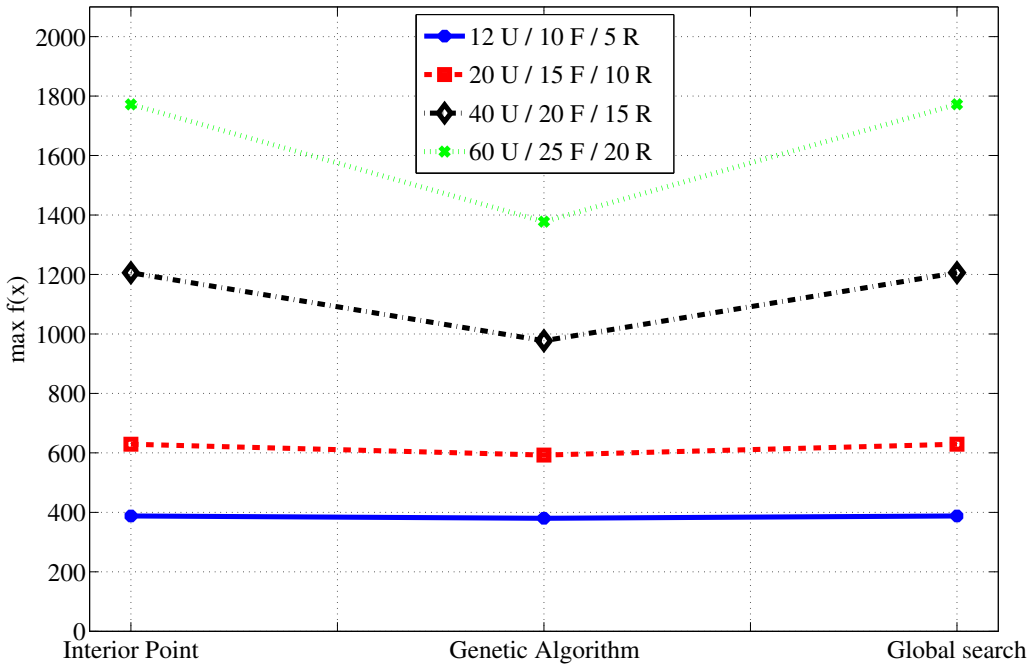


Figure 4.3. *Non linear solutions found with varying techniques*

- **Global Search**: Global search optimization uses a scatter-search mechanism for generating different start points, which are analyzed and, those unlikely to improve the best local minimum found so far, rejected [123].

From Fig. 4.3, we derive a couple of interesting insights: i) global search optimization does not introduce any significant improvement, even though we span tens of different solutions; ii) on the other hand, we can observe that the genetic algorithm without predetermined starting point gets farther from the solution found with the interior point method, which makes it less efficient.

4.1.5 MAX-SINR based association

Here, we describe the algorithm used to compare our proposed proportional approach. Association schemes based on the SINR represent the typical solutions employed in traditional cellular networks. In the next Section, we will show the main limitations of such approach, which, in most heterogeneous scenarios, leads to an inefficient use of the relays.

This association technique, whose details are provided in Algorithm 2, operates in two phases.

1) Cell Selection: First of all, for each user, we need to find the best serving station based on the maximum perceived SINR, which is straightforward.

2) Resource Allocation: Then, we accordingly split the resources based on such information. Please note that: i) users connected only to macrocells divide their rates equally with the other UEs connected to the same base station, ii) users connected to femtocells split their rates according to the number of other UEs connected to the same femto plus the number of users connected to relays which connect to the same femtocell, and iii) users connected to relays get a rate which is divided by the number of other UEs connected to the same relay if the access link is the bottleneck, otherwise they achieve a proportional slice of the backhaul rate.

4.1.6 Performance evaluation

In the following, we aim at comparing and evaluating the proposed cell selection techniques. Our main goal is twofold: (1) we study the performance profiles of the log-based association scheme as opposed to a MAX-SINR approach; (2) we compare the results obtained through the simulation of a UE relay-assisted heterogeneous network against a traditional 2-tier network, where the macrocell is overlaid with a set of femtocells.

Upper bound: In Section 4.1.3, we have introduced a continuous relaxation of the optimization variable in order to avoid the combinatorial nature of the problem. In other words, we have allowed each UE to simultaneously connect to more than a serving station on the same carrier. This might require additional overhead, and an increased complexity of the system. For this reason, we present a set of results where each UE selects its best candidate BS after solving the overall optimization problem. Thus, the multiple BS association may be seen as an upper bound on the overall system performance. We show in Fig. 4.4 how such rounding only slightly affects the overall throughputs on the second carrier, which is the resource shared by the macrocell and relays. This is due to the fact that it is much more likely for a UE to be surrounded by a set of potential candidate relays rather than femtos. For this reason, rounding the associations on f_0, which is the carrier used for macro and femto transmissions, shows a negligible gap among the curves. Conversely, we can observe

Algorithm 2 We first fill the association vector (*association*) based on maximum SINR; then, we apply the resource allocation (*rate*) based on the performed cell selection.

$B_1 \rightarrow$ base stations on carrier 1 (macro, femtos)
$B_2 \rightarrow$ base stations on carrier 2 (macro, relays)
$SINR_{[N_{UES},N_{BSS}]} \rightarrow$ matrix containing UEs geometries
for $c = 1$ to $N_{carriers}$ **do**
  for $u = 1$ to N_{UES} **do**
    $association \leftarrow max(SINR(u, B_c))$
  end for
end for
for $c = 1$ to $N_{carriers}$ **do**
  for $u = 1$ to N_{UES} **do**
    $b \leftarrow association(u, c)$
    if $b \in M$ **then**
      $rate(u, c) \leftarrow C_{ub} / \sum_{k \in U} U_{kb}$
    else if $b \in F$ **then**
      $rate(u, c) \leftarrow C_{ub} / (\sum_{k \in U} U_{kb} + \sum_{r \in R_b} \sum_{t \in U} U_{tr})$
    else if $b \in R$ **then**
      $f \leftarrow$ backhaul femtocell
      $d = \sum_{k \in U} U_{kf} + \sum_{r \in R_f} \sum_{t \in U} U_{tr}$
      $backhaul \leftarrow (C_{bf} \cdot \sum_{k \in U} U_{kb})/d$
      $sum\_cap \leftarrow \sum_{t \in U} C_{tb} / \sum_{k \in U} U_{kb}$
      if $sum\_cap > backhaul$ **then**
        $portion \leftarrow C_{ub}/sum\_cap$
        $rate(u, c) \leftarrow backhaul \cdot portion$
      else
        $rate(u, c) \leftarrow C_{ub} / \sum_{k \in U} U_{kb}$
      end if
    end if
  end for
end for

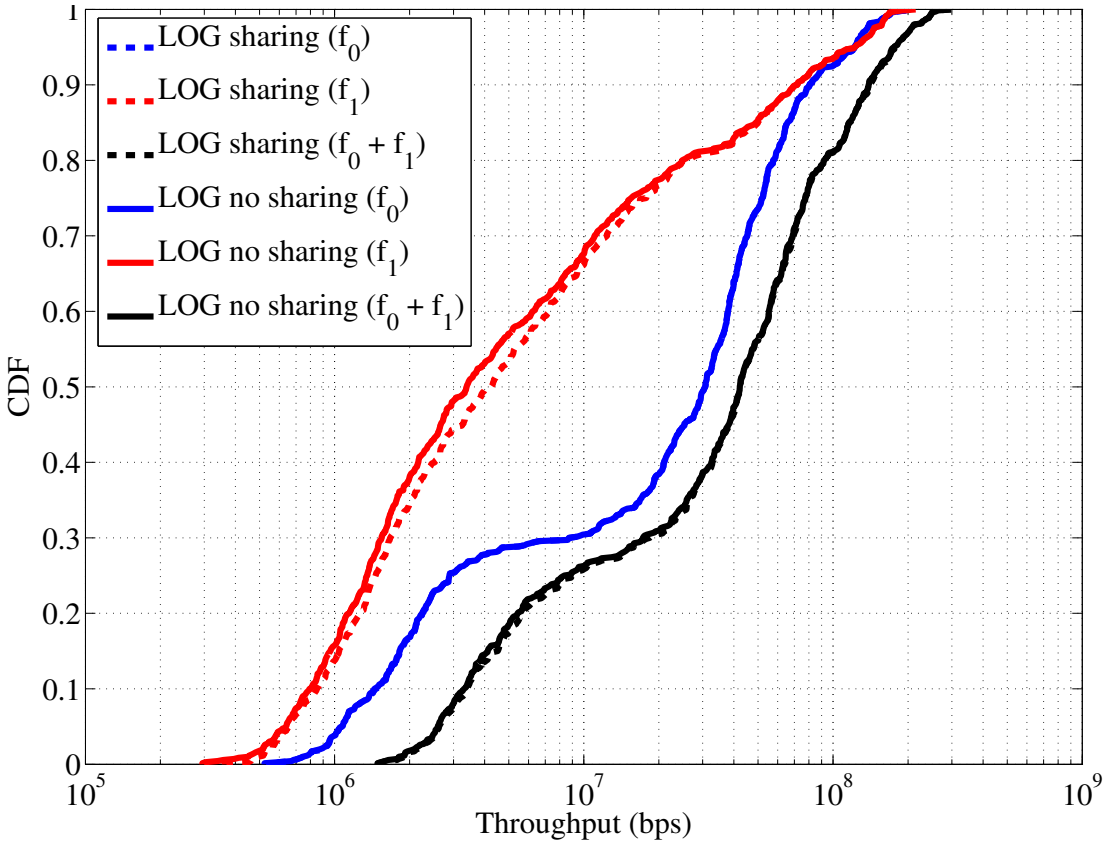


Figure 4.4. *Upper Bound: effects of rounding the multiple BS association relaxation towards single cell selection.*

a slight throughput reduction on f_1, which is the frequency carrying direct communication among UEs and relays.

Lower bound: Association is always, somehow, related to the SINR: i) indirectly, as in our log sum-rate optimization problem, or ii) directly, as in the case of MAX-SINR based association. In our heterogeneous scenarios, a cell selection choice will affect the overall SINR. In fact, some devices might be activated, thus generating interference, whereas some other BSs might be disregarded and kept IDLE, not causing any interference on the existing connections. It is a chicken and the egg problem, which we conservatively face by solving the optimization problem while considering the interference coming from all the potential BSs. Once the association is performed, we recompute each link geometry by taking into account only the interference coming from the set of active serving base stations. We can observe in Fig. 4.5 how such approach, namely I_{AWARE}, slightly outperforms the overall throughput corresponding to the resulting conservative approach, I_{CONS}. In particular, we

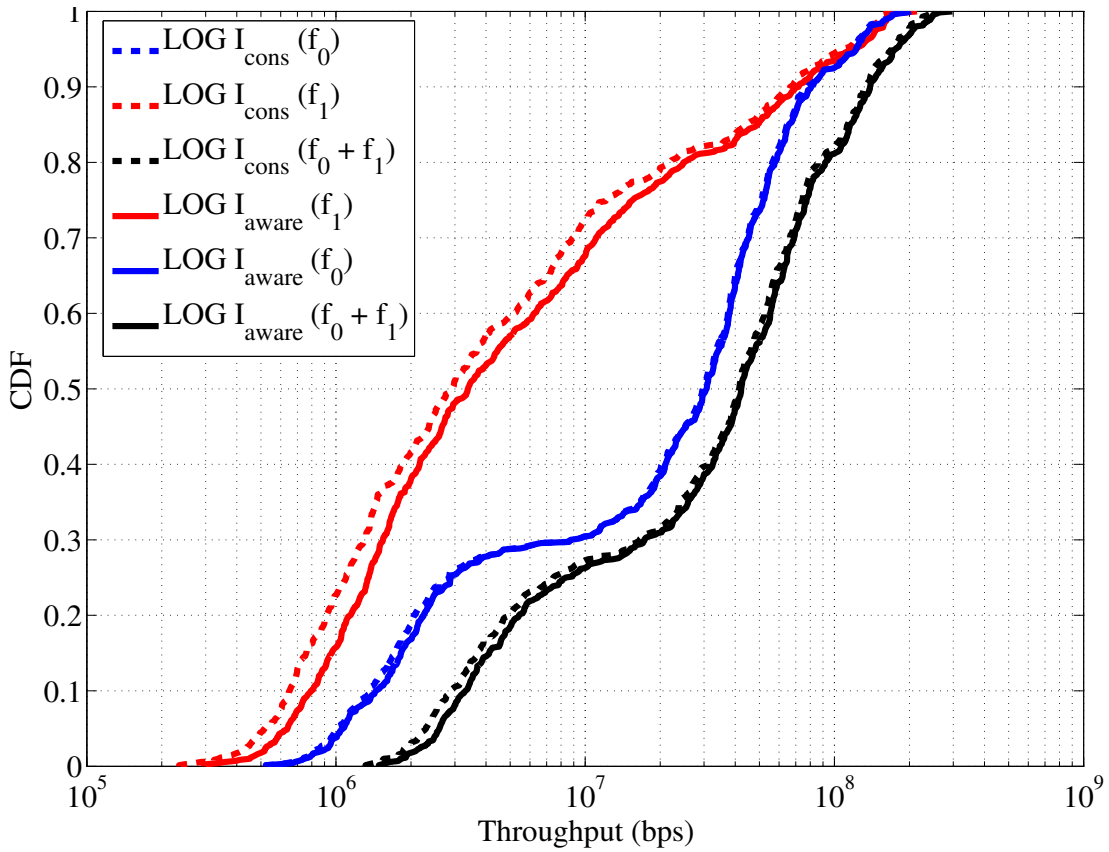


Figure 4.5. *Lower bound: effects of removing the not-active (thus not-interfering) BSs on the SINR computation.*

can note that f_1 shows a wider gap, which is mainly related to outdoor UE-Relays whose interference contribution on the shared spectrum is thus removed.

Our simulation scenario settings, reported in Table 4.1, are mainly based on 3GPP documentation. [1]

In Figs. 4.6 and 4.7, we show the overall throughput trends which result from the resolution of our nonlinear optimization problem. Here, the CDF represents the empirical cumulative distribution function of the per-user throughput values achieved by the users connected to each class of devices (macro, femto, relay), on each carrier (f_0, f_1). In addition, we provide the aggregate throughput trends, i.e., the distribution of the per-user sum throughput $(f_0 + f_1)$. We can clearly observe the gains obtained when potential relays overlay the 2-tier heterogeneous topology. In particular, we can note the following:

[1] From 3GPP TR 36.814: Table A.2.1.1.2-3, Heterogeneous system simulation baseline parameters.

Parameter	Value
Macrocell TX power	46 dBm
Femtocell TX power	20 dBm
UE-Relay TX power	23 dBm
Macrocell Pathloss	$15.3 + 37.6\log_{10}(d)[m]$
Small cells Pathloss	$38.46 + 20\log_{10}(d)[m]$
Femtocells density	70 per cell
Active UEs density	140 per cell
IDLE UEs density	70 per cell
Small cells & UE distribution	Uniform
Indoor UEs (active & IDLE)	70%

Table 4.1. *Scenario configuration*

- f_0: As shown in Fig. 4.6, the use of relays deteriorates the throughput achieved by UEs connected to femtos; this is due to the fact that relays, which establish a femto-based backhaul connection, subtract resources from UEs previously connected to femtocells. The bimodal throughput distribution, as shown in Fig. 4.7, represents the 70% indoor UEs distribution. In fact, users that directly connect to femtocells get very high throughput values. As expected, UEs connected to the macrocell on this carrier, in the presence of relays, nearly achieve the same throughput. Slight differences relate to minor changes in the optimization solutions, as shown in Table 4.2. [2]

- f_1: On the other hand, we can observe an enhanced throughput corresponding to users connected to the macrocell on carrier f_1. This is because relays promote the offload of UEs previously connected to the macrocell, towards the small cells. Thus, the remaining users share the same amount of resource with fewer devices, which results in an improved throughput.

- $f_0 + f_1$: The sum throughput on both carriers corroborates our intuition. We do not see much gain on the 70% indoor population, which was already satisfied with indoor small cells. Conversely, we can note how the lower tail throughput distribution of

[2]These statistics represent mean values averaged over 10 different seeds.

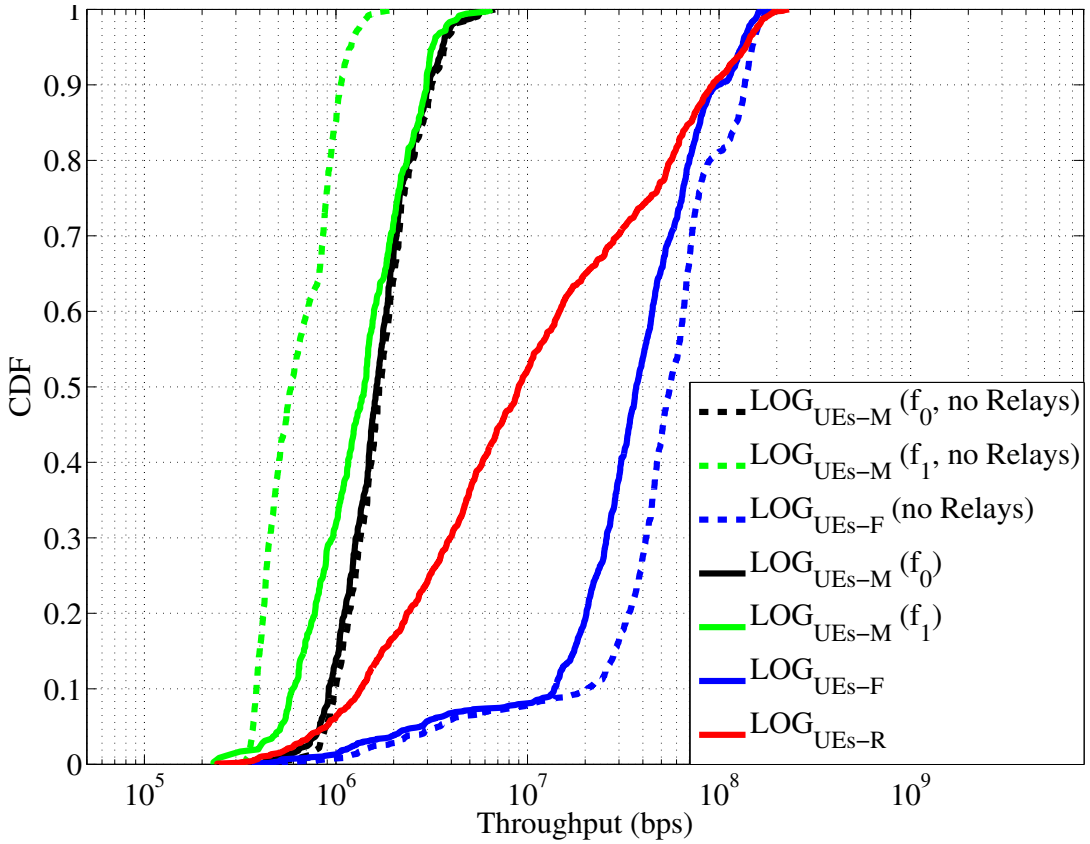


Figure 4.6. *Proportional association, class throughputs.*

scenarios with relays outperforms the case without any relay, thanks to the promoted offload of macro users that are handed off to femtocells via UE-Relays.

Now, if we take a look at the results generated with a MAX-SINR based association scheme, we can note how throughput trends change, leading us to an important conclusion.

- f_0: In Fig. 4.8, as in the case of proportional association, and for the same reasons, we can note how relays deteriorate the throughput achieved by UEs connected to femtos. Users connected to the macrocell on carrier f_0 achieve the same throughput as in scenarios without any active relay, because the association statistics, as shown in Table 4.2, remain the same.

- f_1: As opposed to the results of our log sum-rate based association, when we introduce relays, we can observe a degradation of the distribution of the lower tail throughput values related to users which connect to the macrocell on carrier f_1, i.e., the worst 50% macro UEs throughputs. The MAX-SINR based cell selection will trigger the activation

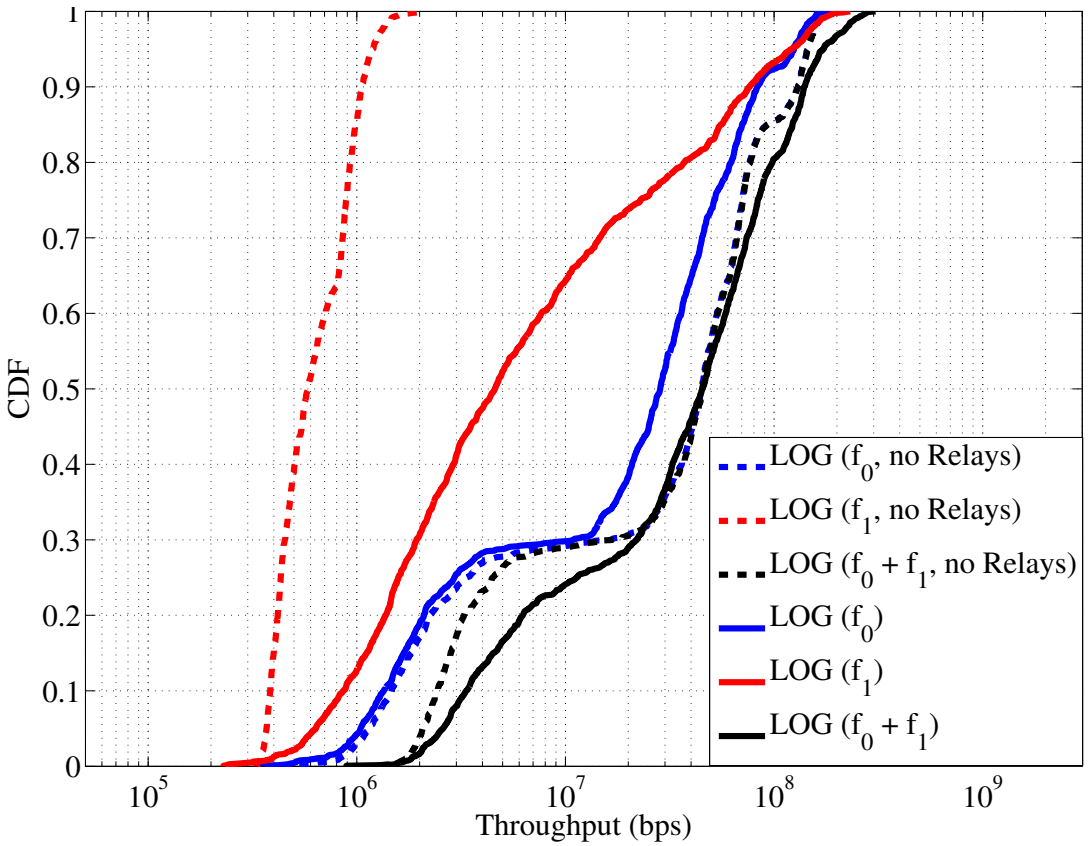


Figure 4.7. *Proportional association, aggregate throughputs.*

of nearly the same number of relays used with the proportional approach; nonetheless, because fewer UEs are handed off from the macrocell, as shown in Table 4.2, the interference generated by active relays results in a degraded performance profile for 50% of the users that remain connected to the macrocell on f_1.

- $f_0 + f_1$: The sum throughput depicted in Fig. 4.9 confirms our description: when adding relays on top of femtocells, i) the upper tail throughput slightly deteriorates, and ii) the lower tail throughput distribution also degrades.

Based on these results, we can draw some important conclusions. A relay-tier which wirelessly connects to femtocells does not necessarily bring throughput improvements. As we observed, a cell selection algorithm based on UEs geometries, namely SINR, actually results in degraded performance. This is explained by the fact that UE-Relays operating on a shared carrier, in our example f_1, come at a price: candidate IDLE users acting as relays will introduce additional interference. Thus, the more relay cooperation is applied,

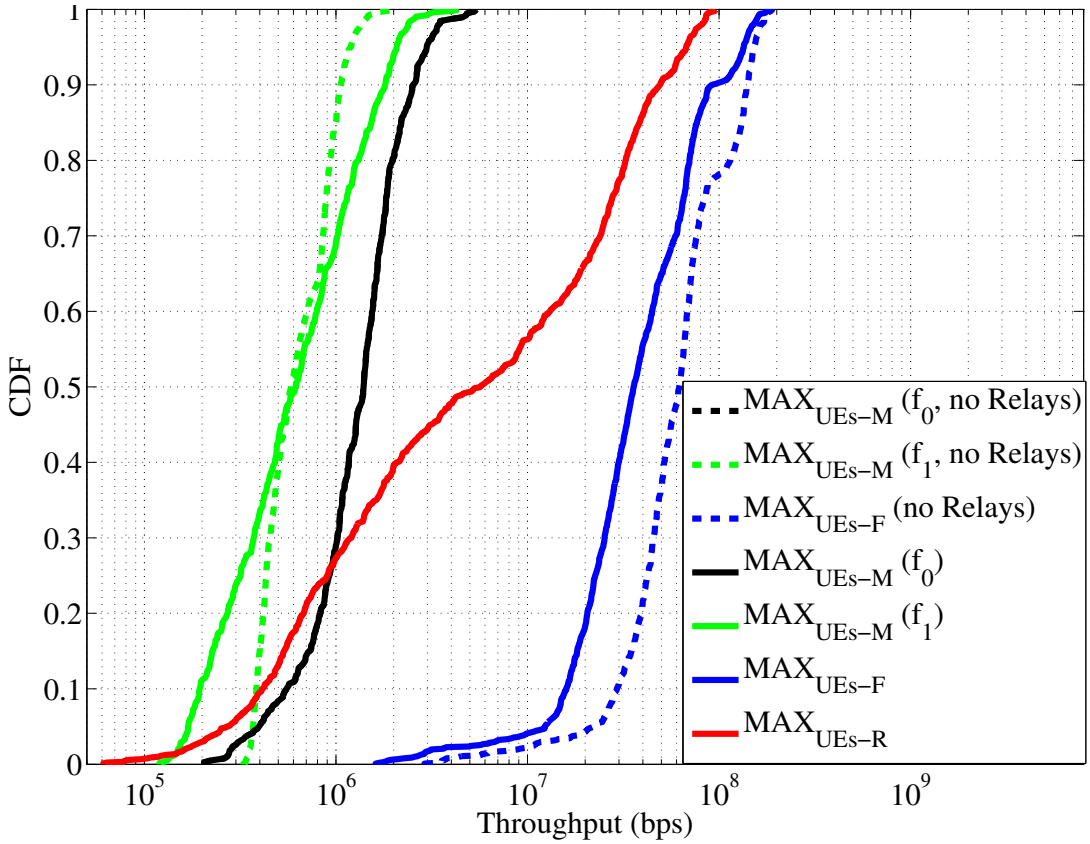


Figure 4.8. *MAX-SINR association, class throughputs.*

obviously limited by the backhaul link capabilities, the more throughput improvements we can observe.

In Fig. 4.10, we show an overall comparison of the throughput distributions obtained with our proportional approach, as opposed to the case without any relay, and the MAX-SINR based association algorithm. These trends confirm our main conclusion. In particular, we can note how the log sum-rate based cell selection outperforms the other approaches except for the median, which, as we have explained above, represents most of the indoor users served by indoor devices. Finally, we can observe the inefficiency of the MAX-SINR approach. In fact, except for the 25% throughput, where users connected to active UE-Relays compensate the resource split of f_0 and the interference generated on f_1, all the other trends confirm the losses previously identified. We provide further details in Table 4.2, where we compare various association statistics, e.g., the mean number of users connected to indoor/outdoor relays, the mean number of UEs per single femtocell, the mean number of

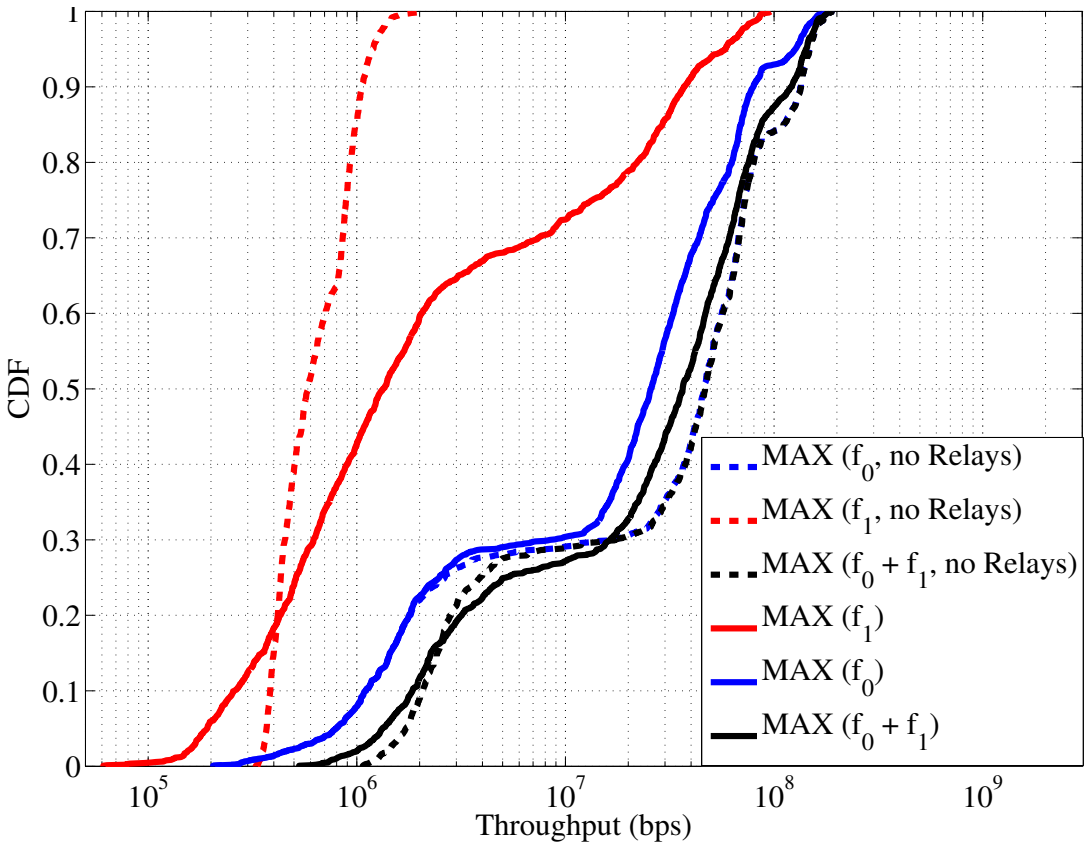


Figure 4.9. *MAX-SINR association, aggregate throughputs.*

UEs connected to a relay whose bottleneck is the backhaul, and so forth.

In addition, we have run a wider set of simulations with different scenario settings. We varied the indoor UE percentage, because such value provides general information about the load of the femtocells. A small amount of indoor UEs generally indicates a scenario where femtos, which are indoor devices, are less loaded. Conversely, when the indoor UE percentage is high, it is more likely to have more users connected to the indoor femtocells. As expected, in a scenario without any indoor user, UE-Relays show higher backhaul capacities because the femtocells to which they connect are unloaded. For this reason, a MAX-SINR approach nearly shows the same throughput improvements as those that we get by solving our nonlinear optimization problem. In fact, the load-unaware (SINR-based) offload performed through this algorithm is beneficial because of the high probability of connecting to unloaded femtocells. By increasing the indoor UE penetration, we observed how these gains deteriorate, and how adopting a rate-based cell selection becomes more convenient.

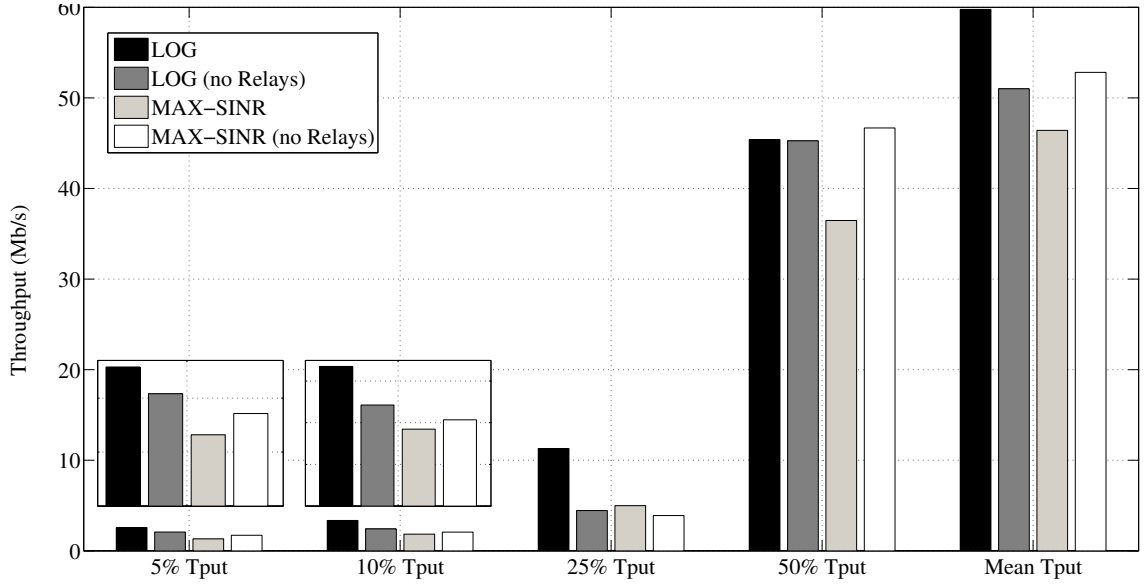


Figure 4.10. *Throughput distributions*

4.2 Admission Control in Wireless Point-to-Multipoint and Multi-Hop Networks

Wireless networks are a key element of today's society to communicate, access and share information. The ever increasing wide range of services and applications building on wireless network capabilities results in a diverse range of Quality of Service (QoS) requirements that needs to be fulfilled to ensure user satisfaction. Third (3G) and Fourth-generation (4G) broadband wireless technologies such as 3GPP Long Term Evolution (LTE) [124], LTE-Advanced [125], IEEE 802.16-2009 [126] and IEEE 802.16m [127] have already defined flexible mechanisms to be able to support a large number of different QoS requirements. In the Wireless Local Area Network (WLAN) domain, a similar path was followed with the standardization of IEEE 802.11e [128] and IEEE 802.11-2007 [129]. Complying with the QoS requirements of granted service demands is mandatory for service providers and requires accurately estimating the available system capacity when deciding whether new service requests can be accepted. Precise capacity estimation techniques allow for the design of efficient admission control algorithms in order to maximize the utilization of networks while ensuring a satisfactory Quality of Experience (QoE) for users.

In this Section we propose a peak allocated capacity estimation algorithm, *E-Diophantine*,

	RELAYS		**NO RELAYS**	
	log sum-rate	MAX-SINR	*log* sum-rate	MAX-SINR
#Active F (f_0)	55.4	54.3	55.4	54.3
#Active R (f_1)	49.5	44.2	X	X
#Active R (in)	30.9 (62.5%)	26.4 (60%)	X	X
#Active R (out)	18.6 (37.5%)	17.8 (40%)	X	X
#UEs-M (f_0)	33.1	38.4	32.4	38.4
#UEs-M (f_1)	35.4	51.7	140	140
#UEs-F	106.9	101.6	107.9	101.6
#UEs-R	104.6	88.3	X	X
#UEs-R (in)	48 (46%)	40.6 (46%)	X	X
#UEs-R (out)	56.6 (54%)	47.7 (54%)	X	X
#UEs per F	1.93	1.87	1.95	1.87
#UEs per R (in)	1.56	1.55	X	X
#UEs per R (out)	3	2.7	X	X
Access Bottleneck	24.1 (24%)	0	X	X
Backhaul Bottleneck	80.5 (76%)	88.3 (100%)	X	X

Table 4.2. *Average association statistics for different scenarios (with/without relays, and using log sum-rate or MAX-SINR in the optimization)*

and evaluate its performance. This new algorithm improves the accuracy of prediction versus complexity trade off when compared with currently available approaches. The results here presented extend our previous work in [130] by: i) analyzing the *E-Diophantine* complexity as compared to competing approaches, ii) designing and evaluating two heuristics of polynomial complexity: *E-diophantine-W* and *E-diophantine-UW*, iii) extending the design and analysis to the wireless multihop case and iv) evaluating the performance of the algorithm in an 802.16j-2009 [131] wireless multihop scenario.

This article is structured as follows. In Section 4.2.1 we define a model for QoS resource reservations in wireless networks and describe solutions available in the literature. Our proposed *E-Diophantine* approaches are explained in Section 4.2.2 along with their mathematical foundations and corresponding complexity analysis. A multihop extension of the

proposed solution is provided in Section 4.2.3. The allocated capacity estimation accuracy and computational load of the different solutions is compared and a realistic wireless system evaluation performed in Section 4.2.4. The evaluation used OPNET's simulator and considers both IEEE 802.16-2009 and 802.16j-2009 networks.

4.2.1 Estimation of Peak Allocated Capacity for QoS Guaranteed Services

Wireless networks support QoS reservation of resources by allowing new flows to apply for admittance in the system through request messages indicating their specific requirements. Such requests contain a set of QoS parameters which include different information depending on the service type. In the following we review the different QoS parameters defined by the currently most deployed wireless technologies in order to find out their commonalities.

We start our analysis considering 3GPP's LTE, the wireless cellular technology expected to become the most widely deployed worldwide in the near future. QoS reservations in LTE's evolved packet system (EPS) are based on bearers which correspond to packet flows established between the packet data network gateway (PDN-GW) and the mobile stations. The bearer management and control follows the network-initiated QoS control paradigm and defines two types of bearers: Guaranteed bit rate (GBR) and Non-Guaranteed bit rate (Non-GBR). Bearers are assigned a scalar value referred to as a QoS class identifier (QCI). Several standardized QCI values with specific characteristics have been defined to allow for successful multivendor deployment and roaming. Table 4.3 summarizes these standardized QCIs as described in [132].

Second, we consider WiMAX networks as the major LTE competing technology. IEEE 802.16-2009 supports QoS reservation of resources by allowing a new flow to apply for admittance in the system through a Dynamic Service Addition REQuest message (DSA-REQ). Such requests contain a QoS parameter set which includes different mandatory information depending on the data delivery service requested in the downlink direction (DL), Base Station (BS) to Subscriber Station (SS), or the scheduling service requested in the uplink direction (UL). Five different QoS services are supported: Unsolicited Grant Service (UGS), Extended Real-Time Variable Rate Service (ERT-VR), Real-Time Variable Rate Service (RT-VR) and Best Effort Service (BE). Table 4.4 summarizes the required QoS parameter set per

3GPP LTE Standardized QCI Parameters					
Res. Type	**QCI**	**Min. Rate**	**Delay**	**Loss Rate**	**Service Example**
GBR	1	•	100 ms	10^{-2}	Conv. Voice
GBR	2	•	150 ms	10^{-3}	Live Streaming
GBR	3	•	50 ms	10^{-3}	Real-time Gaming
GBR	4	•	300 ms	10^{-6}	Buffered Streaming
Non-GBR	5		100 ms	10^{-6}	IMS Signaling
Non-GBR	6		300 ms	10^{-6}	TCP-based apps

Table 4.3. *3GPP LTE Standardized QCI Parameters*

IEEE 802.16-2009 QoS Parameters per Data Delivery Service					
Data Delivery Services	**UGS**	**ERT-VR**	**RT-VR**	**NRT-VR**	**BE**
Min. Resv. Tr. Rate (MRTR)	•	•	•	•	
Max. Sust. Tr. Rate (MSTR)		•	•	•	•
SDU size	•				
Maximum Latency	•	•	•		
Tolerated Jitter	•	•			
Traffic Priority		•		•	•
Req./Trans. Policy	•	•	•	•	•

Table 4.4. *WiMAX QoS Parameters per Data Delivery Service*

Data Delivery Service according to the IEEE 802.16-2009 standard [126]. A similar set of parameters is required in the uplink direction.

Finally, we consider the Wireless LAN technology as the most popular wireless technology in homes and hotspots nowadays. In Wireless LAN QoS reservation of resources has been also enabled as defined by IEEE 802.11e [128] and 802.11-2007 [129]. These standards introduce the Traffic Specification (TSPEC) mechanism which defines a set of mandatory QoS parameters in admissible TSPECs for different service types. In Table 4.5 we summarize a subset of these admissible TSPECs.

IEEE 802.11-2007 Subset of Parameters in Admissible Traffic Specifications				
TSPEC	**MSDU Size**	**Mean Rate**	**Delay Bound**	**Burst Size**
Continuous	•	•	•	
Controlled	•	•	•	
Bursty				•
Contention	•	•		

Table 4.5. *IEEE 802.11-2007 Subset of Parameters in Traffic Specifications*

4.2.1.1 Model for Peak Allocated Capacity Estimation

Based on the aforementioned set of parameterized QoS guarantees for 3GPP LTE, WiMAX and Wireless LAN, we identify a minimum common subset of QoS parameters composed by i) starting time of granted QoS reservation request, ii) required data rate and iii) periodicity at which the guaranteed data rate should be fulfilled (delay bound). Based on this minimum common subset of information available in the QoS reservations, in the following we define a reservation model which will be used for estimating the peak of allocated capacity when considering admitting a new reservation.

Let us consider an arbitrary reservation r_i for which a minimum set of requirements can be defined for services with QoS guarantees as:

- Reservation start time: t_i (ms)
- Reservation required capacity: B_i (bits)
- Periodicity of reserved resources: T_i (ms)

Then, based on these requirements, our QoS reservation model can be described as follows: given a reservation starting time t_i, a certain amount of capacity B_i (bits) is reserved periodically for transmitting reservation r_i data within a time interval T_i. Based on this reservation model, a resource reservation request r_i can be expressed as a periodic discrete sequence of Kronecker deltas with amplitude B_i in the following way

$$r_i(t) = B_i \cdot \delta_{t,t_i+n\cdot T_i} = \begin{cases} B_i & \text{if } t = t_i + n \cdot T_i \ ; \ n \in \mathbb{Z}_{\geq 0} \\ 0 & \text{otherwise} \end{cases} \tag{4.8}$$

Once a new reservation request is received, an allocated capacity estimation algorithm needs to evaluate the impact of accepting it on the currently existing aggregated allocated capacity peak. Assuming a wireless system with N reservations already granted, we define $A(t)$ as the aggregation (as a function of time) of the N flows already in the system plus the one requesting admittance. See Figure 4.11 for a graphical representation of the model for a 5 reservations example.

The objective of our estimation algorithm is thus, starting from a given point of time, t_0, find the allocated capacity peak, $max(A(t))$, when considering the new reservation request. This can be expressed as an optimization problem as follows:

$$\max_{t} \quad A(t) = \sum_{i=1}^{N+1} B_i \cdot \delta_{t,t_i+n\cdot T_i}$$

$$\text{s.t} \qquad t \geq t_0$$

In the next Section we will consider different possible approaches available in the literature to estimate the peak allocated capacity at any given point of time, $max(A(t))$, according to the reservation model defined in Eq.4.8 and their dependence with the *granularity* defined. We define *granularity* (Gr) as the minimum unit that can be used for setting t_i and T_i. This limitation is introduced to account for the time-slotted scheme used in the wireless technologies under consideration which define admitted frame/slot durations.

4.2.1.2 Related Work

An extensive body of literature exists in the area of reserved capacity estimation solutions for admission control purposes in wireless networks. In the following we review some of the main approaches related to the objectives of our work which focuses on providing deterministic QoS guarantees based on explicitly signaled reservation information. Note that, given the large amount of related works, this summary is not meant to be exhaustive but representative of the available solutions. Well-known literature not specifically designed for providing deterministic QoS guarantees as *Effective Bandwidth based* approaches, e.g., [133, 134] or *Measured Throughput based* solutions, e.g., [135], have been omitted for the sake of brevity.

4.2.1.2.1 Throughput-based *Throughput-based* approaches, e.g., [136], consider the mean or peak data rate requirements specified by an application in order to estimate the peak

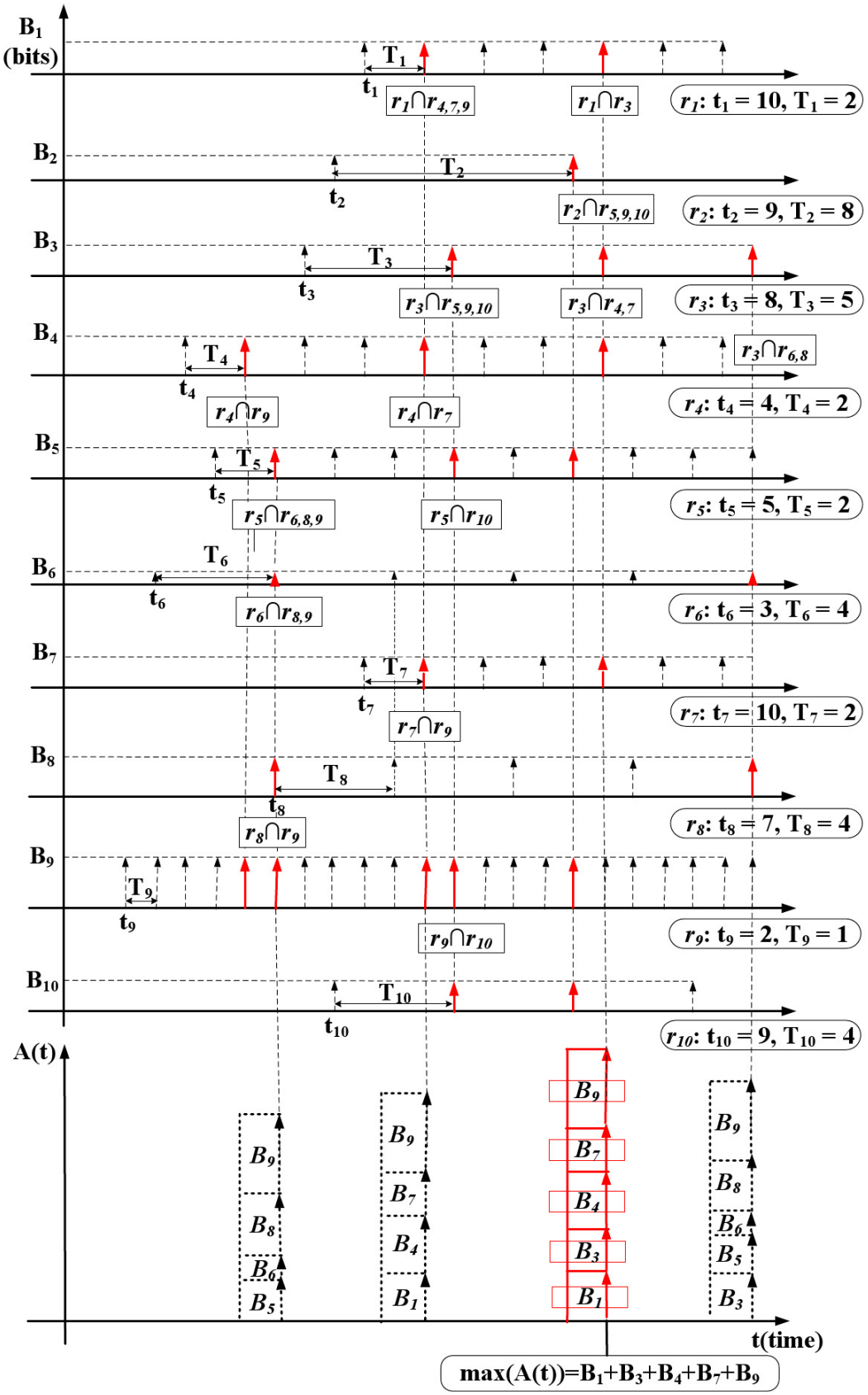


Figure 4.11. *Illustration of QoS reservation model for a 5 reservations example considering granularity* 1 *and identification of reservations needs overlap* ($r_i \cap r_j$) *in time.*

allocated capacity. Such approaches require few computational resources; however, they are neither well suited to take into consideration the bandwidth-varying nature of typical applications such as video, nor the times at which resources are required. Thus, in peak-based implementations, the actual available resources are likely to end up unused while, in mean-based implementations, QoS guarantees might be jeopardized.

In order to determine $max(A(t))$ the *Throughput-based* approach assumes that all admitted reservations need to be served simultaneously, i.e., without taking into account the time at which flows actually need to be served. The following equation corresponds to the *Throughput-based* approximation of $A(t)$.

$$A_{Thr} = \sum_{i=1}^{N+1} B_i \tag{4.9}$$

Throughput-based Complexity The complexity of this solution increases linearly with the number of reservations considered and therefore, can be expressed as $O(N)$.

4.2.1.2.2 Throughput and Periodicity-based *Throughput and Periodicity-based* approaches, e.g., [137] and [138], take into account not only the rate requirements of the reservations but also the delay constraints and/or periodicity of requests obtained from a traffic description included in the admittance request. These solutions make use of the knowledge of the least common multiple (LCM) of the periods of the different accepted reservations and consider it for estimating the peak allocated capacity.

The *Throughput and Periodicity-based* approach obtains an accurate solution for $max(A(t))$ by computing all values of $A(t)$ within a T_{LCM} period. Note that since $A(t)$ is composed of *N+1* periodic reservations, its period T_{LCM} corresponds to the Least Common Multiple (LCM) of the periods of the reservations in the system plus the one under consideration. The following optimization problem formulation corresponds to the $max(A(t))$ computation in this case.

$$\begin{aligned} \max_{t} \quad & A(t) = \sum_{i=1}^{N+1} B_i \cdot \delta_{t,t_i+n\cdot T_i} \\ \text{s.t} \quad & t \geq t_0 \\ & t \leq T_{LCM} \end{aligned}$$

The *Throughput and Periodicity-based* approach though has a dependence with the LCM

of the reservations in the system which, depending on the granularity allowed Gr, might increase exponentially with the number of reservations and thus, become too expensive in computational terms. Therefore, such a solution might not be feasible in practice unless a limitation in the granularity of periods is imposed. This issue will be studied in Section 4.2.4.1.

Throughput and Periodicity-based Complexity The complexity of the *Throughput and Periodicity-based* algorithm increases according to the number of reservations $N+1$ and their corresponding LCM as $(N+1) \cdot LCM$. In the worst case, the LCM increases with the number of flows as $\frac{1}{Gr}\prod_{i=1}^{N+1} T_i$ resulting in a complexity of $O(\frac{N+1}{Gr}\prod_{i=1}^{N} T_i) \rightarrow O((T_{max})^N)$, where T_{max} corresponds to the maximum period of the reservations under consideration.

4.2.1.2.3 Diophantine In order to remove the LCM dependency of the *Throughput and Periodicity-based* approach, another solution is considered based on *Diophantine* theory which, in general, deals with indeterminate polynomial equations and allows variables to be integers only. In the rest of the Section this approach will be referred as *Diophantine* and it has already been considered in [139] as a solution for estimating the peak allocated capacity in wireless networks.

We define this solution as follows. Considering a flow already accepted in the system described with the resource reservation $r_i(t) = B_i \cdot \delta_{t,t_i+n_i \cdot T_i}$ and a new flow requesting admittance characterized by $r_j(t) = B_j \cdot \delta_{t,t_j+n_j \cdot T_j}$, the peak $B_i + B_j$, will occur for the set of n_i and n_j combinations which fulfill

$$\{n_i, n_j \in \mathbb{Z}_{\geq 0} : t_i + n_i \cdot T_i = t_j + n_j \cdot T_j\} \tag{4.10}$$

Based on the previous result, we define the term *diophantine set* as that composed of non-negative integer solutions for n_i and n_j satisfying the condition in Eq.4.10. In order to find this set of solutions Eq.4.10 can be expressed as the following linear diophantine equation with two variables

$$\{n_i, n_j \in \mathbb{Z}_{\geq 0} : n_i \cdot T_i - n_j \cdot T_j = t_j - t_i\} \tag{4.11}$$

Then, based on linear diophantine equations theory, we know that there will be a set of integer solutions for n_i and n_j if

$$\frac{t_j - t_i}{d} \in \mathbb{Z} \tag{4.12}$$

where $d = gcd(T_i, T_j)$ and gcd stands for greatest common divisor.

When the previous condition holds, the *diophantine set* of solutions corresponding to a specific pair of reservations can be found with the extended Euclidean algorithm which will find a and b such that

$$a \cdot T_i + b \cdot T_j = d \; ; \text{ where } a, \; b \; \in \; \mathbb{Z} \tag{4.13}$$

The diophantine set corresponding to the intersection of reservations i and j, indicated with the subindex ij, can be then expressed as follows

$$\{n_{ij} \in \mathbb{Z}_{\geq 0} : t_{ij} + n_{ij} \cdot T_{ij}\} \; ; \text{ where } T_{ij} = lcm(T_i, T_j) \tag{4.14}$$

where the smallest n_i and $n_j \; \in \mathbb{Z}_{\geq 0}$ satisfy

$$t_i + n_i^{min} \cdot T_i = t_j + n_j^{min} \cdot T_j \equiv t_{ij} \tag{4.15}$$

By applying the $Diophantine$ solution to all pairs of reservations in the system, as well as to their found *diophantine sets* in a recursive manner, an exact solution for $max(A(t))$ can be found which is independent of the LCM length.

Diophantine Complexity This solution requires to compute the gcd for all pairs of reservations in the system as well as the sets of intersections, *diophantine sets*, found. Therefore, the total number of gcd executions in the worst case corresponds to

$$\sum_{i=1}^{N+1} \binom{N+1}{i} \tag{4.16}$$

Eq. 4.16 presents a maximum at $i = \lceil \frac{N+1}{2} \rceil$ because of its symmetric distribution and, for a fixed value of $N + 1$, it can be expressed as 2^{N+1}. Thus, since the number of operations to compute the gcd is upper bounded by $log_2(T_{max})$, where T_{max} corresponds to the maximum period of the pairs of reservations or intersections under consideration, the complexity of the *Diophantine* solution is $O(2^{N+1} \cdot log_2(T_{max})) \rightarrow O(2^N)$.

Note that since the computational complexity increases significantly as the number of reservations grows, it might become unfeasible in practice as in the *Throughput and Periodicity-based* case.

	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}
r_1	1	0	1	1	0	0	1	0	1	0
r_2	0	1	1	0	1	0	0	0	1	1
r_3	1	1	1	1	1	1	1	1	1	1
r_4	1	0	1	1	0	0	1	0	1	0
r_5	0	1	1	0	1	1	0	1	1	1
r_6	0	0	1	0	1	1	0	0	0	0
r_7	1	0	1	1	0	0	1	0	1	0
r_8	0	0	1	0	1	0	0	1	1	0
r_9	1	1	1	1	1	0	1	1	1	0
r_{10}	0	1	1	0	1	0	0	0	0	1

Table 4.6. *Example of matrix of intersections of reservations*

4.2.2 E-Diophantine

Based on the performance issues identified for the approaches described in the previous Section, an enhancement to the *Diophantine* solution is proposed, hereinafter referred as *E-Diophantine*. The objective is to achieve the same accuracy when estimating the aggregated allocated peak capacity but at a lower computational cost.

The *E-Diophantine* solution proposed consists in first, exactly as in the original case, finding the sets of intersections, *diophantine sets*, between all pair of reservations under consideration. After this step, instead of repeating the process in a recursive manner for all diophantine sets found, the results are structured in what we will refer to in the rest of the Section as *matrix of intersections* of reservations. We define *matrix of intersections* as a matrix where the information regarding the intersections found between all pairs of reservations in the system are indicated with a 1 when there is an intersection and 0 otherwise. See Table 4.6 for a 10 reservations example where the values for t_i and T_i of each reservation were randomly drawn from a uniform distribution between 1 and 10. Based on this matrix, the rest of the sets are derived based on the information obtained regarding the reservations involved in each intersection. In the following we provide the theorems and proofs that enable the designed *E-Diophantine* algorithm.

4.2.2.1 Intersection of 2 Diophantine Sets

Theorem 1. For any reservation intersecting with two different reservations according to the matrix of intersections, a joint intersection between the three reservations will exist if the two different reservations also intersect with each other

Proof. Consider that for reservations i, j and k two diophantine sets exist based on the intersections between reservations i and j, indicated with the subindex ij, as well as j and k, indicated with the subindex jk. According to the notation introduced in Equation 4.14 they are defined as follows

$$\{n_{ij} \in \mathbb{Z}_{\geq 0} : t_{ij} + n_{ij} \cdot T_{ij}\} \ ; \text{ where } T_{ij} = lcm(T_i, T_j) \tag{4.17}$$

$$\{n_{jk} \in \mathbb{Z}_{\geq 0} : t_{jk} + n_{jk} \cdot T_{jk}\} \ ; \text{ where } T_{jk} = lcm(T_j, T_k) \tag{4.18}$$

An intersection between both sets will occur if a set of n_{ij} and n_{jk} exists such that

$$\{n_{ij}, n_{jk} \in \mathbb{Z}_{\geq 0} : t_{ij} + n_{ij} \cdot T_{ij} = t_{jk} + n_{jk} \cdot T_{jk}\} \tag{4.19}$$

Considering that t_{ij} and t_{jk} can be expressed as $t_j + n_j^{min} \cdot T_j$ and $t_j + n_j^{\prime min} \cdot T_j$ respectively[3], Eq. 4.19 can be rewritten as follows

$$\left\{t_j + n_j^{min} \cdot T_j + n_{ij} \cdot T_{ij} = t_j + n_j^{\prime min} \cdot T_j + n_{jk} \cdot T_{jk}\right\} \tag{4.20}$$

Which recalling that $T_{ij} = lcm(T_i, T_j)$, $T_{jk} = lcm(T_j, T_k)$ and the relation between lcm and gcd becomes

$$\left\{n_j^{min} + n_{ij}\frac{T_i}{gcd(T_i, T_j)} = n_j^{\prime min} + n_{jk}\frac{T_k}{gcd(T_j, T_k)}\right\} \tag{4.21}$$

Then, since

$$\left\{n_j^{min} + n_{ij}\frac{T_i}{gcd(T_i, T_j)}\right\} \subseteq \{t_i + n_i \cdot T_i\} \tag{4.22}$$

and

$$\left\{n_j^{\prime min} + n_{jk}\frac{T_k}{gcd(T_j, T_k)}\right\} \subseteq \{t_k + n_k \cdot T_k\} \tag{4.23}$$

we can affirm that a solution will exist for n_{ij} and $n_{jk} \in \mathbb{Z}_{\geq 0}$ such that the condition in Eq. 4.21 holds if reservations i and k intersect. The resulting set for reservations i, j and k would then be defined as

$$\{n_{ijk} \in \mathbb{Z}_{\geq 0} : t_{ijk} + n_{ijk} \cdot T_{ijk}\} \tag{4.24}$$

[3]Note that the superindex $'$ for $n_j^{\prime min}$ has been introduced to differentiate the n_j^{min} in t_{ij} and t_{jk} which are not necessarily the same.

where $T_{ijk} = lcm(T_i, T_j, T_k)$

□

4.2.2.2 Intersection of N+1 Diophantine Sets

Theorem 2. For any set of intersections corresponding to N diophantine sets, an intersection with another diophantine set will exist if and only if all reservations involved in both sets intersect with each other

Proof. Assuming a set of intersections of N diophantine sets defined as[4]

$$\{n_{1..N} \in \mathbb{Z}_{\geq 0} : t_{1..N} + n_{1..N} \cdot T_{1..N}\} \equiv I_N \tag{4.25}$$

where $T_{1..N-1} = lcm(T_1, .., T_N - 1)$

For a set $\{n_{N+1} \in \mathbb{Z}_{\geq 0} : t_{N+1} + n_{N+1} \cdot T_{N+1}\} \equiv I_{N+1}$ to intersect with I_N, the following must hold

$$\begin{aligned}\{n_{1..N}, n_{N+1} \in \mathbb{Z}_{\geq 0} : t_{1..N} + n_{1..N} \cdot T_{1..N} = \\ t_{N+1} + n_{N+1} \cdot T_{N+1}\}\end{aligned} \tag{4.26}$$

Considering that

$$I_N = \{t_1 + n_1 \cdot T_1\} \cap ... \cap \{t_N + n_N \cdot T_N\} \tag{4.27}$$

Then, the set I_{N+1} will intersect with I_N if and only if

$$I_{N+1} \cap \{t_1 + n_1 \cdot T_1\} \cap ... \cap \{t_N + n_N \cdot T_N\} \notin \emptyset \tag{4.28}$$

□

4.2.2.3 E-Diophantine Algorithm

Algorithm 3 details the steps followed by the *E-Diophantine* solution. The first part of the algorithm, which finds the diophantine sets, is identical to the *Diophantine* approach. Once the sets have been obtained, a matrix of intersections is constructed by simply structuring

[4]Note that the notation for a set of intersections has been simplified for readibility reasons such that a set of intersections involving several reservations is referred with a single subindex instead of with the indexes of the reservations involved.

the diophantine sets results regarding the intersections between each pair of reservations in a matrix form. This operation corresponds to the function $compute\_matrix\_inters(.)$ in the algorithm and an example is shown in Table 4.6 for a set of 10 reservations.

Applying Theorem 1 and 2, the problem of finding the aggregated peak allocated capacity is similar to the *clique problem* which consists in finding particular complete subgraphs ($cliques$) in a graph, i.e., sets of elements where each pair of elements is connected [140], [141]. The clique problem is based on an adjacency matrix where the edges between vertices are given for a specific graph indicating their interconnection. This adjacency matrix can be mapped in our case to the matrix of intersections where edges represent intersections between reservations. Our goal is then finding a complete subgraph containing the *weighted maximum clique*. This problem has been extensively researched in the past and several solutions are available in the literature, see for instance [142].

The *E-Diophantine* algorithm designed is based on the general backtracking solution [143]. It consists in a recursive algorithm which incrementally builds, on a reservation by reservation basis, a set of candidate intersections keeping for each reservation the aggregated peak resource reservation in a vector called $solutions$. The solutions obtained starting the backtracking process at each different reservation are kept in $solutions$ and once all have been explored the function $find\_maximum(solutions)$ simply selects the maximum value contained in this matrix. In order to improve the efficiency of the $backtracking\_search(.)$ function, we introduce the condition $weight + solution(next) \leq max$ in Algorithm 3 where $weight$ corresponds to the aggregate bandwidth in the current candidate set while max corresponds to the aggregated allocated capacity peak found so far. Based on this condition we can determine if the consideration of a new reservation in our current candidate set will result in a larger aggregated resource reservation peak. If the condition is true we can skip backtracking based on this reservation and move on to the next one. The same logic applies to the condition $weight + max\_potential(start) \leq max$ where we evaluate at each iteration whether it is still possible for the current set considered to be above the peak found in previous iterations.

Finally, before starting the backtracking algorithm based on the matrix of reservations, we order the reservations based on their potential aggregated peak resource reservation $B_{pot}(resv)$. The operation corresponds to the function $order(.)$ and its advantage is that it

Algorithm 3 **E-Diophantine** algorithm to find out the peak resource requirement for a new reservation r_{N+1} with starting time t_{N+1}, period T_{N+1} and requirement B_{N+1} considering the set of N reservations already accepted in the system with their corresponding starting times $t = (t_1...t_N)$, periods $T = (T_1...T_N)$ and requirements $B = (B_1...B_N)$

Call executed for each new reservation request
for $i = 1$ to $N + 1$ **do**
  for $j = i + 1$ to $N + 1$ **do**
    if $solution\_exists(t_i, t_j, T_i, T_j)$ **then**
      $intersections \leftarrow find\_inters\_dioph(t_i, t_j, T_i, T_j)$
    end if
  end for
end for
$m\_inters \leftarrow compute\_matrix\_inters(intersections)$
$m\_inters \leftarrow order(m\_inters)$
for $root = N + 1$ to 1 **do**
  $solutions(root) \leftarrow backtracking\_search(root, B_{root}, max)$
end for
return $find\_maximum(solutions)$

Function backtracking_search(start,weight,max)
$next \leftarrow find\_next\_inters(start, m\_inters)$
if $next == NULL$ **then**
  if $weight > max$ **then**
    $max = weight$
  end if
end if
while $next \neq NULL$ **do**
  if $weight + solutions(next) \leq max$ **then**
    return
  end if
  if $weight + max\_potential(start) \leq max$ **then**
    return
  end if
  $backtracking\_search(next, weight + B_{next}, max)$
  $next \leftarrow find\_next\_inters(next, m\_inters)$
end while
end function

increases the probability of finding the peak solution during the first backtracking operations. Thus, reducing the need of backtracking iterations to complete the full exploration of

possible solutions.

Figure 4.12 illustrates the tree of solutions found based on the matrix of intersections in Table 4.6 and applying the *E-Diophantine* algorithm. As an example, in the case of reservation 1, r_1, there is no branch of solutions with r_2 since according to the matrix of intersections r_1 does not intersect with r_2. In the case of the branch of solutions $r_1 \rightarrow r_3 \rightarrow r_4$ the continuation with the branch of r_5 is also discarded since reservations r_1 and r_3 do not intersect with r_5.

E-Diophantine Complexity The first part of the *E-Diophantine* algorithm, which finds the first set of intersections (diophantine sets) between all pairs of reservations, requires in the worst case the same number of operations as the first part of the *Diophantine* algorithm: $\binom{N+1}{2} \cdot log_2(T_{max})$.

After this, both algorithms differ. In the the *E-Diophantine* case, the set of intersections found are used to construct a matrix of intersections. Based on this matrix, the rest of the intersections are derived by exploring the tree of possible solutions, $solutions\_tree$ in Algorithm 3, without requiring the *gcd* computation. This is possible due to Theorems 1 and 2.

As aforementioned, the exploration of the rest of possible intersections in order to find the peak resource requirement is similar to solving the *weighted maximum clique* problem for an arbitrary graph. In our particular case, this problem can be translated to the *unweighted maximum clique* problem (all weights are equal) by defining a reference resource requirement unit B_{ref} and then mapping the resource reservations requests in the matrix of intersections according to $\lceil B/B_{ref} \rceil$. Thus, resulting in an increase of the number of reservations of:

$$N+1 \rightarrow (N+1)^* = \sum_{i=1}^{N+1} \left\lceil \frac{B_i}{B_{ref}} \right\rceil \tag{4.29}$$

$(N+1)^*$ can be then expressed as $(N+1)^* = k*(N+1)$, where k is a constant controlled by the network operator.

Efficient exact algorithms for the *unweighted maximum clique* problem have been extensively researched in the literature. [144] represents one of the fastest solutions known today which incurs in a complexity of $O(2^{\frac{N}{4}})$. Considering the complexity of both parts of the *E-Diophantine* algorithm together, the overall complexity can be expressed as $O(2^{\frac{k \cdot N}{4}})$.

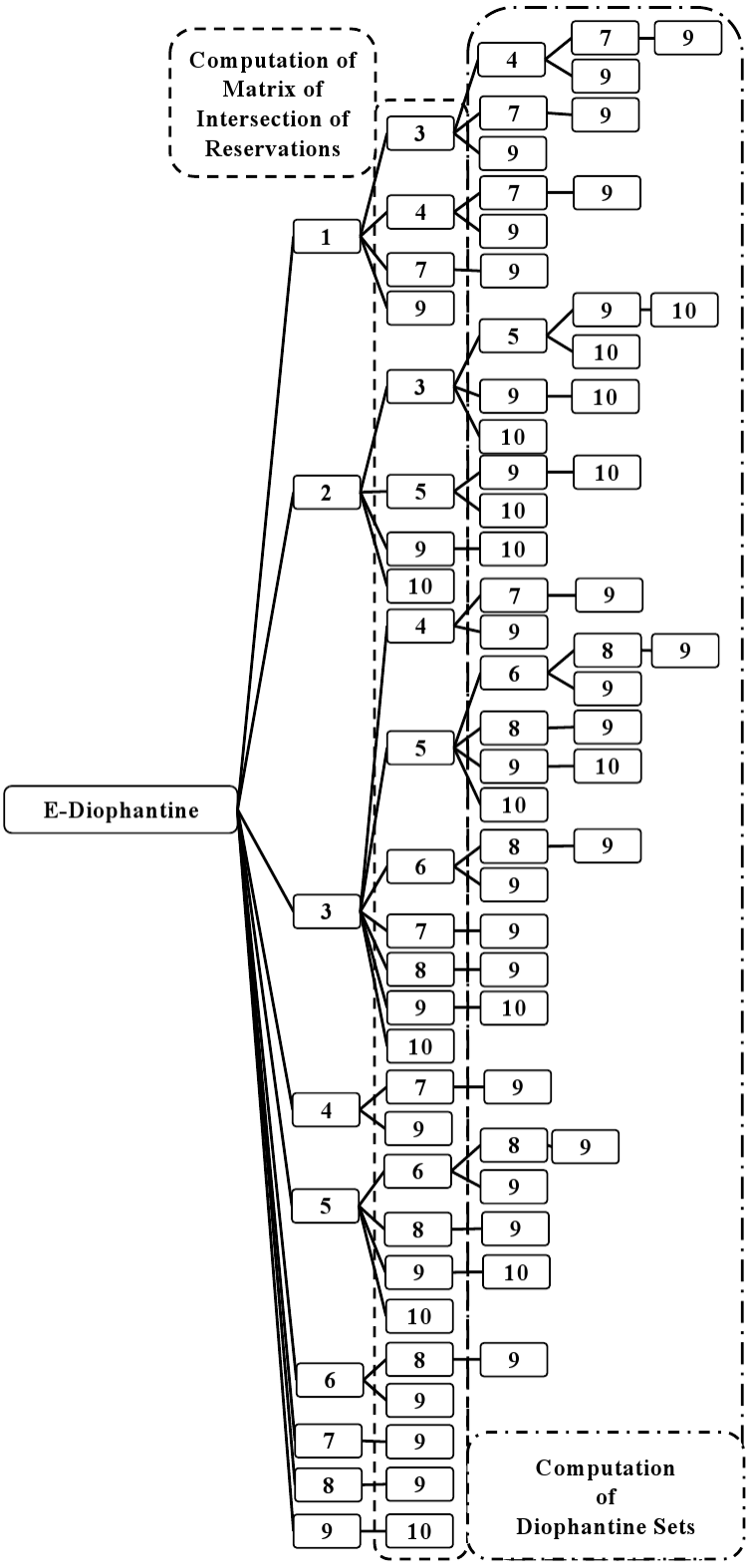


Figure 4.12. *Illustration of tree of solutions of intersections for the E-Diophantine algorithm based on the matrix of intersections shown in Table 4.6*

The worst case *E-Diophantine* complexity thus can be potentially worse than the *Diophantine* one depending on the k value determined by the operator configuration. However, in contrast to the generic *maximum clique* case, in our particular problem we do not deal with *random* edges between vertices and this can be exploited to obtain better performance. These differences will be explained in more detail in the next Section and their actual performance compared in Section 4.2.4.1.

4.2.2.4 E-Diophantine Heuristics

In the previous Section we have shown that the *E-Diophantine* exact solution might still become unfeasible in practice due to its exponential growth with the number of reservations. In the following we describe two non-exact heuristics designed to overcome this limitation by performing a *partial* exploration of the tree of solutions instead of an *exhaustive* one. The proposed approaches achieve a polynomial computational load increase (cubic) based on the number of reservations at a bounded probability estimation error cost.

Note that, as explained in Section 4.2.2.3, the *E-Diophantine* exact solution can be mapped to the unweigthed maximum clique problem which is NP-hard and therefore, no polynomial solutions are expected to be found. The difference of this problem to our particular case is that we can exploit the fact that, in our cliques, we do not deal with *random* edges between vertices but with edges determined by common properties based on our matrix of intersections construction.

The *E-Diophantine* heuristics are based on the following observation: Let us assume that three reservations r_i, r_j and r_k result in two diophantine sets $\{t_{ij} + n_{ij} \cdot T_{ij}\}$ and $\{t_{ik} + n_{ik} \cdot T_{ik}\}$ having reservation r_i in common. In such a case, since both diophantine sets are a contained inside the set $\{t_i + n_i \cdot T_i\}$ it can be intuitively observed that the probability of reservations r_j and r_k of intersecting with each other is higher than if they would not have a reservation in common. Moreover, the larger the number of reservations in common between diophantine sets, the larger the probability of intersecting with each other. The actual probability though can not be computed unless specific assumptions are made on the way the starting times and periods of reservations are chosen.

E-Diophantine Weighted (E-diophantine-W)

As aforementioned, a reduction of the computational load of *E-diophantine* can be achieved by performing a partial exploration of the tree of solutions instead of an exhaustive one. The approach taken in order to select which branches to explore and which ones to discard is the following and is described in pseudo-code in Algorithm 4.

First, the matrix of intersections is built in the same way as described for the *E-Diophantine* algorithm. Then, for each reservation we start the process of selecting which branch from the possible solutions tree is evaluated next until no more reservations can be considered. The selection process consists in first creating, for each reservation, a *candidate list* with the rest of the reservations with which it could potentially intersect and a *black list* with the reservations which should be discarded. This operation corresponds to the function $initiate(.)$. After this, for each reservation in the candidate list the function $find\_next\_candidate(.)$ computes its *Weight* considering the potential aggregated peak allocated capacity that would be required by this reservation and removing the reservations present in the *black list*. Once the *Weight* for all reservations has been computed, the reservation with the largest *Weight* is selected as the next branch to be explored and the *black list* updated accordingly by the function $update(.)$. The process is then iterated until no more reservations can be considered.

E-Diophantine Unweighted (E-diophantine-UW)

A complementary solution to *E-Diophantine-W* can be designed by converting our *Weighted Maximum Clique* problem to an *Unweighted* one by applying a transformation to the matrix of intersections as indicated in Eq. 4.29. Compared to the *Weighted* problem the advantage of this transformation is that our algorithm can concentrate in finding the branch of the tree of solutions with the largest length without having to consider the actual *Weight* of them. In the downside though a larger number of reservations might need to be considered depending on the k value determined by the network operator B_{ref} configuration.

We apply this principle to design an alternative to *E-diophantine-W* which we will refer to in the rest of the Section as *E-Diophantine Unweighted (E-diophantine-UW)*. In this case the same Algorithm 4 can be used by transforming the matrix of intersections used as input according to Eq. 4.29 and simplifying the $find\_next\_candidate(.)$ function such that it neglects the *Weight* information.

Algorithm 4 **E-Diophantine-W** algorithm to find out the peak resource requirement for a new reservation r_{N+1} with starting time t_{N+1}, period T_{N+1} and requirement B_{N+1} considering the set of N reservations already accepted in the system with their corresponding starting times $t = (t_1...t_N)$, periods $T = (T_1...T_N)$ and requirements $B = (B_1...B_N)$

Call executed for each new reservation request
for $i = 1$ to $N + 1$ **do**
  for $j = i + 1$ to $N + 1$ **do**
    if $solution\_exists(t_i, t_j, T_i, T_j)$ **then**
      $intersections \leftarrow find\_inters\_dioph(t_i, t_j, T_i, T_j)$
    end if
  end for
end for
$m\_inters = compute\_matrix\_inters(intersections)$
for $i = 1$ to $N + 1$ **do**
  $initiate(black\_list, candidate\_lists)$
  while $cnt + size(black\_list) < N + 1$ **do**
    $candidate\_lists \leftarrow find\_next\_candidate(m\_inters, i)$
    $update(black\_list, candidate\_lists)$
    $cnt = cnt + 1$
  end while
end for
return $find\_maximum(candidate\_lists, B)$

E-Diophantine Heuristics Complexity The first part of the *E-Diophantine-W* and *E-Diophantine-UW* algorithms is identical to the *Diophantine* one and requires $\binom{N+1}{2} \cdot log_2(T_{max})$ operations to obtain the matrix of intersections.

In addition to this operation, Algorithm 4 explores for each reservation $(N + 1)$ the rest of potential intersecting reservations (N). Then, for each of them a comparison is performed considering the candidate and black lists (N) in the function $find\_next\_candidate(.)$. Thus, the worst case complexity of the optimization algorithms can be approximated as $O(N^3)$ for *E-Diophantine-W* and $O((k \cdot N)^3)$ for *E-Diophantine-UW*.

E-Diophantine Heuristics: Probability of Estimation Error Since the *E-Diophantine* heuristics described in the previous Section do not perform an exhaustive search through the matrix of intersections but a partial one, it could happen that these solutions underestimate the actual aggregated allocated capacity peak. In this Section we analyze the probability of this

case to happen. We focus in *E-Diophantine-UW* for analysis simplicity reasons. A similar analysis can be applied to *E-Diophantine-W* case.

In order to calculate the probability of *E-Diophantine-UW* selecting a branch of solutions ($selected$) different than the one containing the actual peak (max), we consider a matrix of intersections of $N \times N$ reservations and two possible branches containing different sets of intersecting reservations $S_{selected} = I_{s_1}, ...I_{s_N}$ and $S_{max} = I_{m_1}, ...I_{m_N}$; where I takes the corresponding binary value indicating whether an intersection between two flows occur based on the matrix of intersections. Then, we define the probability of a reservation of intersecting with another as P_x and i and j as the total number of intersections in $S_{selected}$ and S_{max}, respectively. Thus, assuming that the intersections between reservations are independent from each other[5], the probability of the S_{max} branch containing the maximum number of intersections, k, instead of $S_{selected}$ can be expressed as follows

$$P_{\epsilon_{sel}} = (P_x)^j \cdot (P_x)^{2(k-1)} \cdot (P_x)^i \sum_{l=i-k+1}^{i-2} (1-P_x)^{2l} \tag{4.30}$$

where $i \geq j \geq k$, $P_x < 1,\ i < N$ and $k < N$

Since our objective is to find an upper bound for the error probability, Equation 4.30 can be simplified by assuming $j = k$ because the lower the value of j, the higher the value of $P_{\epsilon_{sel}}$

$$P_{\epsilon_{sel}} = (P_x)^{3k+i-2} \sum_{l=i-k+1}^{i-2} (1-P_x)^{2l} \tag{4.31}$$

where $i \geq k$, $j = k$, $P_x < 1,\ i < N$ and $k < N$

Equation 4.31 provides the error probability of selecting a branch from the tree of solutions that does not contain the maximum. Since an error when selecting the next branch could occur in k to i branch selection operations in the worst case, the overall error probability can be expressed as

$$P_\epsilon = \sum_{s=k}^{i} (P_x)^{3k+s-2} \sum_{l=s-k+1}^{s-2} (1-P_x)^{2l} \tag{4.32}$$

[5]Note that in general the instant at which a set of users in a system start using services (t_i, t_j, ...) and the periodicity in the need of resources of each service (T_i, T_j, ...) is uncorrelated and thus, the intersection between their reservations as well.

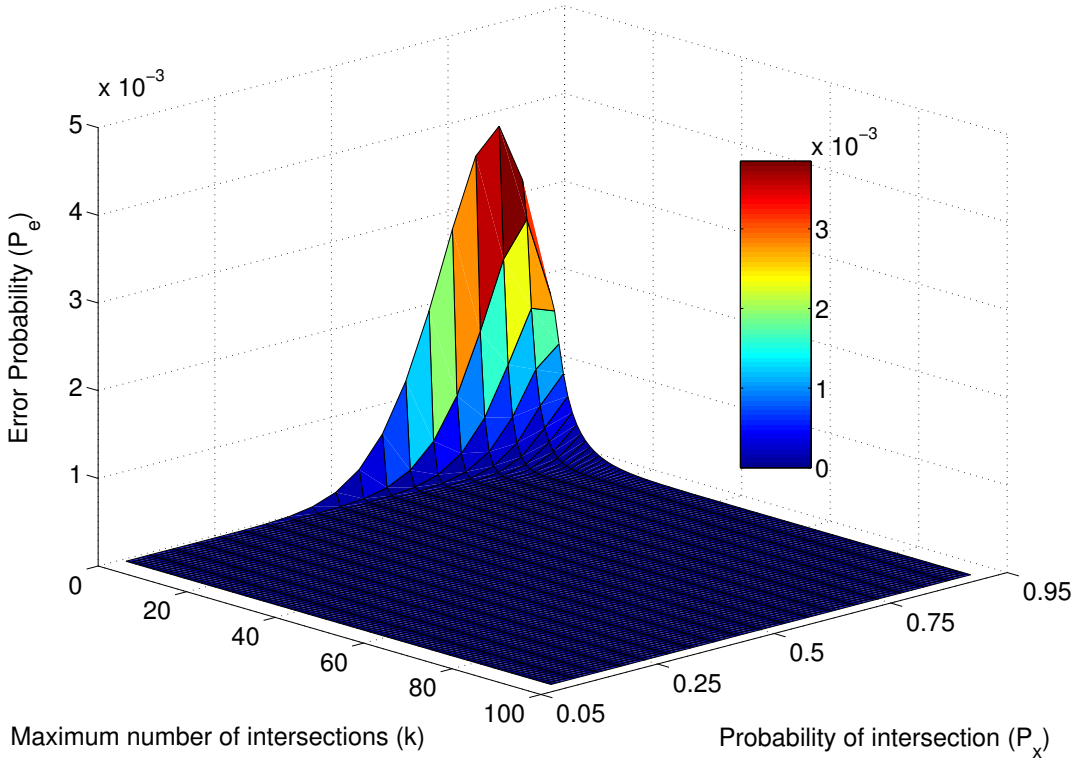


Figure 4.13. *Illustration of the error probability according to Eq.4.32 for $j = i = k$ and a wide range of P_x and k values.*

where $i \geq k$, $j = k$, $P_x < 1$, $i < N$ and $k < N$

Based on Eq.4.32 we plot in Figure 4.13 for illustration purposes the value of P_ϵ for a wide range of maximum number of intersections (k) and probabilities of intersections (P_x) considering $j = k = i$. Note that this case represents an upper bound since the larger the values of i and j above k, the lower the value of P_ϵ. Moreover, in order to compute the actual probability of different reservations of intersecting with each other, specific assumptions regarding the way their starting times and periods would be required. Thus, a sweep on different possible values has been performed to cover most cases.

As it can be observed from the figure, the *E-Diophantine* heuristics introduce a low error probability ($P_\epsilon < 0.1\%$) and present the interesting property that, as the maximum number of intersections increases (relevant case for admission control decisions), P_ϵ decreases, independently of the probability of intersection value.

	r_1	r_2	r_3	r_4	~~r_5~~	r_6	r_7	r_8	r_9	r_{10}
r_1	1	0	1	1	~~0~~	0	1	0	1	0
r_2	0	1	1	0	~~1~~	0	0	0	1	1
r_3	1	1	1	1	~~1~~	1	1	1	1	1
r_4	1	0	1	1	~~0~~	0	1	0	1	0
~~r_5~~	~~0~~	~~1~~	~~1~~	~~0~~	~~1~~	~~1~~	~~0~~	~~1~~	~~1~~	~~1~~
r_6	0	0	1	0	~~1~~	1	0	0	0	0
r_7	1	0	1	1	~~0~~	0	1	0	1	0
r_8	0	0	1	0	~~1~~	0	0	1	1	0
r_9	1	1	1	1	~~1~~	0	1	1	1	0
r_{10}	0	1	1	0	~~1~~	0	0	0	0	1

Table 4.7. *Example of removal of a reservation from the matrix of intersections*

4.2.2.5 Additional E-Diophantine Design Considerations

Dynamic Insertion and Removal of Reservations E-Diophantine solutions can adapt to entering and leaving reservations by just removing or adding new columns in the matrix of intersections.

In the case of removing reservations, the only operation needed is the deletion of the corresponding reservations in the matrix of intersections and the substraction of the corresponding reservation from the resource reservation peak previously computed if applicable. See Table 4.7 for an example.

In the case of adding a new reservation, our solution requires to find out the potential intersections between the new reservation and the ones already present in the system. The rest of the matrix of intersections remains unchanged and does not require new computations. See Table 4.8 for an example.

Channel Conditions Variation Channel conditions will vary in time due to different reasons, e.g., multipath propagation effects, co-channel interference, signal blocking by new obstacles, etc. Such interference variations might result in modulation and coding schemes changes and thus, maximum allocated resources fluctuations. Accounting for this kind of uncertainty can be achieved by the *E-Diophantine* solution, if desired, by including the chan-

	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{new}
r_1	1	0	1	1	0	0	1	0	1	?
r_2	0	1	1	0	1	0	0	0	1	?
r_3	1	1	1	1	1	1	1	1	1	?
r_4	1	0	1	1	0	0	1	0	1	?
r_5	0	1	1	0	1	1	0	1	1	?
r_6	0	0	1	0	1	1	0	0	0	?
r_7	1	0	1	1	0	0	1	0	1	?
r_8	0	0	1	0	1	0	0	1	1	?
r_9	1	1	1	1	1	0	1	1	1	?
r_{new}	?	?	?	?	?	?	?	?	?	1

Table 4.8. *Example of insertion of a new reservation in the matrix of intersections*

nel variation expected impact, α_{ch}, for a specific network and/or user profile in the reservation model described in Eq. 4.8 as follows

$$r_i = B_i \cdot \alpha_{ch} \cdot \delta_{t,t_i+n\cdot T_i} \tag{4.33}$$

User Mobility In wireless networks changes of location of users are to be expected which might result in significant channel condition fluctuations and their corresponding modulation and coding scheme variations. The reason for this channel variation is different to the one considered in the previous case and therefore, its estimated likelihood and range might also be different. If considering these potential channel conditions variations due to mobility is desired, the *E-Diophantine* solutions can be extended by including the user mobility expected impact, β_{mob}, into our reservation model as follows based on user mobility models and/or profiles

$$r_i = B_i \cdot \beta_{mob} \cdot \delta_{t,t_i+n\cdot T_i} \tag{4.34}$$

Handover Mobility of users in cellular networks might result in the need of handing over service flows from one cell to another. Handover situations can be handled by the E-diophatine solutions in a transparent manner by considering the leaving and entering reservations in a cell due to handover as ending/new reservations. In this way the same method previously described for the insertion and removal of reservations can be directly applied.

4.2.3 E-Diophantine: Multihop Extension

So far the different solutions considered for estimating the peak allocated capacity have focused in wireless point-to-multipoint networks. In this Section we consider the changes required by the *E-Diophantine* solutions presented in the previous Section to apply them in wireless multihop networks.

Two cases need to be differentiated. If a new reservation request requires resources from a single Base Station (BS), the *E-Diophantine* solutions described in Section 4.2.2 can be directly applied. On the other hand, if additional BSs are involved, the increase in the maximum capacity requirement needs to be checked for all BSs involved in the reservation path from origin till destination.

In the latter case, starting from the first Base Station in the new reservation data path and ending at the last BS involved, the new maximum capacity requirement will be sequentially computed. In order to compute the new maximum capacity requirement at each BS involved, the set of reservations already accepted in the system at each BS plus the new one need to be considered taking into account that the reservation arrival time to each next BS will be increased by a BS specific delay, D. This value needs to be configured by the network operator according to technology specific capabilities and actual BS configuration.

In the following we describe the *E-Diophantine* changes required for the downlink and uplink direction.For illustration purposes in Fig.4.14 we provide an example of an IEEE 802.16j-2009 multi-hop relay scenario showing an uplink and downlink communication involving one multihop-relay base station (MR-BS) and one or more relay stations (RS). Note that in the rest of this Section we will use the term BS to refer to all wireless network elements providing network connectivity directly to users for the sake of generality since, depending on the specific technology considered, different naming conventions would apply.

Downlink In the downlink case two potential sets of reservations need to be considered by the algorithm at every BS in addition to the new reservation requesting admission. First, a set of already admitted reservations with QoS requirements, R, arriving in the downlink direction, DL, to BS i through an interface different than the one corresponding to the wireless communication with other BSs (typically a wired interface to the core network), i.e., $R_{DL_{BS_i}} = \{r_{DL_{BS_{i_1}}}, ..., r_{DL_{BS_{i_N}}}\}$. Second, a set of already admitted reservations with QoS

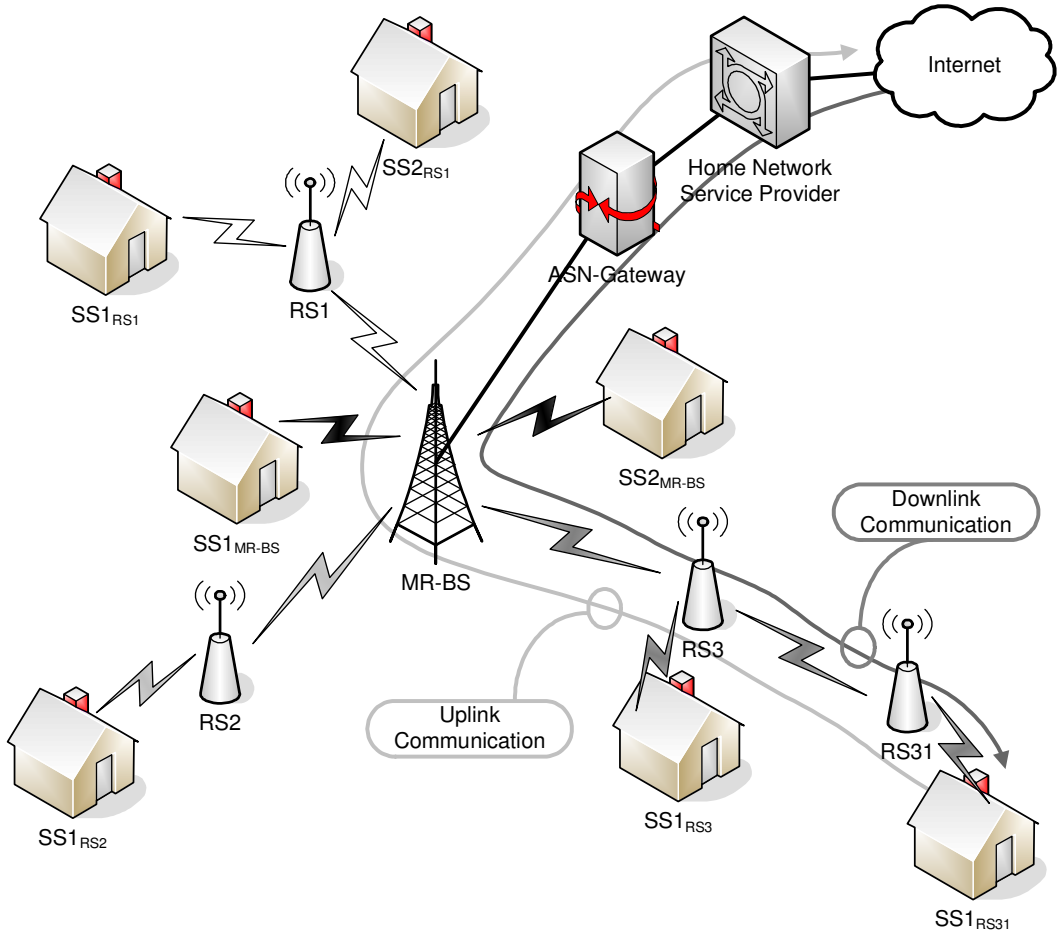


Figure 4.14. *IEEE 802.16j Multi-Hop Relay Scenario Example*

requirements arriving to BS i through the interface corresponding to the wireless communication with another BS j, i.e., $R_{DL_{BS_{j\to i}}} = \{r_{DL_{BS_{j\to i_1}}}, ..., r_{DL_{BS_{j\to i_M}}}\}$. Considering a new reservation requesting admission defined as $r_{DL_{new}}$ and L BSs wirelessly comunicating to BS_i, the set of reservations to be considered in the downlink for BS_i can be expressed as

$$R_{Total_{DL_{BS_i}}} = R_{DL_{BS_i}} \cup R_{DL_{BS_{1\to i}}} ... \cup R_{DL_{BS_{L\to i}}} \cup r_{DL_{new}} \tag{4.35}$$

While for the first set of reservations, $R_{DL_{BS_i}}$, no additional considerations need to be made compared to the previously described point-to-multipoint solutions, the arrival time of the rest of reservations needs to be modified by a BS specific delay, $D_{DL_{BS}}$, according to the BSs traversed by each of them until their arrival to the current BS under consideration, BS_i. As an example, if reservation $r_{DL_{new}}$ would need to traverse BS_1 until BS_j before being forwarded to BS_i, the modification introduced in our original model according to Eq.

4.8 would be defined as

$$r_{DL_{new}}(t) = B_{DL_{new}} \cdot \delta_{t,t_{DL_{new}}+\sum_{k=1}^{j} D_{DL_{BS_k}}+n_{DL_{new}} \cdot T_{DL_{new}}} \tag{4.36}$$

Uplink Similar to the downlink case, in the uplink direction, UL, two potential sets of reservations need to be considered by the algorithm at every BS in addition to the new reservation requesting admission. The main difference here is that the first set of reservations are originated by the Mobile Stations (MSs) associated with the Base Station under consideration and thus, Eq. 4.35 needs to be modifed as follows

$$R_{Total_{UL_{BS_i}}} = R_{UL_{MS_{BS_i}}} \cup R_{UL_{BS_{1\rightarrow i}}} ... \cup R_{UL_{BS_{L\rightarrow i}}} \cup r_{UL_{new}} \tag{4.37}$$

Similarly Eq. 4.36 becomes

$$r_{UL_{new}}(t) = B_{UL_{new}} \cdot \delta_{t,t_{UL_{new}}+\sum_{k=1}^{j} D_{UL_{BS_k}}+n_{UL_{new}} \cdot T_{UL_{new}}} \tag{4.38}$$

E-Diophantine Multihop Complexity Based on the aforementioned changes required by the algorithm to be extended to the wireless multi-hop case, we can conclude that the same complexity expressions derived in Section 4.2.2, i.e., $O(N^3)$ in the *E-Diophantine* heuristics case, can be applied by modifying the value of N accordingly. Note that in this case the main additional operation is the potential consideration of additional reservations traversing multiple Base Stations.

4.2.4 Performance Evaluation & Discussion

4.2.4.1 Estimation Algorithms of Peak Allocated Capacity

In Sections 4.2.1 and 4.2.2 we have described different options to compute the peak allocated capacity based on QoS reservation information as defined by the currently most deployed wireless standards. In this Section we compare the performance of these algorithms both in estimation accuracy and computational load terms to provide an insight on the trade-offs to be considered when deciding which option is more appropriate for a specific network deployment.

In order to evaluate the performance differences between the *Throughput-based*, *Throughput and Periodicity-based*, *Diophantine* and *E-Diophantine* approaches, we implemented these algorithms in $Matlab$ and designed the following experiment. We considered a system with

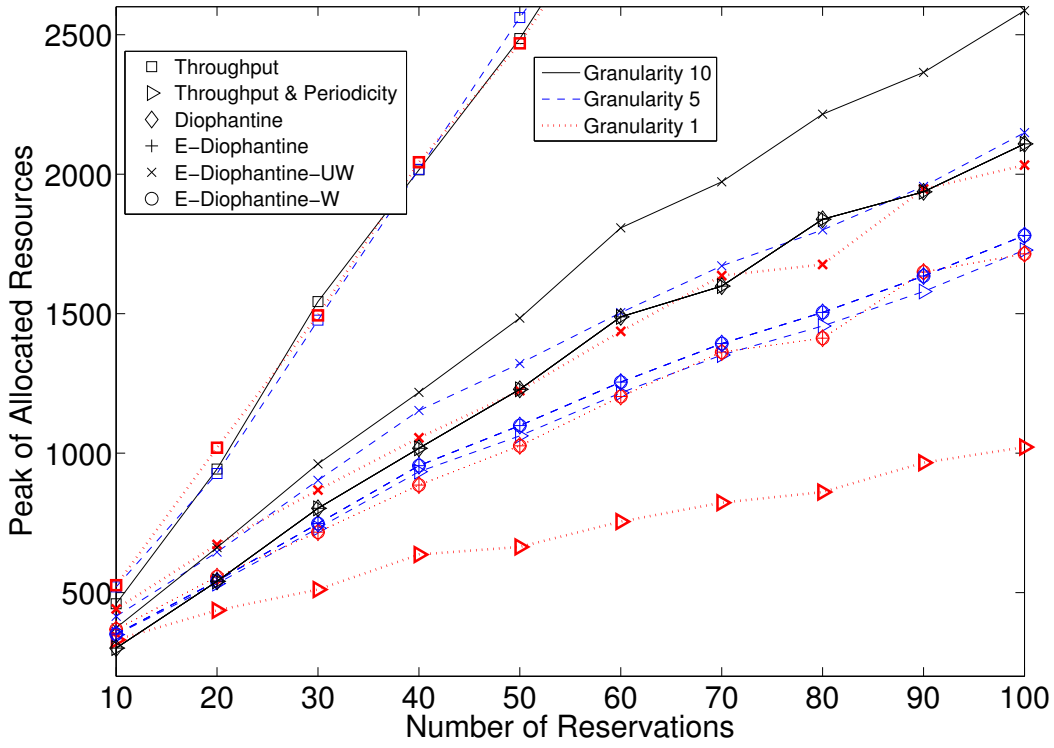


Figure 4.15. *Peak of resources allocated for services with QoS guarantees*

10 to 100 reservations where, for each one, t_i and T_i were randomly drawn from a uniform distribution ranging from the $granularity$ value chosen up to 100 (in multiples of the granularity). Three different granularity values were evaluated: 1, 5 and 10. B_i was randomly drawn from a uniform distribution ranging from 1 to 100. Figures 4.15 and 4.16 summarizes the results of the experiment after running a minimum of 100 seeds for each value. Note that the confidence intervals are not displayed in this case because their superimposition with the symbols of the different approaches would significantly degrade the readability of the Figures.

In Figure 4.15 the difference between the estimated maximum number of resources required by each of the approaches can be observed. Taking the *Diophantine* results as reference since it represents the exact solution, as expected, the *Throughput-based* approach is the one presenting the largest difference to the actual values; reaching differences of more than 100% in some of the cases. Such a large estimation deviation to the actual value would

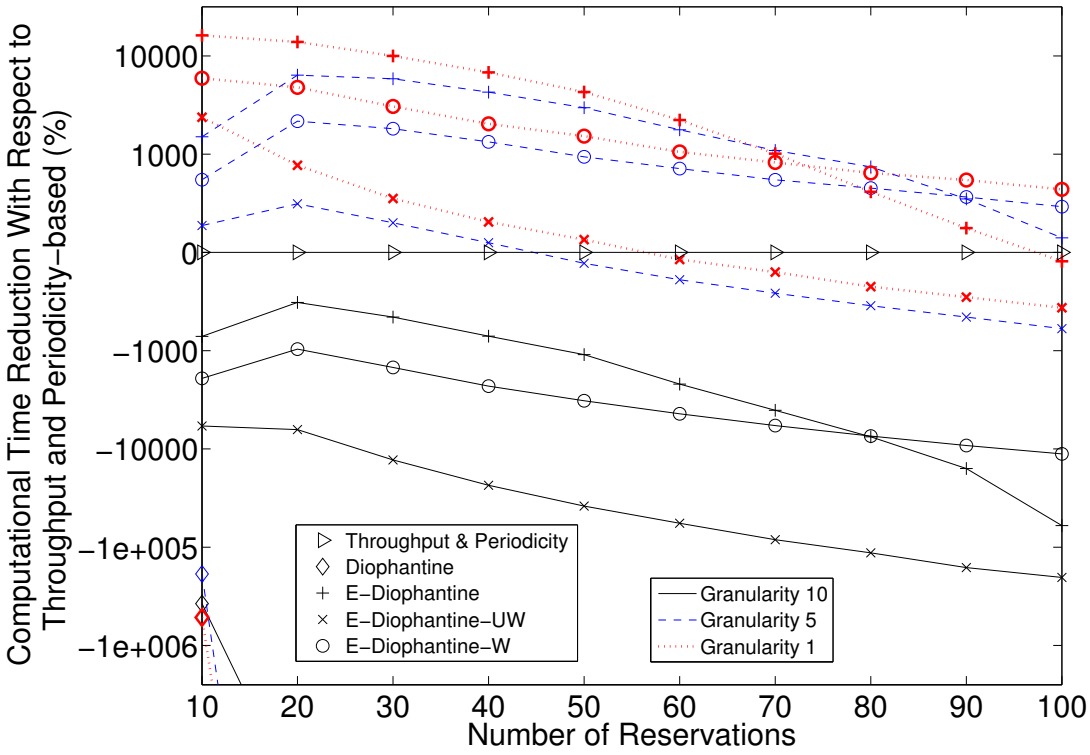


Figure 4.16. *Relative computational time of the different estimation algorithms*

obviously result in available resources being underutilized and thus, in a lower revenue for a network operator.

In the *Throughput and Periodicity-based* case, the smaller the granularity, the larger the difference to the correct value due to a limitation in the maximum T_{LCM} value that can be considered in a real implementation (10^7 in our system). Furthermore, the estimation is below the actual value and therefore, its usage for admission control purposes could compromise the QoS guarantees. Regarding the *E-Diophantine* estimation, as expected, it is always equal to the exact value (*Diophantine*). Finally, the *E-Diophantine* heuristics present a different accuracy depending on the solution considered. While *E-Diophantine-W* closely matches the correct prediction, *E-Diophantine-UW* overestimates because of the B_{ref} chosen (25 in this experiment). This B_{ref} value was chosen as an intermediate value to illustrate the trade-off to be considered when configuring the *E-Diophantine-UW* algorithm. The larger the B_{ref} value, the smaller the computational load increase as compared to the *E-Diophantine-W* so-

lution but at the cost of accuracy loss.

Regarding computational complexity, Table 4.9 summarizes the worst case complexity of the different solutions considered as derived in Sections 4.2.1 and 4.2.2. Since the upper bounds provided in Table 4.9 correspond to the worst case performance, in this experiment we analyze the differences to be expected when considering a wide range of possibilities regarding the number of flows in the system. In Figure 4.16 the computational load differences are shown, computed as the percentage of reduction achieved with respect to the *Throughput and Periodicity-based* approach, taken here as reference due to its implementation simplicity[6].

As it can be observed, the *Diophantine* solution, although exact, exceeds by far the computational load of the alternative solutions considered and thus, it might not be feasible in practice [7]. For the largest granularity considered (*granularity 10*), the *Throughput and Periodicity-based* approach clearly outperforms in computational time the *E-Diophantine* solutions with no loss of accuracy. However, as the granularity considered decreases, the *E-Diophantine* solutions reduction with respect to the *Throughput and Periodicity-based* one increases. In particular, the *E-Diophantine* algorithm presents the largest computational load advantage as compared to the *Throughput and Periodicity-based* but this advantage eventually starts decreasing as the number of reservations grows. With respect to the *E-Diophantine* heuristics, while *E-Diophantine-W* manages to keep a very considerable advantage even for a large number of reservations (100), the *E-Diophantine-UW* advantage starts lower and eventually gets outperformed by the *Throughput and Periodicity-based* algorithm. The reason is the increase in the number of reservations to be considered by the the *E-Diophantine-UW* approach which grows according to Eq. 4.29.

4.2.4.2 Wireless Network Performance Evaluation

In the previous Section we have analyzed the performance of the proposed *E-Diophantine* solution as compared to its alternatives considering a generic scenario. In this Section, we complete this evaluation by considering additional elements in the performance compar-

[6]The *Throughput-based* case is not considered since its computational load is low but its estimation accuracy very poor (see Fig. 4.15).

[7]For instance, in the 30 reservations case with granularity 10, the computation time in a 2 · Quad Core simulation server took >1000 seconds.

	Complexity
Throughput-based	$O(N)$
Throughput and Periodicity-based	$O((T_{max})^N)$
Diophantine	$O(2^N)$
E-Diophantine	$O(2^{\frac{k \cdot N}{4}})$
E-Diophantine-W	$O(N^3)$
E-Diophantine-UW	$O((k \cdot N)^3)$

Table 4.9. *Worst Case Complexity of the Different Solutions Considered*

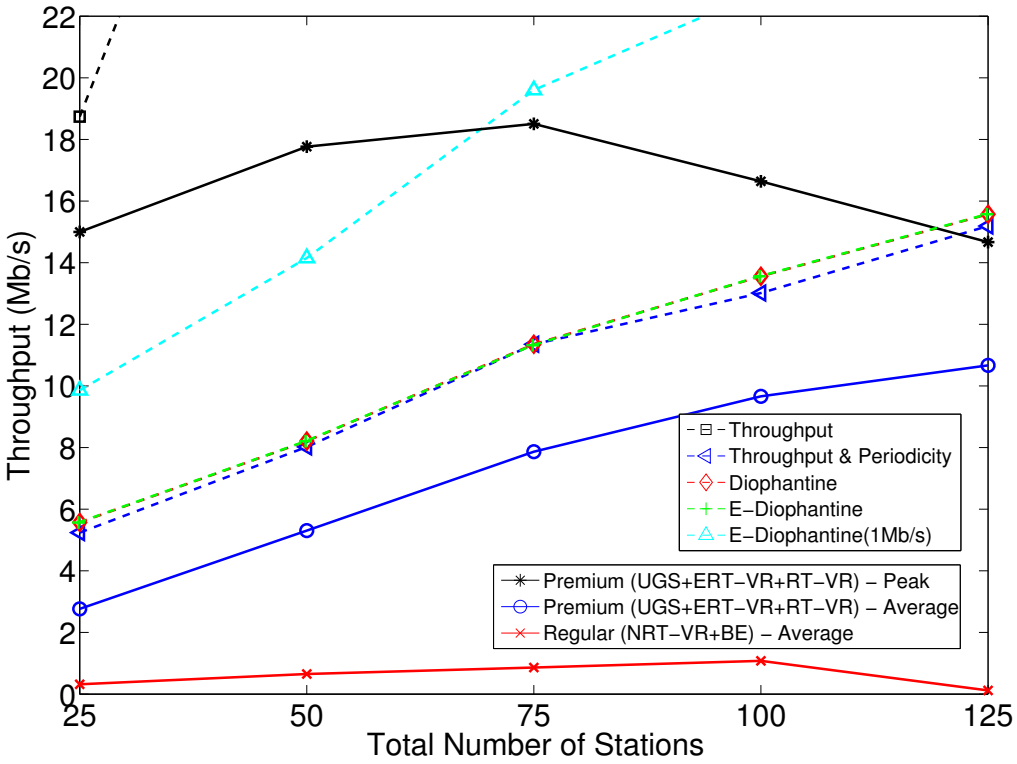


Figure 4.17. *Downlink Throughput*

ison that could have an impact in the allocated peak capacity estimation of the different approaches. Examples of these elements are: wireless physical channel, Transport layer, Network layer, MAC layer, control plane signaling, realistic applications, QoS scheduler, number of stations, etc. In order to achieve this, from the different wireless technologies mentioned in Section 4.2.1 supporting services with QoS guarantees, WiMAX is selected as

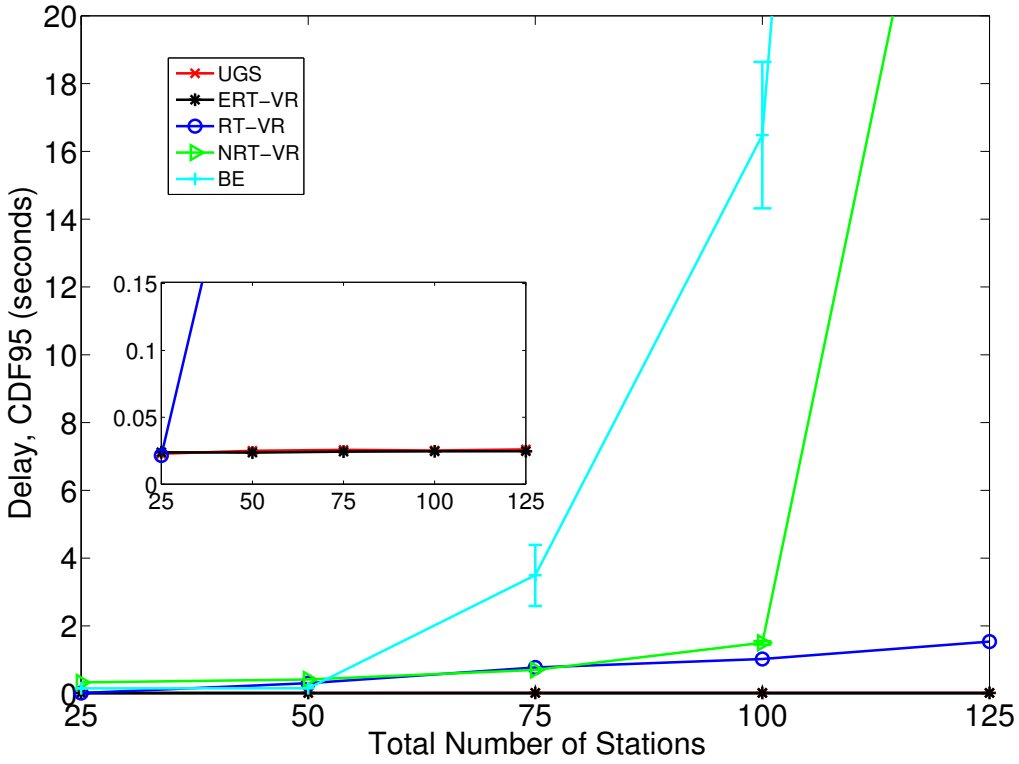


Figure 4.18. *Downlink Delay*

a representative example and OPNET's simulator [145] as evaluation tool for analyzing an IEEE 802.16-2009 and 802.16j-2009 system. The *E-Diophantine* solution used is *E-Diophantine-W* because of its better scalability properties.

An 802.16-2009 simulation scenario is setup according to Table 4.10 consisting of one Base Station (BS) and multiple Subscriber Stations (SS) where each station is configured to send and receive traffic from their corresponding pair in the wired domain of its type of application, i.e., one station sends and receives Voice traffic (without silence suppression), a second station sends and receives Voice traffic (with silence suppression), a third one receives a Video stream, a fourth one does an FTP download and the last one does Web browsing. The number of subscriber stations is increased in multiples of five stations up to 125 in total, always keeping the relation of $1/5$ of stations of each application type. The QoS scheduling policy chosen is Strict Priority applied first to fulfill the Minimum Reserved Traffic Rates (MRTRs) and then, the Maximum Sustained Traffic Rates (MSTRs). The length of

the simulations performed is 120 seconds with a warm-up phase of 10 seconds. In the case of the delay performance metric, the values represent the 95th percentile of the delay (CDF95) considering all simulation runs. Regarding the confidence intervals, they are shown only in the cases where their superimposition with the symbols of the different approaches does not degrade their readability. The configuration used for the different applications is detailed below:

- *Voice* G.711 Voice codec. Data rate: 64kb/s. Frame length: 20ms. Mapped to the *UGS* service in the downlink (BS $\rightarrow$ SS) and uplink direction.
- *VAD (Voice with Activity Detection)* G.711 Voice codec. Data rate: 64kb/s. Frame length: 20ms. Talk spurt exponential with mean 0.35 seconds. Silence spurt exponential with mean 0.65 seconds. Mapped to *ERT-VR* in the downlink and to *ertPS* in the uplink.
- *Video* MPEG-4 real traces [146].Target rate: 450 kb/s. Peak: 4.6 Mb/s. Frame generation interval: 33ms. Mapped to *RT-VR* in the downlink and to *rtPS* in the uplink.
- *FTP* Download of a 20MB file. Mapped to *NRT-VR* in the downlink and to *nrtPS* in the uplink.
- *Web Browsing* Page interarrival time exponentially distributed with mean 60s. Page size 10KB plus 20 to 80 objects of a size uniformly distributed between 5KB and 10KB [147]. Mapped to the BE service both in the downlink and uplink direction.

IEEE 802.16-2009: System Performance Results In Figure 4.17 we show the peak and average throughput experienced in the downlink by the different application types as compared to the peak capacity estimations of the different approaches described in Sections 4.2.1 and 4.2.2. The throughput of the differents applications is aggregated according to whether it is *Premium* traffic (UGS+ERT-VR+RT-VR) or *Regular* traffic (NRT-VR+BE).

From the performance results in Fig.4.17 the first remarkable result is that the peak of *Premium* traffic is in some cases above the peak estimated with the algorithms considered but the *Throughput-based* one. The reason for this result is the 2Mb/s MSTR configured for RT-VR which allows video applications to get more than its 500 Kb/s MRTR if there is leftover capacity after serving all MRTRs. Note that the *Throughput-based* estimation is too

IEEE 802.16-2009 Configuration			
Base Freq. (GHz)	2.5	DL Subfr. # Subch.	30
Bandwidth (MHz)	10	UL Subfr. # Subch.	35
Frame Duration (ms)	5	# Data Subc./Subch	24
Symbol Duration (μs)	102.86	# SSs 64 QAM (3/4)	60%
Number of Subcarriers	1024	# SSs 16 QAM (3/4)	30%
DL Subfr. # Symbols	35	# SSs QPSK (1/2)	10%
UL Subfr. # Symbols	12	# SSs per Scenario	25-125
Multipath Channel	ITU Ped-B	Pathlos Model	Sub. Fix (Erceg)

IEEE 802.16j-2009 Configuration			
	Multihop Relay BS	Relay St. 1	Relay St. 2
Downlink Res. Alloc.	17 Symbols	12 Symbols	6 Symbols
Uplink Res. Alloc.	6 Symbols	4 Symbols	2 Symbols

Data Delivery Services			**Scheduling Services**		
UGS	MRTR	80 Kb/s	UGS	MRTR	80 Kb/s
	Max. Lat	20 ms		Max. Lat	20 ms
ERT-VR	MRTR	80 Kb/s	ertPS	MRTR	80 Kb/s
	Max. Lat	20 ms		Max. Lat	20 ms
RT-VR	MSTR	2 Mb/s	rtPS	MSTR	800 b/s
	MRTR	500 Kb/s		Max. Lat	33 ms
	MRTR	800 b/s		Max. Lat	20 ms

E-Diophantine Configuration					
UGS	B_{UGS}	1600 bits	UGS	B_{UGS}	1600 bits
	T_{UGS}	20 ms		T_{UGS}	20 ms
ERT-VR	B_{ERT}	1600 bits	ertPS	B_{ERT}	1600 bits
	T_{ERT}	20 ms		T_{ERT}	20 ms
RT-VR	B_{RT}	16500 bits	rtPS	B_{RT}	800 bits
	T_{RT}	33 ms		T_{RT}	1 s

Table 4.10. *Performance Evaluation Parameters*

conservative and therefore, it will not be considered in the remainder of this Section. Also note that the difference between the *Throughput and Periodicity-based* and *Diophantine* result is caused by the LCM upper limit configured for the former, set to half of the actual required one in order to illustrate its effect on the estimation accuracy.

As the number of stations increases, the difference between the allocated peak capac-

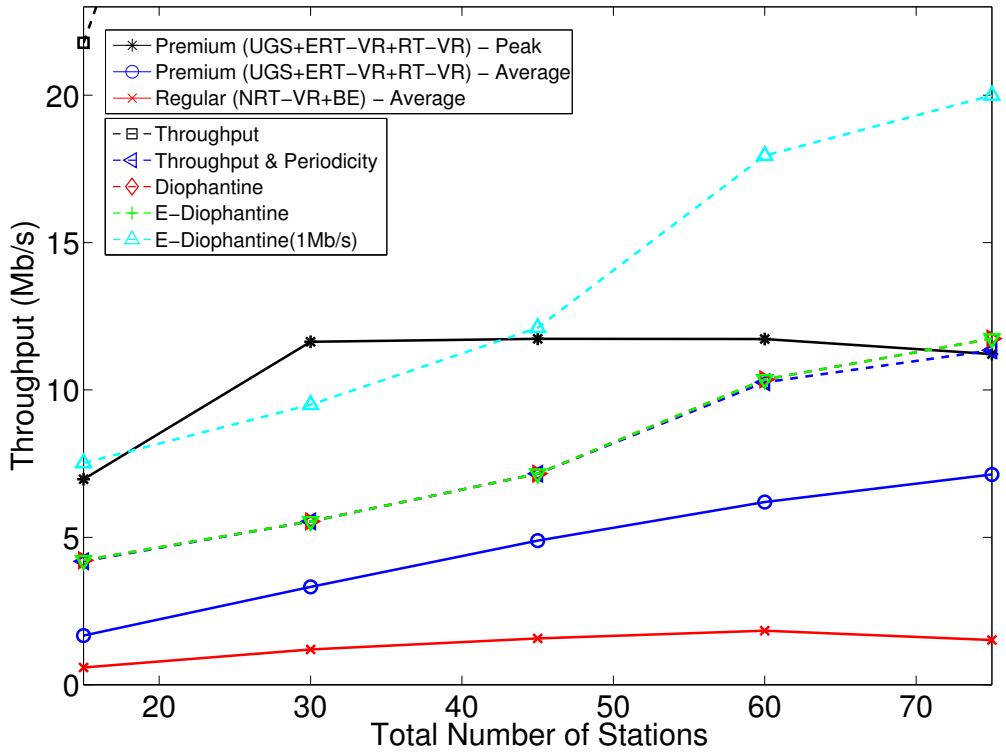


Figure 4.19. *Multihop Relay Base Station (MR-BS): Downlink Throughput*

ity estimations and the throughput peak of *Premium* traffic decreases. Note that the larger the amount of *Premium* traffic in the network, the lower the opportunities to go above the MRTR value. Eventually a point is reached where even the MRTR guarantees can not be satisfied, see crossing point between 100 and 125 stations. Moreover, as the number of flows in the system increases, the signaling overhead required for the DL-MAP increases as well, resulting in a lower *Premium* average throughput. For illustration purposes, an additional *E-diophantine* case has been added, *E-Diophantine (1Mb/s)*, where the MRTR for RT-VR has been configured to 1 Mb/s instead of 500 Kb/s. This case provides an example of how the allocated capacity peak estimation would vary by allowing bursty traffic to transmit significantly above their average.

With respect to the delay performance, the results are shown in Fig. 4.18. As expected, when the wireless resources become scarce, the delay experience degrades according to the traffic priority. In the case of RT-VR traffic, in contrast to UGS and ERT-VR, the delay experi-

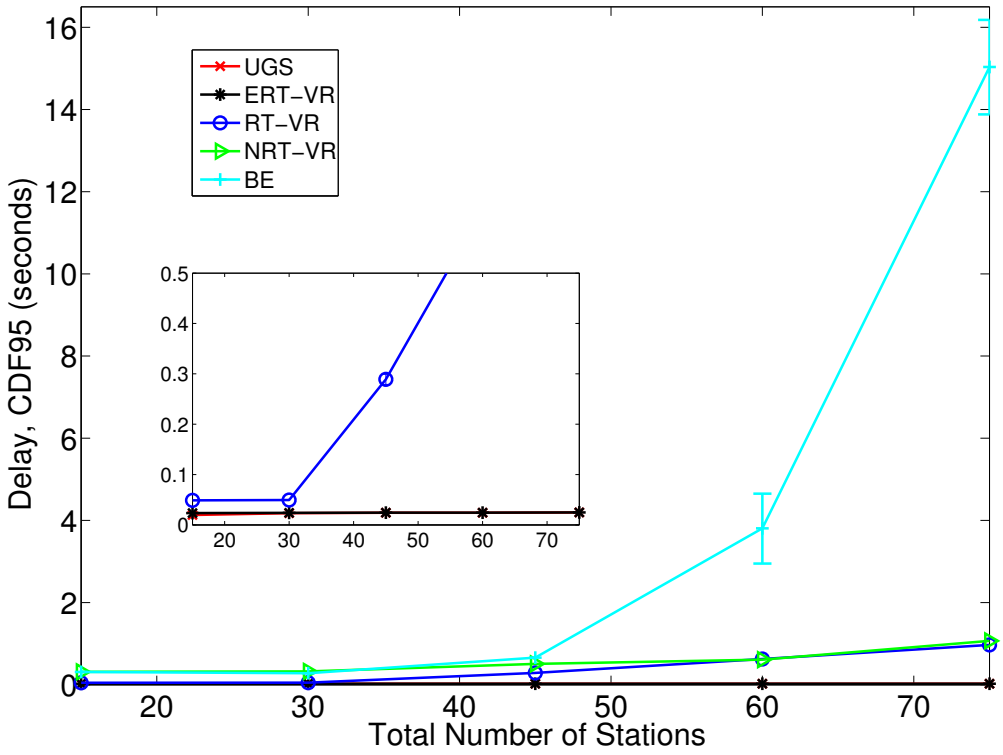


Figure 4.20. *MR-BS: Downlink Delay*

enced increases constantly. This is due to the performance metric chosen, 95% percentile of the delay (CDF95), which yields a close to worst case delay for each application traffic and thus, as the number of flows grows, it increasingly represents the Video peaks that can not be absorbed because there is not enough remaining capacity after serving all MRTRs. The delay performance of BE, which increases very rapidly, is due to the simple QoS scheduling policy used, Strict Priority, resulting in BE traffic being served only if the rest of the available traffic has already been served. Finally, the NRT-VR delay performances experience an extreme degradation after the 100 stations point. Note that this is where the peak estimation of the different algorithms but the *Throughput-based* crosses the *Premium* peak throughput and therefore, the probability of NRT-VR traffic to be served decreases significantly.

The performance results corresponding to the uplink direction have been omitted due to space restrictions since in this case there is always enough capacity to satisfy the needs of all application flows.

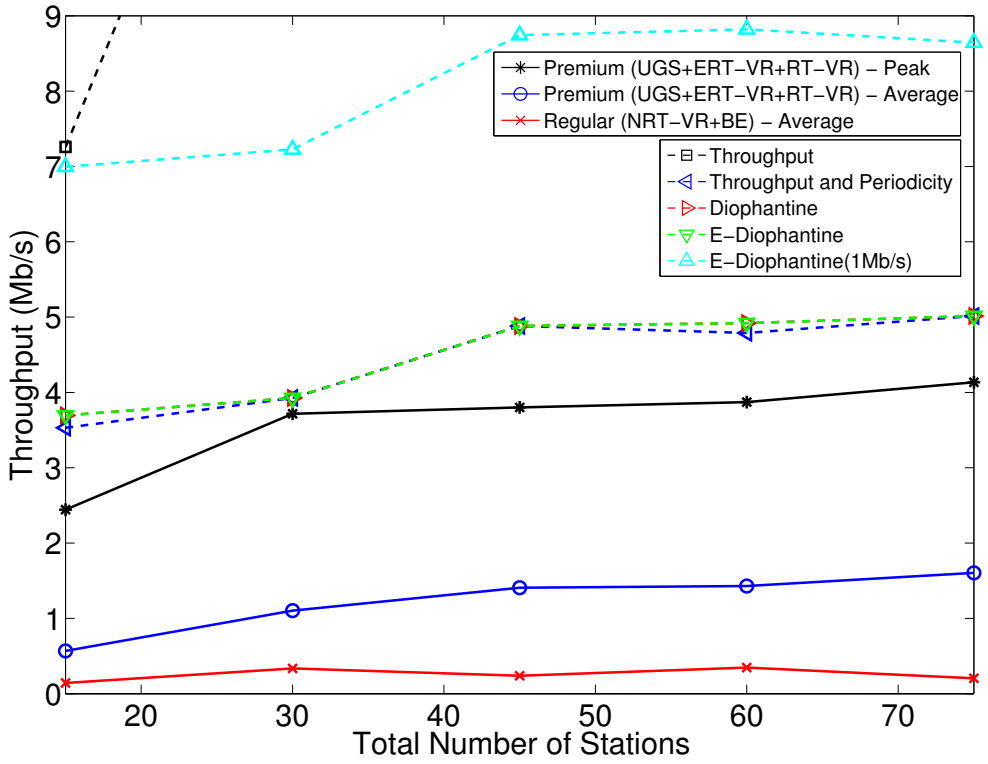


Figure 4.21. *Relay Station 2 (RS2): Downlink Throughput*

IEEE 802.16j-2009: System Performance Results In this Section we evaluate the performance of the different approaches when used in a WiMAX 802.16j-2009 Multihop Relay network [131]. The scenario used is the same as the one described in the previous Section plus adding two relay stations such that the 1-hop and 2-hop cases are considered. The Multihop Relay BS (MR-BS) and the Relay Stations (RSs) share the same spectrum resources distributed as indicated in Table 4.10. The selected mode of 802.16j operation for the relays is centralized transparent mode. Finally, the capacity estimation algorithms have been modified to consider the multihop case as described in Section 4.2.3.

In Figures 4.19 and 4.21 the throughput results for MR-BS and RS2 are shown. The performance results corresponding to the uplink direction and Relay Station 1 have been omitted due to space restrictions. Note that, as in the IEEE 802.16-2009 scenario, in the uplink direction there is always enough capacity to satisfy the needs of all application flows.

As expected, when the number of stations increases, the capacity in RS2 is the first one

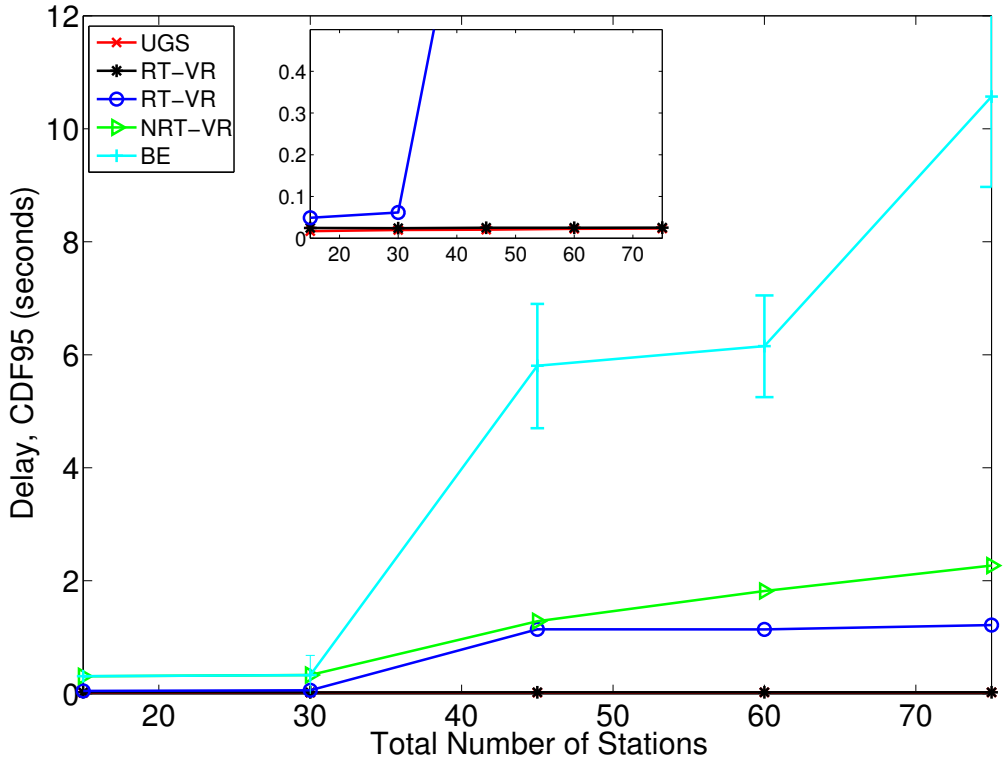


Figure 4.22. *RS2: Downlink Delay*

to reach its maximum because of the resource distribution configured, see Table 4.10. In the RS2 case of 15 stations per scheduling service the different aproaches estimate a peak requirement above the actual measured one (premium peak) which means that the MRTRs for the premium services can not always be honoured. However, some capacity is still left in some of the frames, as it can be seen in the 30 stations point, where the premium peak value increased. After this point though, the maximum peak capacity is reached and it is below the peak required estimated by the *Throughput and Periodicity-based*, *Diophantine* and *E-Diophantine* approaches. As a result, the delay of regular services significantly degrade from this point on as well as the last premium service class, RT-VR. This can be observed in Figures 4.20 and 4.22.

4.3 Summary

In Section 4.1, we have proposed an approach to enhance the offload from macro to femtocells through UE-Relays, i.e., idle users that can opportunistically serve as relays through femto backhauls. In particular, we focused on a multi-carrier deployment scenario where the relay access is orthogonal to the backhaul, and macro and femtocells share the same frequencies. We have defined a user association optimization problem which aims at maximizing the overall log utility of the network. Then, we have shown the inefficiency of a MAX-SINR based approach, thus validating the idea that, in heterogeneous cellular networks, user association should be performed on a load-basis rather than relying merely on link quality indicators.

Wireless networks providing QoS guarantees require an algorithm to estimate the increase in aggregated peak allocated capacity if a new resource reservation is admitted in the system. For this reason, in Section 4.2, we have proposed the *E-Diophantine* solution, along with its mathematical foundations, and evaluated its benefits as compared to alternative approaches of increasing complexity, namely: *Throughput-based*, *Throughput and Periodicity-based* and *Diophantine*. The performance comparison comprised both accuracy and computational load analysis in a generic scenario as well as an evaluation using OPNET's simulator in wireless point-to-multipoint (802.16-2009) and multihop (802.16j-2009) scenarios.

The main conclusions that can be drawn from our results are: i) *E-Diophantine* solutions can be successfully used to predict the allocated peak capacity demand of admitted QoS reservations in point-to-multipoint and multihop wireless networks, ii) the *Throughput and Periodicity-based* approach can outperform *E-Diophantine* in computational terms if limitations in the period between resource allocations can be imposed, iii) the larger the degree of flexibility allowed for defining the resource reservation periods, the larger the potential benefit of the *E-Diophantine* solutions both in accuracy and computational load terms and iv) for systems supporting a large number of reservations, the *E-Diophantine-W* heuristic can be used to reduce the computational load from exponential to polynomial (cubic) at a low estimation error probability cost.

Chapter 5

Conclusion

This Thesis covers a number of issues related to radio resource management in next generation wireless networks. We have proposed solutions based on alternative technologies and architectures, along with the definition of advanced techniques to enhance state-of-the-art scenarios.

In the first part of this Thesis, we have provided an overview of the wireless access components in a novel network architecture, which is designed to better support video delivery in next generation networks. Several cross-layer functions and enhancements have been proposed within this architecture, with particular focus on LTE; to achieve an efficient communication with heterogeneous access, an abstraction layer based on the IEEE 802.21 standard has been adapted to our proposed architecture. Then, to achieve more reliable results, we have introduced a link abstraction model for the simulation of Transport Block errors in the downlink of an LTE system, aiming to provide a lightweight but still accurate procedure for the computation of the residual errors, after PHY layer processing, without having to go through the detailed simulation of LTE PHY procedures. Toward this end, we combined Mutual Information-based multi-carrier compression metrics with Link-Level performance curves matching. This allowed us to obtain pre-calculated lookup tables, which can be used in an online fashion to track residual bit errors after physical layer modulation and coding procedures. In addition, we have proposed a CQI evaluation procedure which can be used as part of the LTE Adaptive Modulation and Coding scheme, showing its superiority in terms of achievable spectral efficiency with respect to current ns-3 solutions. Moreover, based on the proposed error model, we have introduced a lightweight approach to analyze

retransmission errors in LTE-HARQ. Finally, we have presented a simulation framework for the evaluation of the impact of Jumboframes in LTE networks. A comparative observation based on diverse network configuration criteria highlighted some benefits and caveats related to this approach. On the one hand, under certain channel and scheduling conditions, we observed a slight transmission time reduction, and a great overhead saving, mainly due to the reduced amount of higher layer headers. In addition, Jumboframes permit to further reduce the size of the RLC headers by requiring less segmentation/concatenation mechanisms. On the other hand, scarce resources lead to rapid buffer saturation. In order to mitigate this effect, a cross-layer approach has been proposed, allowing the radio access and the core network layers to interact and tune the packet size dynamically. We also tested our evaluation framework through the analysis of realistic video traces. There are many other issues to consider relatively to the impact of Jumboframes in wireless networks, such as error control techniques, ad-hoc allocation schemes, LTE-A Carrier Aggregation exploitation, and so forth. This work represents a first contribution for the Jumboframes feasibility in LTE networks, paving the way for further research leveraging future mobile accesses with these enhancing mechanisms.

In the second part of this Thesis, we have reported the implementation of our proposed cognitive techniques for the optimization of next generation wireless networks. The first main contribution is represented by the design and implementation of a framework for multi-operator spectrum sharing mechanisms within an LTE implementation of the well-known network simulator ns-3. The aim is to provide the scientific community with an effective and flexible simulation tool that can be easily used, and eventually extended, for the investigation of such a challenging research field. Besides an in-depth description, the resulting software has been thoroughly tested to evaluate its correctness and reliability in achieving spectrum sharing functionalities. Two different algorithms for intra-cell allocation have been implemented in order to show the flexibility of the architecture and its importance for performance comparisons. Of course, the focus of this phase was on the simulator itself and not on the algorithms, whose performance are not expected to be optimal. However, the results have been satisfactory under all aspects, showing that our proposed extension can serve as a practical tool to evaluate resource sharing mechanisms in next generation wireless networks. The code has been released and is publicly available [76]. Secondly, we

have adopted the cognitive network paradigm to address the intrinsic inefficiency of standard TCP in mobile ad hoc networks. With our probabilistic approach, we have been able to identify in real time the presence of mobility events, and we have estimated also the prediction accuracy. Through a simulation campaign, we have shown that our approach can significantly outperform the standard TCP with OLSR protocol both in a static and in a mobile scenario, in terms of increased average throughput and decreased outage probability.

Finally, in the last part of this Thesis, we have described the mathematical models introduced to enhance the radio access of heterogeneous networks. On the one hand, we have proposed an approach to enhance the offload from macro to femtocells through UE-Relays, i.e., idle users that can opportunistically serve as relays through femto backhauls. In particular, we focused on a multi-carrier deployment scenario where the relay access is orthogonal to the backhaul, and macro and femtocells share the same frequencies. We have defined a user association optimization problem which aims at maximizing the overall log utility of the network. Then, we have shown the inefficiency of a MAX-SINR based approach, thus validating the idea that, in heterogeneous cellular networks, user association should be performed on a load-basis rather than relying merely on link quality indicators. On the other hand, we have proposed the *E-Diophantine* solution, a peak allocated capacity estimator, along with its mathematical foundations, and evaluated its benefits as compared to alternative approaches of increasing complexity, namely: *Throughput-based*, *Throughput and Periodicity-based* and *Diophantine*. The performance comparison comprised both accuracy and computational load analysis in a generic scenario as well as an evaluation using OPNET's simulator in wireless point-to-multipoint and multihop scenarios. The main conclusions that can be drawn are: i) *E-Diophantine* solutions can be successfully used to predict the allocated peak capacity demand of admitted QoS reservations in point-to-multipoint and multihop wireless networks, ii) the *Throughput and Periodicity-based* approach can outperform *E-Diophantine* in computational terms if limitations in the period between resource allocations can be imposed, iii) the larger the degree of flexibility allowed for defining the resource reservation periods, the larger the potential benefit of the *E-Diophantine* solutions both in accuracy and computational load terms and iv) for systems supporting a large number of reservations, the *E-Diophantine-W* heuristic can be used to reduce the computational load from exponential to polynomial (cubic) at a low estimation error probability cost.

List of Publications

The work presented in this thesis has appeared in the articles reported below.

Journal papers

[J1] L. Anchora, **M. Mezzavilla**, L. Badia, M. Zorzi, "A Performance Evaluation Tool for Spectrum Sharing in Multi-Operator LTE Networks", *Elsevier Computer Communications*, vol. 35, no. 18, pp. 2218–2226, November 2012.

[J2] X. Pérez-Costa, Z. Wu, **M. Mezzavilla**, J. Roberto B. de Marca and J. Arauz, "E-Diophantine: An Estimation Algorithm of Peak Allocated Capacity for Wireless Point-to-Multipoint and Multihop Networks", under review.

Conference papers

[C1] X. Pérez-Costa, **M. Mezzavilla**, J. R. B. de Marca, J. Aráuz, "E-Diophantine: An Admission Control Algorithm for WiMAX Networks", in *Proc. of IEEE WCNC*, April 2010.

[C2] **M. Mezzavilla**, M. Wetterwald, L. Badia, D. Corujo, A. de la Oliva, "Wireless Access Mechanisms and Architecture Definition in the MEDIEVAL Project", in *Proc. of IEEE ISCC*, June 2011.

[C3] **M. Mezzavilla**, "Communication Protocols and Simulation Tool Development for Multimedia Traffic Optimization in LTE Networks", in *Proc. of IEEE WoWMoM*, June 2011.

[C4] L. Anchora, **M. Mezzavilla**, L. Badia, M. Zorzi, "Simulation Models for the Performance Evaluation of Spectrum Sharing Techniques in OFDMA Networks", in *Proc. of ACM MSWiM*, November 2011.

[C5] **M. Mezzavilla**, M. Miozzo, M. Rossi, N. Baldo, M. Zorzi, "A Lightweight and Accurate Link Abstraction Model for the System-Level Simulation of LTE networks in ns-3", in *Proc. of IEEE MSWiM*, October 2012.

[C6] **M. Mezzavilla**, D. Chiarotto, D. Corujo, M. Wetterwald, M. Zorzi, "Evaluation of Jumboframes Feasibility in LTE Access Networks", in *Proc. of IEEE ICC*, June 2013.

[C7] **M. Mezzavilla**, G. Quer, M. Zorzi, "On the Effects of Cognitive Mobility Prediction in Wireless Multi–hop Ad Hoc Networks", accepted at *Proc. of IEEE ICC 2014*.

[C8] **M. Mezzavilla**, K. Somasundaram, M. Zorzi, "Joint User Association and Resource Allocation in UE-Relay Assisted Heterogeneous Networks", submitted to *Proc. of IEEE ICC 2014 Workshop on Small Cell and 5G Networks*.

Bibliography

[1] "3GPP TS 36.300: Evolumeved Universal Terrestrial Radio Access (E-UTRA) and Evolumeved Universal Terrestrial Radio Access Network (E-UTRAN)," July 2012.

[2] "3GPP TS 24.312: Access Network Discovery and Selection Function (ANDSF) Management Object (MO)," June 2012.

[3] S. Sesia, I. Toufik, and M. Baker, *LTE: The UMTS Long Term Evolumeution - from Theory to Practice*, 2nd ed. John Wiley & Sons, 2011.

[4] A. D. et al., "A survey on 3gpp heterogeneous networks," *IEEE Wireless Communications*, vol. 18, no. 3, June 2011.

[5] Ericsson, "Heterogeneous networks," *White Paper*, February 2012.

[6] "The network simulator - ns-3." [Online]. Available: http://www.nsnam.org/

[7] L. Badia, R. L. Aguiar, A. Banchs, T. Melia, M. Wetterwald, M. Zorzi, "Wireless Access Architectures for Video Applications: the Approach Proposed in the MEDIEVAL Project," in *Proc. of IEEE ISCC*, June 2010.

[8] "FP7 EU project: MultimEDia transport for mobIlE Video AppLications (MEDIEVAL)," Grant agreement number 258053.

[9] Y. Shan, "Cross-Layer Techniques for Adaptive Video Streaming over Wireless Networks," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 2, pp. 220–228, 2005.

[10] D. S. T. M. van der Schaar, "Content-Based Cross-layer Packetization and Retransmission Strategies for Wireless Multimedia Transmission," *IEEE Transactions on Multimedia*, vol. 9, no. 1, pp. 185–197, January 2007.

[11] "IEEE 802.21: Standard for Local and Metropolitan Area Networks– Part 21: Media Independent Handover Services," 2008.

[12] G. Piro, N. Baldo, and M. Miozzo, "An LTE Module for the ns-3 Network Simulator," in *Proc. of SIMUTOOLS*, March 2011.

[13] "ns-3 Lena Project [Online]. Available:." [Online]. Available: http://code.nsnam.org/nbaldo/ns-3-lena-trunk/

[14] L. Badia, N. Baldo, M. Levorato, and M. Zorzi, "A Markov Framework for Error Control Techniques Based on Selective Retransmission in Video Transmission over Wireless Channels," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 3, pp. 488–500, April 2010.

[15] "IEEE P802.21: IEEE Standard for Local and Metropolitan Area Networks: Media Independent Handover Services," January 2009.

[16] "3GPP TS 36.214: (E-UTRA) Physical Layer and Measurements."

[17] "3GPP TS 36.213: E-UTRA Physical Layer Procedures."

[18] Y. Blankenship, P. Sartori, B. Classon, V. Desai, and K. Baum, "Link Error Prediction Methods for Multicarrier Systems," in *Proc. of IEEE VTC*, vol. 6, September 2004.

[19] E. Tuomaala and Haiming Wang, "Effective SINR Approach of Link to System Mapping in OFDM/Multi-Carrier Mobile Network," *IEE Conference Publications*, vol. 2005, pp. 140–140, 2005.

[20] M. Moisio and A. Oborina, "Comparison of Effective SINR Mapping with Traditional AVI Approach for Modeling Packet Error Rate in Multi-State Channel," *Lecture Notes in Computer Science: Next Generation Teletraffic and Wired/Wireless Advanced Networking*, vol. 4003, pp. 461–473, 2006.

[21] WiMAX Forum, "WiMAX System Evaluation Methodology Document. Version 1.0," January 2007.

[22] "The Network Simulator - ns-2." [Online]. Available: http://www.isi.edu/nsnam/ns/

[23] A. Sayenko, O. Alanen, H. Martikainen, V. Tykhomyrov and O. Puchko, "WINSE: WiMAX NS-2 Extension," in *Proc. of ICST SIMUTOOLS*, March 2009.

[24] M. Miozzo and F. Bader, "Accurate Modeling of OFDMA Transmission Technique using IEEE 802.16m Recommendations for WiMAX Network Simulator Design," in *Proc. of ICST MONAMI*, September 2009.

[25] B. Sadiq, R. Madan and A. Sampath, "Downlink Scheduling for Multiclass Traffic in LTE," in *EURASIP Journal on Wireless Communications and Networking - 3GPP LTE and LTE Advanced*, vol. 2009, no. 14, 2009.

[26] J. Fan, Q. Yin, G.Y. Li, B. Peng and X. Zhu, "MCS Selection for Throughput Improvement in Downlink LTE Systems," in *Proc. of ICCCN*, August 2011.

[27] C. Mehlfuhrer, J. Colom Ikuno, M. Simko, S. Schwarz, M. Wrulich and M. Rupp, "The Vienna LTE simulators: Enabling Reproducibility in Wireless Communications Research," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, July 2011.

[28] G. Piro, L.A. Grieco, G. Boggia, F. Capozzi and P. Camarda, "Simulating LTE Cellular Systems: An Open-Source Framework," *IEEE Transactions on Vehicular Technology*, vol. 2011, pp. 498–513, February 2011.

[29] "The lte-epc network simulator (lena) project." [Online]. Available: http://iptechwiki.cttc.es/LTE-EPC_Network_Simulator_(LENA)

[30] N. Baldo, M. Requena-Esteso, J. Nin-Guerrero, and M. Miozzo, "A New Model for the Simulation of the LTE-EPC Data Plane," in *Proc. of ICST SIMUTOOLS*, March 2012.

[31] SmallCell-Forum, "LTE MAC Scheduler Interface Specification v1.11," October 2010. [Online]. Available: http://www.smallcellforum.org/resources-technical-papers

[32] N. Baldo, M. Miozzo, M. Requena-Esteso, and J. Nin-Guerrero, "An Open Source Product-Oriented LTE Network Simulator based on ns-3," in *Proc. of ACM MSWiM*, November 2011.

[33] "LENA documentation, http://lena.cttc.es/manual/." [Online]. Available: http://lena.cttc.es/manual/

[34] N. Baldo and M. Miozzo, "Spectrum-aware Channel and PHY Layer Modeling for ns-3," in *Proc. of NSTOOLS*, October 2009.

[35] "3GPP TS 36.101: "E-UTRA User Equipment (UE) radio transmission and reception"."

[36] "3GPP TS 36.104: "E-UTRA Base Station (BS) radio transmission and reception"."

[37] K. Brueninghaus, D. Astely, T. Salzer, S. Visuri, A. Alexiou, S. Karger, and G.-A. Seraji, "Link Performance Models for System Level Simulations of Broadband Radio Access Systems," in *Proc. of IEEE PIMRC*, September 2005.

[38] "The Vienna LTE Simulators." [Online]. Available: http://www.nt.tuwien.ac.at/about-us/staff/josep-colom-ikuno/lte-simulators/

[39] C. Mehlfuhrer, M. Wrulich, J. C. Ikuno, D. Bosanska, and M. Rupp, "Simulating the Long Term Evolumeution Physical Layer," in *Proc. of EUSIPCO*, August 2009.

[40] "IEEE 802.16: Broadband Wireless Access Working Group, Evaluation Methodology Document (EMD)."

[41] IST-2000-30116 FITNESS Project, "MTMR Baseband Transceivers Needs for Intra-system and Inter-system (UMTS/WLAN) Reconfigurability," Deliverable 3.3.1.

[42] "IEEE802.11-04/0269: "PHY abstraction based on PER prediction". March 2004."

[43] "3GPP TS 25.892: "Feasibility Study for OFDM for UTRAN enhancement (release6)". v1.1.0, (2004-03)."

[44] Ericsson, "Effective SNR Mapping for Modeling Frame Error Rates in Multiple-State Channels," 3GPP2-C30-20030429-010, April 2003.

[45] J. Colom Ikuno, M. Wrulich and M. Rupp, "Performance and Modeling of LTE-HARQ," in *Proc. of ITG WSA*, February 2009.

[46] Hanbyul Seo and Byeong Gi Lee, "A Proportional-Fair Power Allocation Scheme for Fair and Efficient Multiuser OFDM Systems," in *Proc. of IEEE GLOBECOM*, December 2004.

[47] "3GPP R1-081483: "Conveying MCS and TB size via PDCCH."

[48] "WiMAX Forum White Paper: WiMAX System Evaluation Methodology," July 2008.

[49] "ITU-T G.984.1: Gigabit-Capable Passive Optical Networks (GPON): General Characteristics," March 2008.

[50] M. Mathis, J. Semke, J. Mahdavi, and T. Ott, "The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm," *Computer Communication ACM SIGCOMM*, vol. 27, no. 3, July 1997.

[51] J. Mogul and S. Deering, "Path MTU Discovery," IETF RFC 1063, November 1990.

[52] N. Garcia, P. Monteiro, and M. Freire, "Burst Assembly with Real IPv4 Data - Performance Assessment of Three Assembly Algorithms," in *Proc. of IEEE NEW2AN*, June 2006.

[53] P. Koopman, "32-Bit Cyclic Redundancy Codes for Internet Applications," in *Proc. of DSN*, April 2002.

[54] A. Iyer, G. Deshpande, E. Rozner, A. Bhartia, and L. Qiu, "Fast Resilient Jumbo frames in wireless LANs," in *Proc. of IWQoS*, July 2009.

[55] "3GPP TS 36.913: Requirements for further advancements for Evolumeved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced)," April 2006.

[56] "3GPP TS 36.322: E-UTRA Radio Link Control (RLC) protocol specification."

[57] "Video traces for network performance evaluation, [online]. available:." [Online]. Available: http://www2.tkn.tu-berlin.de/research/trace/trace.html

[58] A. MacKenzie and L. DaSilva, *Game Theory for Wireless Engineers*. Morgan and Claypool, 2006.

[59] S. Geirhofer, L. Tong, and B. Sadler, "Dynamic Spectrum Access in the Time Domain: Modeling and Exploiting White Space," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 66–72, May 2007.

[60] V. Srinivasan, P. Nuggehalli, C. F. Chiasserini, and R. R. Rao, "Cooperation in Wireless Ad Hoc Networks," in *Proc. of IEEE INFOCOM*, July 2003.

[61] J. M. III and G. M. Jr., "Cognitive Radio: Making Software Radios more Personal," *IEEE Personal Communications Magazine*, vol. 6, no. 4, pp. 13–18, 1999.

[62] J. Zhang and Q. Zhang, "Stackelberg Game for Utility-Based Cooperative Cognitive Radio Networks," in *Proc. of ACM MobiHoc*, May 2009.

[63] L. Badia and M. Levorato and F. Librino and M. Zorzi, "Cooperation Techniques for Wireless Systems from Networking Perspective," *IEEE Wireless Communications Magazine*, vol. 17, no. 2, pp. 89 – 96, April 2010.

[64] A. Leshem and E. Zehavi, "Cooperative Game Theory and the Gaussian Interference Channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1078–1088, September 2008.

[65] P. Liu, Z. Tao, S. Narayanan, T. Korakis, and S. S. Panwar, "CoopMAC: a Cooperative MAC for Wireless LANs," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 340–354, February 2007.

[66] E. Larsson and E. Jorswieck, "Competition Versus Cooperation on the MISO Interference Channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1059–1069, September 2008.

[67] "Simlib simulator, [online]. available:." [Online]. Available: http://www.fit.vutbr.cz/~peringer/SIMLIB/

[68] "Matlab, [online]. available:." [Online]. Available: http://www.mathworks.com/products/matlab/

[69] "Omnet++ simulator, [online]. available:." [Online]. Available: http://www.omnetpagesorg/

[70] A. B. Carleial, "Interference Channels," *IEEE Transactions on Information Theory*, vol. 24, no. 1, pp. 60–70, January 1978.

[71] R. Etkin, A. Parekh, and D. Tse, "Spectrum Sharing for Unlicensed Bands," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 517–528, April 2007.

[72] B. Aazhang, J. Lilleberg, and G. Middleton, "Spectrum Sharing in a Cellular System," in *Proc. of IEEE ISSSTA*, August 2004.

[73] G. Middleton, K. Hooli, A. Tölli, and J. Lilleberg, "Inter-Operator Spectrum Sharing in a Broadband Cellular Network," in *Proc. of IEEE ISSSTA*, August 2006.

[74] M. Bennis and J. Lilleberg, "Inter Base Station Resource Sharing and Improving the Overall Efficiency of B3G Systems," in *Proc. of IEEE VTC*, April 2007.

[75] M. Bennis, S. Lasaulce, and M. Debbah, "Inter-Operator Spectrum Sharing from a Game Theoretical Perspective," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 1–12, 2009.

[76] "ns-3 Spectrum Sharing Framework for LTE, [Online]. Available:." [Online]. Available: http://code.nsnam.org/lanchora/ns-3-lte-SpectrumSharing/

[77] L. Anchora, L. Canzian, L. Badia, and M. Zorzi, "A Characterization of Resource Allocation in LTE Systems Aimed at Game Theoretical Approaches," in *Proc. of IEEE CAMAD*, December 2010.

[78] L. Anchora, L. Badia, and M. Zorzi, "Joint Scheduling and Resource Allocation for LTE Downlink Using Nash Bargaining Theory," in *Proc. of IEEE ICC - Workshop on Game Theory for Resource Allocation*, June 2011.

[79] M. Osborne and A. Rubinstein, *A Course in Game Theory*. MIT Press, 1994.

[80] Z. Fu, P. Zerfos, H. Luo, S. Lu, L. Zhang, and M. Gerla, "The Impact of Multihop Wireless Channel on TCP Throughput and Loss," in *Proc. of IEEE INFOCOM*, March 2003.

[81] B. Manoj, R. Rao, and M. Zorzi, "Cognet: a cognitive complete knowledge network system," *IEEE Wireless Communications Magazine*, vol. 15, no. 6, pp. 81–88, December 2008.

[82] G. Holland and N. Vaidya, "Analysis of TCP performance over mobile Ad Hoc networks," *Wireless Networks (Kluwer Academic Publishers)*, vol. 8, pp. 275–288, March 2002.

[83] S. Xu and T. Saadawi, "Performance Evaluation of TCP Algorithms in Multi-hop Wireless Packet Networks," *Wireless Communications and Mobile Computing (Wiley)*, vol. 2, no. 1, pp. 85–100, February 2002.

[84] K. Xu, S. Bae, S. Lee, and M. Gerla, "TCP Behavior across Multihop Wireless Networks and the Wired Networks," in *Proc. of ACM WoWMoM*, September 2002.

[85] B. Sreenivas, G. B. Prakash, and K. Ramakrishnan, "L2DB-TCP: An adaptive congestion control technique for MANET based on link layer measurements," in *Proc. of IEEE IACC*, February 2013.

[86] S. R. Chokhandre and U. Shrawankar, "An Algorithm to Improve Performance over Multihop Wireless Mesh Network," *Journal of Computing*, vol. 3, no. 5, May 2011.

[87] K. Chen, Y. Xue, and K. Nahrstedt, "On setting TCP's Congestion Window Limit in Mobile Ad Hoc Networks," in *Proc. of IEEE ICC*, May 2003.

[88] H. Balakrishnan, V. Padmanabhan, S. Seshan, and R. Katz, "A Comparison of Mechanisms for Improving TCP Performance over Wireless Links," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, December 1997.

[89] K. Sundaresan, V. Anantharaman, H.-Y. Hsieh, and R. Sivakumar, "ATP: A Reliable Transport Protocol for Ad hoc Networks," in *Proc. of ACM MobiHoc*, June 2003.

[90] A. Ahuja, S. Agarwal, J. Sing, and R. Shorey, "Performance of TCP over Different Routing Protocols in Mobile Ad Hoc Networks," in *Proc. of IEEE VTC*, May 2000.

[91] T. Dyer and R. Boppana, "A Comparison of TCP Performance over Three Routing Protocols for Mobile Ad Hoc Networks," in *Proc. of ACM MobiHoc*, October 2001.

[92] J. Liu and S. Singh, "ATCP: TCP for Mobile Ad Hoc Networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 7, July 2001.

[93] T. Rasheed, U. Javaid, M. Jerbi, and K. A. Agha, "Scalable Multi-hop Ad Hoc Routing Using Modified OLSR Routing Protocol," in *Proc. of IEEE PIMRC*, September 2007.

[94] A. Ouacha, N. Lakki, J. E. Abbadi, A. Habbani, and M. E. Koutbi, "OLSR protocol enhancement through mobility integration," in *Proc. of IEEE ICNSC*, April 2013.

[95] L. Barolli, M. Ikeda, F. Xhafa, and A. Duresi, "A testbed for manets: Implementation, experiences and learned lessons," *IEEE Systems Journal*, vol. 4, no. 2, June 2010.

[96] M. Hiyama, E. Kulla, M. Ikeda, L. Barolli, and M. Takizawa, "Investigation of OLSR Behavior for Different Hello Packets Intervals in a MANET Testbed," in *Proc. of IEEE AINA*, March 2013.

[97] G. Anastasi, E. Ancillotti, M. Conti, and A. Passarella, "Experimental Analysis of TCP Performance in Static Multi-hop Ad Hoc Networks," in *chapter 6 of Mobile Ad Hoc Networks: from Theory to Reality (M. Conti, J. Crowcroft, and A. Passarella, Editors - Nova Science Publishr)*, June 2007.

[98] V. Kawadia and P. Kumar, "Experimental investigation into TCP Performance over Wireless Multihop Networks," in *Proc. of ACM SIGCOMM*, August 2005.

[99] G. Quer, H. Meenakshisundaram, B. Tamma, B. S. Manoj, R. Rao, and M. Zorzi, "Using Bayesian Networks for Cognitive Control of Multi-hop Wireless Networks," in *Proc. of IEEE MILCOM*, San Jose, CA, US, November 2010.

[100] G. Quer, F. Librino, L. Canzian, L. Badia, and M. Zorzi, "Inter-Network Cooperation Exploiting Game Theory and Bayesian Networks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 2110–2122, March 2013.

[101] B. Bojovic, G. Quer, N. Baldo, and R. Rao, "Bayesian and neural network schemes for call admission control in lte systems," in *Proc. IEEE GLOBECOM*, Atlanta, GA, USA, December 2013.

[102] T. Clausen and P. Jacquet, "Optimized Link State Routing Protocol," in *RFC 3626*, Oct. 2003.

[103] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

[104] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2003.

[105] F. V. Jensen and T. D. Nielsen, *Bayesian Networks and Decision Graphs*. Springer, 2007.

[106] D. B. da Costa and S. Aissa, "Performance Analysis of Relay Selection Techniques With Clustered Fixed-Gain Relays," *IEEE Signal Processing Letters*, vol. 17, no. 2, February 2010.

[107] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User Association for Load Balancing in Heterogeneous Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, June 2013.

[108] X. Wang and G. B. Giannakis, "Resource Allocation for Wireless Multiuser OFDM Networks," *IEEE Transactions on Information Theory*, vol. 57, no. 7, March 2011.

[109] H. Zhang, W. Zheng, X. Chu, X. Wen, M. Tao, A. Nallanathan, and D. Lopez-Perez, "Joint subchannel and power allocation in interference-limited OFDMA femtocells with heterogeneous QoS guarantee," in *Proc. of IEEE GLOBECOM*, December 2012.

[110] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in Heterogeneous Networks: Modeling, Analysis, and Design Insights," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, January 2013.

[111] A. Rath, S. Hua, and S. S. Panwar, "FemtoHaul: Using femtocells with relays to increase macrocell backhaul bandwidth," in *Proc. of IEEE INFOCOM Workshops*, July 2010.

[112] I. F. Akyildiz, D. M. Gutierrez-Estevez, and E. C. Reyes, "FEMTO-RELAY SYSTEMS AND METHODS OF MANAGING SAME," in *US Patent Application 20120076027*, March 2012.

[113] I. F. Akyildiz, E. C. Reyes, D. M. Gutierrez-Estevez, R. Balakrishnan, and J. R. Krier, "Enabling Next Generation Small Cells through Femtorelays," *Elsevier (Physical Communications)*, vol. 9, April 2013.

[114] T. Elkourdi and O. Simeone, "Femtocell as a Relay: An Outage Analysis," *IEEE Transactions on Wireless Communications*, vol. 10, no. 12, December 2011.

[115] S. Samarakoon, M. Bennis, W. Saad, and M. Latva-aho, "Enabling relaying over heterogeneous backhauls in the uplink of femtocell networks," in *Proc. of IEEE WiOpt*, May 2012.

[116] A. T. Gamage, M. S. Alam, X. Shen, and J. W. Mark, "Joint relay, subcarrier and power allocation for OFDMA-based femtocell networks," in *Proc. of IEEE WCNC*, April 2013.

[117] H. Alves, M. Bennis, R. D. Souza, and M. Latva-aho, "Enhanced performance of heterogeneous networks through full-duplex relaying," *EURASIP Journal on Wireless Communications and Networking*, vol. 12, no. 365, December 2012.

[118] P. Jacob and A. S. Madhukumar, "Interference reduction through femto-relays," *IET Communications*, vol. 6, no. 14, September 2012.

[119] D. Zhou and W. Song, "Interference-Controlled Load Sharing with Femtocell Relay for Macrocells in Cellular Networks," in *Proc. of IEEE GLOBECOM*, December 2011.

[120] I. Basturk, B. Ozbek, C. A. Edemen, A. S. Tan, E. Zeydan, and S. Ergut, "Radio Resource Management for OFDMA-Based Mobile Relay Enhanced Heterogenous Cellular Networks," in *Proc. of IEEE VTC*, June 2013.

[121] R. H. Byrd, M. E. Hribar, and J. Nocedal, "An interior point algorithm for large scale nonlinear programming," *SIAM Journal on Optimization*, vol. 9, 1997.

[122] M. Gen and R. Cheng, *Genetic Algorithms and Engineering Optimization*. Wiley-Interscience, 1999.

[123] Z. Ugray, L. Lasdon, J. Plummer, F. Glover, J. Kelly, and R. Martí, "Scatter search and local nlp solvers: A multistart framework for global optimization," *INFORMS Journal on Computing*, vol. 19, no. 3, July 2007.

[124] "3GPP TS 36.300: Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Terrestrial Radio Access Network (E-UTRAN). Overall Description Stage 2," Dec 2011.

[125] "3GPP TS 36.913: Requirements for further advancements for Evolved Universal Terrestrial Radio Access (E-UTRA, lte-advanced," March 2011.

[126] "IEEE Standard 802.16-2009: Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Broadband Wireless Access Systems," May 2009.

[127] "IEEE 802.16 Working Group: DRAFT Amendment to IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Broadband Wireless Access Systems," March 2011.

[128] "IEEE 802.11E: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements," November 2005.

[129] "IEEE 802.11-2007: Standard for Local and metropolitan area networks. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications." June 2007.

[130] X. Pérez-Costa, M. Mezzavilla, J. de Marca, and J. Aráuz, "E-Diophantine: An Admission Control Algorithm for WiMAX Networks," in *Proc. of IEEE WCNC*, April 2010.

[131] "IEEE 802.16j-2009: Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Broadband Wireless Access Systems Amendment 1: Multiple Relay Specification," May 2009.

[132] T. G. P. P. (3GPP), "Technical Specification Group Services and System Aspects. Policy and charging control architecture," 3GPP TS 23.203 v.9.10.0, Dec 2011.

[133] A. Elwalid and D. Mitra, "Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks," *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pp. 329 – 343, 2002.

[134] J. Evans and D.Everitt, "Effective Bandwidth Based Admission Control for Multi-Service CDMA Cellular Networks," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 1, pp. 36 – 46, 2002.

[135] S. Jamin, S. Shenker, and P. Danzig, "Comparison of Measurement-Based Admission Control algorithms for Controlled-Load Service," in *Proc. of IEEE INFOCOM*, April 1997.

[136] H. Wang, W. Li, and D. Agrawal, "Dynamic Admission Control and QoS for 802.16 Wireless MAN," in *Proc. of IEEE WTS*, April 2005.

[137] C.Ciconetti, L.Lenzini, E.Mingozzi, and G.Stea, "Design and Performance Analysis of the Real-Time HCCA Scheduler for IEEE 802.11e WLANs," Elsevier Computer Networks Journal (CN), Volume 51, Issue 9, pp. 2311 – 2325, June 2007.

[138] S. Chandra and A. Sahoo, "An efficient call admission control for IEEE 802.16 networks," in *Proc. of IEEE Workshop on Local and Metropolitan Area Networks*, June 2007.

[139] O.Yang and J.Lu, "Call Admission Control and Scheduling Schemes with QoS support for Real-time Video Applications in IEEE 802.16 Networks," In IEEE Journal of Multimedia, pp. 21 – 29, May 2006.

[140] R. D. Luce and A. D. Perry, "A Method of Matrix Analysis of Group Structure," In Psychometrika Journal, pp. 95 – 116, June 1949.

[141] F. Harary and I. C. Ross, "A Procedure for Clique Detection Using the Group Matrix," In Sociometry Journal, pp. 205 – 215, 1957.

[142] D.-Z. Du and P. M. Pardalos, "Handbook of Combinatorial Optimization - The Maximum Clique Problem," Kluwer Academic Publishers, 1999.

[143] T. Cormen, C. Leiserson, R. Rivest, and C.Stein, "Introduction to Algorithms," The MIT Press, 2001.

[144] J. M. Robson, "Finding a maximum independent set in time $O(2^{n/4})$," 2001. [Online]. Available: http://www.labri.fr/perso/robson/mis/techrep.html

[145] "OPNET Simulator." [Online]. Available: http://www.opnet.com

[146] F. Fitzek and M.Reisslein, "MPEG-4 and H.263 Video Traces for Network Performance Evaluation," *IEEE Network*, vol. 15, no. 6, pp. 40 – 54, November 2001.

[147] "Website Optimization." [Online]. Available: http://www.websiteoptimization.com/speed/tweak/

Acknowledgments

To Alessandra, Maurizio, and Stefano. My forever home.

To you, Giulia. My best buddy, my only love.

A te, Lidia, che sei presente in ogni mio successo.

To Isa, Silvano, and all my supporting relatives.

To Clara, Renato, Licia, Giordana, and Chiara, for you mean a lot to me.

To Alessandro, Riccardo, and all my crazy friends.

To Ema+Fran, for you are such an inspiration.

To Tim, Gedas, and all the amazing people I have met in Barcelona.

To Lorenzo and Davide, for we are such a damn special team.

To Marco, Alfred, and all the beautiful friends in SoCal.

To Xavier, Kiran, Aleksandar, and all the great researchers who guided me in the industry.

To Davide, Matteo, Daniele, Marco, Francesco, Irene, and all. We made a fantastic team.

To Andrea, Leonardo, Michele, Paolo, and the SIGNET staff. It was great working with you.

To Prof. Michele Zorzi, for you have wisely inspired both my professional and human growth.