Quality Of Service and MObility driven cognitive radio Systems

FP7-ICT-2009-4/248454

QoSMOS

D5.3

Final description and specification of cognitive manager and corresponding QoS support mechanisms with performance evaluation

Contractual Date of Delivery to the CEC:	31-Jul-2012
Actual Date of Delivery to the CEC:	31-Jul-2012
Editor(s):	Ulrico Celentano (UOULU)
Author(s):	Hicham Anouar (TCF), Péter Bakki (BME), János Bitó (BME), Keith Briggs (BT), Ulrico Celentano (UOULU), Olasunkanmi Durowoju (UNIS), Martina Fuentevilla (TST), Tao Guo (UNIS), Péter Horváth (BME), Ingo Karla (ALD), Zsolt Kollár (BME), Stéphanie Leveil (TCF), Miguel López-Benítez (UNIS), Richard MacKenzie (BT), Geneviève Mange (ALD), Arturo Medela (TST), Juan Rico (TST), Christophe Rosik (NTUK), Isameldin Suliman (UOULU), Johanna Vartiainen (UOULU)
Internal reviewer(s):	Vincent Berg (CEA), Arturo Medela (TST), Rainer Wansch (Fraunhofer IIS)
External reviewer(s):	Shyamalie Thilakawardana (BBC)
Workpackage:	WP5
Est. person months:	100
Security:	PU
Nature:	R
Version:	1.1
Total number of pages:	**118**

Abstract:

The present deliverable D5.3 describes the final structure of the cognitive manager for resource management, the mapping of its internal blocks according to topologies, and its interworking with the cognitive manager for spectrum management. It also presents tools for design and performance evaluation for cognitive radio networks. Solutions for resource management in all the three main target scenarios of QoSMOS (cellular extension, femtocell, ad hoc network) are also included.

Keyword list: Access control, ad hoc network, coexistence, cognitive manager, cognitive radio network (CR, CRN, CRS), femtocell, functional architecture, incumbent protection, interference management, LTE, MAC, malicious users (DoS), mobility, modelling, power control, quality of service (QoS), reference model, resource allocation, resource management, whitespace (TVWS).

Executive Summary

This deliverable reports on the cognitive manager for resource management, on tools for design and performance evaluation for cognitive radio networks (CRNs), and on resource management solutions for CRNs in all the three main target scenarios of QoSMOS.

The development of the cognitive manager for resource management (CM-RM) was tracked with D5.1, including its first and basic concepts, and with D5.2, where it was further detailed. In the present deliverable D5.3, the final state of development is presented. The mapping of CM-RM internal block according to the different topologies, thus interacting with the work done in WP2, is also covered. The adaptation layer (AL) is another key functional block in the QoSMOS system allowing managed communication of remote entities and its validation with performance and complexity analysis is included in this report.

Key aspects affecting protocol design and assessment are also presented. The details of the physical layer developed under WP4 are here abstracted and modelled with low-complexity. The applicability of two existing relevant MAC protocols to peculiarities of CRN is analysed here. The effects of various malicious users on resource allocation process, the effects of mobility on spectrum sensing, and of imperfect sensing on the quality of service (QoS) of both incumbent and opportunistic users are also discussed. The concept of mobility in CRNs includes spectrum mobility and optimal strategies for QoS provision under spectrum mobility are also presented in this report.

Solutions for resource management in the three main target scenarios of QoSMOS include a cognitive access control applicable to the cellular extension in TV whitespace, described and analysed discussing simulation results for various QoS metrics under different system configurations. For the femtocell scenario a downlink power control algorithm is presented and analysed. For the third scenario regarding a mobile cognitive ad hoc scenario, algorithms for operating channel cognitive selection and acquisition are presented and the opportunistic user performance as well as the incumbent protection are analysed by simulations under different system configurations.

Finally, this report provides an overview of all the activities relevant to QoS and mobility support produced in the project, including the results presented here as well as those previously reported.

Abbreviations

2-D	two-dimensional
2G	second generation
3G	third generation
3GPP	3rd Generation Partnership Project
4G	fourth generation
AC	access control, access categories
ACC	adjacent cluster combining
ACK	acknowledgement
AL	adaptation layer
ARP	allocation and retention priority
AWGN	additive white Gaussian noise
BE	best effort
BLER	block error rate
BPSK	binary phase-shift keying
BRAN	broadband radio access network
BS	base station, beacon slot
CAC	cognitive access control
CAF	channel access function
CAN	cognitive ad hoc network
CBR	connection blocking rate
CDR	connection dropping rate
CH	cluster head
CM	cognitive manager
COORD	coordination
CORBA	common object request broker architecture
CR	cognitive radio
CRN	cognitive radio network

CSA	channel selection algorithm
CSAP	channel selection and acquisition protocol
CTMC	continuous-time Markov chain
CTS	clear to send
CW	contention window
DA	demand adaptation
DAS	distributed antenna system
DB	database
DL	downlink
DoS	denial of service
DPC	distributed power control
DS	data slot
DSL	digital subscriber line
E-UTRAN	evolved universal terrestrial radio access network
ECMA	European association for standardizing information and communication systems (originally, European Computer Manufacturers Association)
EESM	exponential effective SINR mapping
eNB	e node B
EPC	evolved packet core
EPS	evolved packet system
ErtErt	extended real-time
ETSI	European Telecommunications Standards Institute
FAP	femtocell access point
FBMC	filter bank multicarrier modulation
FCC	Federal Communications Commission
FGW	femtocell gateway
FIFO	first in, first out
FTP	file transfer protocol
FUE	femtocell user equipment

GIOP	general IOP
GLPK	GNU linear programming kit
GNU	Gnu's not UNIX
GPRS	general packet radio service
GPS	global positioning system
HO	handover
HTTP	hypertext transfer protocol
HTTPS	HTTP secure
ICI	inter-carrier interference
ID	identification
IDL	interface definition language
IEEE	Institute of Electrical and Electronics Engineers
IIOP	Internet IOP
IOP	inter-ORB protocol
IP	Internet protocol
ISI	inter-symbol interference
ISO	International Organization for Standardization
IU	incumbent user
LAD	localisation algorithm based on double-thresholding
LAN	local area network
LTE	long term evolution
Lx	x-th protocol layer of the ISO-OSI reference model
M2M	machine-to-machine
MAC	medium access control
MBR	maximum bit rate
MBS	macrocell base station
Mcast	multicast
MCS	modulation and coding scheme
MMIB	mean mutual information per bit

mn	minutes
MSC	message sequence chart
MSDU	MAC service data unit
MSG	message
MU	malicious user
MUE	macro user equipment
NACK	negative ACK
NBS	Nash bargaining solution
NC	networking domain cognition
NET	network
OC	operating channel
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
OMG	object management group
ORB	object request broker
OSI	open systems interconnection
OU	opportunistic user
P2P	point-to-point
PC	power control, personal computer
PCA	power control algorithm
PCRF	policy and charging rules function
PDN	packet data network
PDU	protocol data unit
P-GW	PDN gateway
PHY	physical layer
PRB	physical resource block
PU	public
QAM	quadrature amplitude modulation
QCI	QoS class identifier

QoE	quality of experience
QoS	quality of service
QPSK	quadrature phase-shift keying
RA	resource allocation
RAA	resource allocation algorithm
RAS	random-access slot
RB	resource block
RBIR	received bit mutual information
RC	resource control
RE	restricted
RECM>2mn	rate of successfully established communication and maintained during at least 2 minutes
REQ	request
RLC	radio link control
RM	resource management
RNC	radio network controller
RPC	remote procedure call
RS	resource control support
RSS	received signal strength
RT	real-time
RU	resource use
RTS	request to send
Rx, RX	receiver
SD	spectrum databases
SGSN	serving GPRS support node
SINR	signal to interference-plus-noise ratio
SLA	service level agreement
SM	spectrum management
SMDP	semi Markov decision process

SNR	signal to noise ratio
SS	spectrum sensing
SSDF	sensing data falsification attack
SSR	systematic sum rate
TCP	transmission control protocol
TDD	time-division duplexing
TDMA	time-division multiple access
TV	television
TVWS	TV whitespace
Tx, TX	transmitter
TXOP	transmit opportunity
UDP	user datagram protocol
UE	user equipment
WiFi	wireless fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	wireless LAN
WP	work-package
WSLA	weighted sum link approach
WSPRT	weighted sequential probability ratio test
www	worldwide web
XML	extensive markup language

Table of contents

List of figures

List of tables

1 Introduction

The present deliverable D5.3 is the final report about the activities carried out in WP5 and is composed of three parts.

The first part, chapter 2, gives an overview of the relevant QoSMOS goals (section 2.1). Together with the outline of the QoSMOS functional architecture reported in section 2.2, this chapter provides the requirements and constraints put for the design of the cognitive manager relevant in this work-package and of the solutions for resource management. These requirements and constraints determine the functional architecture of the cognitive manager for resource management (CM-RM). Its development was already tracked by [D5.1], which included its first and basic concepts, whereas [D5.2] provided more details on its functionalities. The final state of the development of the CM-RM is presented in section 2.3 together with the mapping of its internal functional blocks according to the various supported topologies.

The second part of the deliverable, chapter3, discusses key aspects affecting protocol design and tools for performance assessment of cognitive radio systems. Apart from the interface between the CM-RM and the CM-SM, all QoSMOS interfaces exploit the services of its adaptation layer. Such a crucial functional block is analysed in section 3.1 about its performance and complexity. The peculiarities of the physical layer developed under WP4 are abstracted in section 3.2 making possible the design of its upper layers. Section 3.3 presents the analysis of selected existing medium access control protocols discussing their applicability to a cognitive radio network. As seen in section 2.1, robustness to the attack of malicious users is a requirement and the effects of them on resource allocation are discussed in section 3.5. The effects of mobility on spectrum sensing, whereas the impact of imperfect spectrum sensing on the quality of service (QoS) of both incumbent and opportunistic users is presented in section 3.4. The concept of mobility in QoSMOS includes spectrum mobility and optimal strategies for QoS provision under spectrum mobility are presented in section 3.6.

The third part of this report, chapter 4, presents resource management solutions enabling QoS support for all the three main target scenarios of QoSMOS reminded in section 2.1. In particular, a cognitive access control applicable to the cellular extension in TV whitespace is described and analysed discussing simulation results for various QoS metrics under different system configurations in section 4.1. For the femtocell scenario, a downlink power control algorithm is presented and analysed in section 4.2. Finally, section 4.3 presents algorithms for operating channel selection and acquisition in a mobile cognitive ad hoc scenario and analyses them by simulations under different system configurations for both the opportunistic user performance and the incumbent protection.

In the concluding chapter 5 an overview of the results presented in this report is given together with a summary of all the activities relevant to QoS and mobility support produced in the project.

2 Cognitive manager for quality of service and mobility support

This chapter describes the final functional architecture of the cognitive manager for resource management. This fundamental block of the QoSMOS system interworks with another important block, the cognitive manager for spectrum management..Also, the interaction between the two is discussed here. The role of those blocks within the overall QoSMOS system architecture, outlined in the following, is also explained. At the roots of the design of the QoSMOS architecture and its functional blocks are the requirements and constraints imposed by the nature of the system and the relevant regulations, as well as by those set as targets by the project itself. The QoS classes applicable in this context derived from the analysis of these preconditions are also discussed here. The aforementioned requirements and constraints are presented first, and then the functional architecture is described.

2.1 Constraints for cognitive manager and resource management design

The design of the QoSMOS system and of its functional blocks depends on the requirements and constraints set by both the applicable regulations and the exploitability conditions. One side of those requirements comes from the flexibility imposed by the project on its system architecture, which must be able to cope with the target scenarios identified within the project and should also be able to cover the possible needs. Another set of constraints comes from the system requirements that have been identified. Those two aspects are discussed in the following sections 2.1.1 and 2.1.2.

2.1.1 The QoSMOS target scenarios

The QoSMOS project identified, after a rationalisation process, six scenarios [D1.2] [MacEtal2011]. These scenarios have undergone an additional analysis as potential business cases [LehEtal2012] and resulted in the three key target scenarios summarised in Figure 2-1.

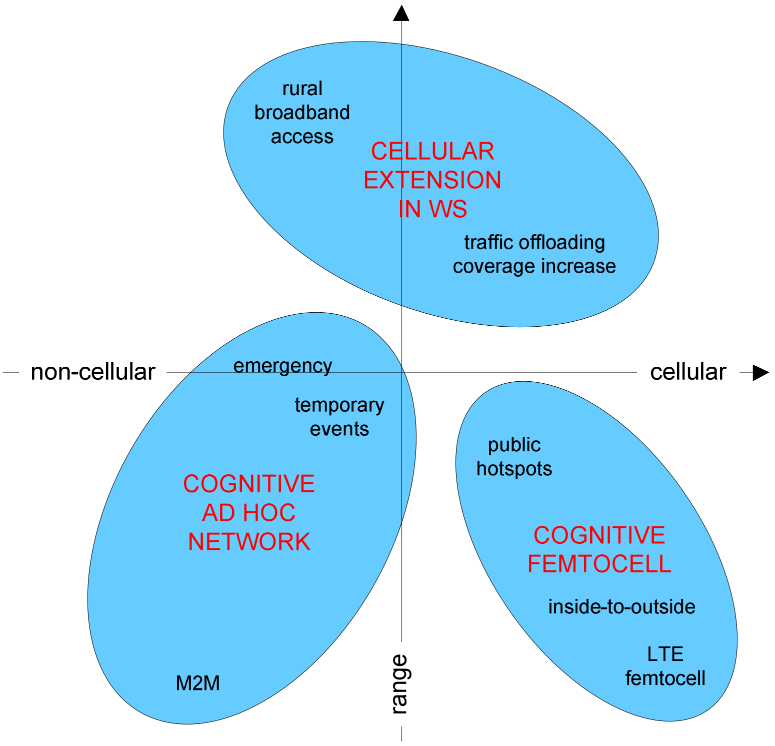


Figure 2-1: The final QoSMOS scenarios.

While fuller details are available from the above references, the scenarios are briefly outlined in the following.

2.1.1.1 Cellular extension in the whitespace

The cellular extension in whitespace scenario allows mobile telecommunication operators to improve coverage and/or capacity of their networks by using whitespace spectrum in addition to their own licensed spectrum. This scenario allows improving link quality and offering more flexible services, and therefore not only covers LTE-like business cases, but also rural broadband access.

2.1.1.2 Cognitive femtocells

The cognitive femtocell scenario makes use of small access points, for example to distribute broadband access in the home or its neighbourhoods, or to provide internet access in public areas, e.g., by so-called hot-spots. This scenario covers extensions of both concepts: femtocells in cellular networks and WLAN-type deployments. Example deployments include WLAN extension, indoor-to-outdoor coverage extension and public hot-spots.

2.1.1.3 Cognitive ad hoc networks

In the cognitive ad hoc networks scenario, whitespace is used to connect terminals as the operation of these networks is limited in both time and space. In order to comply with regulatory requirements, cognitive ad hoc networks are likely to require access to relevant repositories, whether direct or indirect, static or temporary... Without such a connection the regulatory requirements may have to rely on much stricter spectrum sensing capabilities. The cognitive ad hoc network scenario covers emergency operations, service during temporary events and machine-to-machine communication (M2M).

2.1.2 The QoSMOS system requirements

Requirements on a system level are detailed in [D1.4] whereas those specific to WP5 are indicated in [D5.2] [ManEtal2011]. Those relevant to the content of this deliverable are reminded in the following part of this section (and are highlighted with italic font).section

First, the QoSMOS system shall be capable of adaptation in order to comply with regulations. To be able to do that, the QoSMOS system shall be able to collect *environmental information* depending on the QoS requirements and shall enable access to regulation information and policies. In particular, it shall allow collecting information from a *geolocation database* according to the requirements and shall allow updating it as required. In addition, it should support *spectrum sensing* for incumbent detection. Anyway, when (spectrum sensing is needed), the system shall be capable of scheduling quiet periods for spectrum sensing purpose without QoS degradations for the opportunistic system(incumbent user?) below acceptable levels.

Then, the QoSMOS system shall be able to *react to the changes* in the environment. In particular, it shall avoid interference to incumbent communications. This may imply *vacating* the operating channel upon appearance of an incumbent user; alternatively, when this is authorised by regulations and policies, the system shall define a limitation on the *transmit power* in order to avoid interfering with simultaneous incumbent transmissions. In addition, it also shall coexist with other opportunistic systems. Moreover, it shall be able to detect *attacks of malicious users* and be robust to those misbehaviours.

As an implication to its design, the QoSMOS system shall be *flexible* enough to comprise different architectures which can support a variety of diverse use cases. In particular, it shall support multiple radio access technologies and be able to select the most appropriate. Moreover, it shall support different frequency bands for opportunistic use.

The performance of the QoSMOS system should be good enough to meet *expectations of the users* about the delivered service, even in presence of variations in the available spectrum resources. As a consequence, the system should maintain the agreed level of QoS and should be able to re-establish a disrupted service within an agreed time. This implies that an appropriate number of *reserve channels* shall be maintained based on the QoS needs. In any case, it shall provide data rates and latencies needed to satisfy QoS needs of the supported services in order to provide quality of experience (QoE) comparable to the one offered by other access technologies.

The QoSMOS system also shall support mobile users. However, the level of *mobility* varies among the target scenarios. The coverage increase and traffic offloading cases for the cellular extension possess high levels of mobility, together with the cognitive ad hoc network, whereas the rural broadband access case has virtually no mobility at all. The mobility for the femtocell cases somehow fall in the middle.

2.2 Overall functional architecture

In order to respond to the requirements and constraints outlined in sections 2.1.2 and 2.1.1 and references therein, the QoSMOS system architecture (Figure 2-2), as documented and specified in [D1.2] [D2.1] [D2.2] [D2.3] and also presented in [CelEtal2011] [MacEtal2011] [LevEtal2012], defines a two-fold cognitive manager at its core: the cognitive manager for resource management (CM-RM), developed in WP5 and presented in section 2.3, and the cognitive manager for spectrum management (CM-SM), developed in WP6. In addition to those, a spectrum sensing (SS) block developed in WP3, a flexible transceiver (TRX) developed in WP4, and an adaptation layer (AL), developed as a cross-WP activity.

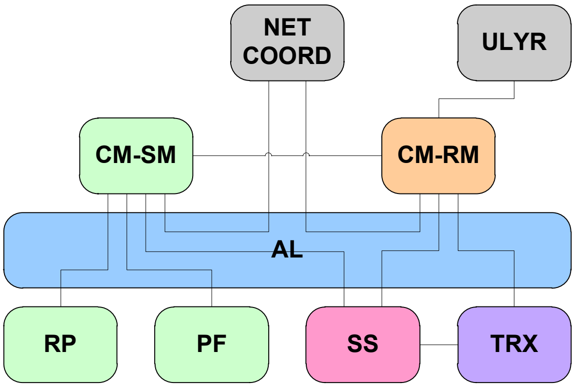


Figure 2-2: The QoSMOS reference model.

The spectrum resources are opportunistically used by the TRX and exploited for serving the needs of the upper layers (ULYR), with the SS providing radio context information. The AL facilitates communication among all the remote entities, but it is not involved in the interaction between the CM-RM and the CM-SM. Before that, these two blocks are briefly presented. More details are available from the aforementioned references, but section 2.3 is dedicated to the CM-RM.

Related to the activities of CM-RM and CM-SM is the external block providing network coordination (NET COORD). This block is optional and for the example case of the cellular scenario, it is part of the core network (CN).

At the user equipment, or in case of missing core network at any network device, the ULYR is for example the application, whereas in case of presence of a core network, the ULYR is at the transport layer. The split of SS functionalities depends on the spectrum sensing topologies [D3.4] and is out of scope here. The split of CM-RM functionalities is addressed later in section 2.3.2.

The CM-SM is responsible for the management of the frequency spectrum allocated to the QoSMOS entities for dedicated use. To this end, the CM-SM first acquires the relevant context information, including environment information obtained from spectrum sensing results provided by the spectrum sensing block and performance reports of current spectrum usage provided by the CM-RM, which may include filtered status information such as average supported load levels or average interference levels experienced in the system. Based on this information and on external constraints such as regulatory policies, operator policies and operator frequency planning, the CM-SM then builds up the spectrum portfolio containing spectrum usage information and spectrum usage policies and putting constraints on the decisions that can be taken by other entities of the QoSMOS system. To this end, the CM-SM accesses the regulatory repository (RP), which includes constraints and requirements about the spectrum use, and the common portfolio repository (PF)used to store and exchange context information.

The CM-RM is the main user of the spectrum portfolios generated by the CM-SM and allocates radio resources from the assigned spectrum portfolio to the end-users in order to provide service to the application layer according to an agreed level of QoS. The CM-RM is also responsible for the management of the user mobility as well as the implementation of functionalities needed to protect the incumbent users, relying in a close cooperation with the CM-SM, which in turn implements incumbent protection on a spectrum management level.

In brief, the CM-RM provides the CM-SM with updated network status information and spectrum usage reports, and this information is exploited by the CM-SM in order to decide which parts of the spectrum can be used by the QoSMOS entities and under which conditions/constraints, based on the information collected from the CM-RM and other QoSMOS entities. The CM-SM thus relies upon context information provided by the CM-RM and responds to requests of the CM-RM to adjust portfolio allocations in accordance with the current radio resource management needs.

Section 2.3 presents the functional architecture of the CM-RM. While the interface involved in the tight interworking between the CM-RM and the CM-SM is direct, as it is clear from Figure 2-2, almost all the other interactions happen through the adaptation layer. The validation of the AL is presented in section 3.1.

Both cognitive managers are decision-making entities in charge of handling spectral and radio resources in an efficient manner in order to meet the users' QoS needs and mobility profiles and, as such, there is a tight and close interworking between them. There are however relevant differences between the CM-RM and the CM-SM in terms of functions and responsibilities as well as time-scales and frequency granularities, which determine the way in which both elements interact with each other.

2.3 CM-RM functional architecture

The functionalities required by the CM-RM [LevEtal2012] are divided among the CM-RM-internal functional blocks, see Figure 2-3, described in the section 2.3.1 below and grouped as explained in the following section 2.3.2.

2.3.1 Functional blocks

The QoS maintenance and mobility management functions of the CM-RM are assigned to the admission control (AC) and mobility control (MC) blocks. The resource allocation (RA) block is in charge of the allocation of the resources, which are then made available to the resource exploitation (RE) block that offers the communication services to the upper layers, optimising the use of the lower layers. In this, for topological reasons explained shortly below, the resource control support (RS) block acts as an intermediate block[1]. Similarly, gathering of cognitive information to be used for system reconfiguration is split between two blocks: the networking domain cognition (NC) and the terminating domain cognition (TC) blocks, which manage performance measurements, sensing results and inputs from repositories.

[1] The RS controls the transceiver configuration locally to the terminating domain, implementing resource management actions performed at networking domain. The topological domains [CelEtal2011] are explained in section 2.3.2.

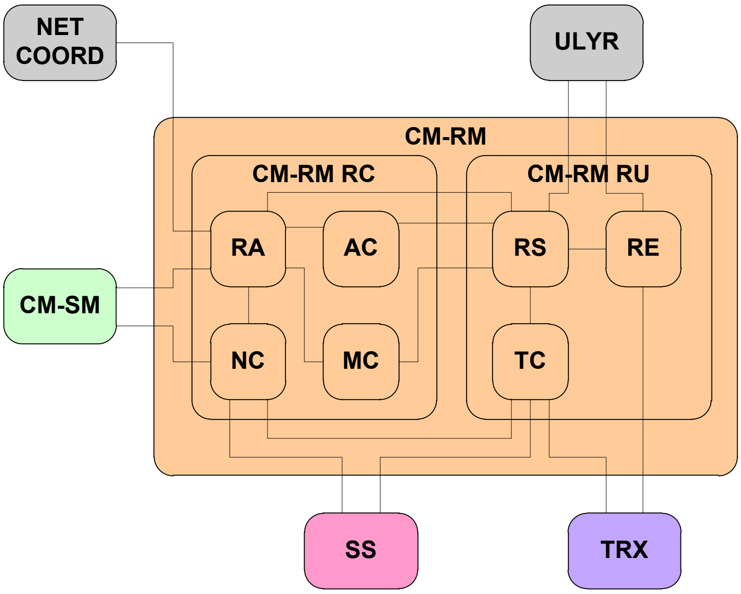


Figure 2-3: The functional architecture of the CM-RM.

2.3.2 Topology mapping

The CM-RM is designed to be applicable to a range of diverse scenarios, such as those outlined in section 2.1 and potentially even beyond those. The various architectural options corresponding to spectrum sensing and resource control topologies are discussed in [D2.2] [D2.3]. In particular, from the resource control perspective, a QoSMOS system, and therefore the CM-RM, supports a centralised or a distributed resource control and may exploit the aid of a core network or not. To this end, the functional blocks of the CM-RM are further grouped. A first grouping of the functions into a resource control[2] (RC) and a resource use (RU) group was introduced by [D5.1] making use of the four topological domains introduced by [CelEtal2011] as explained below.

The *terminating domain* pertains to the wireless border of the system, thus including those parts of the UEs as well as of the network nodes they are attached to, for example a base station, an access point or any other central controller. The *networking domain* covers the functions related to the control of interconnected network devices, and hence included in a central controller as the above or alternatively in all nodes of a network with distributed control. The *coordination domain* is dedicated to the coordination with neighbouring and related networks, while the *coexistence domain* concerns larger-scale coexistence, therefore including the repositories mentioned in section 2.3.1 above.

The resource control group includes the functions belonging to the networking domain. The RC is therefore present in a base station, an access point, or a cluster head in the case of centralised resource control, while it is found in any peer node of a network with distributed resource control. The resource use is by its nature always found in any QoSMOS wireless node. Both RC and RU groups are shown in Figure 2-3.

[2] In [D5.1] it was called resource allocation, but the naming was adjusted when the internals of the CM-RM were further developed.

3 Tools for design and performance evaluation

System design and its performance assessment rely on suitable tools. A set of such tools covering the protocol modelling, design, analysis and evaluation is presented in this chapter.

3.1 Adaptation layer validation

The adaptation layer (AL) has a remarkable role in the QoSMOS architecture; its functionalities allow enhancing overall system performance as it is monitoring the status of each entity connected to it, thus triggering the needed alarms in case of misbehaviour of any piece of the system. The analysis of the performance of the AL and of its complexity is presented in this section.

3.1.1 Introduction

The AL comprises a set of components responsible for carrying out the different functions of the entity. [D2.3] has already presented the different activities carried out by the AL and also a detailed description of all its components. Figure 3-1 depicts the internal AL architecture. Next, a brief description of them is presented (see also section 3.1.3 and figures therein):

- AL_CORE: it is the module in charge of performing the most critical activities in the AL. It also has a database, called contactable entities DB, where all the information about registered QoSMOS entities is stored.
- AL_END: the interfaces between the AL_CORE and the QoSMOS entity they are connected to.

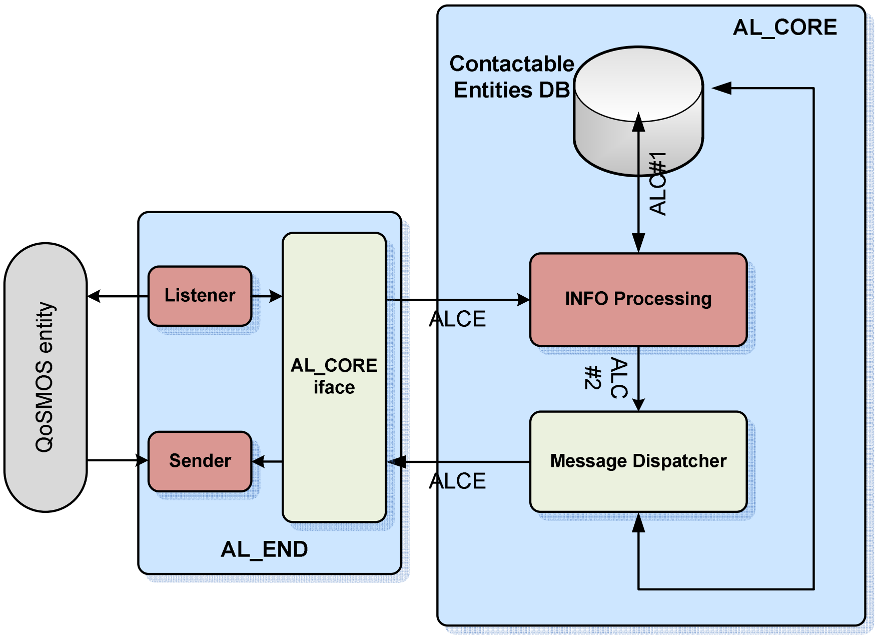


Figure 3-1: Internal AL architecture

The present deliverable focuses on the impact of including the AL in the architecture. The AL not only brings improvements to the system but potential delays are also introduced and should be accurately defined. The following subsections present the results obtained in different tests for validating functionalities and characterise the AL.

3.1.2 Considered technologies

The AL presents an architecture which allows many different communication paradigms and technologies. Considering the interactions with other blocks in the QoSMOS system, two of these technologies have been chosen due to the fact that they are the most widely used. The following subsections present the key aspects of them both.

3.1.2.1 XML-RPC

XML-RPC is a remote procedure call that uses HTTP as the transport protocol and XML to encode its calls. It allows transmitting, processing and returning complex data structures, but it has been designed to be as simple as possible. XML-RPC owns a specification [XML99] and a set of implementations that enable software to run in different operating systems and environments to make procedure calls over the Internet (Figure 3-2).

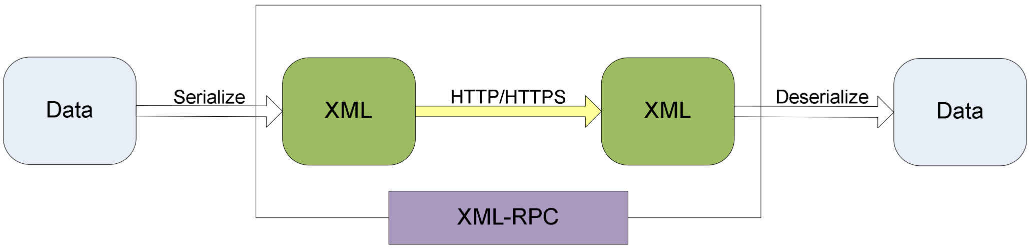


Figure 3-2: Client's request example.

To establish a communication through XML-RPC protocol the client, which uses a software wanting to call a method of a remote system, sends a XML-RPC call. Then, XML-RPC sends an HTTP request to a server implementing this protocol, too. The remote method can accept multiple input parameters and return only one value. However, it can transport larger structures, since the parameter types allow nesting parameters into maps and lists. Therefore, it can transport structures or objects both as input and as output parameters.

XML-RPC also implements client's identification for authorisation purposes. It can be realised using HTTP/HTTPS security methods.

3.1.2.2 CORBA

The Object Management Group (OMG) has developed a standard called common object request broker architecture (CORBA) [COR11] in order to provide interoperability among distributed objects. CORBA is a middleware solution that enables the exchange of information, independent of hardware platforms, programming languages and operating systems.

The object request broker ORB is the essential concept in CORBA. It provides the mechanism required to achieve that a client programme can request services from a server programme without having to understand where the server is in a distributed network or what the interface to the server programme looks like. So as to make requests or return replies between the ORBs, programmes use the general inter-ORB protocol (GIOP) and, in the case of the Internet, the Internet inter-ORB protocol (IIOP), whose work is mapping GIOP requests and replies to the Internet's transmission control protocol layer in each computer.

CORBA applications are composed of objects, individual units of running software which combine functionality and data. For each object type, an interface in OMG interface definition language (IDL) is defined. This IDL allows the development of language and location-independent interfaces to distributed objects. The interface is the syntax part of the contract that the server object offers to the clients that invoke it. That is, if a client wants to invoke an operation on the object, it must use this IDL interface to specify what operation it wants to perform and to marshal the arguments that it sends. Then, when the invocation reaches the target object, the aforementioned interface definition is used to unmarshal the arguments and the object can perform the requested operation with them. After this, the results are marshalled for their way back and they are unmarshalling again when they get at their destination.

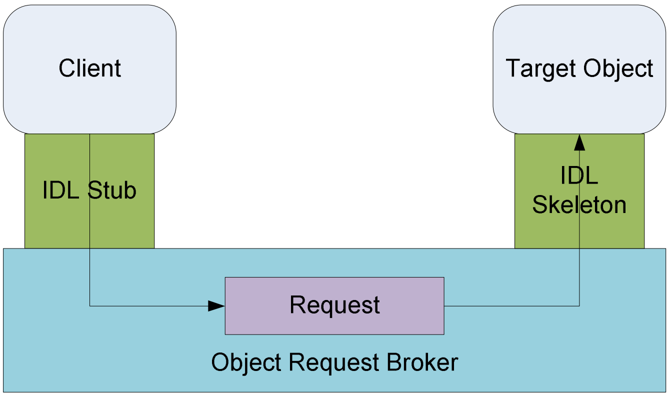


Figure 3-3: Client's request example.

Figure 3-3 shows an example of a client request. First, it has to compile the IDL into client stubs and object skeletons and write an object and a client for it. Stubs and skeletons serve as proxies for clients and servers, respectively. Then, because the IDL defines the interfaces so strictly, the stub on the client side has no problem meshing perfectly with the skeleton on the server side, although the two are compiled into different programming languages or running on different ORBs from different vendors.

3.1.3 AL functionalities validation and characterisation

The validation of the different functionalities included in the AL is done following two approaches. On the one hand, there are some initial procedures that must be validated in order to assess AL capabilities as presented in previous documents, such as [D2.3] [D5.1] [D5.2]. This has been done based on basic functional tests where the aim was just checking that the information flow is exchanged as it is supposed to be, and the entities are able to exploit that information in a short period of time. The second part focuses on providing specific figures for the impact of the AL on the system. Several test scenarios have been deployed stressing the AL and measuring the performance of some of the AL key functionalities. Furthermore, in some cases it is difficult to evaluate some features by stressing the system, so a theoretical approach has been performed to provide figures for the AL performance.

3.1.3.1 Entity registration/deregistration

Two of the basic functionalities of the AL are the registration and deregistration of the QoSMOS blocks into the adaptation layer.

Before QoSMOS entities are able to take advantage of the AL capabilities it is necessary to register them to the AL. This action should be taken in order to request/provide information, not only data, but also context, through the AL.

The process starts when a QoSMOS entity triggers a registration request. This request is formulated as a specific packet sent to the AL_END. There the packet is detected and forwarded to the AL_CORE, and once there, the information contained in the request packet is extracted and the contactable entities database is updated.

The validation of the functionality is done by assuring the insertion of the data associated to the entity in the database and the correct reception of the acknowledgement sent by the AL to the source of the request.

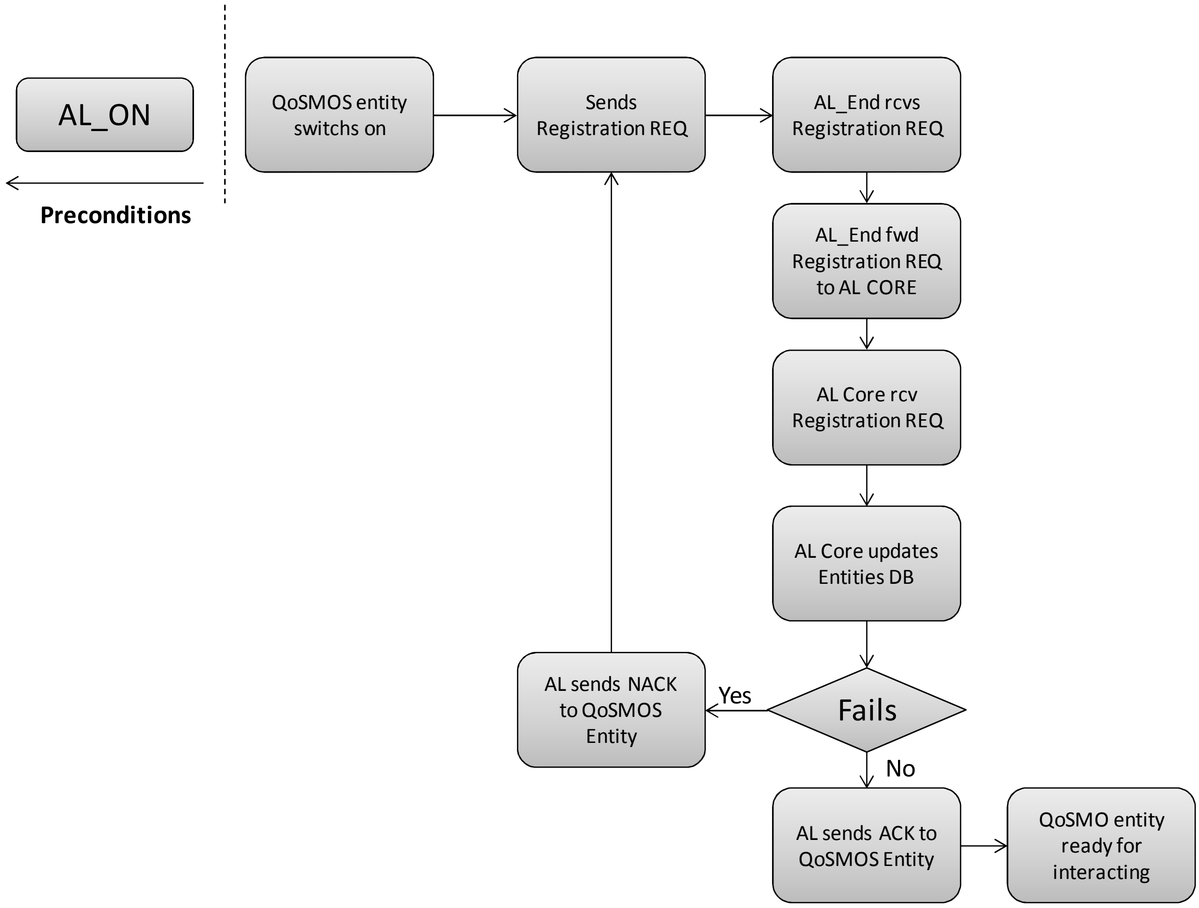

AL_ON
Preconditions
QoSMOS entity switchs on
Sends Registration REQ
AL_End rcvs Registration REQ
AL_End fwd Registration REQ to AL CORE
AL Core rcv Registration REQ
AL Core updates Entities DB
Fails
Yes
No
AL sends NACK to QoSMOS Entity
AL sends ACK to QoSMOS Entity
QoSMO entity ready for interacting

Figure 3-4: Validation of registration functionality flow.

The procedure used to validate this functionality is detailed in Figure 3-4. All intermediate steps have been confirmed before assuring the whole functionality. The validation process comprises not only the reception of data packets but also confirming that the content is properly processed and inserted in the contactable entities database. Finally, the creation of the response and the reception by the source QoSMOS entity is also checked in order to validate the whole information flow of the functionality. This procedure also includes the insertion of errors that are translated into wrong updates and no confirmation of registration in the AL.

In order to carry out the characterisation of this AL operation, some key points have been selected for evaluating the impact of the different actions done through the whole execution of the functionality, the validation flow helps identifying the most critical points in the system.

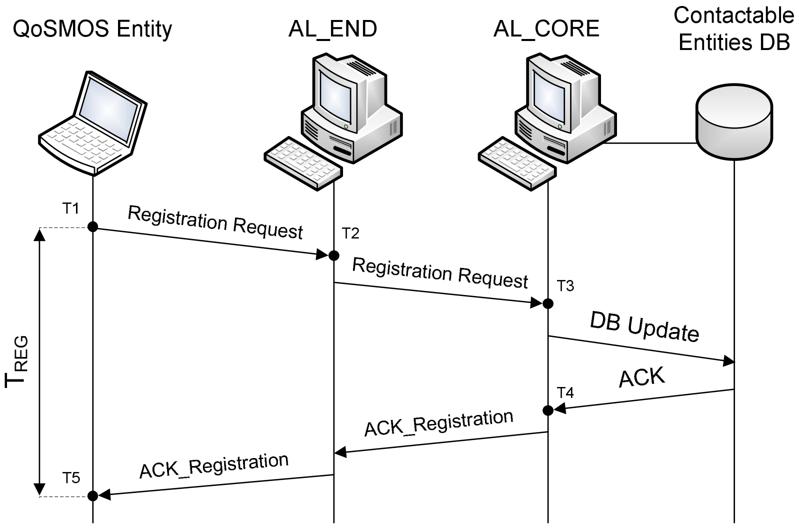


Figure 3-5: Entity registration.

As presented in Figure 3-5, the first test point is set at the time the QoSMOS entity starts the registration. Then, when the request arrives at the AL, another mark is established thus identifying the amount of time needed to reach the AL_END, that time is considered as transmission time. The next is situated when the packet arrives at AL_CORE. The following one is located after the AL_CORE database is properly updated. Finally, it is mandatory to monitor the time when the confirmation from the AL is received by the source QoSMOS entity.

Complementary to the registration procedure there is a deregistration one. This method involves the removal of the entity's data from the contactable entities DB in the AL_CORE in order to unsubscribe that specific entity from the AL services.

In this process, a QoSMOS entity asks for deregistration sending a deregistration packet to the appropriate AL_END, which dispatches it to the AL_CORE. At that time, the AL_CORE deletes the entity's data from the contactable entities DB and retransmits an acknowledgement to the entity through the aforementioned AL_END.

To validate this functionality, it is necessary to ensure the erasure of the corresponding data from the database and the right reception of the confirmation delivered by the AL to the entity which has sent the request. The flow followed in the process is similar to the registration procedure.

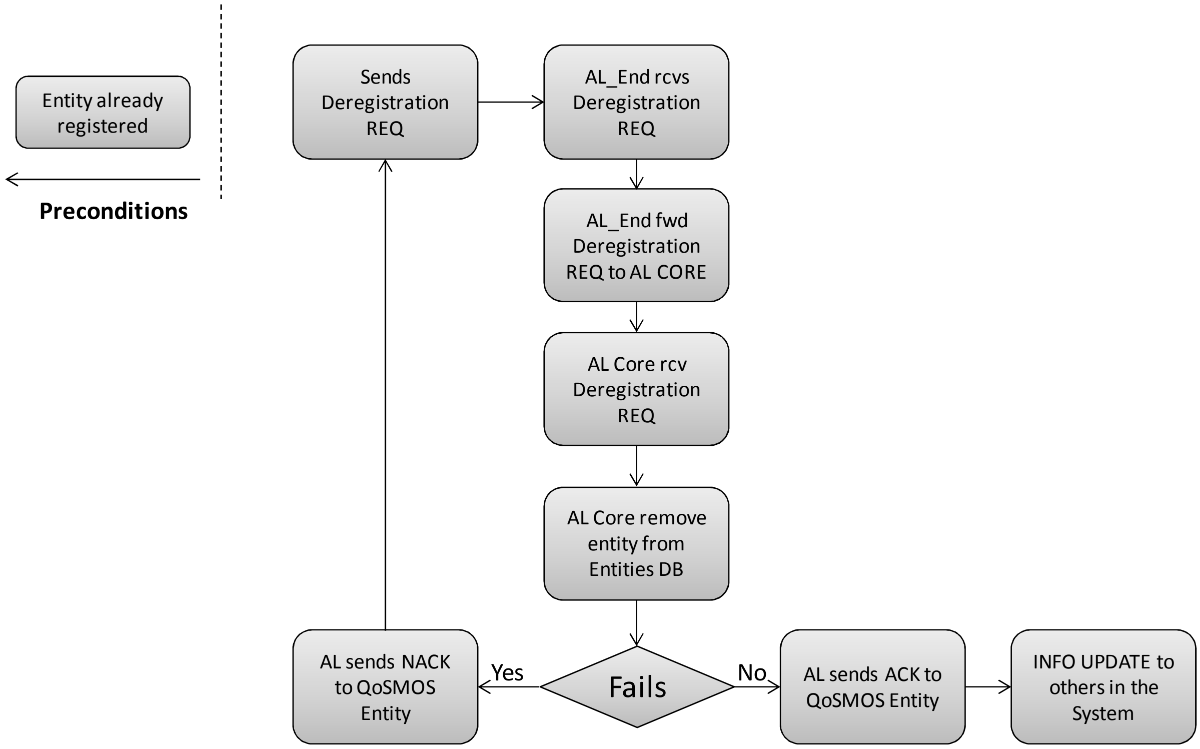

Entity already registered
Preconditions
Sends Deregistration REQ
AL_End rcvs Deregistration REQ
AL_End fwd Deregistration REQ to AL CORE
AL Core rcv Deregistration REQ
AL Core remove entity from Entities DB
Fails
Yes
No
AL sends NACK to QoSMOS Entity
AL sends ACK to QoSMOS Entity
INFO UPDATE to others in the System

Figure 3-6: Deregistration flow.

The flow of the deregistration process can be seen in Figure 3-6. The approach of validating functionality is the same as the one done with the registration feature to confirm the correct reception of the messages and their process, and also robustness against errors.

Considering that the action will be completed when the entity gets the AL_CORE acknowledgment, five key points, represented in Figure 3-7, were settled to characterise the AL's behaviour.

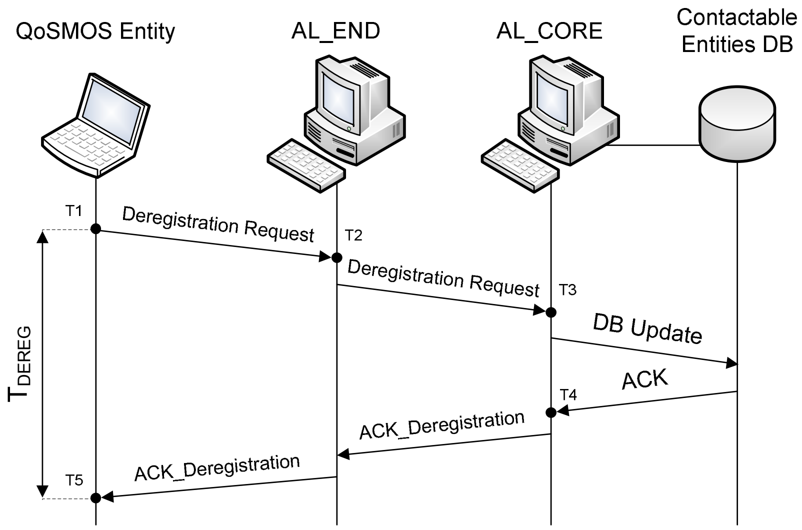


Figure 3-7: Entity deregistration.

The first of them consists in setting when the entity wants to deregister and deliver a deregistration packet to the AL. The moment the message reaches the AL_END is considered as the next test point and the following mark is located when the request arrives at the AL_CORE. After deleting data from the database another test point is set. Lastly, the instant the confirmation sent by the AL reaches the QoSMOS entity should be checked.

3.1.3.2 Information update

In order to keep entities updated about different events happening in the system, the AL will send messages to the different entities registered in its database informing about specific issues that they previously marked as interesting for them. Thus, every registered block can interact with updated information of the environment it is working in.

The validation of this functionality has been done following the flow depicted in Figure 3-8.

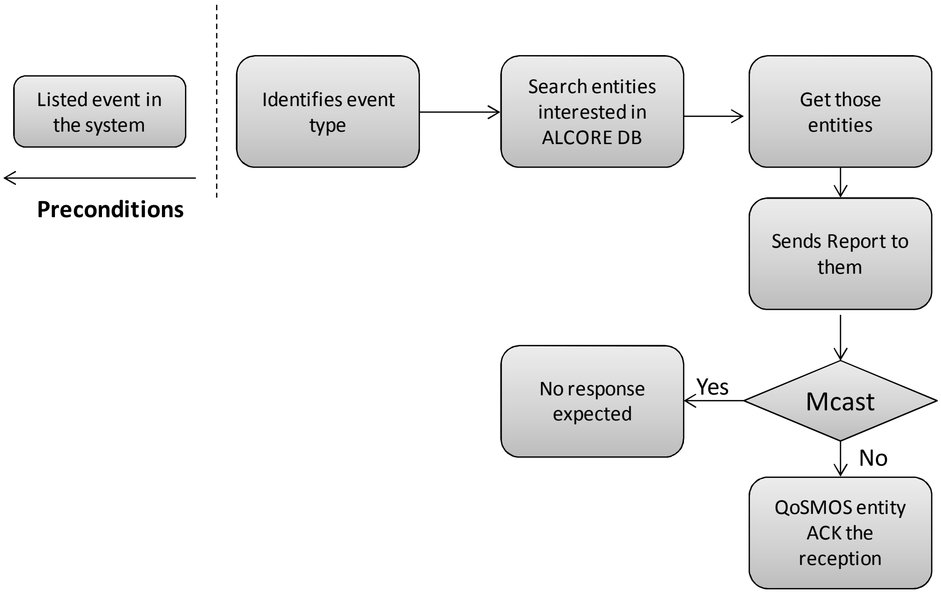


Figure 3-8: Information update flow.

Regardless if the final message is sent in a multicast packet or by the creation of multiple unicast, the procedure is the same. The first key part is not part of the process although it is the trigger, which is the detection of an event. Once the event is identified, the AL must search in the contactable entities DB which are the entities that are interested in this specific event and send all of them, either multicast or unicast a packet informing about it. In the case a unicast message is sent, the destination QoSMOS entity must send back a confirmation of the reception of the packet. This is not the case for multicast where the system trusts that all entities have received the message.

In order to characterise this procedure, the methodology followed is based on the entry of a new entity in the system as trigger. The AL_CORE will send information update message to the corresponding AL_ENDs when the contactable entities DB is updated. After that, AL_ENDs deliver this message to their corresponding block that should reply with an update ACK packet. Then, the confirmation message will arrive at the AL_CORE through said AL_END and the information update process can be considered finished.

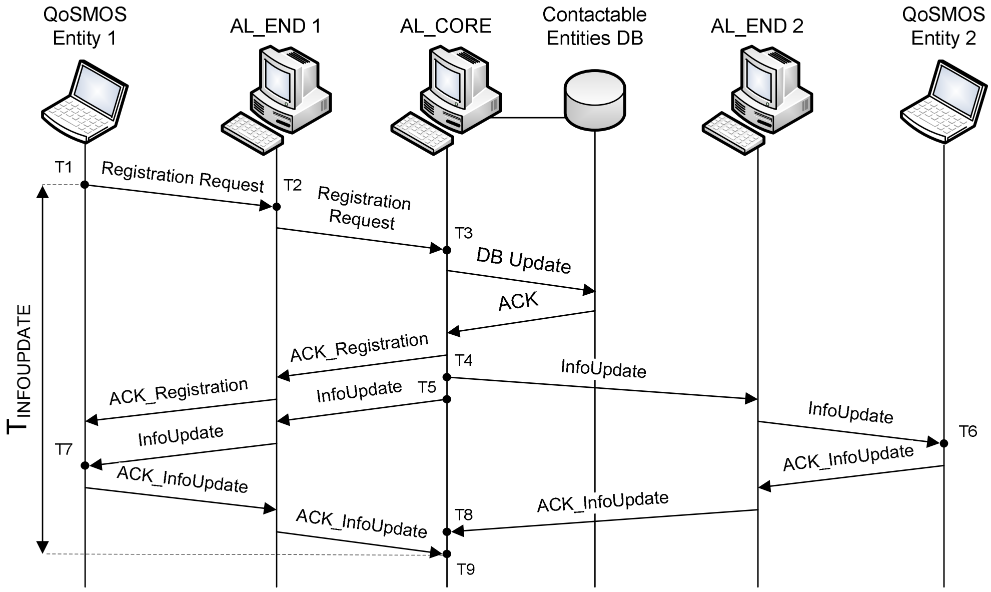


Figure 3-9: Information update.

To characterise this AL's functionality, Figure 3-9 shows the nine checkpoints that are proposed. Starting the registration is the initial point. The next one is the moment when the request reaches at the AL_END, followed by the time when the packet arrives at the AL_CORE. The instants when said AL_CORE sends information update messages are the next test points. After these ones, when the update come to the entities, one checkpoint is settled in each case and the latter two points are the times when acknowledgments of the different blocks arrive at the AL_CORE.

On the other hand, there is the possibility of sending the information update messages using IP multicast. In this process, the functionality will be considered valid when the multicast packet gets to the interested entities and the AL_CORE receives the acknowledgement from the block has just registered, as seen in Figure 3-10. Confirmation from the other entities is not expected due to multicast transmissions are based on UDP transport protocol. It was decided to use this protocol to improve the performance of the adaptation layer and to minimise the number of sent packets.

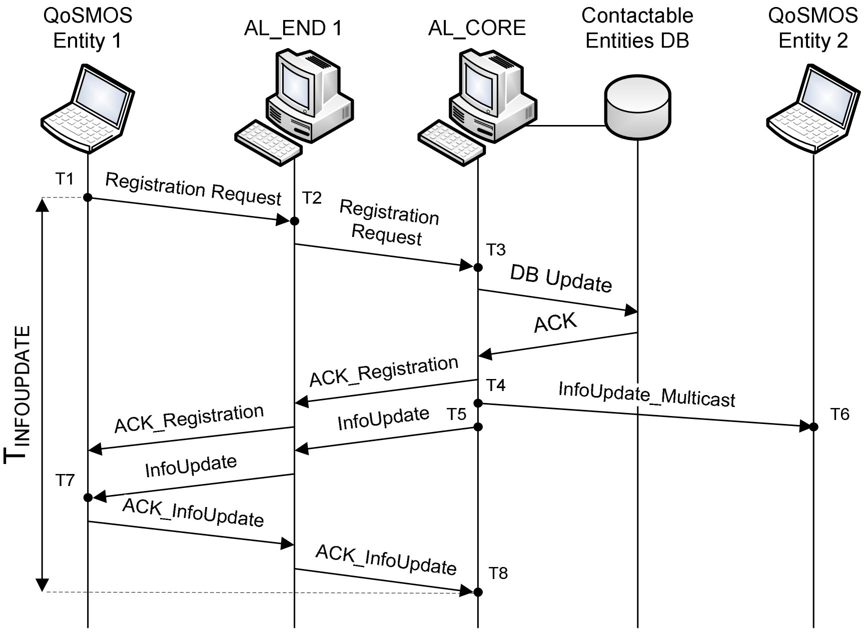


Figure 3-10: Information update using IP multicast.

For the multicast case it is not needed to include a timestamp for the ACK message of updated entity as multicast is based on UDP messages and no confirmation is required.

In the case of the XML-RPC protocol, the information update functionality will be considered valid when the AL_CORE receives the acknowledgement from all QoSMOS entities with common information registered in the database. When multicast messages are used, the AL_CORE does not expect the confirmation.

3.1.3.3 Information message

As mentioned in D2.3, the AL also provides the possibility of sharing information among QoSMOS blocks by using the capabilities it provides. In this case, the AL will assure all security aspects regarding communication, integrity and reliability.

In order to use the AL to send a message, a QoSMOS block will deliver a dispatch to its AL_END. This AL_END sends it to the AL_CORE, which remits the message to the corresponding destination AL_END. Finally, this last AL_END will forward the information to the QoSMOS destination entity, which replies with a confirmation through AL to QoSMOS source block.

This functionality is validated by ensuring the arrival of the information message at the destination entity and checking the right reception of the acknowledgement at the source entity following the flow presented in Figure 3-11.

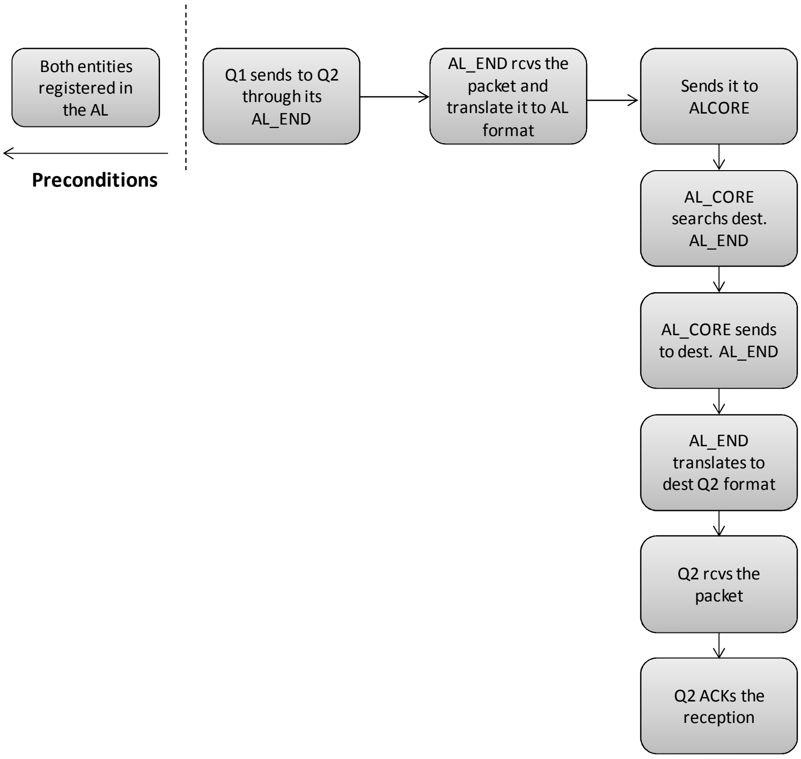

Both entities registered in the AL
Preconditions
Q1 sends to Q2 through its AL_END
AL_END rcvs the packet and translate it to AL format
Sends it to ALCORE
AL_CORE searchs dest. AL_END
AL_CORE sends to dest. AL_END
AL_END translates to dest Q2 format
Q2 rcvs the packet
Q2 ACKs the reception

Figure 3-11: Data message sent through the AL.

The most important points to check in this functionality are to keep the information through the adaptation layer and the correct dispatch of the message at its destination, as well as receiving the confirmation at the source entity. All this is necessary to assure a correct operation and, as demonstrated in tests, that the AL is capable of making it without mistakes.

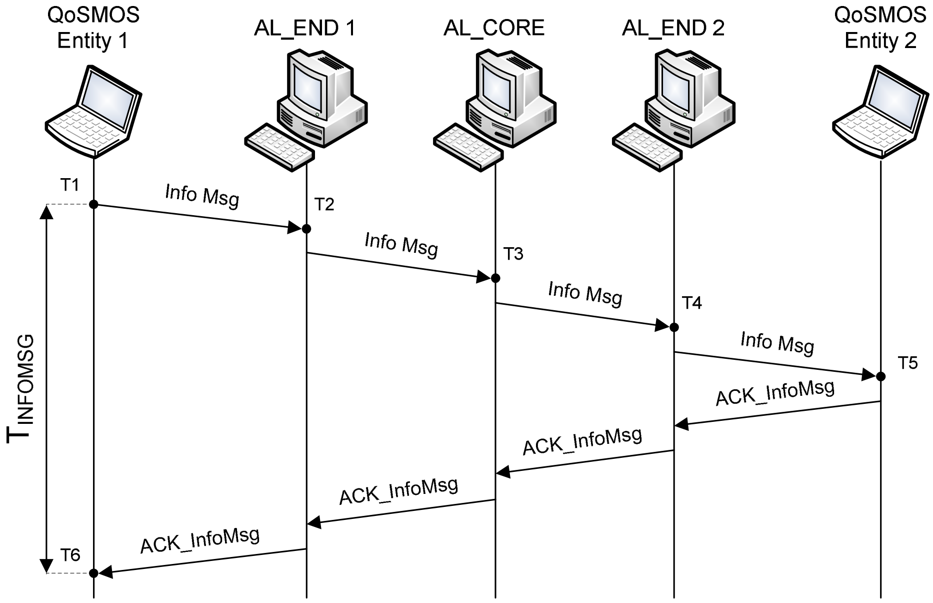


Figure 3-12: Information message test points.

To characterise this, it is considered that a sending information process begins when the QoSMOS source block delivers the message and concludes when the confirmation from the destination entity arrives at the source entity.

Taking this into account, it was decided to choose six test points, as shown in Figure 3-12. The test points has been set in order to identify the most critical activities, that are not only those involving regular AL_CORE activities, but also the translation feature presented in the AL_END in order to adapt a message to the destination. Considering this, the sending procedure can be seen as a store and forward procedure where some basic activities must be done in each step, with translation or seeking for the next hop in the communication path.

3.1.3.4 Keep-alive

As it has been stated in the definition of the adaptation layer in previous documents, up-to-date information is a must in its management. The AL should maintain the contactable entities DB updated in order to ensure that the QoSMOS entities are informed about all changes that occur in the system. To achieve this, the AL will send to the registered blocks in its database a keep-alive message based on a regular basis at certain intervals of time (keep-alive period), waiting for response from QoSMOS entities letting the system know about their availability.

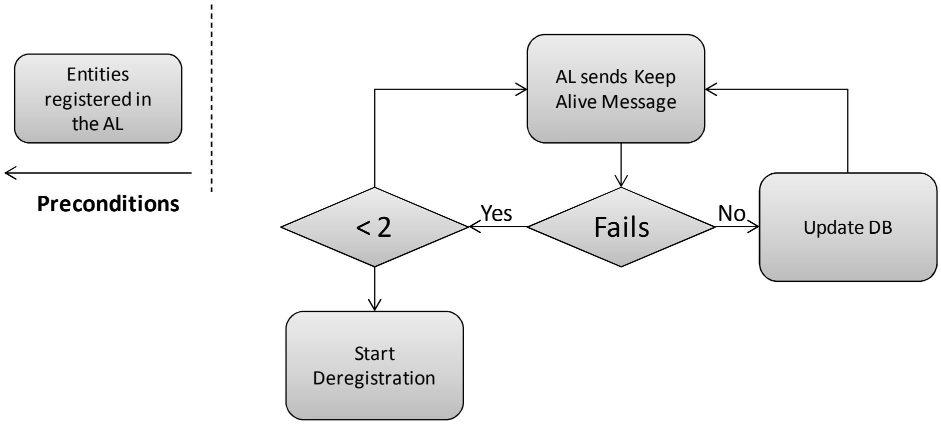


Figure 3-13: Keep-alive flow.

As presented in Figure 3-13, the system allows two fails from an entity, after that second failed try, the "hard" deregistration procedure is started. In this case, as the QoSMOS entity is no longer available, the deregistration procedure is initiated by the AL, removing the entry from the database and informing all entities sharing activities with it about the disappearance. The process is presented in a simplified way in the message sequence chart (MSC) in Figure 3-14.

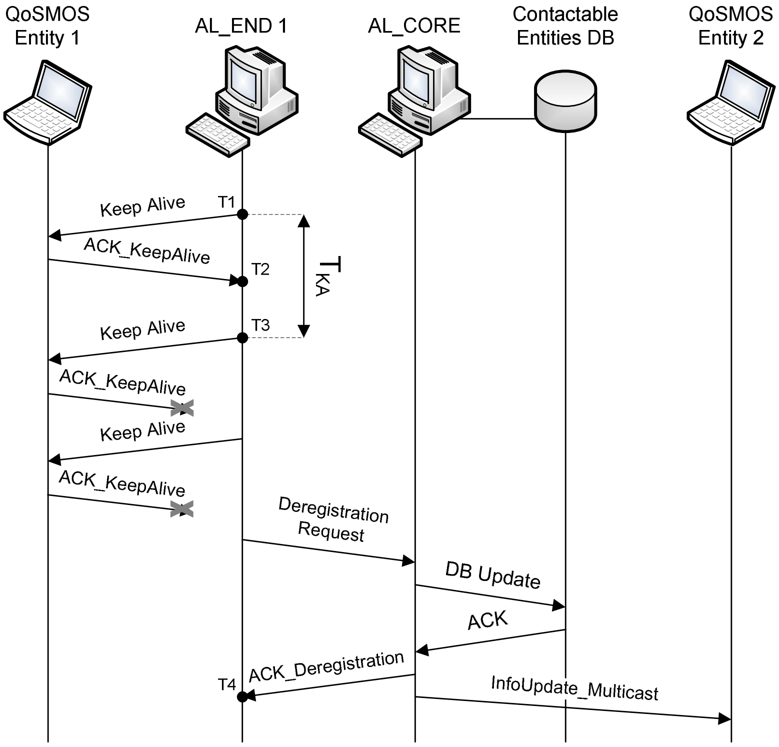


Figure 3-14: Keep-alive behaviour.

3.1.4 AL performance evaluation

In order to properly evaluate the AL features, a thorough theoretical analysis has been carried out, paying special attention to the overhead caused by its presence in the architecture.

Impact of AL in the system overhead

The role and impact of the AL has been also analysed from a theoretical point of view. The improvements brought by the AL in the system are especially remarkable in the discovery of the entities process.

The registration between entities can be done directly without the interaction of any intermediate element that centralises information and delivers it in an efficient manner. A proceeding like this is presented in Figure 3-15.

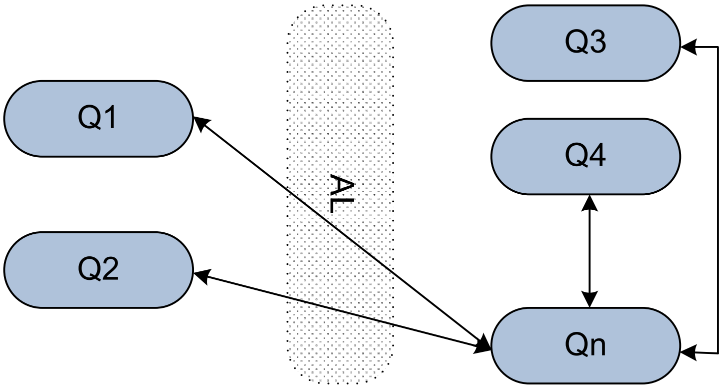


Figure 3-15: Registration between entities without AL.

The amount of messages that are needed for introducing a new entity into the system are dependant of the already present ones. Evaluating the general case for the *n*-th entity:

- one message to each of the *n*-1 requesting information
- one message from the *n*-1 responding the request

$$\#MSG_n = 2(n-1)$$

Based on the previous equation, it is possible to calculate the total amount of messages created in a system with *M* entities as:

$$\#MSG = \sum_{i=1}^{M} 2\times(i-1) = 2\times((M-1)+1)\times\frac{M-1}{2}$$

$$\#MSG = M\times(M-1) = M^2 - M$$

So, the order of magnitude of the messages exchanged in this case is $O(M^2)$.

Taking the previous analysis as a starting point, the following Figure 3-16 represents one of the two possibilities for operating with the AL. Here the case is presented when the AL sends unicast messages to the entities that possible interact with the new entity registered in the AL.

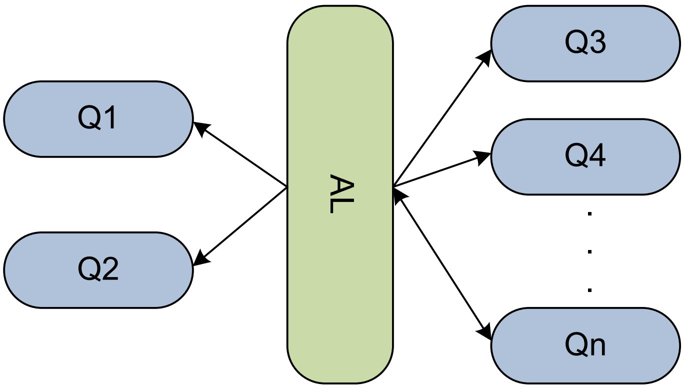


Figure 3-16: Registration with AL unicast messages.

Following the same approach to the one used in the case of no AL included in the system, the amount of messages needed for making all entities aware of others is calculated as follows.

For the *n*-th entity:

- 1 message to the AL
- 1 message to all the elements in the system that shares information with the new one.

This can be formulated as:

$$\#MSG_n = 1 + n$$

In a system with *M* entities connected the maximum amount of messages is:

$$\#MSG = \sum_{i=1}^{M} i+1 = ((M+1)+1)\times\frac{(M+1)-1}{2}$$

$$\#MSG = (M+2)\times\frac{M}{2} = \frac{M^2}{2} + M$$

In this second case the benefits of the AL are clearly visible due to the amount of messages is reduced to the order of $O\left(M^2\!/_2\right)$.

The third case considers that the AL sends a multicast message to the group of devices that is sharing the same interests with the recently registered node. In this last scenario the amount of messages sent are quite easy to calculate.

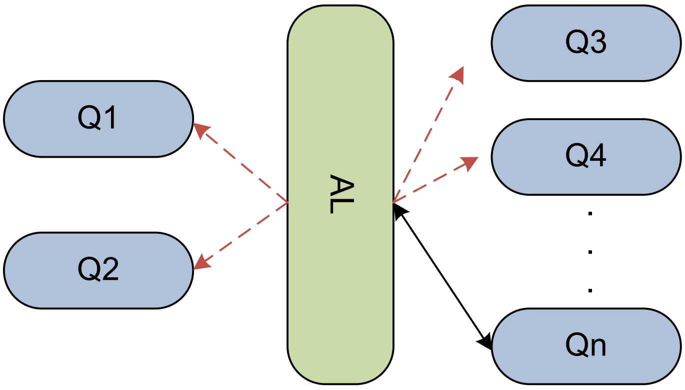


Figure 3-17: Registration with multicast AL feature.

In Figure 3-17, the red dotted lines represent one multicast message received by all entities sharing the same interests within the network. Considering this, the total amount of messages can be calculated as:

$$\#MSG = 3M$$

So the order of magnitude in this last case is $O(M)$.

This theoretical approach will be confirmed in the realisation of the different tests done and presented in the following sections.

Detection time and overhead based on keep-alive activities

Keep-alive messages are intended to track the status of all entities connected to the AL. Depending on the periodicity of the messages, the signalling overhead within the system could vary a lot. Depending on the value of the period between two consecutive keep-alive messages, the system will tend to be more reactive or proactive, being signalling overhead higher in the second case.

The theoretical analysis of the impact of the message could be seen as follows.

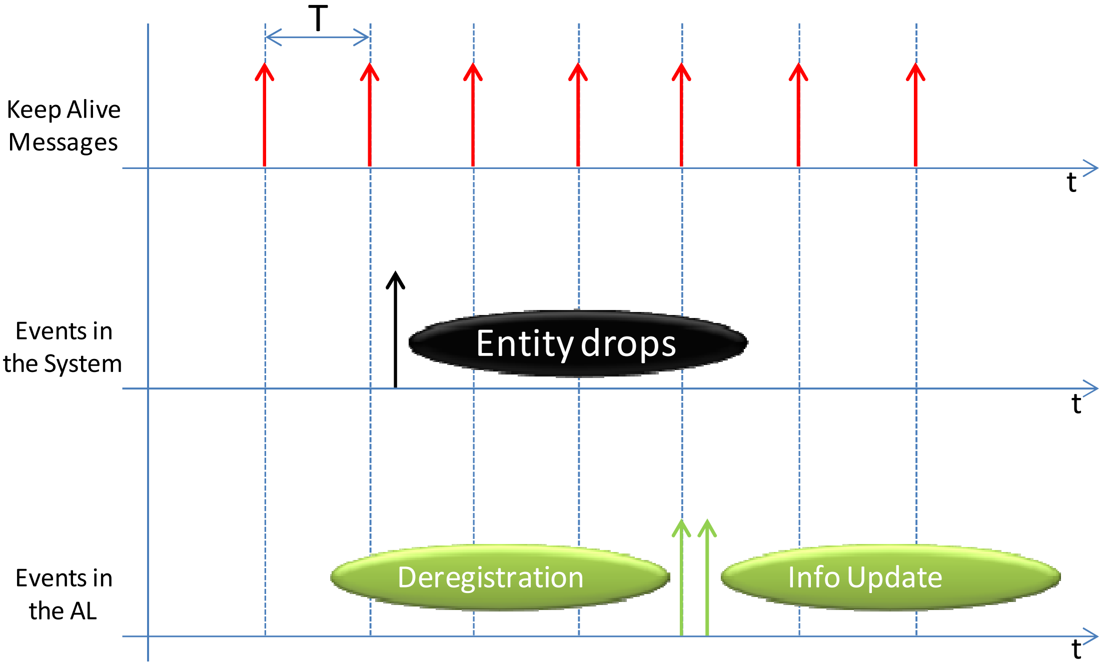


Figure 3-18: Time evolution of detection of a dropped entity.

As seen in Figure 3-18, every *T*, one keep-alive message is sent to each entity; all of them will ACK this message in order to be updated as “alive” in the AL. Since the removal of an entity from the AL has a remarkable impact on the overall system, it has been defined a procedure based on the missed acknowledgement reception after two consecutive messages. This representation can be seen in the figure producing a deregistration, initiated by the AL, and the correspondent Info Update message sent to all entities sharing some kind of information with the one that has disappeared.

$$\overline{t_{\det ection}} = 2T + \overline{P_{drop}}$$

The dropping of an entity is a discrete event without memory, so the drop probability (P_{drop}) can be represented as a constant for the whole time period comprised between two consecutive keep alive messages (Figure 3-19).

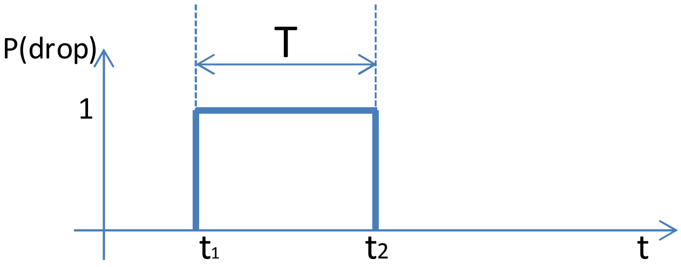


Figure 3-19: Drop probability of an entity between two consecutive keep-alive messages.

$$\overline{t_{\det ection}} = 2T + \overline{P_{drop}}$$

$$\overline{t_{\det ection}} = 2T + \frac{t_2 - t_1}{2}$$

$$\overline{t_{\det ection}} = 2T + \frac{T}{2} = {}^{5}\!/_{2}T$$

The mean detection time has a linear response. By the way, the amount of messages needed for tracking the status of the entities in the system is exponential as presented in Figure 3-20.. It is displayed the number of messages that are inserted into the network due to the improvement of detection time.

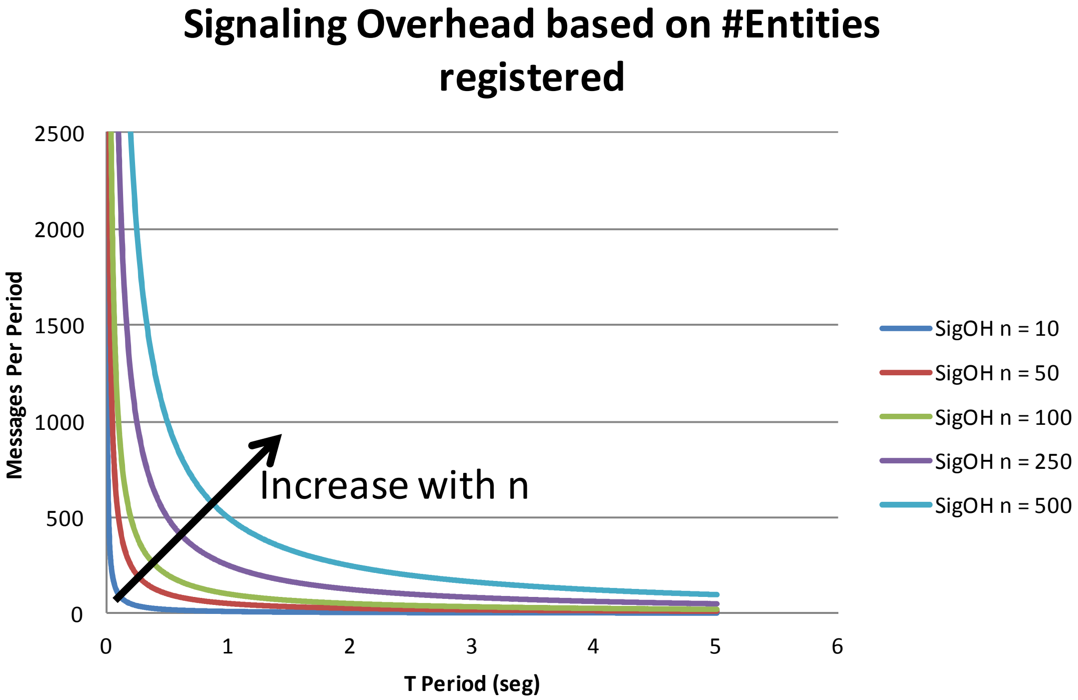


Figure 3-20: Signalling overhead in the system.

As it is shown, the more accurate the detection is the higher the number of messages introduced in the network will be. This represents a trade-off between load in the network and reaction time versus entities dropping.

3.1.5 Test scenarios

Scenario definition

The AL provides the rest of QoSMOS entities the chance to know up-to-date information of other blocks or to send messages without having to contact them directly. To check how the AL affects the performance of the QoSMOS system and to characterise AL capabilities, a set of tests has been set up.

These tests are based on the functionalities and validation presented in section 3.1.3 and they are running in the scenarios which are presented below in Figure 3-21 and in Figure 3-22.

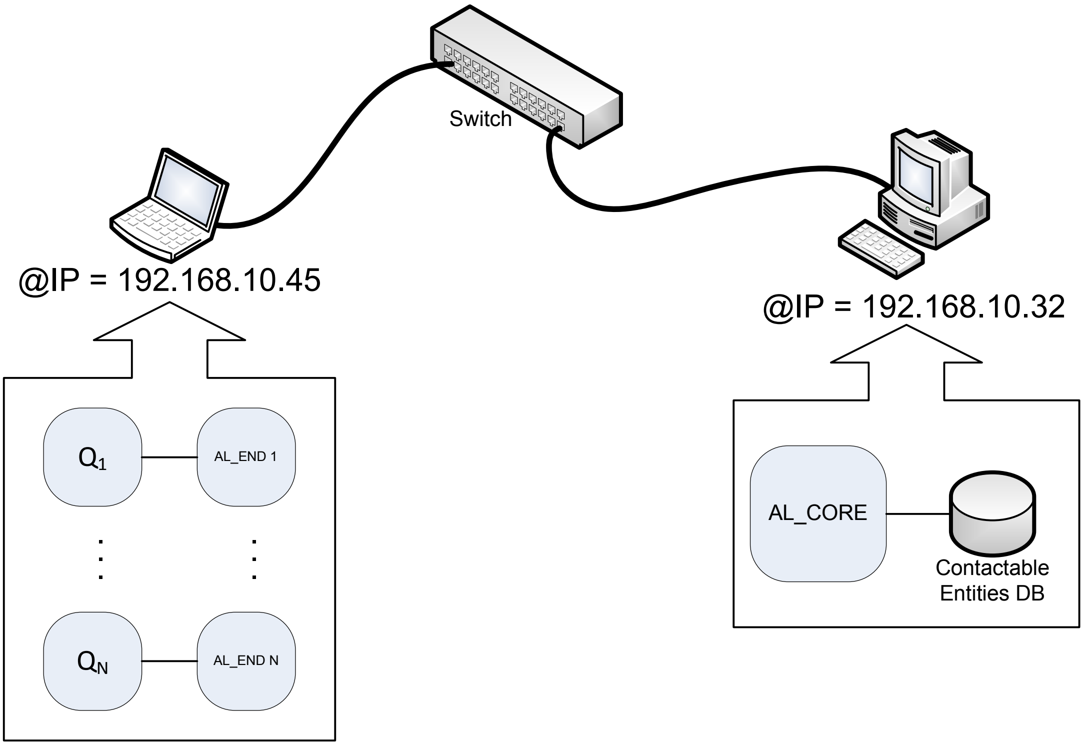


Figure 3-21: Single network test scenario.

In this scenario, the AL_CORE and the contactable entities database are in a personal computer (PC) while the QoSMOS entities and their corresponding AL_ENDs are situated in a laptop in the same sub-network. The two computers communicate through a wired LAN.

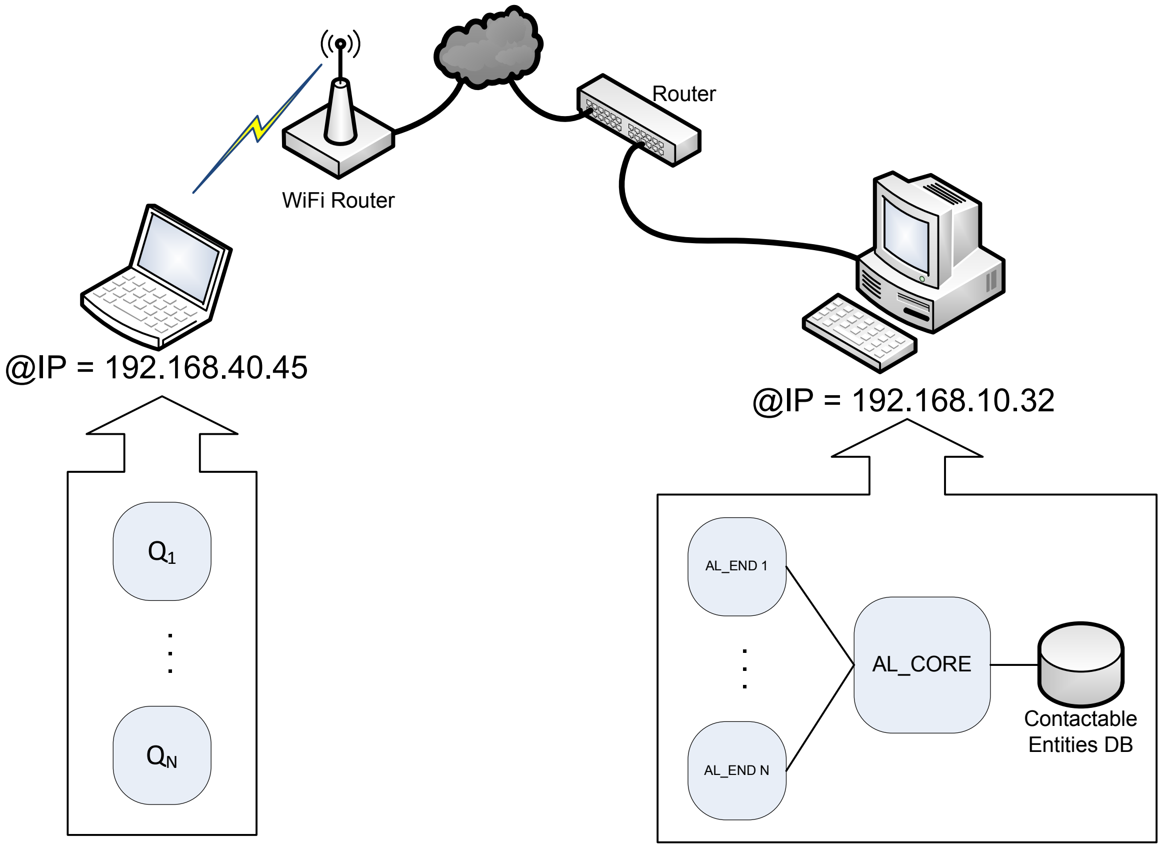


Figure 3-22: Multi-network test scenario.

In the case of the different sub-network scenario, the PC and the laptop communicate with each other through a wireless LAN. As seen in Figure 3-22, the AL_CORE and the AL_ENDs are centralised on the personal computer while the QoSMOS entities are on the laptop.

The aim of these tests is to analyse the impact of the AL within the QoSMOS system. To achieve this, it was decided to measure the delay that could be introduced by the adaptation layer and the network load caused by sending messages from the AL.

Results obtained

In order to get and understand the real performance of the adaptation layer, the results obtained in the different tests are presented in this section. It was decided to measure the delay in the message dispatch in registration, deregistration, information update and information message. Besides, it will be tested how the number of registered entities in the database can affect to the information update process.

The outcomes for the first test in the same sub-network scenario are shown in Figure 3-23. It presents the results for various processes of the AL. As can be seen, the time employed to perform these actions by the AL is less than the time used by the entity, except for the information update cases. It is due to the fact that the AL must wait until it receives the acknowledgment packets from the entities.

For performing an information update using the XML-RPC protocol, the AL will receive confirmations from the new entity and from the registered entity, whereas for carrying out an information update using IP multicast it only has to expect the XML-RPC acknowledgement from the new entity. For this reason, information update with IP multicast option is faster than the one which only uses XML-RPC.

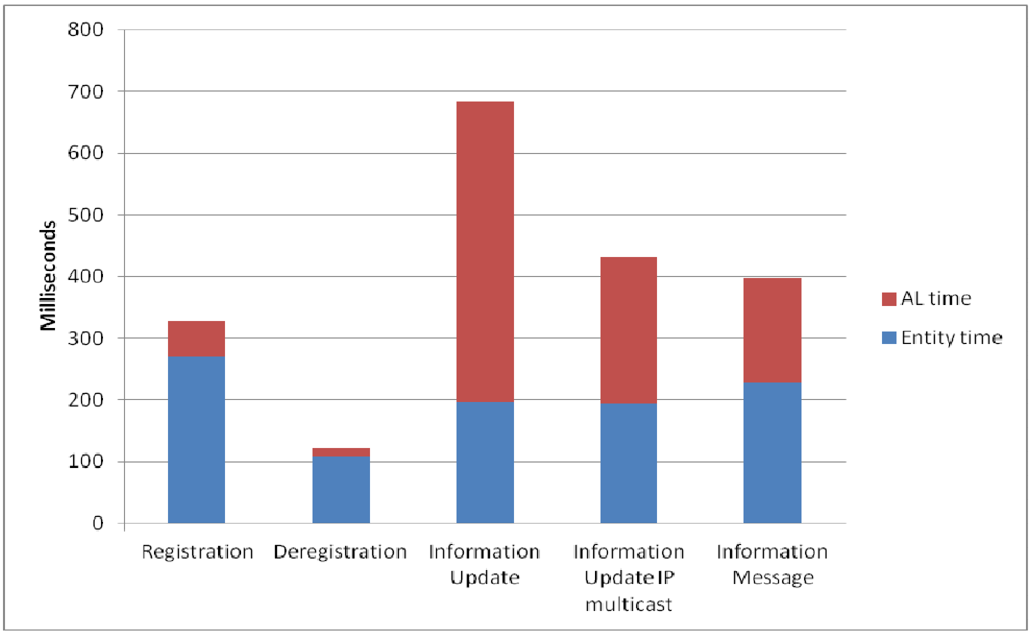


Figure 3-23: Delay introduced by AL the system.

The delay derived from the registration of a new entity in the system is tightly related to the existing ones already connected. In order to evaluate the most efficient solution, three different methods already presented in the theoretical analysis have been implemented.

The information update process can be carried out through the dispatch of a message for each registered entity in the database. That is, if there are *N* registered entities and a new entity enters the system, 2*N* information update packets will be sent from the AL, *N* to the new entity and *N* to the registered blocks. The following process is also executed on the XML-RPC protocol, but this time the AL will dispatch *N*+1 messages, one to the new entity with all common information and *N* to the others.

It is worth noting that in the case of using IP multicast, only two messages will be sent, one XML-RPC packet to the new entity and a multicast message to the rest.

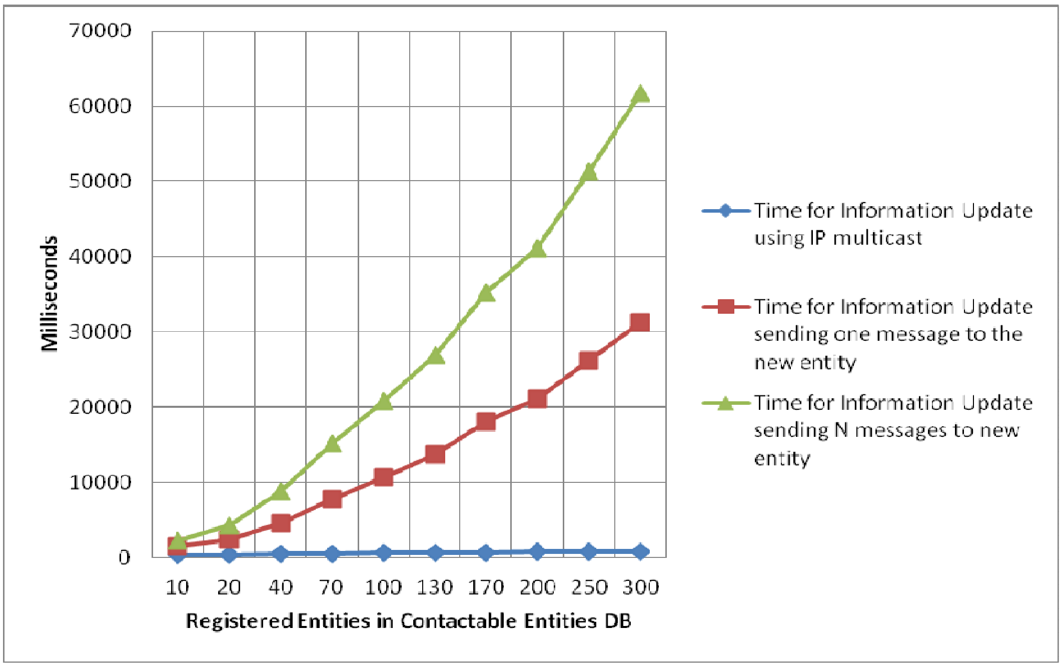


Figure 3-24: Comparison between update strategies.

As shown in Figure 3-24, the IP multicast method introduces alarge improvement in terms of time delay when the contactable entities database has many records. This is due to the fact that, not having to dispatch confirmations, the network load is considerably reduced.

Regarding the keep alive process, it has been found that the duration of send a message and receive the confirmation is always in the order of fifteen milliseconds. The time it takes the AL to realise that the entity is gone is around two times *T*. In terms of network load, this process introduces 2*N* messages every *T* seconds, where *N* is the number of entities registered in the database.

3.2 Physical layer abstraction for fast system-level performance prediction

The main purpose of physical layer abstraction is to predict the link performance of the communication system based on a small number of measureable metrics. It can be applied for simplified system performance evaluation and also for dynamic adaption of the parameters in order to match the predefined performance limits. In this way, a fast system level evaluation can be performed without the need for exhaustive search or detailed, extremely time consuming link-level simulations within the system-level simulator.

3.2.1 Overview

For an additive white Gaussian noise (AWGN) channel the signal to noise ratio (SNR) clearly determines the error rate (bit, block, frame or packet error rate). In the case considered here, the system level performance will be investigated based on the block error rate (BLER).

During a communication, in order to adapt the link performance, there are numerous parameters that can be adjusted, such as modulation type, coding rate and packet size. There are other parameters which cannot be altered but which have a strong influence on the performance, such as the channel profile.

The task of the PHY layer abstraction within the system-level simulation is to compose a simple technique, which can compress a large set of parameters and predict the final BLER of the communication link. One possible way of performing the steps of the abstraction can be seen in Figure 3-25. It consists of three major blocks strongly dependent on the physical layer parameters.

First, quality measurements for all of the physical resources are gathered (average SNR values, e.g. the per-subcarrier SNRs in the case of multicarrier modulation; antenna configuration; additional transmitter and receiver (Tx/Rx) processing; channel information; etc.) and the signal to interference-plus-noise ratio (SINR) values for each sub-channel are calculated. Then, in a second part, a suitable compression of the multiple SINR values to one single effective quantity SINReff is performed for further processing. Finally, this single value is mapped to BLER measures to predict the instantaneous block error probability.

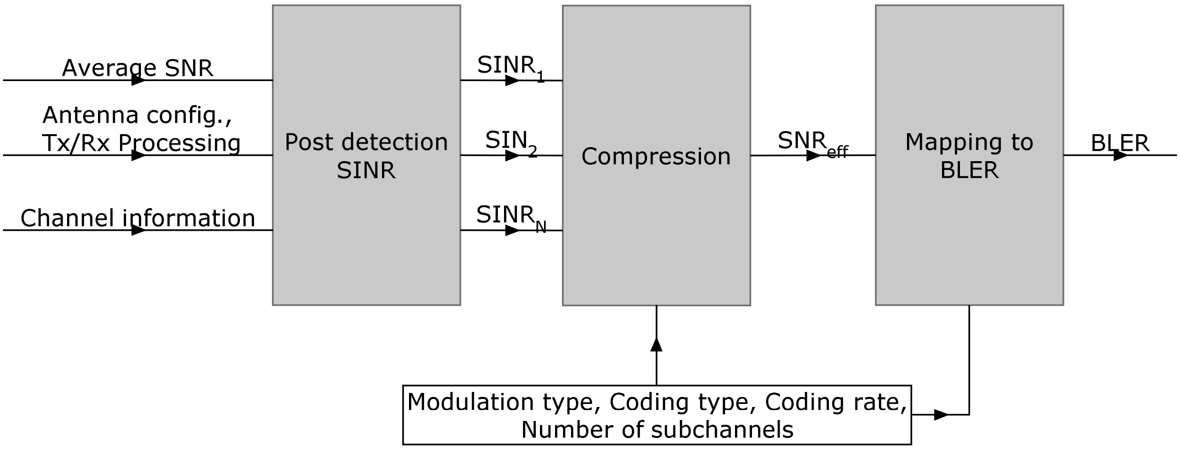


Figure 3-25: Link layer abstraction procedure.

The following steps need to be performed in order to apply the abstraction technique:

- define/acquire a set of communication scenario parameters (channel, modulation, coding);
- perform link-level simulations to get the abstraction parameter table;
- use an abstraction method to determine system performance (BLER);
- validate the abstraction through test data set with different channel realisations.

For all effective SNR mapping techniques, the aim is to obtain a function for which all independent SNR values can be mapped to one single effective SNR value which yields to an accurate estimate of the BLER value over the AWGN channel:

$$\mathrm{SINR}_{\mathrm{eff}} = \frac{1}{\mathrm{N}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \Phi(\mathrm{SINR}_{\mathrm{n}}) \qquad (3\text{-}1)$$

There exist well-known and powerful abstraction techniques for orthogonal frequency-division multiplexing (OFDM)-based mobile communication systems. In the following section some of these techniques for the mapping procedure are briefly reviewed and their applicability to filter bank multicarrier modulation (FBMC) based systems will be discussed. In the last section the most suitable one based on hard decision will be evaluated.

3.2.2 PHY abstraction techniques

The following abstraction methods described in [KliEtal2009] for OFDM based systems, are here investigated:

- exponential effective SINR mapping (EESM);
- mutual information based effective SINR mapping – received bit mutual information (RBIR);

- mutual information based effective SINR mapping – mean mutual information per bit (MMIB).

The EESM technique is the simplest one as it only requires the knowledge of a per-subcarrier SNR for performance prediction. On the other hand, the RBIR and MMIB techniques require soft information which makes the computation and algorithm complexity higher.

Recent results of WP4 show [DatEtal2011] that in case of FBMC the inter-symbol interference (ISI) and inter-carrier interference (ICI) caused by the multipath channel can strongly degrade the performance of the system and soft decision calculations of the received bits can be rather complicated. As a result the simplest technique, which is suitable for FBMC, the EESM was selected for further investigations. In the case of the EESM the per-subcarrier SNR values are mapped using an exponential function and a weighting constant β as:

$$SINR_{eff} = -\beta \ln\left(\frac{1}{N}\sum_{n=1}^{N} e^{\left(-\frac{SINR_n}{\beta}\right)}\right) \qquad (3\text{-}2)$$

The goal is to determine the optimal β, which is a non-linear curve-fitting problem.

3.2.3 Parameter estimation tool and conclusions

The basic steps of the parameter calculation and optimisation can be seen in Figure 3-26 below. First, based on the channel profile, a random channel realisation is generated. Subsequently, Monte Carlo simulations are performed for the given set of modulation/coding parameters in order to determine the BLER versus SNR values. Then, based on the SNR of each subcarrier and the simulated BLER curves, a curve fitting is performed in order to find the optimal β parameter to fit the curves to the AWGN curves.

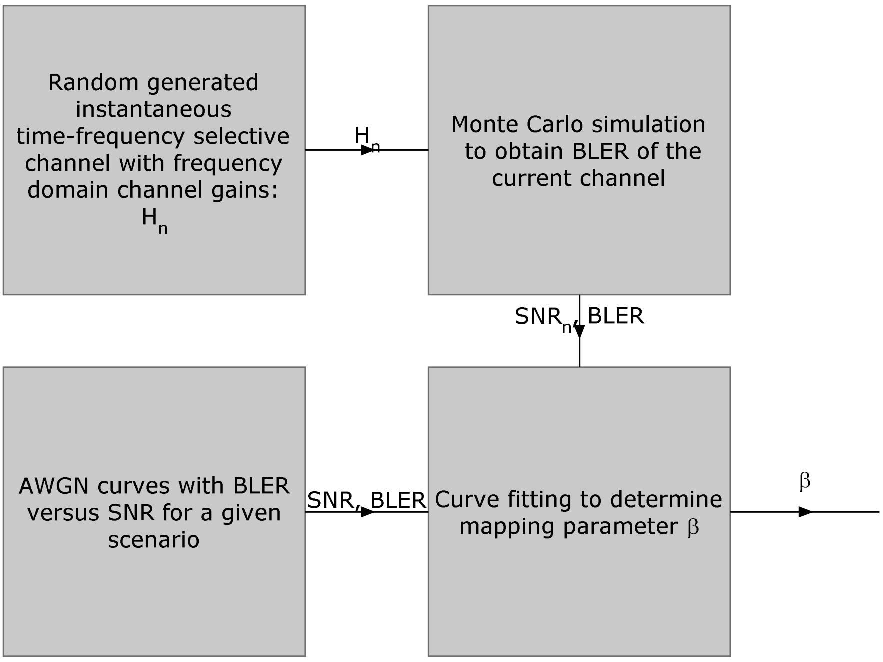

Random generated instantaneous time-frequency selective channel with frequency domain channel gains: Hn
Hn
Monte Carlo simulation to obtain BLER of the current channel
SNRn, BLER
AWGN curves with BLER versus SNR for a given scenario
SNR, BLER
Curve fitting to determine mapping parameter β
β

Figure 3-26: Parameter optimisation for PHY abstraction.

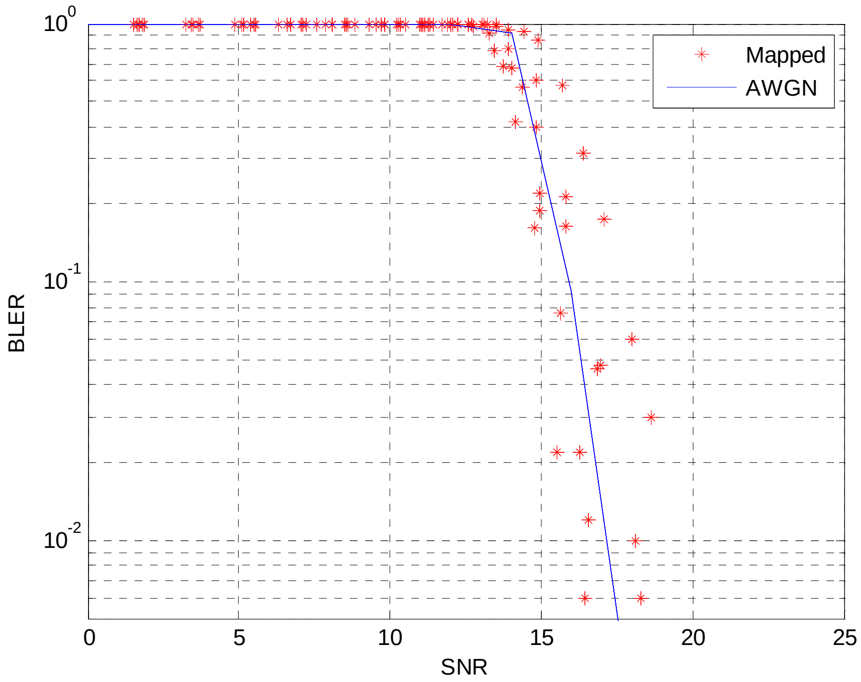


Figure 3-27: Validation of the abstraction parameter β.

For demonstration of the abstraction technique, the parameters presented in Table 3-1 have been chosen.

Table 3-1: Simulation parameters.

block size	5 symbols = 1205 bytes
modulation	16-QAM
number of subcarriers	1024
number of used subcarriers	482
coding	convolutional code, R = ½
channel	IEEE 802.22 channel profile B

After performing Monte Carlo simulations, a value of 4.5 for β was achieved. The parameter was then validated for 10 channel realisations. This result can be seen in Figure 3-27. The mapped SNR curves with red stars fit well to the AWGN curve presented in blue.

A Matlab programme was developed for parameter optimisation. The input parameters of the programme are defined in collaboration with WP4.

As a final result, a table is to be built with all β values for all possible channel profiles and transmission parameters. Based on this table a system can instantaneously adapt the transmission to match the desired BLER based on the measured SINR pro subcarrier. This table can also be used for simplified system-level simulations at a higher level to predict system performance.

3.3 MAC performance evaluation

A comparison of contention-based MAC layers for TV whitespace systems was described in [D5.2]. The two protocols compared were those specified in the IEEE 802.11 and ECMA-392 standards. One of the reasons for this comparison was to demonstrate the difference in the behaviour of two similar methods of channel access which can provide significantly different performance from one another in terms of throughput and also the ability to coexist with other systems effectively. A QoSMOS system can use this knowledge in a number of ways, as described in [D5.2]. A system could use its knowledge of how other systems access the channel to negotiate fair sharing. A QoSMOS system may actually access a shared channel using one of these contention-based methods if contention-based methods are used among systems sharing the same channel. A QoSMOS system can further improve channel utilization and effective coexistence by issuing parameter set updates to systems using the same channel.

The work in [D5.2] focused on the behaviour of saturated systems and highlighted the key differences in the IEEE 802.11 and ECMA-392 behaviours. This work has been further investigated by QoSMOS in [MacEtal12] where simple mechanisms for parameter optimisation of these saturated systems are shown. This section summarises the key findings from this previous work and extends on it by providing further results on the optimization of saturated systems and then also non-saturated systems.

The key difference between the contention-based channel access mechanisms of IEEE 802.11 and ECMA-392 is how their backoff procedures reset the minimum contention window (CWmin)[3]. In the description of the backoff procedure of ECMA-392 there is a general rule which matches the behaviour of IEEE 802.11 that a contention window (CW) is reset to CWmin following a successful frame transmission. This rule applies to each channel access function (CAF). Each CAF at a station will only deal with traffic from one access category. Several rules are also applied in ECMA-392 which, when in conflict with the general rule, are adhered to instead. Two of these rules apply to a CAF when a successful transmission has taken place which is the final transmission in a transmit opportunity (TXOP). One rule, which applies when the CAF has no further packets in its buffer, states that a new backoff is selected with CW reset to CWmin. This is in agreement with the general rule. The other rule applies when the CAF still has packets in its buffer and states that a new backoff is selected but CW is not reset and remains at its current value. This conflicts with the general rule. This ability to change behaviour based on buffer status is a simple and effective way to reduce congestion on the wireless channel when a system becomes heavily loaded.

For all results shown in this section the MAC service data unit (MSDU) packet size is 1500 bytes and the physical layer rate is 31.65 Mbit/s. Note that both ECMA and IEEE802.11 are contention-based MAC protocols and the performance improvement in ECMA is due to the optimising of the contention window by varying CW values based on buffer status. The comparison between contention-based and reservation-based (scheduled) MAC are not part of this study. One example of reservation-based MAC is HIPERLAN/2. Also, whitespace radio based on WiMAX technology uses TDD/TDMA-based scheduled MAC protocol, whereas IEEE802.11-based Ubiquiti whitespace radio uses a contention MAC protocol.

[3] The rules for invoking the backoff procedure for ECMA-392 are defined in section 7.5.1.7 of the standard. The authors found a little ambiguity in understanding these rules. The description provided here was confirmed with discussions with the ECMA-392 editor.

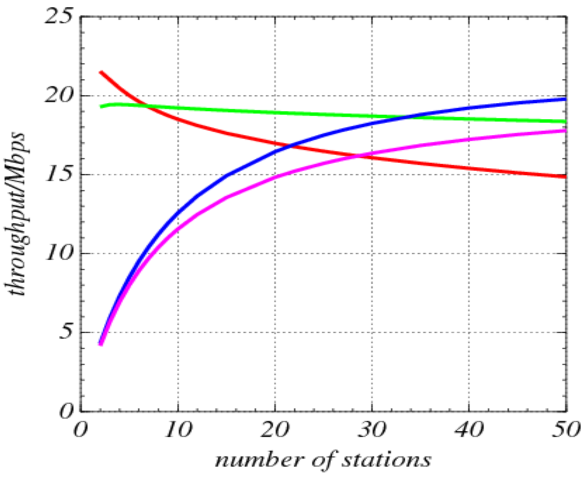


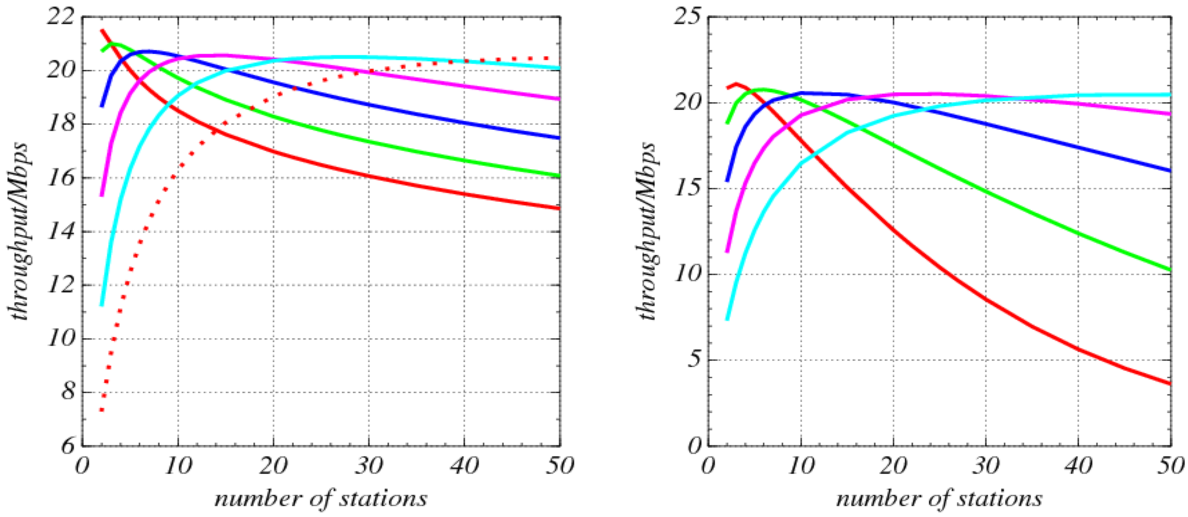
Figure 3-28: IEEE 802.11 (basic=red, rts=green) and ECMA 392 (basic=blue, rts=magenta) system performance.

Figure 3-28 shows that saturated IEEE 802.11 and ECMA-392 systems behave differently, even using the same parameter values. In short, the IEEE 802.11 system is more aggressive. This means that it can gain high throughput when a network is small, but when the system is larger the high probability of collisions causes the throughput to fall. ECMA-392 on the other hand is more conservative. This means that it does not utilize the channel very effectively when the network is small, but as the system increases in size its utilization increases. This also makes ECMA-392 a better protocol for sharing a channel with another system.

The use of the RTS/CTS (request to send, clear to send) mechanism is also shown (marked as “rts” in the figure). For the IEEE 802.11 system, where a significant proportion of time can be wasted on collisions, RTS/CTS can improve the system performance. The RTS/CTS mechanism applied to ECMA-392, however, simply increases overheads and so doesn’t offer the same benefits as IEEE 802.11 sees[4].

It has been found that a simple and effective way to optimise an IEEE 802.11 system is to fix CWmax and tune CWmin based on the system load. This is shown in Figure 3-29a. For ECMA-392 it has been found that a simple and effective way to optimise is to fix CWmin and tune CWmax based on the system load as shown in Figure 3-29b. These results show that the optimised system throughput for both systems can stay approximately constant (for this network settings ≈ 21Mbit/s). What is impressive with these results is that only one parameter is being tuned at a time. By allowing the parameter to be tuned over 5 or 6 values a near constant system capacity can be achieved for a network size of up to 50 stations. There is even the scope to reduce the number of values used for optimising the system further while allowing the system capacity to still remain relatively constant. For example, if the IEEE 802.11 system in Figure 3-29a used only three values for CWmin (15, 63 and 255) the system throughput would only vary from 20 to 21Mbit/s up to a network size of 50.

[4] In [MacEtal12] it is explained that, by understanding the backoff rules of ECMA-392, there are several ways to make an ECMA-392 device behave more like an IEEE 802.11 device. One suggestion includes the use of RTS/CTS in which case the backoff behavior is the same as 802.11. In this work we only consider ECMA-392 performance when its backoff behavior differs from IEEE 802.11, as explained above (i.e. following a successful transmission CW is only reset to CWmin if the buffer is empty). The ECMA-392 results with RTS/CTS, shown above, show how the curve for basic access would look if the same backoff rules were applied when the RTS/CTS feature was used.



(a) (b)

Figure 3-29: Parameter optimisation. (a) IEEE 802.11 with CWmax is fixed at 1023 while CWmin varies: CWmin=15 (red), 31 (green), 63 (blue), 127 (magenta), 255 (cyan) and 511 (dotted) and (b) ECMA with CWmin fixed at 7 while CWmax varies: CWmax=31 (red), 63 (green), 127 (blue), 255 (magenta), 511 (cyan).

The optimisation results can be better understood by looking at collision behaviour of these systems. The results shown so far are based on Bianchi's model for IEEE 802.11 [Bianchi04]. The analysis in this model uses τ to represent the probability of a station transmitting in a particular timeslot for a system that has *n* stations. This analysis can be extended to calculate Pr[NTX=*x*] which represents the probability that when a transmission does occur (NTX=no transmit), there are *x* stations that start their transmissions at the same time. For values of $x \geq 2$ a collision will occur. Pr[NTX=*x*] can be calculated as

$$\Pr[NTX = x] = \frac{\binom{n}{x}\tau^{x}(1-\tau)^{n-x}}{1-(1-\tau)^{n}}$$

A QoSMOS system, or any other opportunistic system that may use contention-based channel access, has another reason to minimise the probability of collisions on the channel. As well as reducing the collision probability to maintain a high system utilisation, the collision behaviour can also pose a problem with aggregate interference. The maximum transmit power of opportunistic systems is likely to be restricted so that the total interference into the area of an incumbent user is kept below a certain threshold. If the probability of collisions is high then the transmit power of any individual transmission will be reduced in order to make sure the aggregate interference stays below the interference threshold.

Figure 3-30 shows the collision behaviour for the IEEE 802.11 system in Figure 3-29a (CWmin=15 and 63). The basic trend for describing any of these parameter settings is that when the network is small, so is the probability of collision. As the network size increases so does the probability of a collision. Also, as the probability of a collision increases, the average number of transmissions involved in a collision increases. The first set of results, for CWmin=15, are the most aggressive 802.11 parameter settings used for the optimisation shown in Figure 3-29a. These are only used for an optimised system when the network size is 3 or less. For networks of this size, where this parameter setting is optimal, the probability of a collision with just 2 stations (NTX=2) is quite low, while the probability of NTX>2 is almost negligible. As seen in Figure 3-30a, when Pr[NTX=2] starts to rise above 10% and Pr[NTX>2] becomes noticeable, these settings are no longer considered as optimal. For the second set of 802.11 parameter settings shown here, for CWmin=63, these settings are optimal for a network size from 6 up to 11. The collision probabilities rise slower than for the CWmin=15 results confirming that these settings are less aggressive. When the network size is less than 6 these parameter settings are not aggressive enough, meaning that the channel utilisation would be low. When the network size reaches 6, the Pr[NTX=2] has risen to around 5% in Figure 3-30b. At this point the parameter settings are aggressive enough to prevent the channel from being unused while the collision probabilities are still low enough for most transmissions of most stations to be successful.

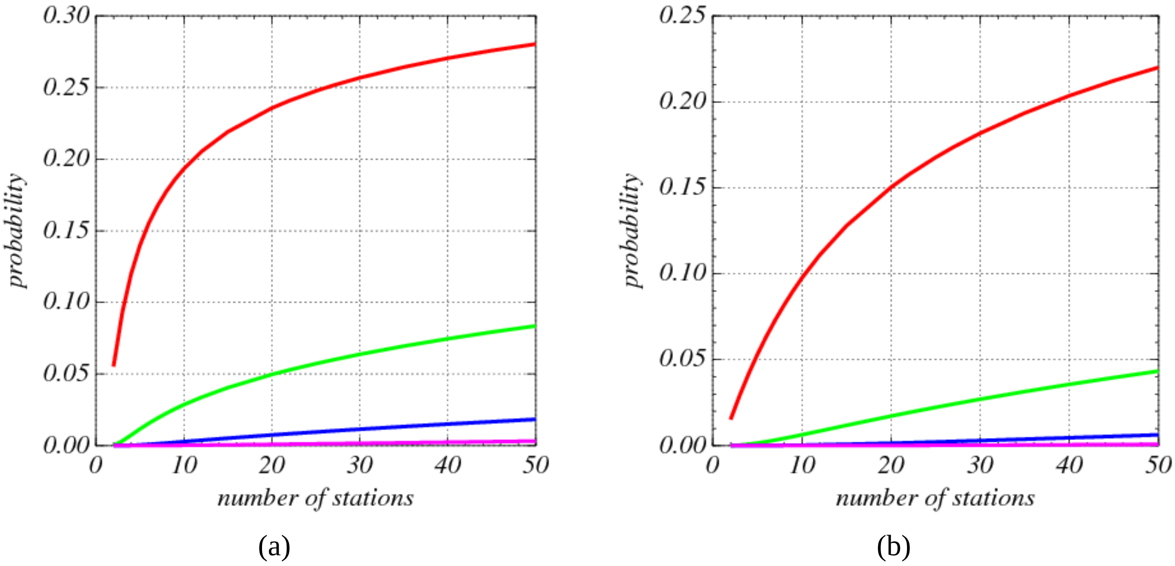
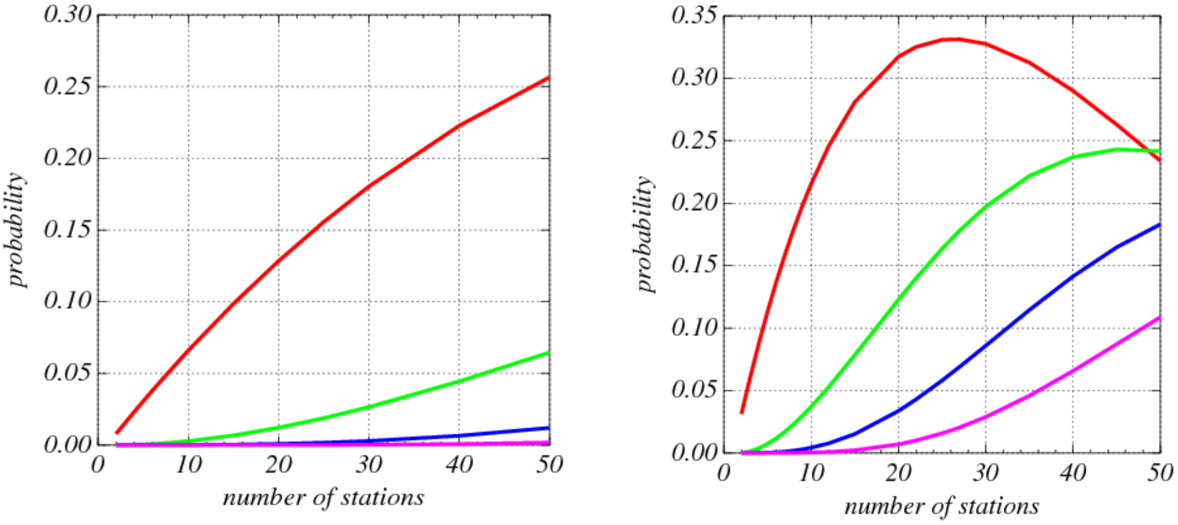


Figure 3-30: Collision behaviour for IEEE 802.11 (a) CWmin=15 (b) CWmin=63. NTX= 2 packets (red), 3 (green), 4 (blue) and 5 (magenta)

Figure 3-31 shows the collision behaviour for the ECMA-392 system in Figure 3-29b (CWmax=31 and 127). Very similar trends are seen as with the 802.11 collision behaviour results.

These optimisation results also show that when a system is optimised collisions should not occur too often, but when they do occur it is likely to only involve two transmissions. This shows that optimising these networks is also beneficial for minimising the effects of aggregate interference which would otherwise cause the transmit powers of opportunistic systems to be reduced further.

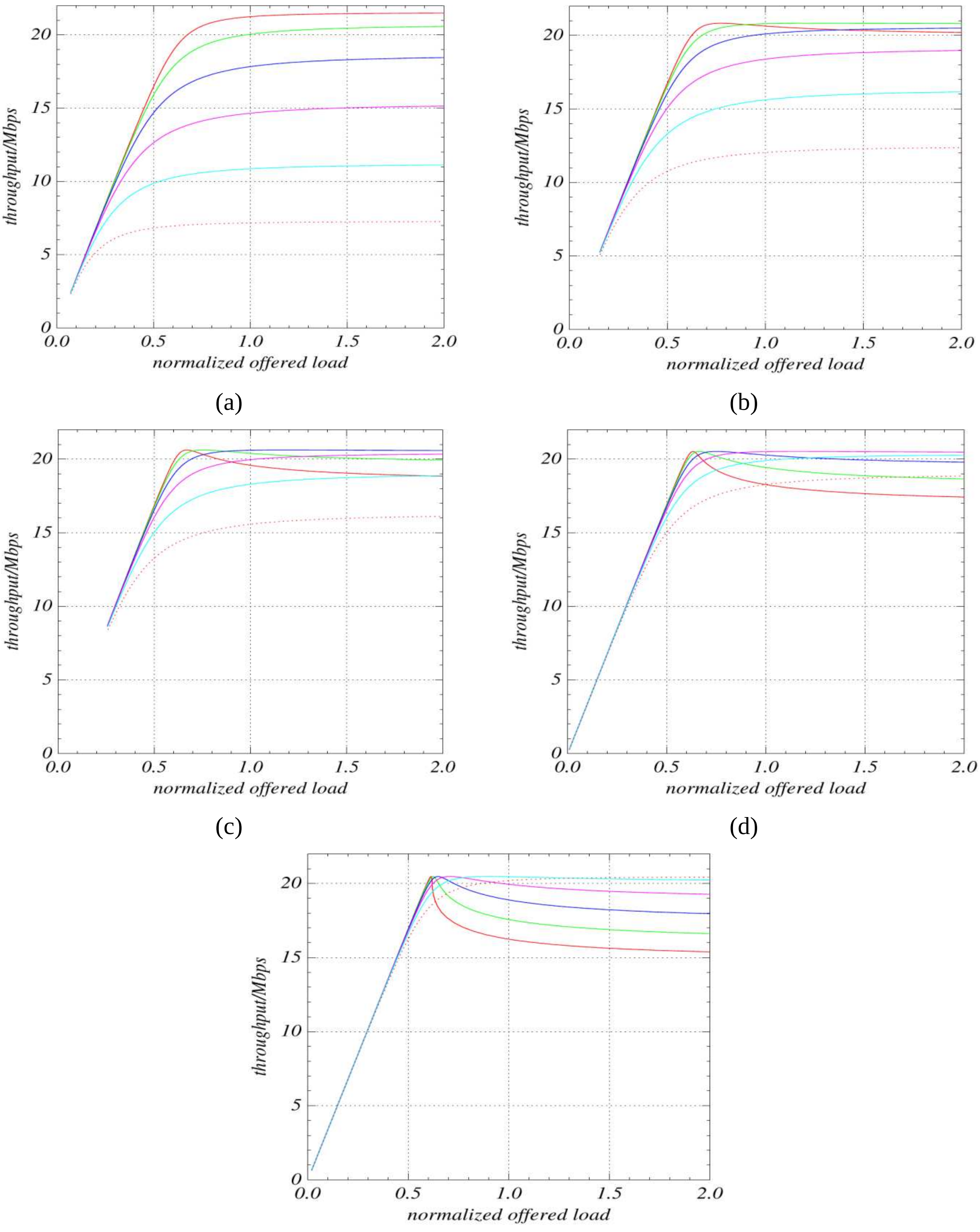
It is found in this work that different parameter settings are used to optimize these two different protocols. However, the optimised performance of these two systems is effectively the same. So, in a system which optimises its parameters in the fashion just described, the choice of protocol has little effect on the throughput performance of the system.



(a) (b)

Figure 3-31 Collision behaviour for ECMA (a) CWmax=31 (b) CWmax=127. NTX= 2 packets (red), 3 (green), 4 (blue) and 5 (magenta)

The final set of results, shown in Figure 3-32, looks at the optimisation of a network with stations that are not saturated. The results are for an IEEE 802.11 system using the Duffy model [DufEtal05]. The five subfigures are for a network of 2, 5, 10, 20 and 50 stations respectively. The results show that the optimisation of the parameter settings is also a function of the system offered load, not just the network size. At the highest offered load in these graphs (i.e. as the system approaches saturation) you will see that the performance shows a good match with the saturated results in Figure 3-29a.



(a)

(b)

(c)

(d)

(e)

Figure 3-32: Parameter optimisation for IEEE 802.11 for non-saturated stations. Number of stations = 2 (a), 5 (b), 10 (c), 20 (d) and 50 (e). CWmin=15 (red), 31 (green), 63 (blue), 127 (magenta), 255 (cyan) and 511 (dotted)

For a network of size 2 we see that the only parameter value needed is CWmin=15, which is the most aggressive set of parameters tested for IEEE 802.11 here. The reason this setting is always optimal for such a small network is that there are very few stations that can transmit at any one time, so large contention windows, which are used for collision avoidance, are not necessary and a small contention window is better. For a network size of 5, the CWmin=15 setting is still optimal while the system load is low. As the offered load in the network rises, the CWmin=31 setting takes over as the optimal parameter value. If we look at the results for the large networks (20 or 50) we see that the aggressive parameters might not be needed at all. A setting such as CWmin=63 can give close to optimal performance even when the offered load is very low. As the load increases the optimized system requires further increases in CWmin.

In summary, this section has demonstrated the expected behaviour of opportunistic systems that use contention-based channel access. It is shown that optimising a single parameter over a small parameter set can offer very high channel utilization over a small set of parameter choices. The optimisation of the system also limits the likelihood of multiple stations transmitting at the same time. This reduces the aggregate interference that these systems could cause to neighbouring systems.

The optimised results in this work tend to offer a system capacity (MAC layer throughput) of around 21Mbit/s with a 31.65Mbits/s physical layer. This is a utilization of around 65%. With the same MAC and PHY header overheads a perfect scheduler would be able to achieve a utilization of around 77% (as may be easily checked by a simple ration of timeslot durations). This shows that the potential to use contention-based access in opportunistic systems can still achieve high performance in terms of channel utilization. This may be a particularly preferable way to share a channel with other users, especially if they are already using contention-based access.

3.4 Impact of sensing accuracy on QoS

The effects of sensing accuracy on the quality of service of both incumbent and opportunistic users are presented in this section.

3.4.1 Introduction

One of the common problems associated with opportunistic radio access based on spectrum sensing is the spectrum sensing accuracy. Collision will occur if an existing opportunistic user could not detect the arrival of an incumbent user or a newly arriving opportunistic user could not detect the presence of the incumbent user on a channel. Different types of spectrum sensing mechanisms have been proposed for detecting the presence and absence of incumbent users. The most widely used sensing mechanism is the energy detector [Urk1967].

In energy detection [Leh2005], the detector measures the energy of the received signal during some time period and in some frequency channel. Sensing methods have been discussed in WP3 deliverable D3.3. In energy detection, the measured value of energy is compared with a threshold. If the threshold is exceeded, the detector decides that a signal was present. The probability of correctly detecting the signal is called the probability of detection P_{D}. If just noise or other interfering signals cause the threshold to be exceeded, this is called a false alarm. False alarms are generated by internal receiver noise and/or external interference. The false alarm probability P_{FA} refers to the probability that a free channel is classified as being occupied. If the threshold is not exceeded when the incumbent user is really present, this is called a missed detection P_{M} (i.e. the probability that an occupied channel is classified as vacant). Another factor influencing the performance of cognitive radios is spectrum handoff capability that enables opportunistic users who have to vacate their current channel due to the arrival of incumbent users in order to perform spectrum handoff to other unoccupied channels.

3.4.2 Network model

This section presents an analytical model to analyse the performance of a cognitive radio network by use of a continuous-time Markov chain (CTMC) model. The model supports multi-channel multi-user cognitive radio network with imperfect sensing. We consider a cognitive radio network as illustrated in Figure 3-33 with N number of channels. The network provides wireless access over a geographical area. These channels can be shared between incumbent and opportunistic users in an opportunistic manner with incumbent users having priority over opportunistic users. Opportunistic users are allowed to access the channels that are not occupied by incumbent users. We assume that incumbent and opportunistic users have distinct arrival rates (λ_1, λ_2) and distinct service rates (μ_1, μ_2). The arrival and service rates are modelled by a Poisson process. New incumbent user call requests will be blocked if all channels are occupied by incumbent users while new opportunistic user call requests will be blocked if all channels are occupied by incumbent and/or opportunistic users. Figure 3-34 illustrates the timing and channel occupancy diagram for a four-channel network.

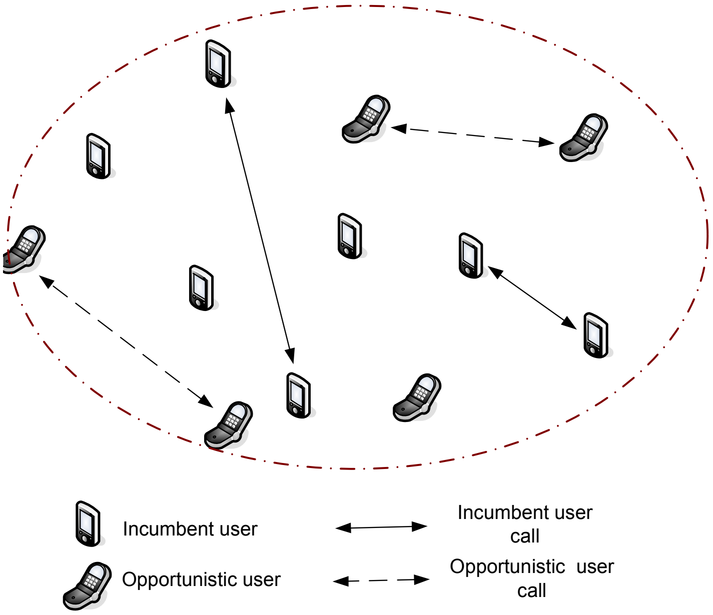


Figure 3-33: Cognitive radio network.

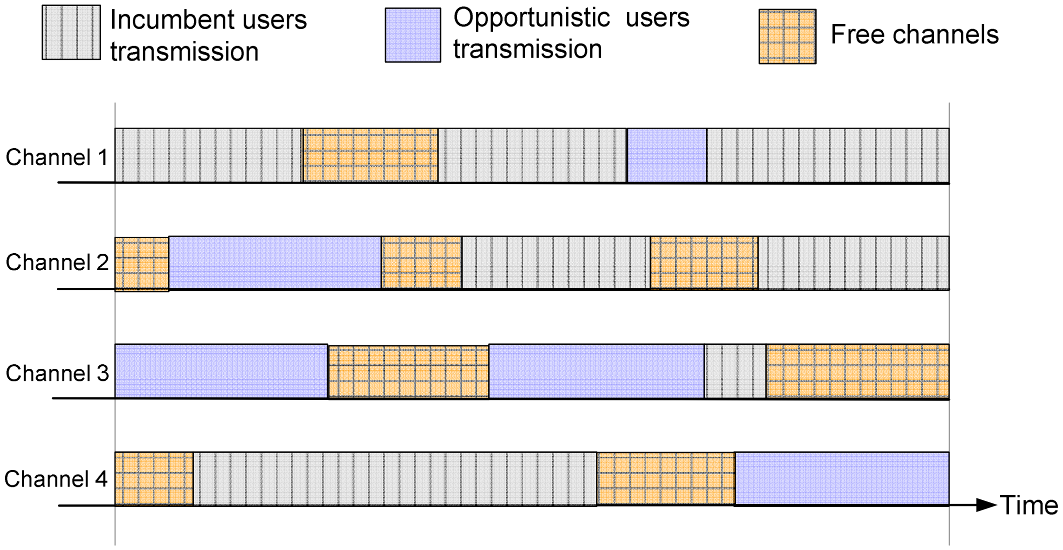

Incumbent users transmission
Opportunistic users transmission
Free channels
Channel 1
Channel 2
Channel 3
Channel 4
Time

Figure 3-34: Timing and channel occupancy diagram in four-channel network.

A two-dimensional Markov chain is used to model the cognitive radio network system. The system states are given by two-tuples (*i*,*j*) where *i* is number of channels used for incumbent users' calls and *j* is number of channels used for opportunistic users' calls. For example, state (1, 2) refers to state with one channel for incumbent user and two channels for opportunistic users. The total number of channel occupied by incumbent and opportunistic users cannot exceed N. Therefore, the following restrictions appear: $0 \leq i \leq N$, $0 \leq j \leq N$, $i+j \leq N$. *Let* $Q_{(i,j)}$ denote the steady state probability that the system is in state (i,j), which can be interpreted as the proportion of time that the system spends in state (i,j).

During the channel searching process, the opportunistic user searches for a free channel by randomly selecting one channel to detect whether it is free or not using the specified false alarm and detection probabilities. If the selected channel is found to be occupied by an incumbent user, the opportunistic user performs detection on the remaining channels with random order until it finds a free channel or all channels are determined to be busy. The time it takes to find a free channel (i.e. channel acquisition time) is assumed to be negligible.

By using the Markov chain and state transitions, formulas for different performance metrics are derived. The following performance metrics are used to evaluate the performance of the cognitive radio network.

Incumbent user termination probability - The term incumbent termination probability is used to refer to the probability that an incumbent user call, which has not been blocked initially, is terminated due to collisions with opportunistic users because of misdetections. There are two cases in which the incumbent user calls can be terminated. First, when an opportunistic user arrives at a channel occupied by incumbent users and it could detect the presence of the incumbent user ending up colliding with the incumbent user. The second case for collision is when an incumbent user arrives at a channel occupied by an opportunistic user, the opportunistic user has to leave the channel and move to a new free channel. In this case, the opportunistic user ends up colliding with another incumbent user.

Opportunistic forced termination probability - The opportunistic forced termination probability is the probability of dropping an active opportunistic user call due to the arrival of an incumbent user to a channel occupied by an opportunistic user. When a new incumbent user call arrives at a channel occupied by opportunistic user, the opportunistic user leaves that channel and starts the channel searching process. If the opportunistic user could not find a free channel, its call will be terminated.

Opportunistic successful probability - The opportunistic successful probability denotes the probability that an opportunistic user call is normally terminated (successful call completion).

Opportunistic blocking probability - The opportunistic blocking probability is the probability that a newly arrived opportunistic user call cannot be accepted due to insufficient radio resources, collision with an already existing incumbent user, and inability of the opportunistic user to find free channels. Opportunistic user's calls can be blocked for several reasons depending on the state of the system: when all channels are occupied by incumbent users, state (N,0), or due to miss detection (in such case the incoming opportunistic user collides with an existing incumbent user and the call will be completely lost). In the case where all channels are occupied by opportunistic users, state (0,N), an incoming opportunistic user will immediately be denied service since it knows about existence of other opportunistic by the opportunistic user controller/access point. An opportunistic user call can also be blocked even though all channels are free, state (0,0), if the opportunistic user could not classify any one of them as being free due to a certain P_{FA}.

3.4.3 Results

Extensive simulations using an event-based approach and Poisson arrival processes are performed to validate the analytical model. Results from the analysis and from the simulation are compared and they match very well validating the analysis. Figure 3-35 illustrates the degradation of the opportunistic successful probability (i.e. the probability that an opportunistic call is started and terminated normally) due to the increase in the incumbent arrival rate λ_1. This indicates that at high incumbent arrival rate the channels are more often occupied by incumbent users reducing opportunities for opportunistic users to access the network. It is clear that the opportunistic successful probability degrades quickly for small number of channels (e.g. *N*=3). It can be seen that as expected the success probability goes down when λ_1 increases.

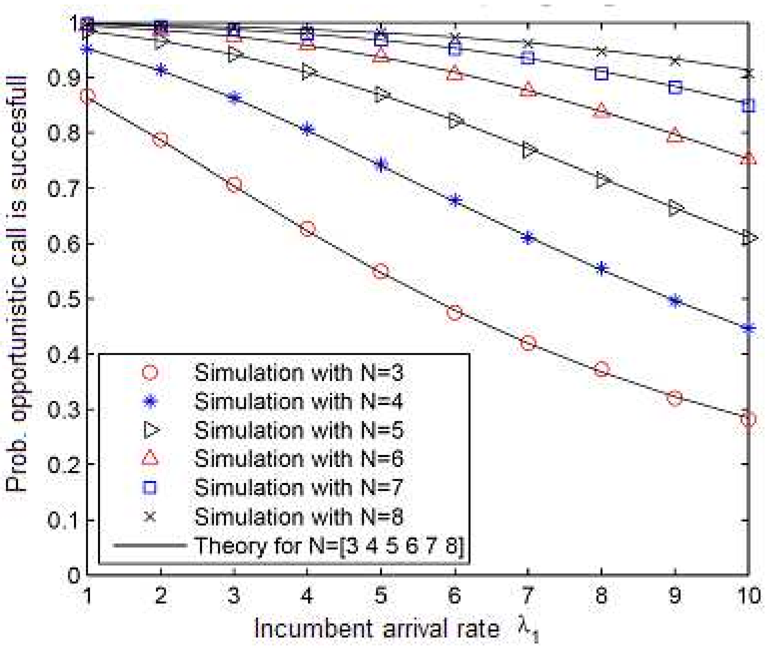


Figure 3-35: Opportunistic successful probability (normally terminated).

Figure 3-36 presents the opportunistic forced termination probability against the probability of detection for different number of channels. As the number of channels increases, the opportunistic forced termination declines. This is because more radio resources are available to handle incumbent calls. Hence the probability that new incumbent calls assigned to channels occupied by opportunistic users is reduced. On the other hand, those opportunistic users who force to handoff their call because they detected arrived incumbent users will more likely find new empty channels. However, as the probability of detection increases, incoming opportunistic detect the presence of incumbent users more accurately, and with P_D=1 opportunistic users will never access the network in the first place and therefore there would be no opportunistic users’ force terminated calls.

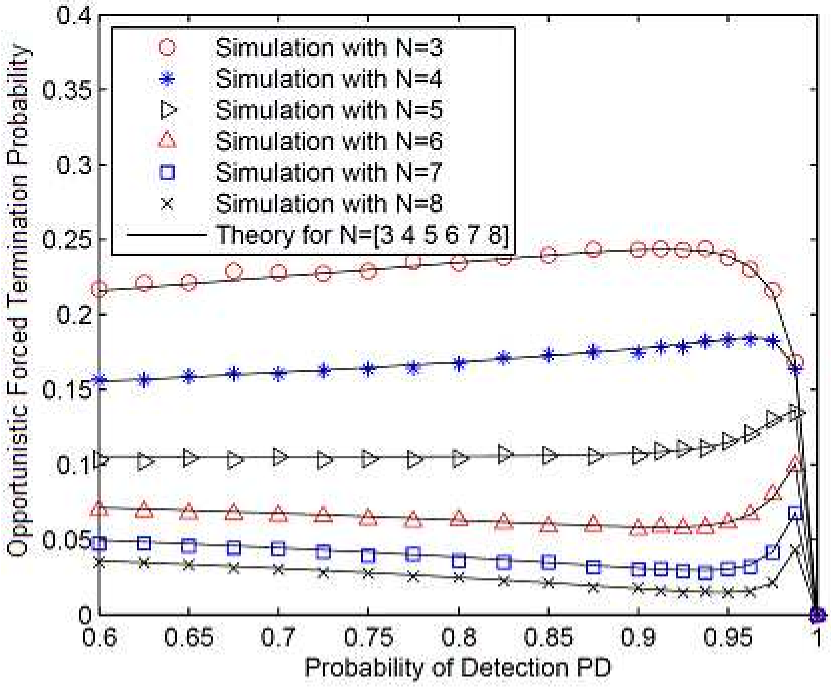


Figure 3-36: Opportunistic termination probability versus probability of detection.

3.5 Malicious users in resource allocation and effects of mobility on sensing

As seen, the robustness to the attack of malicious users is among the desired features. The effects of malicious users on resource allocation and the effects of mobility on spectrum sensing are discussed in the following.

3.5.1 Malicious users in resource allocation

A typical resource in cognitive radio (CR) networks is the bandwidth, for which opportunistic users (OU) are competing between each others. Resources vary over time, frequency and space because of channel variations, mobility of users or fluctuations on wireless traffic. However, it is very important to guarantee some expected performance level to all users. In OU transmission, the goals to achieve include high rate (depends on channel conditions), low power and secure communication. These goals have different weights. For example, video applications need high rate, while in e-mail security is more important. The entity of the damage caused by these attacks depends also on the application.

According to [MuwEtal09], *how*, *when* and *which* CR access to whitespace are key elements in spectrum resource allocation. *How* means that incumbent users (IU) should be able to decide on the transmission rate. *When* implies OU waiting time before going to a whitespace and exit time when they have to leave the white space. *Which* means that IUs should be able to accept/reject the access request. Access to spectrum resources is through opportunistic or negotiation-based methods. The main point is that harmful interference and collisions are not allowed. Game theory can be used to study user’s behaviours, as will be seen shortly below.

Malicious users (MU) may cause serious harm to resource allocation. MUs’ attacks affect in different ways of resource allocation depending on to which part the attacks are targeted to. In spectrum mobility, MUs may fail the handoff process. If a MU pretends to be an IU and forces an OU to change its band, frequency resources are wasted. If MU jams the network, it may totally prevent the transmission in the whole network.

In spectrum management, if MU continuously sends false sensing information, this false information comes part of historical information [WelEtal08]. Thus, historical information that is used in resource allocation becomes incorrect. If historical information is used to help spectrum sensing, the probability of finding free spectrum is decreased. It may happen that its performance is not better than using random selection or its performance may be even worse. Instead, more computation effort and computation time are needed [WelEtal08].

A learning engine does not require pre-programmed policies before starting to operate. Instead, it observes the feedback in order to get the optimal action. It uses both present information and past statistics to predict the future and, thus, to select operations that are optimal. Because it uses long-term experiences, interfering signals caused by MUs may cause also long-term influences. For example, if a MU interferes every time when a faster modulation is used, a learning engine learns that it has to use lower modulation rates always, even though MU interfering signals are not present. Time resources are lost due to too low modulation rates and, thus, slow speed. In the learning phase, a CR observes target radio statistics and tries several input parameters in order to decide which inputs offer best result. If an MU interferes with the channel, data rates are not as high as being without a MUs attack. MUs can also affect to choose non-optimal waveforms, bandwidths, frequencies or higher transmitting power than it is required, if online learning (online optimisation of parameters) is used [Hu11].

In denial of service (DoS) attacks, a MU senses the spectrum and interferes with bands where there is a lot of traffic, i.e., high signal energy, regardless of whose signals (IU/OU) they are. The goal is to maximise the damage when services are denied in those frequency bands. In this kind of attack, many OUs, and possible IUs, are interfered and therefore not able to transmit. Interfered OUs have to go to another band. The risk is that MU may follow where OUs go and interfere also these bands. This corresponds to the so called follow jamming in military communications [Nic88]. This kind of attack lowers communication efficiency and raises synchronisation complexity because of the amount of signalling increases due to requests, channel setups, etc.

In incumbent user emulation (IUE) attack, a MU mimics the IU, thus preventing other OUs to come to the channel (denial of service). Usually, this is done to get unfair benefit, i.e., to get the free space to MUs own usage. Other OUs miss a possibility to share the band, so this kind of attack causes wasting frequencies that could have been used by other OUs. If the IUE attack is performed because of malicious reasons, the goal is to prevent other OUs to use the free space without using the free space itself. In this case, frequency resources are totally wasted, because no-one is using these free frequencies. In both the cases, IUs are not interfered. Instead, the number of OUs that can use the free channels is limited.

MUs may cause harm to spectrum sensing in several ways. MUs may tamper with, flash flood or cheat. In tampering, MU may send false information to the data collector thus leading to false decision. MUs may also cheat and create wrong results. This causes falsely knowledge about environment and leads to ineffective channel allocation, i.e., spectrum underutilisation when all free space is not used or collisions when OUs go to occupied bands. [TanEtal12]

Selfish MUs compete unfairly for resources. Instead of cooperation, selfish MUs want to have as much resources as they need regardless of the harm they cause to other OUs. The benefit that they get is that selfish MUs data rate ray may increase. However, because other OUs loose resources, the total amount of users who get their minimum throughput reduces due to selfish MUs. [NgoEtal11]

In non-cooperative situations, MUs may interfere with other OUs communication preventing the transmitter from learning the states of channels. As a consequence, the interfered OU selects a channel whose transmission quality is poor thus getting a poor data rate. Different non-cooperative games like leader-follower game in the presence of MUs were considered in [TemEtal08]. Due to contradicting requirements, resource management is typically modelled as a non-cooperative game between two parts, the service provider who wants as much users as it is possible and the end user who wants much bandwidth [TemEtal08].].].

MUs may also send signals that mimic IU signals in order to mislead OUs to believe that there is an IU (IUE attack). In this way, MU either uses the free space itself (selfish user), or its purpose is to block the band [WanEtal10].

Several OUs may request transmission slot at the same time. To avoid conflicts and reduce sensing time, available transmission times have to be partitioned. In the case of multiple CR requests, it is possible to give all the free bands to the first CR so that its transmission time is reduced (so called first in, first out, FIFO, system), or to give all the time to the first CR while other OUs are queuing (dynamic system), for example. MUs may use the bands longer time than they are allowed. They may even decline to leave the bands. MUs may transmit some spam which prevents other OUs and also IUs transmission. This leads to long queuing time to other OUs and may cause interference to IUs when preventing their transmission. MUs may also interfere with IU or other OUs transmitting on the same time even though they are not allowed to do that.

Resources are scheduled to OUs. One of the main parts is unoccupied spectrum identification. MUs may cause severe damage by causing erroneous identification of spectrum. If unused spectrum is erroneously classified to be used, resources are lost (in the case of MUs) or resources are not lost but allocated unfair (in the case of selfish MU).

Video quality degradation by malicious users was considered in [ParEtal12]. In 4G systems, MUs may distort the resource allocation at QoS enforcement points if misreporting the parameter values. This leads to suboptimal resource allocation which may give to MUs excessive benefits. Enforcements points include, for example, base stations and service gateways in WiMAX.

Application of game theory

Game theory is used as a help in the study of the effects of malicious users. In overlay and underlay spectrum usage, non-cooperative and congestion game models are suitable. These include, for example, zero-sum/nonzero-sum, cooperative/non-cooperative, congestion and potential games. Usually, Nash equilibrium (balance) is used to measure game models behaviour and stability [Red11].

As cooperative users maximise common benefit, malicious users cause as much damage as possible and selfish users want to maximise their resources to their own benefit. Nash bargaining solution (NBS) is used to achieve optimality in cooperative scenarios. NBS divides the free spectrum to users in a ratio that equals to the rate at which the payoff is transferrable after the users are assigned minimal resources. Cooperative users' group payoff-function advocates their common communication goal. As the payoff-functions of selfish users' illustrate their self-interests, in the case of MUs the payoff-functions describe their damages to dynamic spectrum access networks. For ordinary cooperative users, it defines their common communication goal. [JiEtal07]

The performance of cooperative spectrum sensing is highly depending on the correctness of the local sensing data. MUs may cause that OUs conceive their environment incorrectly. Because of erroneous image of their surrounds, the decision-making process becomes distorted. Thus, the OUs' adaptation can be erroneous. False detection results reported by malicious and selfish users (spectrum sensing data falsification attack, SSDF) lead to inefficient spectrum usage. If occupied spectrum is falsely classified as unoccupied, this causes collisions between IU and OU or between two OUs. If unoccupied spectrum is falsely classified as occupied, free space is lost. As the first one causes interference, the last one cause inefficient spectrum usage which can lead to queuing and rush when opportunity to use unused spectrum is lost. This can be avoided, for example, using AND, OR and majority rules or more effective reputation-based weight methods like weighted sequential probability ratio test (WSPRT) [WanEtal11].

3.5.2 Effects of mobility on sensing

In CR systems, mobility affects network characteristics, e.g., connectivity, capacity, routing and coverage. It is commonly known that in cooperative systems where sensing diversity is used in space domain, gains degrade when the shadow fading correlation among cooperative OUs increases. Mobile

OU may fix this problem. When a mobile OU moves its place and makes measurements in different places, it can be assumed that SNR varies between the places. In some places, SNR may be large as in some other places SNR may be small. Also, the hidden terminal problem in which CR is invisible to other CRs may occur in some places. However, because the mobile OU measures in different places, it is expected that in some places SNR is better than in other ones. In that case, a bad location does not have a high impact to results when there are also better places with larger SNRs included.

Mobile OUs have several properties that differ from stationary OUs and these have to be taken into account. In mobile OUs, which can be assumed to be battery-powered, energy-efficiency is an important aspect. The larger the number of sensing times, the larger the energy consumption is. There is also a detection delay which depends on how many times sensing is performed. The more sensing is performed by each mobile OU, the larger the delay is when deciding if there is an IU signal or not. Also some reporting time is needed during which the sensing report is sent to the base station, and during that reporting time data transmission is interrupted.

In cooperative sensing, one or more OUs can be mobile. According to [MinEtal09], OUs mobility increases spatio-temporal diversity in received IU signal strength when sensing is scheduled several times. This means that the sensing performance is improved. This improvement increases when the speed of OU increases. In [MinEtal09], theoretical analysis as well as numerical results were presented. It was noticed that even with one mobile OU, miss detection performance under cooperative sensing converges to zero. This is because the mobile OU is able to fully exploit spatio-temporal received signal strength (RSS) diversity when moving all around the cell. The trade-off between sensing scheduling and cooperative sensing is that the number of times to sense increases as the number of required sensors decreases.

It is also possible that instead of cooperation when there are several OUs, there is only one, mobile OU that moves its place and makes decision based on its sensing results at different places. In [ArsEtal10], theoretical analysis of energy detection based sensing in the viewpoint of mobility was presented. There was one CR which moved and collected measurements at which were used to decide the presence of IU. It was concluded that mobility of CR is able to improve the sensing performance. Therein, it was also noted that because of scattering of IUs signal, detection performance of CR is better in urban environment when compared to detection in suburban environment. In [ArsEtal10], some errors that occur in [MinEtal09] were noted.

In [VarEtal12], the performance of energy detection-based localisation algorithm based on double-thresholding[5] (LAD) methods in the presence of Doppler and multi-paths were considered. ETSI BRAN/WLAN channel models B and C were studied. Elektrobit (EB) Propsim F8 radio channel emulator was used for realistic channel modelling. There was Doppler corresponding 5 km/h device speed, Rayleigh fading and 17 multipath components. There was 3-8%/0-2% detection loss when compared to AWGN channel, depending on which LAD method (LAD ACC, 2-D LAD ACC) and signal power (-60/-70 dBm) were used. Here, signal powers were fixed, i.e, signal power was the same regardless of the mobility of CR.

Also in mobile CR systems, there may be malicious users. In mobile systems, location of nodes varies, so existing “user trust”-based solutions which are used to define which nodes reports can be taken into account in cooperative sensing cannot be used. In [JanEtal12], detection of MUs for collaborative sensing in mobile cognitive radio networks was studied. Paper [JanEtal12] proposed two new parameters for trust, namely location reliability and malicious intention. As the first one reflects the channel’s channels path loss characteristics, i.e., from which cell the report is generated, the second one captures OUs true intention, i.e., which OU generated the report. The fusion centre evaluates each cell’s trust iteratively using also past reports. To decide which mobile OUs are malicious, Dempster-

[5] The LAD method is a blind, iterative and low-complex outlier detection method that can be used in narrowband signal detection applications like detecting occupied/unoccupied channels [Var10].

Shafer theory [JanEtal12] is used. Fusion centre uses soft-combining technique called equal gain combining to decide if a report was generated by honest OU or malicious user. According to the simulations carried in the aforementioned reference, both MU and IU detection rates were improved. It was also noted that MUs mobility helps their detection. The question which arises was that can MUs exploit mobility to their own advantage and, thus, avoid to be exposed.

FCC specifies two types of devices that are able to use TV white spaces. ‘Mode I’ devices use geo-location spectrum databases and ‘Mode II’ devices do not have to use databases [FC08]. In [MinEtal11], fundamental challenges related to ‘Mode II’ mobile OUs were considered. OUs mobility was studied as for channel availability model, sensing and access strategies. First, OUs mobility affects to spectrum opportunity. In [MinEtal11], two-state continuous-time Markov chain was used to model channel availability. Second, OUs mobility causes challenges when protected IUs from interference caused by OU. It means that, for example, sensing has to perform more often but this causes time overhead. Paper proposes guard distance to solve that problem. Third challenge is that spectrum opportunities are heterogeneous when OUs are mobile. To solve that problem, optimal distributed channel-access strategy was derived in [MinEtal11]. Therein, key factors are spectrum sensing adaptation which is aware about OUs mobility, IU spatial distributions as well as channel-usage patterns, and spectrum sharing among OUs. In [MinEtal11], channel is available for mobile OU when IU is in idle state, or when OU is located outside IUs protection region.

In addition to one or several OUs, also IU may be mobile. When IU is mobile and OUs are stationary, SNRs that OUs measure differs because distance and also channel conditions between IU and all OUs differ. In [CacEtal11], the impact of IUs mobility to sensing was studied. Two mathematical mobility models, namely the random walk mobility model with reflection and random waypoint mobility model, were considered. Both the IU detection capability and mobility-enabled sensing capacities were studied. The latter one describes what OUs expected transmission capacity is when mobile IU exists. According to [CacEtal11], there are five parameters that have influence to the detection capability. These parameters are the size of network region, the mobility model of IU, the protection range of IU, the spatial distribution of CR as well as how much IUs are using the same band. The protection range means the range within OUs have to detect IUs. When considering mobile-enabled sensing capacity, mobility of IUs does increase the sensing capacity even significantly. However, the network region size has to be larger than the IU protection range.

3.6 Optimal decision-making framework for QoS provision under spectrum mobility

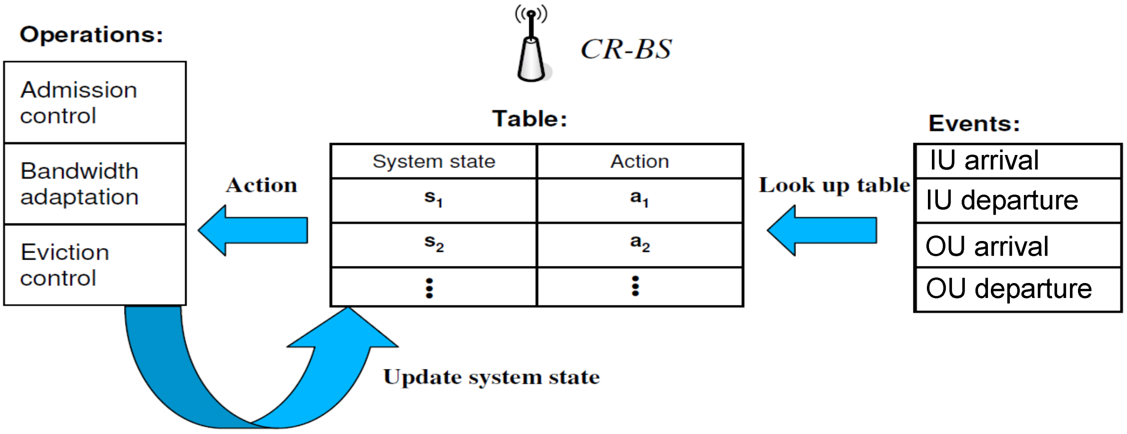
The concept of mobility in QoSMOS, as in cognitive radio networks for opportunistic spectrum use in general, includes spectrum mobility. This section presents optimal strategies for QoS provision under these conditions.

From a user’s perspective, a blocking a new session or a complete termination of an ongoing session is more annoying than compromising the quality of service to a certain extent but still initiating or maintaining the session at an acceptable QoS level. In practice, various technologies have been adopted to adaptively degrade and upgrade the QoS level of a session depending on the varying resource availability. For example, layered encoding schemes have been used for multimedia applications [AVC03]. Depending on the bandwidth availability, different encoding schemes can be adaptively used to keep the session continuity. In a cognitive radio network, bandwidth adaptation mechanism could be especially useful. When few IUs are present, more spectrum bands can be allocated to the OUs to upgrade their QoS levels. When more IUs appear, the OUs may degrade their QoS levels by giving up some spectrum bands but no user is evicted as long as the minimum QoS requirement can be satisfied. In this section, an optimal decision-making framework for joint admission control, eviction control and bandwidth adaptation in cognitive radio networks is proposed. In particular, the problem is formulated as a semi-Markov decision process (SMDP) to derive the optimal decisions for each system state during the lifetime of the system. The objective is to maximise the long-term network revenue as a function of the spectrum utilisation, the OU blocking probability and the bandwidth adaptation cost while keeping the forced dropping probability of the OUs upper-bounded.

3.6.1 System model and decision-making framework

It is assumed that the licensed spectrum consists of M incumbent frequency bands and the capacity (or bandwidth) of each incumbent band is C. The total bandwidths of $M \times C$ are shared by IUs and OUs. The IUshave the priority over the OUs on the utilisation of the spectrum bands and the OUs must vacate the bands when they are reclaimed by the IUs. An IU will use one incumbent band for transmission whereas a OU can use one of the bandwidths among $\{B_1, B_2, \dots, B_K\}$ for transmission, where $B_{\min} = B_1 < B_2 < \dots < B_K = B_{\max}$, based on the spectrum availability. Ideally, an OU will prefer to use the maximum bandwidth for transmission for the best user's quality of experience. However, if this cannot be satisfied due to the lack of spectrum resource, the transmission can still continue as long as the minimum bandwidth requirement can be supported. It is assumed that orthogonal frequency division multiplexing (OFDM) technique is used for OU transmission, and thus, when an incumbent band is reclaimed by an IU, the OUs using any portion of this band can be easily migrated to other bands with idle bandwidths by remapping the OFDM subcarriers. Assuming that the IU and the OU arrival processes follow a Poisson process with arrival rate λ_p and λ_s, respectively. The service time of an IU and an OU follows an exponential distribution with mean $1/\mu_p$ and $1/\mu_s$, respectively. These assumptions have been widely used in the literature and shown to achieve a good balance between the real traffic characteristics and the mathematical tractability.

The decision making framework is presented in Figure 3-37. A state-action mapping table is pre-calculated based on the following SMDP (semi Markov decision process) modelling and is stored in a cognitive radio base station (CR-BS). When a specific event occurs (e.g. an IU arrival), the CR-BS will look up the table with the current system state as the index and find the proper action from the table. The actions may include admission control, bandwidth adaptation and eviction control. After the selected action is implemented, the system state will be updated based on the output of the action.



System state	Action
s_1	a_1
s_2	a_2
⋮	⋮

Figure 3-37: Decision-making framework.

3.6.2 Semi-Markov decision process formulation

The decision-making process for joint admission control, eviction control and bandwidth adaptation is modelled as a SMDP. SMDPs are widely used to model stochastic control problems in dynamic systems satisfying the Markov property. At a decision epoch in a dynamic system, the system is in one of the states in a finite state space. An action is chosen from a finite action space and moves the system to a new state. A corresponding reward/cost is incurred due to the action taken. Markov property means that the time until, the new state at, and the reward/cost incurred thereafter. The reward/cost incurred until the next decision epoch depends only on the current state and the action taken. Given the assumptions above, the process considered here possesses Markov property and can be modelled as an SMDP.

3.6.2.1 System states

At any time instant, the state of the system is represented as **s** = (m, **n**), where m is the number of active IUs in the system, **n** = (n_1, n_2, . . . , n_K), with n_i ($1 \leq i \leq K$) denoting the number of OUs using the bandwidth B_i for transmission. Given the total capacity constraint of the spectrum, the state space is:

$$\mathbf{S} = \{\mathbf{s} : 0 \leq m \leq M, \sum_{i=1}^{K} n_i \leq (M-m)C\} \tag{3-3}$$

3.6.2.2 Actions

Similar to [Xia01],[],[Kim09] and [[Ros89], a decision at a state is made before the occurrence of the next event, i.e. the decision maker should pre-decide the actions for all the possible events at the next decision epoch. At a decision epoch, one of the following four events can occur: (1) an OU arrival; (2) an IU arrival; (3) an OU departure; (4) an IU departure. Thus, an action at a decision epoch can be denoted as:

$$\mathbf{a} = (a_s, a_p, e, \mathbf{b_{sa}}, \mathbf{b_{pa}}, \mathbf{b_{sd}}, \mathbf{b_{pd}}) \tag{3-4}$$

where a_s and a_p stands for the admission control policy for the arrival of a OU and a IU respectively (admit ↔ a_s, a_p = 1; reject ↔ a_s, a_p = 0), e denotes the number of OUs to be evicted when an IU arrives, $\mathbf{b_{sa}}$ denotes the action of bandwidth allocation of the new OU and bandwidth reallocation of the existing OUs when a OU arrives, $\mathbf{b_{pa}}$ denotes the action of OU eviction and bandwidth reallocation when a IU arrives, $\mathbf{b_{sd}}$ and $\mathbf{b_{pd}}$ denote the action of bandwidth reallocation when a OU departs and when a IU departs respectively.

The action a_s = 0 is possible for every state and a_s = 0 is the only possible action when all the spectrum bands have been occupied by the IUs. Thus, we have:

$$a_s = \{a_s \in \{0,1\} : \ a_s = 0 \text{ if } m = M\} \tag{3-5}$$

Since the IUs have the priority over the OUs for spectrum access, the admission control policy for the IUs is deterministic:

$$a_p = \{a_p = 1 \text{ if } m <= M-1, \ a_p = 0 \text{ if } m = M\} \tag{3-6}$$

The action space of e is: $e \in [0, \sum_{i=0}^{K} n_i]$.

The form of $\mathbf{b_{sa}}$ and its action space is defined as:

$$\begin{aligned}
\mathbf{b_{sa}} = \{a_s(b_{sa}^1, b_{sa}^2, \ldots, b_{sa}^K) : \sum_{i=1}^{K} b_{sa}^i = 1,\\
- n_i \le b_{sa}^i \le \lfloor \frac{(M-m)C}{B_i} \rfloor - n_i \text{ for } 1 \le i \le K,\\
\sum_{i=1}^{K} B_i(n_i + b_{sa}^i) \le (M-m)C,\\
\text{if } \exists j \ (1 \le j \le K, b_{sa}^j < 0) \text{ then } b_{sa}^i \le 0 \text{ for } \forall i > j\}
\end{aligned} \tag{3-7}$$

where b_{sa}^i $(1 \le i \le K)$ denotes the variation of the number of the OUs using the bandwidth B_i when a new OU arrives. The first constraint indicates that if an OU is admitted the total number of OUs in the system should be increased by 1. The second constraint gives the range of the OU number variation. The third one is the total system capacity constraint. In order to admit the new OU, some existing OUs in the system may need to give up some spectrum resources they are currently using. However, the vacated spectrum should not be used by other OUs to upgrade their QoS level. And the bandwidth allocated to the new OU should not exceed the bandwidth allocated to any existing OU. This requirement is enforced by the last constraint.

Similarly, the form of $\mathbf{b_{pa}}$ and its action space is defined as:

$$\begin{aligned}
\mathbf{b_{pa}} = \{a_p(b_{pa}^1, b_{pa}^2, \ldots, b_{pa}^K) : \sum_{i=1}^{K} b_{pa}^i = -e,\\
- n_i \le b_{pa}^i \le \lfloor \frac{(M-m-1)C}{B_i} \rfloor - n_i \text{ for } 1 \le i \le K,\\
\sum_{i=1}^{K} B_i(n_i + b_{pa}^i) \le (M-m-1)C,\\
\text{if } \exists j \ (1 \le j \le K, b_{pa}^j < 0) \text{ then } b_{pa}^i \le 0 \text{ for } \forall i > j\}.
\end{aligned} \tag{3-8}$$

where b_{pa}^i $(1 \le i \le K)$ denotes the variation of the number of the OUs using the bandwidth B_i when a new IU arrives. The first constraint indicates that the total number of the evicted OUs is equal to e. The rest constraints have the similar meaning as above.

The form of $\mathbf{b_{pd}}$ and its action space is defined as:

$$\begin{aligned}
\mathbf{b_{pd}} = \{(b_{pd}^1, b_{pd}^2, \ldots, b_{pd}^K) : \sum_{i=1}^{K} b_{pd}^i = 0,\\
- n_i \le b_{pd}^i \le \lfloor \frac{(M-m+1)C}{B_i} \rfloor - n_i \text{ for } 1 \le i \le K,\\
\sum_{i=1}^{K} B_i(n_i + b_{pd}^i) \le (M-m+1)C,\\
\text{if } \exists j \ (1 \le j \le K, b_{pd}^j > 0) \text{ then } b_{pd}^i \ge 0 \text{ for } \forall i > j\}.
\end{aligned} \tag{3-9}$$

where b_{pd}^i $(1 \le i \le K)$ denotes the variation of the number of the OUs using the bandwidth B_i when an IU departs. The first constraint indicates that the total number of the OUs in the system is unchanged. The second and third constraints are similar to the above. The last constraint limits that only the spectrum released by the departed IU should be reallocated among the OUs.

The operation of bandwidth reallocation upon anan OU departure event depends on the amount of the spectrum released by the departed OU. Thus, $\mathbf{b_{sd}}$ is further defined as $\mathbf{b_{sd}} = (\mathbf{b^1_{sd}}, \mathbf{b^2_{sd}}, \ldots, \mathbf{b^K_{sd}})$, where $\mathbf{b^k_{sd}}$ $(1 \leq k \leq K)$ denotes the bandwidth reallocation operation when a OU using the bandwidth B_k departs. Let $\mathbf{n'_k} = (n'_{1,k}, n'_{2,k}, \ldots, n'_{K,k})$ denote the new state of the number of OUs using different bandwidths after a OU using the bandwidth B_k departs. We have $n'_{i,k} = n_i - 1$ for $i = k$ and $n'_{i,k} = n_i$ for i $\neq$ k. Similar to the definition of $\mathbf{b_{pd}}$, the form of $\mathbf{b^k_{sd}}$ and its action space is defined as:

$$
\begin{aligned}
\mathbf{b^k_{sd}} = \{(b^{1,k}_{sd}, b^{2,k}_{sd}, \ldots, b^{K,k}_{sd}) : & \sum_{i=1}^{K} b^{i,k}_{sd} = 0, \\
& -n'_{i,k} \leq b^{i,k}_{sd} \leq \lfloor \frac{(M-m)C}{B_i} \rfloor - n'_{i,k} \text{ for } 1 \leq i \leq K, \\
& \sum_{i=1}^{K} B_i(n'_{i,k} + b^{i,k}_{sd}) \leq (M-m)C, \\
& \text{if } \exists j \ (1 \leq j \leq K, b^{j,k}_{sd} > 0) \text{ then } b^{i,k}_{sd} \geq 0 \text{ for } \forall i > j\}.
\end{aligned}
\tag{3-10}
$$

It is worth noting that the actions above do not need to consider how to adapt the bandwidth for each individual OU, thus leaving the flexibility to the algorithm design. More fine-tuned bandwidth allocation algorithms can be developed on top of the actions defined in this paper to achieve their specific objectives. For instance, the throughput fairness among OUs can be enforced by adopting an algorithm to select which OU to upgrade or degrade its QoS level in each round.

3.6.2.3 State transition probability

Let $\tau_{\mathbf{s}}(\mathbf{a})$ be the expected sojourn time in state $\mathbf{s}$ when action $\mathbf{a}$ is chosen. The exponential distribution of the inter-arrival and service time of the IUs and OUs yields:

$$
\tau_{\mathbf{s}}(\mathbf{a}) = (a_s \lambda_s + a_p \lambda_p + (\sum_{i=1}^{K} n_i)\mu_s + m\mu_p)^{-1} \tag{3-11}
$$

The probability that the system wills transit to the state $\mathbf{s'} = (m', \mathbf{n'})$ at the next decision epoch given the current state of the system is $\mathbf{s} = (m, \mathbf{n})$ is:

$$
p_{\mathbf{s},\mathbf{s'}}(\mathbf{a}) = \begin{cases}
a_s \lambda_s \tau_{\mathbf{s}}(\mathbf{a}), & \mathbf{n'} = \mathbf{n} + \mathbf{b_{sa}},\ m' = m, \\
 & \text{(OU arrival)} \\
a_p \lambda_p \tau_{\mathbf{s}}(\mathbf{a}), & \mathbf{n'} = \mathbf{n} + \mathbf{b_{pa}},\ m' = m + 1, \\
 & \text{(IU arrival)} \\
n_k \mu_s \tau_{\mathbf{s}}(\mathbf{a}), & \mathbf{n'} = \mathbf{n'_k} + \mathbf{b_{sd}},\ m' = m, \\
 & \text{(OU with } B_k \text{ departure)} \\
m \mu_p \tau_{\mathbf{s}}(\mathbf{a}), & \mathbf{n'} = \mathbf{n} + \mathbf{b_{pd}},\ m' = m - 1, \\
 & \text{(IU departure)} \\
0 & \text{(otherwise)}
\end{cases}
\tag{3-12}
$$

3.6.2.4 Reward and cost

Different actions chosen in a state will result in different rewards and costs. On one hand, the reward earned by the network operator is generally proportional to the number of the admitted OUs and the bandwidths allocated to them. On the other hand, bandwidth adaptation operations require extra signalling overhead for service level and radio access level negotiation, incurring the non-trivial operational cost. Whereas the detailed definition of the cost function depends on the specific network environment, one general intuition is that the cost should be proportional to the number of bandwidth adaptation operations. Let $r_{\mathbf{s}}(\mathbf{a})$ and $c_{\mathbf{s}}(\mathbf{a})$ denote the earned reward and the incurred cost until the next decision epoch given action **a** is taken at state **s**, respectively. $r_{\mathbf{s}}(\mathbf{a})$ is defined as:

$$r_{\mathbf{s}}(\mathbf{a}) = \sum_{i=1}^{K} B_k n_k \tau_{\mathbf{s}}(\mathbf{a}) + \gamma a_s \lambda_s \tau_{\mathbf{s}}(\mathbf{a}) \qquad \text{(3-13)}$$

where the first term models the reward generated from the spectrum utilisation over time and the second term models a one-time reward earned by admitting a new OU with an arrival probability $\lambda_s \tau_{\mathbf{s}}(\mathbf{a})$. γ is a weight factor. $c_{\mathbf{s}}(\mathbf{a})$ is given as:

$$c_{\mathbf{s}}(\mathbf{a}) = c_b N_{\mathbf{s}}(\mathbf{a}) \qquad \text{(3-14)}$$

where c_b denotes the cost of one bandwidth adaptation operation and $N_{\mathbf{s}}(\mathbf{a})$ denotes the expected number of bandwidth adaptation operations when action **a** is taken in state **s**, given by:

$$\begin{aligned} N_{\mathbf{s}}(\mathbf{a}) = {} & a_s \lambda_s \tau_{\mathbf{s}}(\mathbf{a}) \sum_{i=1}^{K} I(b_{sa}^{i} > 0) b_{sa}^{i} \\ & + a_p \lambda_p \tau_{\mathbf{s}}(\mathbf{a}) \sum_{i=1}^{K} I(b_{pa}^{i} > 0) b_{pa}^{i} \\ & + \sum_{k=1}^{K} \left(n_k \mu_s \tau_{\mathbf{s}}(\mathbf{a}) \sum_{i=1}^{K} I(b_{sd}^{i,k} > 0) b_{sd}^{i,k} \right) \\ & + m \mu_p \tau_{\mathbf{s}}(\mathbf{a}) \sum_{i=1}^{K} I(b_{pd}^{i} > 0) b_{pd}^{i}, \end{aligned} \qquad \text{(3-15)}$$

where $I(\cdot)$ is an indicator ($I(\cdot) = 1$, if condition in $(\cdot)$ is satisfied; $I(\cdot) = 0$, otherwise). Note that the cost of the eviction operations are not considered here but will be considered as a constraint in the optimisation framework in the following section.

3.6.3 Optimal strategy for QoS provision

To find the optimal action for each possible system state to maximise the long-term network revenue as a function of the reward and cost while keeping the forced dropping probability of the OUs upper-bounded, we formulate a linear programming algorithm [Ros89]:

Maximize revenue:

$$\sum_{\mathbf{s}\in\mathbf{S}}\sum_{\mathbf{a}\in\mathbf{A}(\mathbf{s})}(r_{\mathbf{s}}(\mathbf{a})-c_{\mathbf{s}}(\mathbf{a}))z_{\mathbf{s},\mathbf{a}},$$

Subject to:

$$\begin{aligned}
&\sum_{\mathbf{s}\in\mathbf{S}}\sum_{\mathbf{a}\in\mathbf{A}(\mathbf{s})}\tau_{\mathbf{s}}(\mathbf{a})z_{\mathbf{s},\mathbf{a}}=1,\\
&\sum_{\mathbf{a}\in\mathbf{A}(\mathbf{s}')}z_{\mathbf{s},\mathbf{a}}-\sum_{\mathbf{s}\in\mathbf{S}}\sum_{\mathbf{a}\in\mathbf{A}(\mathbf{s})}p_{\mathbf{s},\mathbf{s}'}(\mathbf{a})z_{\mathbf{s},\mathbf{a}},\ \forall\mathbf{s}'\in\mathbf{S},\\
&\sum_{\mathbf{s}\in\mathbf{S}}\sum_{\mathbf{a}\in\mathbf{A}(\mathbf{s})}ea_p\lambda_p\tau_{\mathbf{s}}(\mathbf{a})z_{\mathbf{s},\mathbf{a}}\\
&\qquad\le P_d^{Bound}\sum_{\mathbf{s}\in\mathbf{S}}\sum_{\mathbf{a}\in\mathbf{A}(\mathbf{s})}a_s\lambda_s\tau_{\mathbf{s}}(\mathbf{a})z_{\mathbf{s},\mathbf{a}},
\end{aligned}\tag{3-16}$$

where variable $z_{\mathbf{s},\mathbf{a}} \geq 0$ can be explained as the long run fraction of the decision epochs at which the system is in state s and action a is chosen. The first two constraints represent the normalisation and balance equation, respectively. The forced dropping probability of the OUs P_d is expressed as:

$$P_d=\frac{\sum_{\mathbf{s}\in\mathbf{S}}\sum_{\mathbf{a}\in\mathbf{A}(\mathbf{s})}ea_p\lambda_p\tau_{\mathbf{s}}(\mathbf{a})z_{\mathbf{s},\mathbf{a}}}{\sum_{\mathbf{s}\in\mathbf{S}}\sum_{\mathbf{a}\in\mathbf{A}(\mathbf{s})}a_s\lambda_s\tau_{\mathbf{s}}(\mathbf{a})z_{\mathbf{s},\mathbf{a}}}\tag{3-17}$$

Enforcing an upper bound p_d^{Bound} for P_d yields the last constraint.

3.6.4 Performance evaluation

In this section, the performance of the proposed strategy based on the SMDP (referred as `SMDP_aBA`) is evaluated and compared with three other schemes: threshold-based channel reservation scheme without bandwidth adaptation (referred as `Threshold_maxBA`), threshold-based channel reservation scheme with bandwidth adaptation (referred as `Threshold_aBA`), and the optimal strategy based on the SMDP but without bandwidth adaptation (referred as `SMDP_maxBA`). In both schemes without bandwidth adaptation, the maximum bandwidth requirement of each OU needs to be satisfied all the time. The operation of `Threshold_aBA` is summarised as follows: when a new OU arrives, the existing OUs subsequently degrade their QoS levels to the next lower level to accommodate the new OU if needed and the new OU is rejected if all the OUs have degraded their QoS levels to the lowest level but the remaining spectrum bandwidth is still less than a threshold; when an IU arrives, if all the OUs have degraded their QoS levels to the lowest level but still cannot find enough spectrum bandwidth to support all of them. Portion or all of the OUs will be evicted subsequently until the remaining OUs can be supported. When an OU or an IU departs, the OUs will subsequently upgrade their QoS levels to the next higher level until approaching the threshold. For fair comparison, the number of bandwidth adaptation operations is counted only based on the initial QoS level and the final QoS level. The parameters fixed in evaluation are listed in Table 3-2. The GLPK tool [GLPK] is used to solve the linear programme in this work.

Table 3-2: Experimental parameters.

M	3	λ_s	0.5
C	3	μ_s	0.5
K	2	μ_p	0.1
B_1	1	P_d^{Bound}	1%
B_2	2	γ	1.0

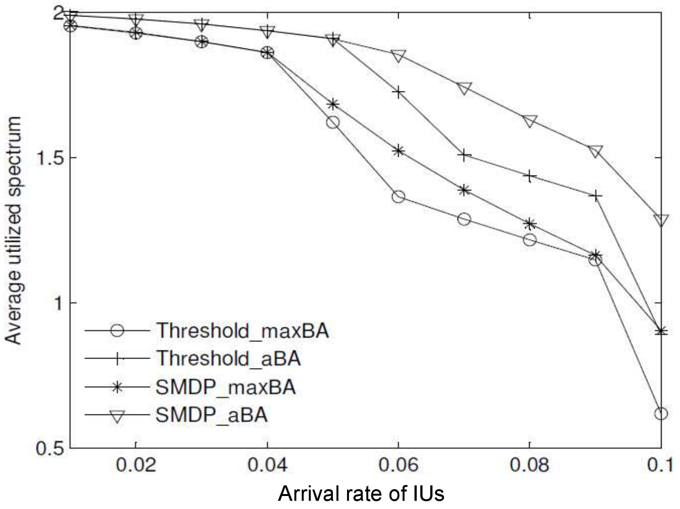


Figure 3-38: Average utilised spectrum vs. arrival rate of IUs.

To fully exploit the benefit of adaptive bandwidth allocation, first it is set the cost of one bandwidth adaptation operation c_b = 0 and then evaluated the average utilised spectrum and OU blocking probability with respect to varying IU arrival rates. For all the schemes, the forced dropping probability of the OUs is required to be below the pre-defined bound. For the threshold-based schemes, the results are derived when the optimal threshold is used for each IU arrival rate. As shown in Figure 3-38 and Figure 3-39, bandwidth adaptation can significantly improve the spectrum utilisation and reduce the OU blocking probability.

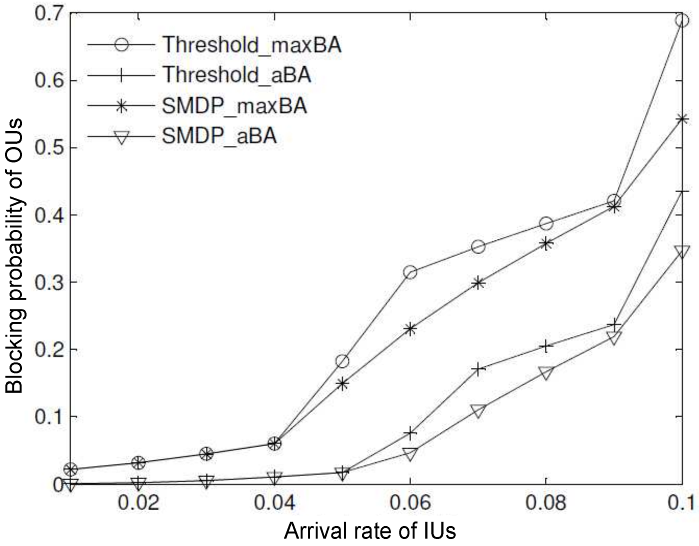


Figure 3-39: Blocking probability of OUs vs. arrival rate of IUs.

The SMDP-based schemes always outperform or at least perform equal to the corresponding threshold-based schemes with varying IU arrival rates. That is due to the fact that the threshold-based schemes cannot adapt the decision for each specific system state.

We then set the IU arrival rate to λ_p = 0.05 and evaluate the impact of the cost per bandwidth adaptation on the average network revenue. As shown in Figure 3-40, the proposed scheme `SMDP_aBA` can adapt to the cost per bandwidth adaptation and decide whether or not a bandwidth adaptation is adopted with the objective to maximise the long-term network revenue. On the other side, `Threshold_aBA` cannot take the bandwidth adaptation cost into account. Its performance drops quickly with the increase of the cost and can be even worse than the schemes without bandwidth adaptation.

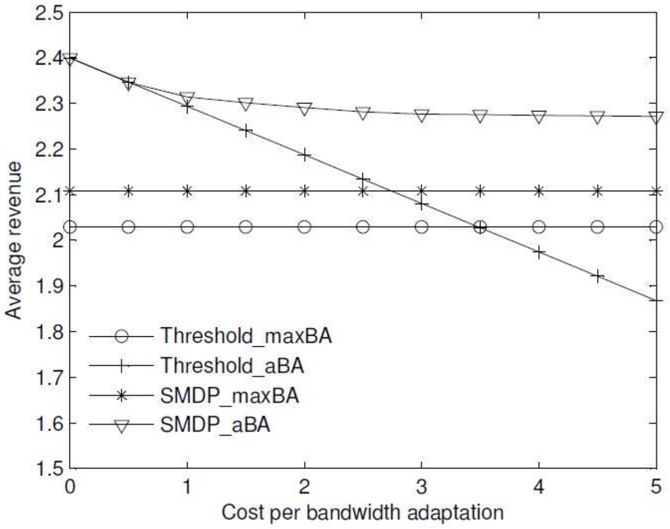


Figure 3-40: Average revenue vs. cost per bandwidth adaptation.

3.6.5 Remark

The optimal decision-making framework for joint admission control, eviction control and bandwidth adaptation to support the QoS of the OUs under spectrum mobility in cognitive radio networks proposed in this section formulates the problem as a SMDP solved via a linear programming-algorithm, where the optimal decision at each system state is derived to maximise the long-term network revenue. Compared to the state-of-the-art schemes, the proposed scheme is shown to improve the spectrum utilisation and reduce the OU blocking probability while keeping the forced dropping probability of the OUs upper-bounded. Furthermore, it is shown that the proposed scheme can adapt to the bandwidth adaptation cost.

4 Resource management solutions for the target scenarios

As reminded in section 2.1.2, a number of requirements [ManEtal2011] [D1.4] drive the design of resource management. As a consequence of the different requirements and constraints set, the resource management should possibly take into account those specificities by the different scenarios [MacEtal2011]. For example, lower frequency whitespace spectrum would be ideal for rural broadband access and coverage enhancement in the cellular extension scenario, as well as for the inside to outside coverage in the cognitive femtocell scenario, due to better through-wall propagation characteristics. On the other hand, higher frequency whitespace spectrum could be preferred for capacity enhancement in the cellular extension scenario, as well as for indoor cognitive femtocells, due to intrinsic better insulation towards the outdoors (other femtocells or macrocells). Similar considerations apply to the cognitive ad hoc network scenario, depending on the specific case.

This chapter presents resource management solutions for all the three main target scenarios of QoSMOS outlined in section 2.1.1: the cellular extension, the cognitive femtocell and the cognitive ad hoc network.

4.1 Cognitive access control for the cellular extension scenario

In this section, it is studied an optimised algorithm for controlling the access to the cells operating in an opportunistic band. The purpose is to prevent network congestion while preserving the QoS of the accepted connections and protecting the incumbent operations. To attain this objective, it has been identified a solution that takes eviction decisions which can be used in the admission control to prevent critical impacts generated by the incumbents' presence.

4.1.1 Introduction

Although demand for high data-rate services grows exponentially, users of heterogeneous wireless networks continue to expect an optimal quality of service depending on their application and service agreement. Due to users' mobility and limited system resources, it is then a complex challenge for network operators to increase their revenue by maximising the number of users, while maintaining a certain degree of satisfaction.

Access control (AC) is a key element in the provision of QoS in conventional networks, as it aims at avoiding network congestion while preserving the committed QoS of already accepted calls. It can even restrict the number of ongoing connection in the system or degrade some user's QoS to satisfy a majority of ongoing user's requirements. However, the decision process is not simple as many factors, sometimes contradictory, have to be taken in consideration, and it becomes even more complex in a context of cognitive radio.

CR is the technical answer from the wireless community to solve the problems resulting from the spectrum underutilisation and the dearth of spectrum bands: an opportunistic system is authorised to transmit in the same frequencies than an incumbent system if minimal interferences are guaranteed. One possible orientation for AC may consist in developing schemes that defines the interference generated onto the incumbent system as the admission criterion. Another one could focus on schemes that are charged to maximise the number of admitted users, taking benefits from the presence of opportunistic resources. Because it addresses both QoS and mobility aspects, the second approach has been privileged in this study.

The literature provides then a plethora of solutions and [Hos05] brings an interesting illustration of the different schemes that could be implemented in a network. Among them, the guard-band policy is used by AC schemes to control the connection dropping/blocking rates. It consists in holding some number of the network resources reserved for handoff connections while the rest of the resources are used for new connections: the threshold can be dynamically set depending on the network/cell status, as illustrated in Figure 4-1.

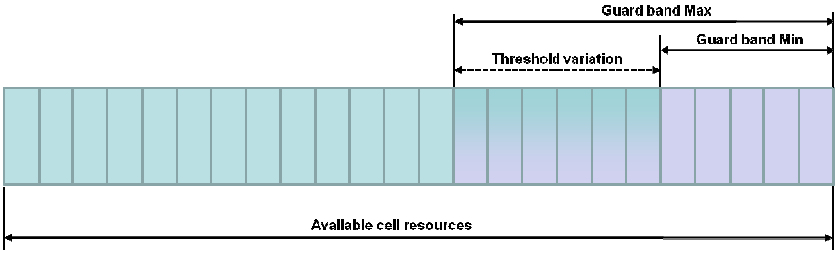


Figure 4-1: Guard band policy for access control.

Because of the preservation of incumbent users, the scarce and highly variable resources of opportunistic systems, and the high users' mobility encountered in wireless networks, access control represents a challenging research item. And since it controls directly the radio resources, the guard band policy is most appropriate scheme to address the challenges formed by the cognitive radio concepts.

Indeed, in such CR network, the frequent apparition of the incumbents may force the system to take eviction decisions that could impact the overall QoS of the opportunistic system. And, forcing UEs that are localised at the edge of the impacted cell(s) to handoff to the adjacent cells is a key measure for maintaining the service to these UEs while vacating the spectrum channels.

However, radio resources reserved in adjacent cells by the access controllers may not be enough to support this sudden increase of handover requests, which may be consequently rejected.

The connection dropping rate should increase significantly upon incumbent apparition and the connection blocking rate should be degraded also afterwards, as the system may not be able to accept new connections, due to the limitations of the cell capacity.

4.1.2 Solution for optimising the access control in opportunistic cellular system

The proposed solution consists in enhancing the access control mechanism deployed in conventional network with cognitive abilities in order to control the impacts of the incumbent's apparition. Thus, the decision-making algorithm will take, in addition to the admission decisions, a set of evictions measures that will be motivated by its observation of the network environment and driven by policies adapted to the context. These measures can be classified, as following, depending on their impact on the QoS:

- (1) forcing the handover of UEs to neighbour cells;
- (2) forcing the handover of UEs to another band (licensed, for instance);
- (3) reconfiguring the base station to operate in a backup channel;
- (4) excluding pre-empted resources from resources allocation patterns;
- (5) reducing the QoS of the connected UEs by allocated less radio resources;
- (6) dropping UE connections.

The main objective of this solution is then to take jointly and in the most efficient way admission and evictions decisions. This process will be done in reactive mode, or in preventive mode depending on the availability of the incumbent detection information.

In the example illustrated in Figure 4-2, the estimation of the additional forced handovers triggered by the incumbent presence (done in BS#1) should benefit to the admission control (in BS#2) which may be able to adapt its guard band to minimise or cancel the degradation of the system QoS and guarantee mobility support.

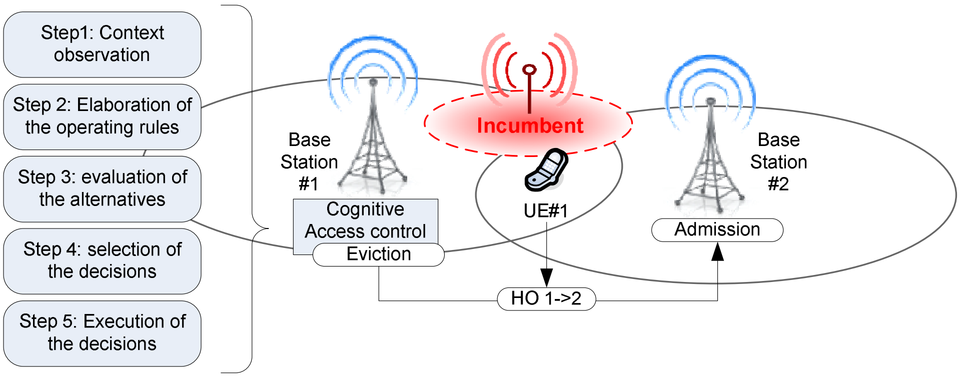

Step1: Context observation
Step 2: Elaboration of the operating rules
Step 3: evaluation of the alternatives
Step 4: selection of the decisions
Step 5: Execution of the decisions
Base Station #1
Cognitive Access control
Eviction
Incumbent
UE#1
Base Station #2
Admission
HO 1->2

Figure 4-2: Illustration of the "cognitive access control" concept.

4.1.3 Solution's assessment

Access control mechanisms are implemented in each base station and their decisions have a direct impact on the behaviour of their adjacent cells: as depicted in Figure 4-3, all base stations, which are under the coverage of the incumbent, may take eviction decisions that impact their adjacent cells and in parallel should support the decisions done by those adjacent cells in their own operations.

It is then challenging in this context to select the set of decisions to be taken in all cells that provide the best protection of the incumbent(s) while optimising the QoS of each established connections. Then, a system approach shall be considered to assess the solution and validate that the system is stable and able to converge to an optimum.

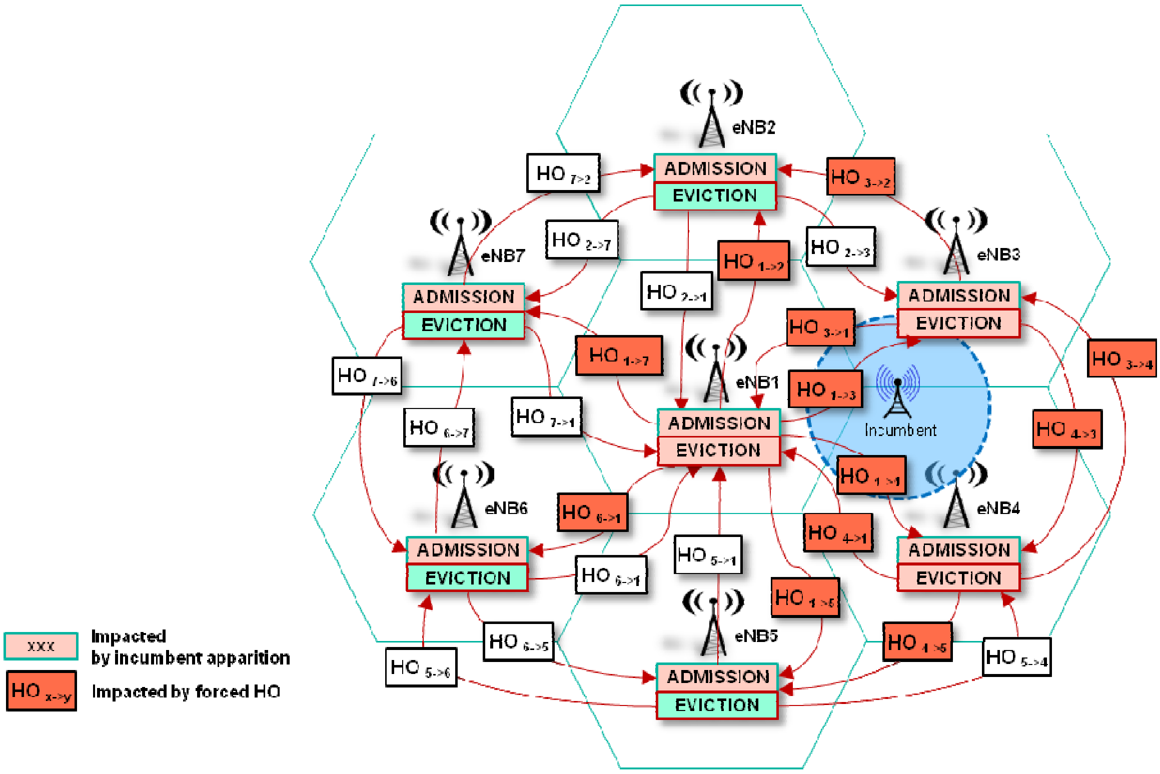


Figure 4-3: System approach.

In addition, one QoSMOS assumption states that the fluctuation of the opportunistic resources should be transparent to the end-users. In consequence, the observed quality of experience should be roughly the same that the one measured in already deployed networks (i.e., 3G).

To address this statement, an ARCEP survey [Arc11] is used as reference. It evaluates the QoS provided by French cellular networks (2G and 3G) and some results, described in Table 4-1, can be exploited as reference to assess QoSMOS solutions. The metrics that have been defined are relative to the service performance, and do not take in account the deployment of the network, nor the used equipments. It consists in the rate of successfully established communications, which have been maintained (RECM) during two and five minutes. A communication is considered as successful if the request has been accepted at the first attempt and if the communication is not disconnected. The rate is calculated based on the total number of requests and the complement consists in the rejection rate after an attempt plus the disconnection rate.

The measurements have been done in major cities, in high speed trains, on highways and also in suburb trains. Some information relative to data transfer (FTP service, smartphone) is also available in the survey but is out of scope of the current analysis.

Table 4-1: QoE metrics (extracted from [Arc11].

metric	city	highway	high speed train	suburb train
RECM>2mn	95,9%	91,1%	69,5%	82,8%
RECM>5mn	94,6%	---	---	---

To assess the previously described solution, the two following network-relative performance metrics shall be considered.

- The connection blocking rate (CBR), which measures the rate of rejected service requests within the system. This metric considers the number of events that occurs in the whole system and is expressed in percentage:

$$CBR = 100 \times \frac{\sum_i NB\_COMM\_REJ}{\sum_i NB\_COMM\_REQ} \tag{4-1}$$

- The connection dropping rate (CDR) measures the rate of established communications that have been prematurely interrupted, whatever the cause:

$$CDR = 100 \times \frac{\sum_i NB\_COMM\_DROPPED}{\sum_i NB\_COMM\_ACCEPTED} \tag{4-2}$$

In addition, the next service-relative performance metric defined by the ARCEP should enable QoE comparison between opportunistic and conventional networks.

- The rate of successfully established communications that have been maintained at least during two minutes, denoted by RECM$_{>2mn}$, is expressed in percentage as:

$$RCEM_{>2mn} = 100 \times \left(1 - \frac{\sum NB\_COMM\_REJ + \sum NB\_COMM\_DROPPED_{Before\_2mn}}{\sum NB\_COMM\_REQ} \right) \tag{4-3}$$

4.1.4 Algorithm description

The cognitive access control (CAC) algorithm firstly collects context information from a local source (e.g. spectrum measurements results provided by sensing sensors) and from a global database such as the common portfolio database. In parallel, the algorithm monitors the UE status as well as the base stations' metrics.

This information is then analysed and any incumbent apparition is logged in order to identify afterwards an eventual periodicity in the incumbent operations (through learning techniques). This input is used by the solution to orientate the action(s) to be taken: preventive action trigs the planning of eviction measures while reactive action forces the system to take decisions for responding immediately to the apparition of the incumbent.

The decision-making algorithm characterises the detected incumbent and lists the different possibilities it can apply to mitigate the impacts. These alternatives are evaluated based on their impacts on the QoS which are pondered depending on the observed context and potentially refined with learning techniques. This problem can be simplified as follows.

- A base station allocates a number of physical resources blocks w_j to a list of connected users u_i (with i ϵ [1, …, N]).
- When an incumbent appears, it pre-empts a spectrum bandwidth W_G from the initial capacity of the cell/sector, managed by the base station, which corresponds to W_N physical resources blocks (W_G encompasses the guard bands, as required by the regulation).
- To compensate the impact of the incumbent presence, the base station has to take a series of eviction decisions D_k (with k ϵ [1, …, M]) on the users u_i , each of them releasing w_j resources and generating a QoS preservation value $V(u,k)$.
- Decisions D_k are prioritised with a weight p_k according to the environment context, the operator policy and the appeared incumbent parameters. Moreover, a condition c_k ϵ {0, 1} is applied to each decision D_k for each user u_i in order to characterise its validity.
- Then, the objective of the problem consists in choosing the set of users u_i that releases at least W_N physical resources blocks, such that the group of eviction decisions D_k applied to these users generates the maximum QoS preservation $Z(u,k)$.

Figure 4-4 illustrates the design of the solution, which can be divided in three steps: the elaboration of the operating rules, the execution of these rules to prioritise both UEs and eviction decisions and the selection of the decisions' set.

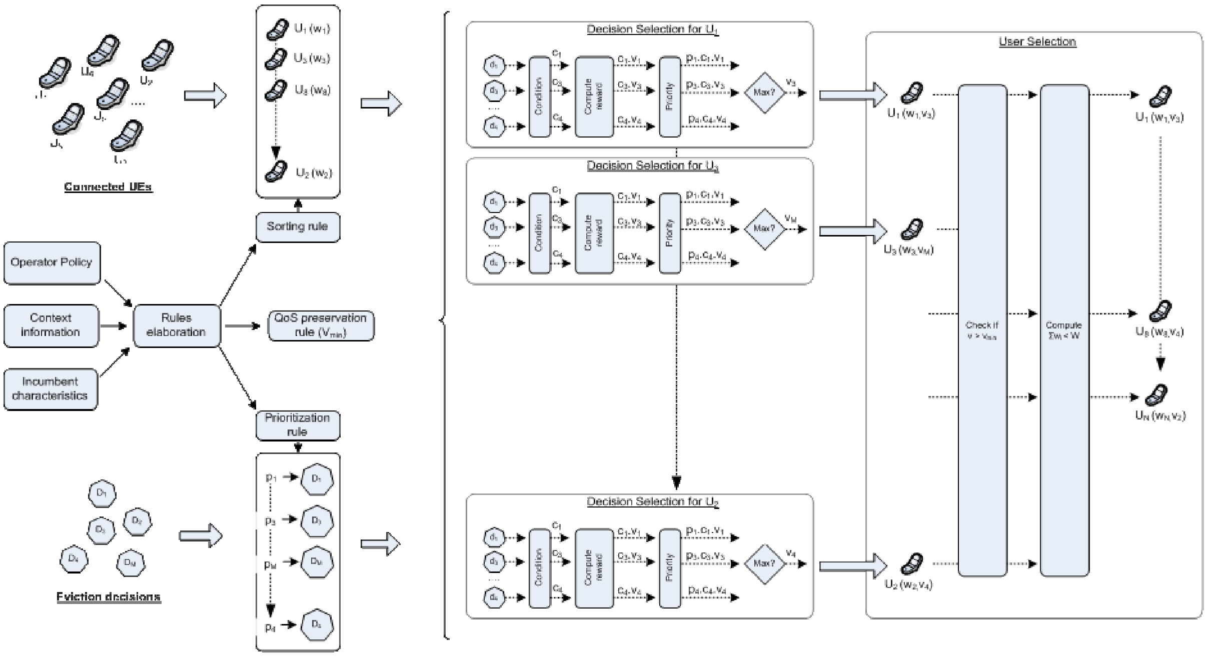


Figure 4-4: Overview of the access control algorithm.

The elaboration of the prioritisation rule is based on the observation of the connection blocking rate (CBR_{MEAS}) and the connection dropping rate (CDR_{MEAS}) as well as on the detection or the prediction of the incumbent (incumbent bandwidth size). These measurements are then compared to the policies defined by the operator (CBR_{MAX} and CDR_{MAX}) to prioritise the decisions (D_k) that address the most the observed context. As an example, Table 4-2 defines a set of weights (p_k) to be used by the decision-making algorithm when evaluating the various alternatives.

In addition, to determine which user equipment has to undergo an eviction decision, several sorting rules are considered, but one is prioritised according to the environment context. Table 4-3 presents an example with two sorting rules: the first one is based on the number of allocated physical resource blocks, and the second one on standardised information specifying if the connection can be pre-empted.

Table 4-2: Example of weights for eviction measures.

policies	D1	D2	D3	D4	D5	D6
$CBR_{MEAS} > CBR_{MAX}$ & $CDR_{MEAS} > CDR_{MAX}$ & INC_{SIZE} = HIGH	10	0	0	10	0	1
$CBR_{MEAS} > CBR_{MAX}$ & $CDR_{MEAS} < CDR_{MAX}$ & INC_{SIZE} = HIGH	10	0	5	5	0	5
$CBR_{MEAS} > CBR_{MAX}$ & $CDR_{MEAS} > CDR_{MAX}$ & INC_{SIZE} = LOW	5	5	5	10	5	1
$CBR_{MEAS} > CBR_{MAX}$ & $CDR_{MEAS} < CDR_{MAX}$ & INC_{SIZE} = LOW	5	0	10	10	5	5
$CBR_{MEAS} < CBR_{MAX}$ & $CDR_{MEAS} > CDR_{MAX}$ & INC_{SIZE} = HIGH	5	5	10	10	5	5
$CBR_{MEAS} < CBR_{MAX}$ & $CDR_{MEAS} < CDR_{MAX}$ & INC_{SIZE} = HIGH	10	0	5	10	0	5
$CBR_{MEAS} < CBR_{MAX}$ & $CDR_{MEAS} > CDR_{MAX}$ & INC_{SIZE} = LOW	5	0	5	10	0	5
$CBR_{MEAS} < CBR_{MAX}$ & $CDR_{MEAS} < CDR_{MAX}$ & INC_{SIZE} = LOW	0	10	5	10	1	1

[D1] Backup channel activation; [D2] Resources re-allocation; [D3] Intra-cell handover; [D4] Inter-cell handover; [D5] QoS degradation; [D6] Connection dropping

Table 4-3: Example of rules for UE sorting (PRB = physical resource block).

Policies	UE sorting rules
$CBR_{MEAS} > CBR_{MAX}$ & $CDR_{MEAS} > CDR_{MAX}$ & INC_{SIZE} = HIGH	Number of allocated PRB
$CBR_{MEAS} > CBR_{MAX}$ & $CDR_{MEAS} < CDR_{MAX}$ & INC_{SIZE} = HIGH	Number of allocated PRB
$CBR_{MEAS} > CBR_{MAX}$ & $CDR_{MEAS} > CDR_{MAX}$ & INC_{SIZE} = LOW	ARP / Priority level
$CBR_{MEAS} > CBR_{MAX}$ & $CDR_{MEAS} < CDR_{MAX}$ & INC_{SIZE} = LOW	ARP / Priority level
$CBR_{MEAS} < CBR_{MAX}$ & $CDR_{MEAS} > CDR_{MAX}$ & INC_{SIZE} = HIGH	Number of allocated PRB
$CBR_{MEAS} < CBR_{MAX}$ & $CDR_{MEAS} < CDR_{MAX}$ & INC_{SIZE} = HIGH	Number of allocated PRB
$CBR_{MEAS} < CBR_{MAX}$ & $CDR_{MEAS} > CDR_{MAX}$ & INC_{SIZE} = LOW	ARP / Priority level
$CBR_{MEAS} < CBR_{MAX}$ & $CDR_{MEAS} < CDR_{MAX}$ & INC_{SIZE} = LOW	ARP / Priority level

Decisions are then evaluated for each user equipment based on their applicability (C_k) and their QoS preservation value *V(u,k)*.

- C_k is conditioned by the environment context: as an example, it is set to null for “intra cell handover” if the UE is under the coverage of a unique cell. In addition, “backup channel activation” is possible only if the system has successfully identified candidate channels.
- *V(u,k)* is calculated by comparing the guaranteed bit rate defined in the QoS class with the user throughput estimated based on SINR measurements.

$$V(u,k) = \frac{(User\_Throughput)_{u,k}}{GBR_u} \tag{4-4}$$

- A QoS preservation policy (V_{MIN}) can be set by the system to exclude the decisions that would impact to much the QoS.

The cognitive access control algorithm previously described contributes to fulfil the WP5 objectives on two aspects:

- by focusing on the CBR and the CDR, the cognitive access control algorithm maintains the QoS of the connected UE within a cell operating with opportunistic resources and aims at providing the same level of QoE than any conventional network;
- in addition, the utilisation of guard band scheme in the admission control guarantees that handovers, and so physical mobility, remain operational even in the presence of an incumbent.

This algorithm is part of the CM-RM QoSMOS entity and its functionalities can be split as presented in the Figure 4-5.

The CM-RM NC is charged to observe the inputs relative to the spectrum environment and to react to any incumbent presence. In parallel, it maintains an internal database that logs the event which is used in parallel to predict any upcoming incumbent apparition.

The CM-RM RA, when triggered by CM-RM NC, gets the network information (CDR, CBR …) from the NET COORD and elaborates the operating rules for the decision-making. It evaluates the impacts of the incumbent and identifies the set of possible eviction decisions. Then, it evaluates the alternatives and selects the set of measures that preserve the most the QoS. Finally, it executes the decisions to protect the incumbent operation.

The CM-RM RA is charged to control the admission of the new connections notified by the upper layers through CM-RM RS, but also to preserve the mobility in the system by reserving resources for handover requests (from CM-RM NC). The admission policy is then updated according to the decisions taken by the CM-RM RA.

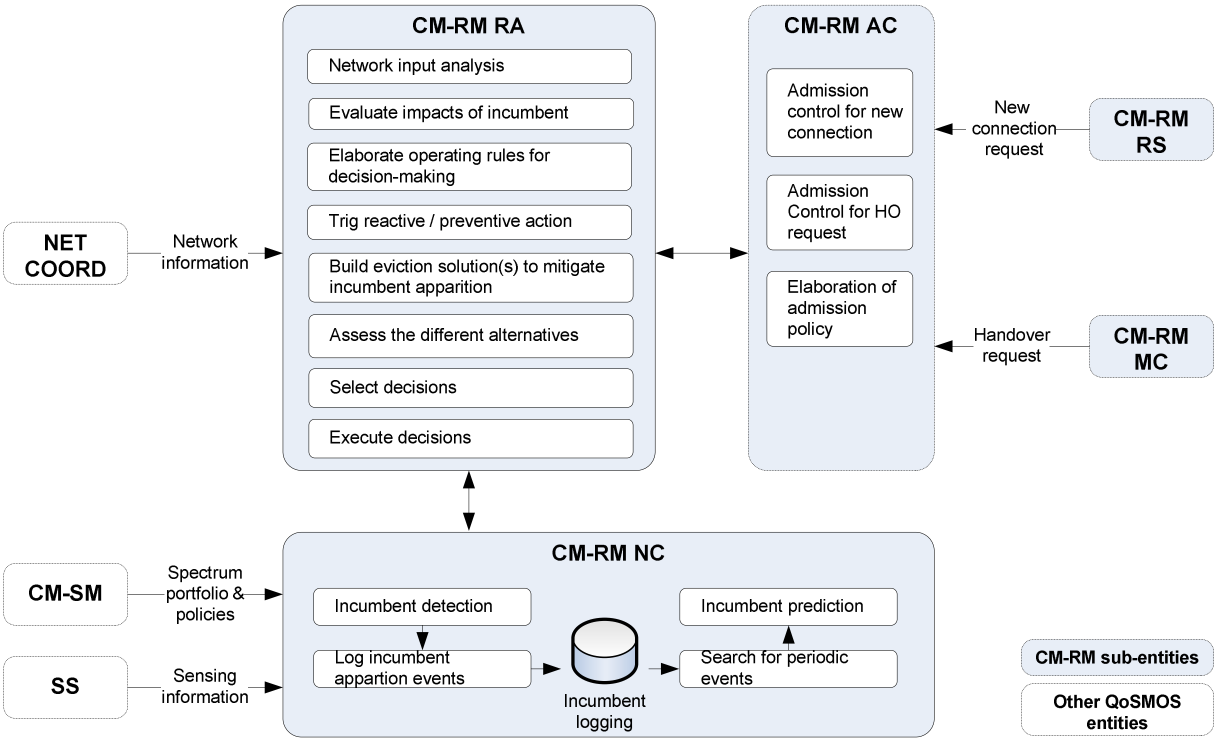

CM-RM RA
Network input analysis
Evaluate impacts of incumbent
Elaborate operating rules for decision-making
Trig reactive / preventive action
Build eviction solution(s) to mitigate incumbent apparition
Assess the different alternatives
Select decisions
Execute decisions
CM-RM AC
Admission control for new connection
Admission Control for HO request
Elaboration of admission policy
New connection request
CM-RM RS
Handover request
CM-RM MC
NET COORD
Network information
CM-RM NC
Incumbent detection
Log incumbent apparition events
Incumbent logging
Incumbent prediction
Search for periodic events
CM-SM
Spectrum portfolio & policies
SS
Sensing information
CM-RM sub-entities
Other QoSMOS entities

Figure 4-5: Cognitive access control in QoSMOS context.

4.1.5 System model

The system model encompasses several base stations operating in an urban area and controlling for each of them three different sectors.

A set of user equipments is deployed uniformly in the simulated region, moving at 5 km/h according to a straightaway direction, randomly defined.

These user equipments connect to the base station which provides the best quality signal and initiate data communication requests according to a Poisson statistical distribution. These requests are then analysed by the admission control algorithm implemented in each sector of the base station which takes the decision to accept or reject the connection depending on the cell load and other cognitive information. These decisions are measured with two performance metrics previously introduced: the CBR and the CDR.

Three cases have been considered: in the first one, no incumbent is present and the system operates as any conventional LTE network. In the second use case, an incumbent appears and covers one sector of cell 5. In the third one, this incumbent impacts seven sectors as illustrated in Figure 4-6.

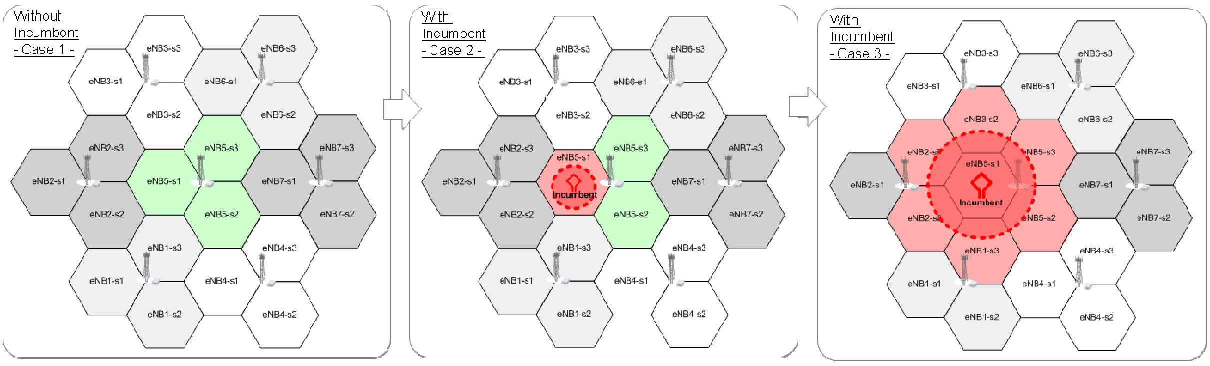


Figure 4-6: System model cases.

4.1.6 Simulation results

4.1.6.1 Observation of the impacts relative to the incumbent apparition

The objective of this analysis is to characterise the impact of the incumbent apparition on the system performance and to quantify the benefits of the CAC algorithm.

The two metrics CBR and CDR have been observed during the simulation time and the following figures illustrate the obtained results for the different cases.

- The CDR and CBR are expressed in percentage;
- The time is expressed in units of 10^5 ms;
- At T_{INC} = T0 + 9 minutes, the incumbent appears (its presence is highlighted by the usage of a grey background). It covers seven sectors (case 3) and has a bandwidth size of 3 MHz.

Figure 4-7 illustrates the case where no incumbent is present and corresponds more or less to the behaviour of a conventional network. This figure represents in consequence the reference to be used to assess the impacts of the incumbent, and also the objective to achieve by the cognitive AC algorithm.

The figure shows that the CDR is quite stable (with a mean value of 5%) while the CBR progresses according to the system load. This demonstrates the behaviour of the admission control which prioritises the handover requests at the expense of the new connection requests, and ensures a dynamic support of the physical mobility within the opportunistic system.

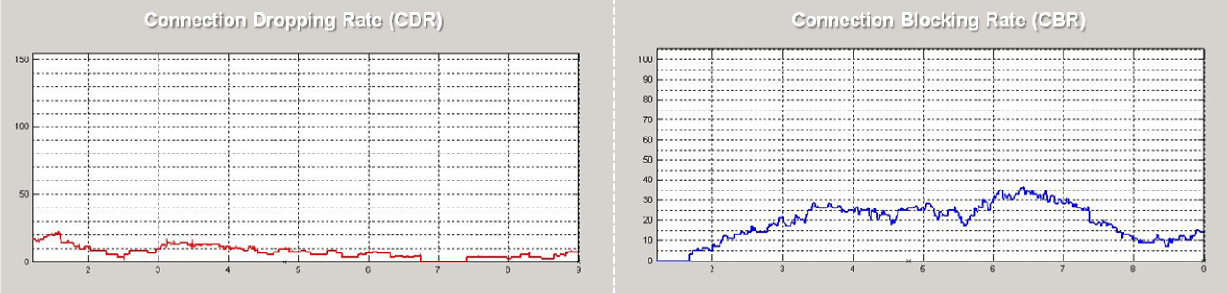


Figure 4-7: CDR/CBR vs time – No incumbent.

The apparition of the incumbent is activated in Figure 4-8, and its impacts are perceptible on both graphs. For CDR, the system has to drop connections to release the radio resources pre-empted by the incumbent: this is materialised by the high peak present at T_{INC} (+ 55% in this context). In parallel, it shall also reject any new connection requests to avoid degrading the QoS even more. This is represented by an amplification of the CBR just after the incumbent apparition.

Moreover, the CDR peak shape seems to indicate that even if connections have been dropped, the system it is still over-loaded and has no other choice than rejecting handover requests (CDR continues augmenting after the incumbent apparition), which challenges the capacity of the network to support mobility in the impacted cells.

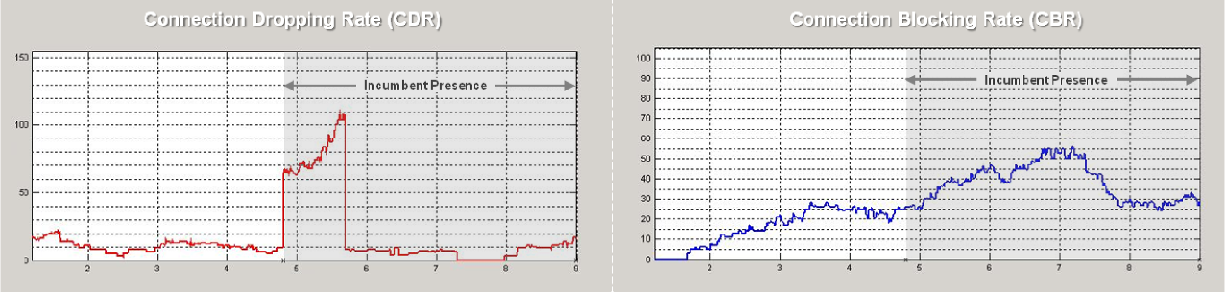


Figure 4-8: CDR/CBR vs time – With incumbent[6].

The cognitive AC algorithm is then activated in Figure 4-9 in order to mitigate the incumbent's impacts. In comparison to Figure 4-8, some improvements can be observed: the system has dropped fewer connections thanks to the eviction control algorithm which has taken "QoS-safe" decisions (for example, intra-cell handover). This has released additional radio resources that have been used by the system to improve the support of mobility in the impacted cells.

Furthermore, a CBR improvement is also noticeable: the decisions taken by the algorithm has facilitated a better distribution of the load among the sectors, which enables the acceptation of additional connections.

[6] The presence of peaks in CDR/CBR graphics is explained by the definition of the metrics which are calculated based on incremented counters. Then, to compare their values between a time T and a time T+1, it is necessary to apply a timing reference, and so to define a sliding timing window. The width of these peaks corresponds to the sliding window value (90 seconds in this configuration). The reason why the CDR goes back to its minimum value is that the incumbent apparition event is no longer taken into consideration in the CDR/CBR calculation.

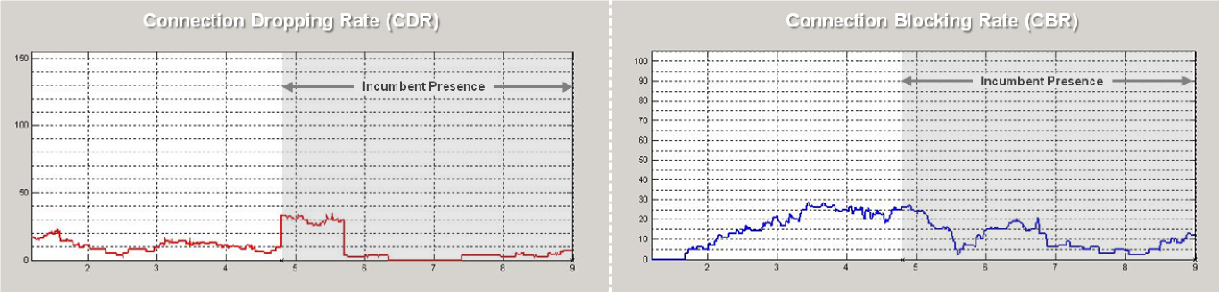


Figure 4-9: CDR/CBR vs time – Activation of the cognitive AC algorithm in reactive mode.

Figure 4-10 represents the evolution of CDR/CBR when the cognitive AC algorithm is configured in preventive mode. This mode consists in predicting the presence of the incumbent for enabling the execution of preventive “QoS-safe” measures before the incumbent appears. In this simulation, the algorithm is aware of the predicting information six minutes before the incumbent apparition, which allows the system to plan five preventive operations triggered every minutes.

Figure 4-10 demonstrates that this mode cancels entirely the impacts of the incumbent apparition, contrary to the reactive mode. Nevertheless, a limited degradation of the CBR/CDR can be expected during the timing period when preventive measures are taken: this degradation is not perceptible for the configuration of this simulation, as the system has found enough candidate UEs to be moved to the neighbour cells.

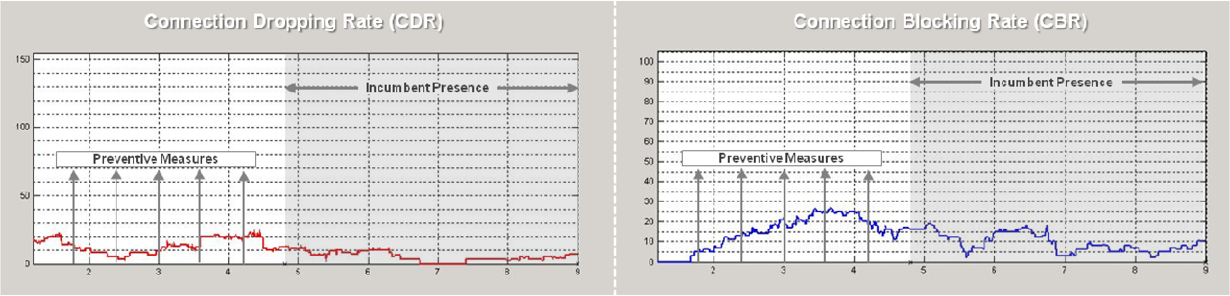


Figure 4-10: CDR/CBR vs time – Activation of the cognitive AC algorithm in preventive mode.

This analysis has characterised the effects of the incumbent apparition on both the CDR and the CBR. But it has principally proved the capacity of the algorithm to cancel these impacts and to contribute to the preservation of the user mobility within the opportunistic system.

However, further investigations are required to address the two following issues. The outcomes are detailed in the subsequent sections.

- Is the algorithm able to support incumbents with higher bandwidth size?
- Has the algorithm the same performance if the prediction window size is reduced?

4.1.6.2 Effect of the incumbent bandwidth size

The objective of this analysis is to characterise the influence of the incumbent bandwidth size onto the system performance (CBR and CDR) for the cases where the incumbent covers one or seven sectors, as described in Figure 4-6. The no-incumbent case has been directly integrated as reference in the graphics.

The metrics have been measured upon incumbent’s apparition (T_{INC}) in order to include in the evaluation the connections that have to be dropped to release the required number of pre-empted radio resources.

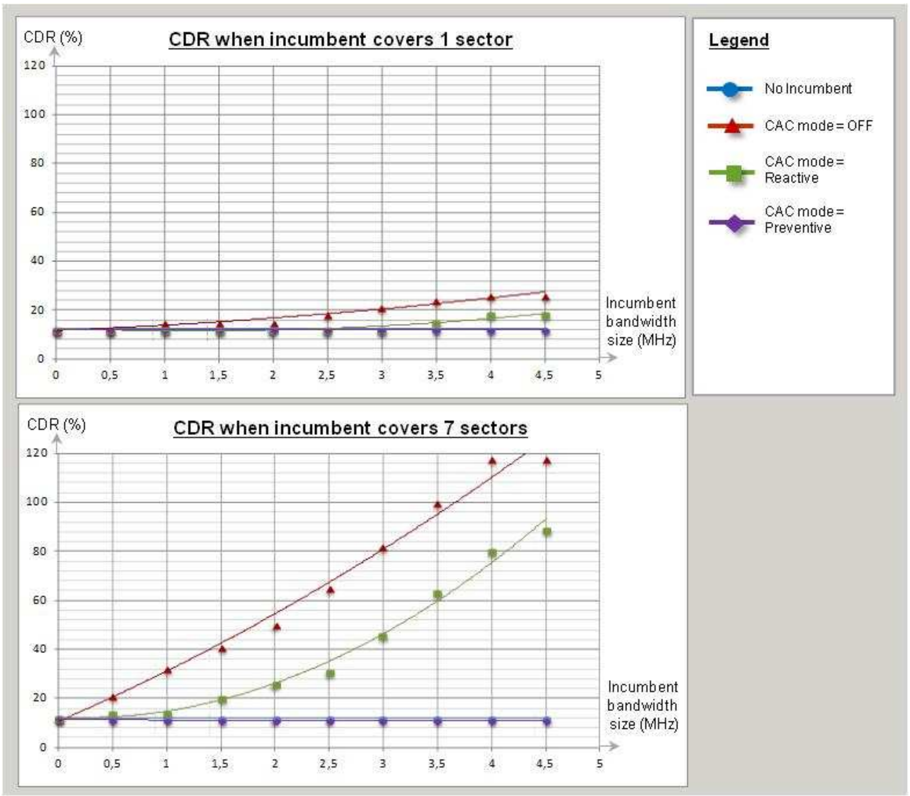


Figure 4-11: CDR vs incumbent bandwidth size (at T_{INC}).

Figure 4-11 shows the significant impact caused by the incumbent apparition onto the CDR: as an example, for a 3 MHz-bandwidth incumbent, the CDR jumps from 12% to 20% for the case 2, and worse from 12% to 80% for the case 3. This simulation clearly confirms the need of a corrective algorithm in order to ensure to the user a similar experience than in a conventional network.

This figure also presents the action of the cognitive access control solution in both reactive (square marks) and preventive (diamond marks) modes and demonstrates an evident benefit compared to the “no-algorithm” case (triangle marks). By considering the previous example,

- The CDR falls, in reactive mode, from 20% to 15% for case 2, and from 80% to 48% for case 3;
- And, in preventive mode, it falls from 20% to 0% for case 2, and from 80% to 12% for case 3.

The simulation also shows the advantages of the preventive mode, compared to the reactive mode, and highlights an optimal efficiency for cases where the incumbent coverage overlaps several sectors and/or for cases where the incumbent pre-empts a large portion of the radio resources used by the opportunistic system.

Figure 4-12 presents the impacts of the incumbent apparition onto the CBR at T_{INC}: no evolution is observed mainly because the effect on the new connections is not immediate. As illustrated in Figure 4-8, the CBR continues to increase after the incumbent apparition.

Figure 4-13 shows the same metrics observed at T_{INC} + 2mn and highlights better noticeable incumbent impacts (the CBR jumps, for a 3 MHz-bandwidth incumbent, from 21% to 27% for the case 3) and solution’s benefits.

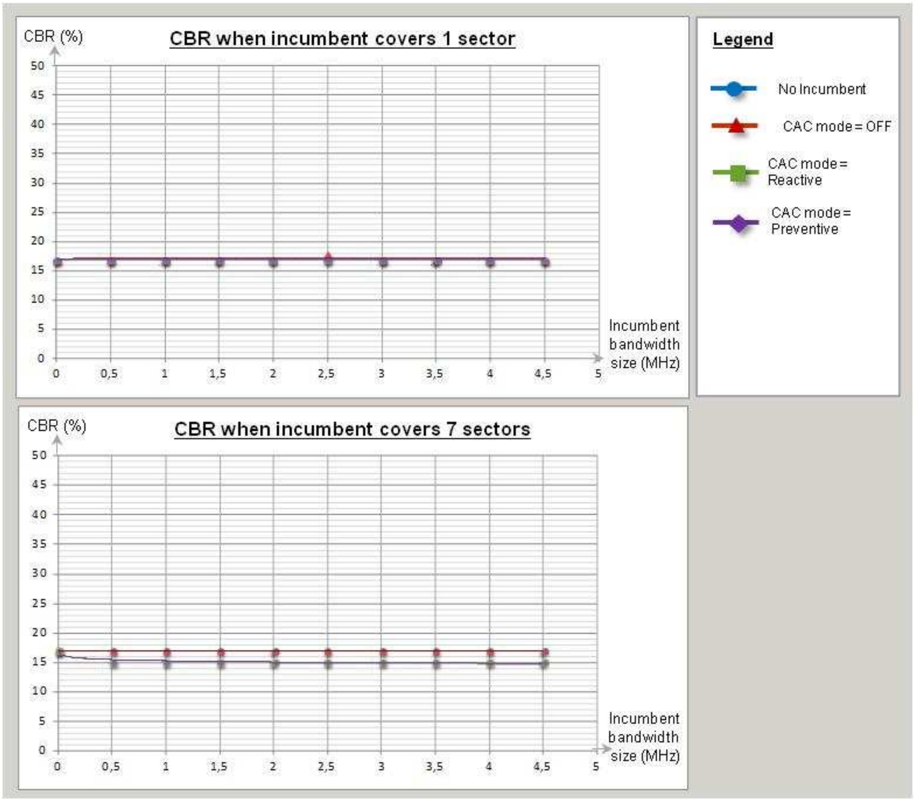


Figure 4-12: CBR vs incumbent size (at T_{INC}).

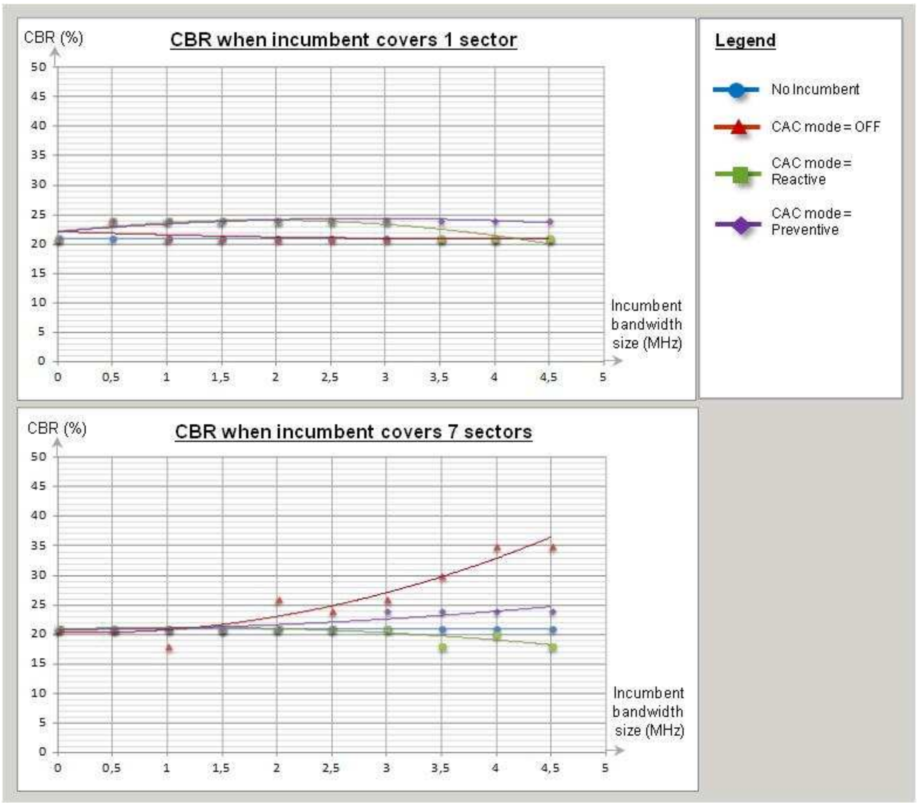


Figure 4-13: CBR vs incumbent bandwidth size (at T_{INC} + 2 mn).

4.1.6.3 Variation of the prediction window size

The objective of this analysis is to identify if the variation of the prediction window size has an effect onto the efficiency of the cognitive AC algorithm. Figure 4-11 has demonstrated that the reactive mode is not sufficient in some cases, and complementary preventive decisions can contribute to improve the system response to the incumbent apparition. Indeed, this mode aims at multiplying the number of UEs that can support QoS-safe eviction decisions. However, reducing the size of the prediction window limits this number of candidates and the risk of dropping UEs upon incumbent apparition mathematically increases.

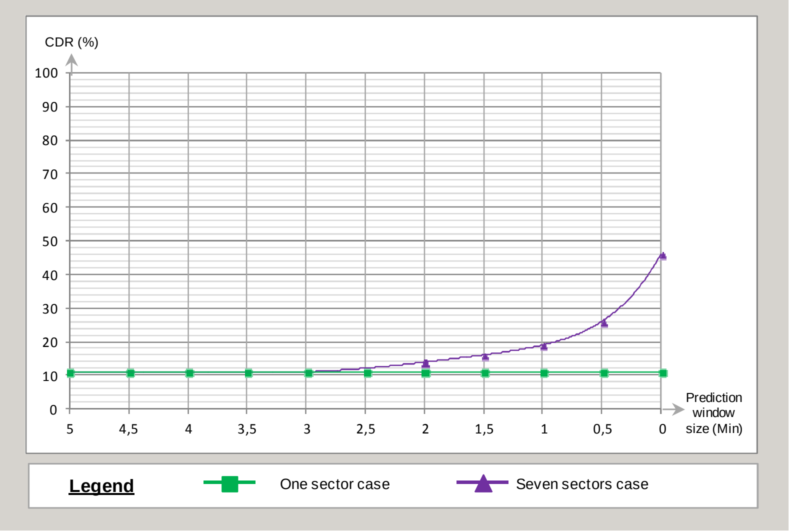


Figure 4-14: Effect of the prediction window size onto the CDR.

Figure 4-14 represents the impacts of this variation onto the CDR and shows that this effect is limited even when the incumbent is predicted one minute before its apparition. This is due to the capacity of the algorithm to adapt its preventive decisions to the window size: the shorter the prediction window is, the rarer the prediction occasions are. But in compensation, the number of eviction decisions per occasion increases.

Then, in this simulation configuration, the algorithm has found enough candidates to perform the required amount of QoS-safe decisions and so, to limit the impacts on the CDR.

4.1.6.4 Comparison with user expectations

As introduced in previous sections, ARCEP has quantified the user expectation regarding the access to the cellular networks. To evaluate the performance of the provided service, $\text{RECM}_{>2mn}$ metric has been used and merges in a certain way both CBR and CDR information.

Figure 4-15 draws the evolution of $\text{RECM}_{>2mn}$ for connections maintained at least two minutes, depending on the incumbent bandwidth size.

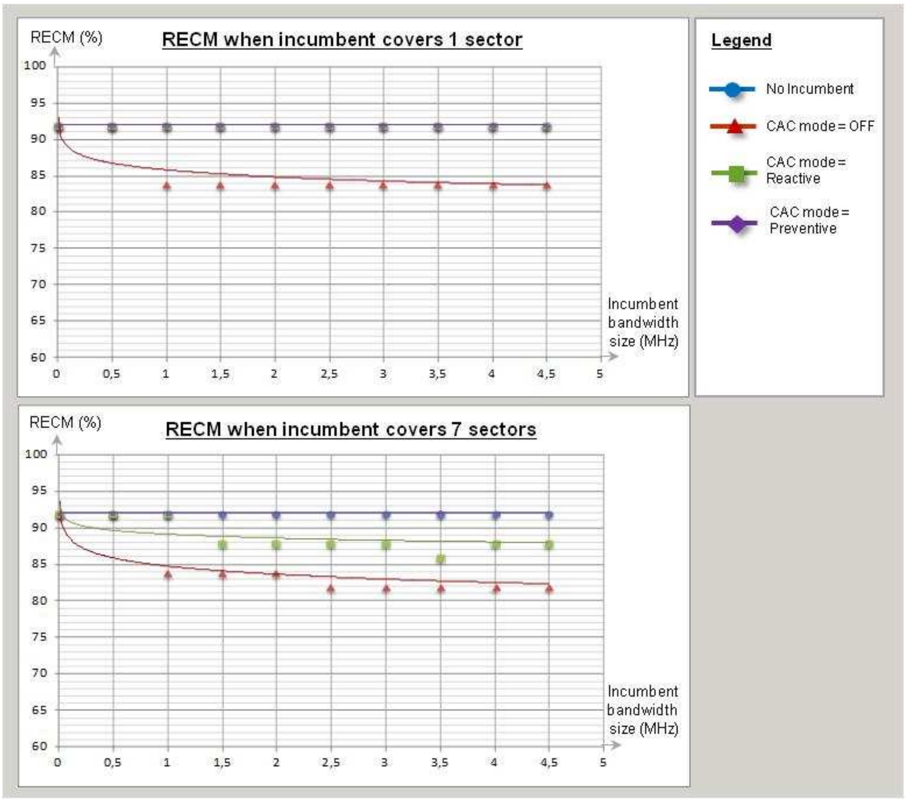


Figure 4-15: RECM>2mn vs incumbent bandwidth size.

“No incumbent” case is represented by the “circle marks” with a constant value of 92%. This is slightly different than the ARCEP measurement, but the small variation can be explained by the semi-realistic configuration used in the simulation platform.

The impacts due to incumbent presence are characterised by the “triangle” marks and the benefits of the cognitive access control algorithm by the “square” marks and “diamond” marks.

The simulation’s results demonstrate that the algorithm is efficient enough to cancel the impacts of the incumbent whatever the considered bandwidth. It also shows that, when preventive actions can be taken, the quality of experience is preserved even in cases where the incumbent coverage overlaps several sectors.

4.1.7 Remark

This section has introduced a new algorithm adding cognitive capabilities to the access control of a LTE cellular network in order to enable its operations in the TVWS band by using opportunistic radio resources. Its design has been described as well as the method for assessing its benefits. Then, simulation results have been provided which lead to the following conclusions:

- the study has demonstrated that the incumbent apparition has a major impact on the system performance (this has been highlighted with CBR/CDR observation), both on the QoS and mobility aspects;
- it has proved the benefits of the solution by reducing/cancelling the observed impacts;
- it has highlighted the advantage to predict the incumbent apparition, even if the window size for taking the preventive actions is short: the algorithm is able to adapt to this situation;
- at last, it has shown the quality of experience for accessing to the service is preserved in comparison to conventional networks.

4.2 Interference management for femtocells

Over the past decade, the demand for higher capacity and data rates has been on the increase. A number of technologies and standards have been developed to cope with this increasing demand i.e. high speed packet access (HSPA), LTE, LTE advanced and WiMAX. These technologies and others have been developed to provide speedy communications to end users. However, with the increasing demand for indoor cellular service, mobile operators simply cannot effectively provide good quality coverage to indoor users given the high in-building penetration loss. A survey [ABI2007] shows that in the near future more than 50% of voice and 70% of data traffic is expected to originate from indoor users. The need for providing good quality indoor voice and data services can therefore not be overemphasised.

Solutions that readily come to mind is for operators to build more base station (BS) sites, which can be prohibitively expensive to meet the needs of a high capacity network which increases the capital and operational expenditure for the operators. There exist other technologies such as pico-cells and distributed antenna systems (DAS) used to extend indoor coverage and capacity in large buildings, which are categorised as hotspots (airports, shopping malls, universities). Such solutions are still too expensive and inadequate to provide mobile services to residential areas and small offices.

With the growing demand of innovative 3G services, most industrial critics see significant potential for the use of technology, so called “femtocells” [Chan2008]. Femtocells, also known as home base station, are small, low power access points and visually look like an ordinary wireless router. These access points are installed by users indoor, which creates a small wireless coverage area and connect user equipment (UE) to the cellular core network through subscribers broadband internet access. The access points known as femtocell access point (FAP) work as BS, enabling high quality voice, data and multimedia services to be delivered to mobile devices in indoor settings without changing the underlining UE radio access font end configuration. The FAP can be connected to the operator’s core network through users DSL, optical fibre or cable broadband connection. Femtocells would require some portion of spectrum from the operators for its operation. This can be a separate portion of spectrum allocated by the operator or the same portion of spectrum as used by macro-cell. The case of same spectrum being used for femtocells (co-channel femtocell deployment) offers the best spectral efficiency, however, with serious interference concerns. This interference can be between neighbouring femtocells (co-tier interference) as well as between femtocells and macrocell (cross layer interference). The main challenge faced by femtocells is interference management. The key techniques that can be used for avoiding and mitigating interference in femtocells are well presented [Chan2008] [Zah2012]. The work herein presented is a power control scheme for interference management in femtocell networks with a focus on reducing the cross tier interference caused by femtocells to macrocell users.

4.2.1 Technical aspects

The femtocells are deployed within a macrocell in an ad hoc fashion. Any user can deploy femtocells in the own home and even can move femtocells from one location to another. Therefore, it is a challenging problem for operators to manage radio resources dynamically. This intensifies the need of more intelligent FAPs. Cognitive femtocells can provide a better solution for the indoor coverage problem. Basically due to small cell radius, the distance between transmitter and receiver is reduced, hence transmitted signal is less attenuated and in turn receiver can receive good signal strength (RSS). Generally, the quality of a signal at the receiver is measured in terms of SINR. The SINR is a function of the transmitted power from the desired BS, transmitted power from interfering transmitters, shadowing, fading and path loss. The penetration losses due to walls cause the interfering signals to be weak. This attenuation is more prominent at higher frequencies that are commonly used in 3G/4G technology for their high bit rate operation. These losses act as insulation to the femtocells, and thus femtocells transmit with low power while maintaining good indoor coverage quality. The good channel conditions enables the femtocells to provide high data rate services to users by using higher modulation and coding schemes. Furthermore, a femtocell usually serves a very small number of users (house residents/office employees) as compared to a macrocell (e.g Vodafone femtocells can support maximum of four users), due to which it can devote a large portion of its resources to the available users. This enables femtocells to provide good QoS to its users as compared to a macrocell, which have to serve larger number of users simultaneously in a large area [Chan2009]. The coverage holes in the footprint of a macrocell can also be eliminated with femtocells. In this aspect, the femtocell can provide coverage to macro-cellular users which are nearby and in the range of the femtocell. This is of more importance in the cell edge areas.

4.2.2 Femtocell system model and problem formulation

A co-existence scenario in which a number of femtocells coexist within a macrocell as shown in Figure 4-16 is investigated. The interference caused by the femtocell downlink to any nearby macrocell user is thus investigated. It is understood that, in a co-channel femtocell deployment, the overall cross tier interference increases with the increase in the number of femtocells in that area. Furthermore, if a macrocell user tends to be near to a femtocell, it can face excessive interference from the femtocell downlink. In such a case, the femtocell should be able to reduce its power, in order to avoid causing interference to macrocell users. The system model of Figure 4-16 shows a random deployment of femtocells within a macrocell. The macrocell contains certain number of macrocell users called macrocell user equipments (MUE) communicating with their microcell base station (MBS) while every femtocell has only one femtocell (opportunistic) user equipment (FUE) communicating with its corresponding FAP. One FUE is considered for simplicity and it is assumed that each femtocell allocates all of its available resources to this one active FUE. On the other hand, the MBS allocates its resources to MUEs by dividing the resources equally among all active MUEs.

It is also assumed that there is a unidirectional common channel between macrocell base station and FAP. This channel is a broadcast channel and the FAP can use this channel only to receive any information from the macrocell base station. On the other hand, MBS-FAP control communication can be achieved through a femtocell gateway (FGW) located in MBS core network. Through this channel, the FAP is able to know the location of any MUEs around it and thus can find the distance to a specific MUE. Due to this information, the FAP is able to predict its interference impact on the MUE. The femtocell has geo location capabilities as well and hence they can also be aware of their own location within the macrocell.

Whenever a MUE is near to a femtocell it would face interference from the femtocell as well as from any other femtocells. The impact of each femtocell on the MUE depends on the distance between MUE and FAP. When the MUE faces an interference level from the surrounding femtocells that is greater than a pre-defined threshold, it informs its base station about the increased level of interference. This pre define threshold is the level of interference that a MUE can handle. If the amount of total interference received from all femtocells is greater than this threshold, the MUE will not be able to communicate with its base station. The macrocell base station similarly receives interference information from any active MUEs under interference. After receiving the information about interference, the serving MBS uses the unidirectional broadcast channel to inform the femtocells about the users that are facing interference along with their GPS coordinates. The FAP receives this information and reduces its power. This technique can cause all the femtocells to reduce its power and the system would not be efficient [Zah2011]. Therefore, in this research, a group of “main aggressors” is defined for each MUE. These groups of main aggressors are the femtocells whose distance “t” from the MUE is equal to or less than a pre-defined value of distance in meters. This distance threshold value is termed as the “aggressor distance threshold”, t_{th}. Whenever the MBS broadcasts information about MUEs under interference, the receiving FAP would first check if it is one of the main aggressors to those MUEs or not. If the femtocell is not in the group of main aggressors to a particular MUE, it would not react to the broadcast information; however, if it is one of the aggressors, it will reduce its power by a step size of Δ. The reduction in power would reduce the interference at the MUE.

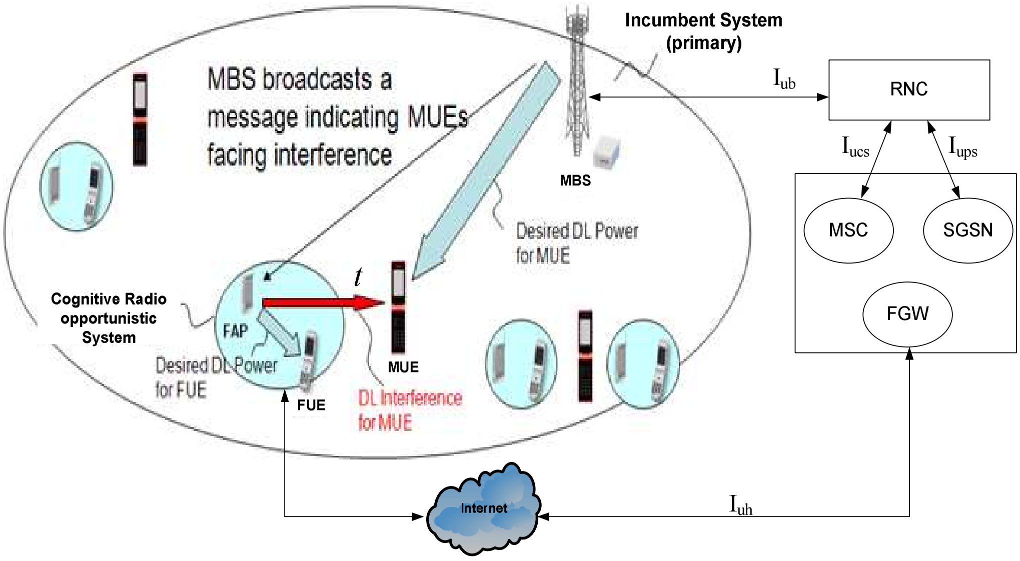


Figure 4-16: System model for femtocell network

For analysis, it is assumed that there are J active MUEs and N is the number of disturbed channels among them. The downlink power vector of macrocell base station is then represented as $\mathbf{P}^m = [p_1^m, p_2^m, \ldots, p_j^m]$, where $p_j^m (j \in J)$ represents the downlink transmit power for the j^{th} MUE in the n^{th} channel $(n \in N)$; assuming a one to one corresponding user-channel mapping. The number of femtocells in the system is K, so the downlink transmit power of the femtocell can be given via a matrix representation as in equation (4-5).

$$\mathbf{P}^f = \begin{bmatrix} p_{11}^f & p_{12}^f & \cdots & p_{1N}^f \\ p_{21}^f & p_{22}^f & \cdots & p_{2N}^f \\ \vdots & \vdots & \ddots & \vdots \\ p_{K1}^f & p_{K2}^f & \cdots & p_{KN}^f \end{bmatrix}^{\mathrm{T}} \tag{4-5}$$

The superscript *f* denotes femtocell, *m* for macrocell, *fm* for femto to macro and *mf* denotes macro to femto. p_{kn}^f in (4-5) therefore represents the downlink transmit power of the k^{th} FAP $(k \in K)$ in the n^{th} channel. It is assumed that each FAP allocates all of its resources through a single channel for FAP to FUE.

The aggregate interference I_n^{fm}, faced by the j^{th} MUE from all K femtocells in the n^{th} channel is given by equation (4-6):

$$I_n^{fm} = \sum_{k \in K} \frac{p_{kn}^f}{L_{kj}^{fm}} \tag{4-6}$$

where L_n^{fm} represents propagation loss from the k^{th} FAP to j^{th} MUE. The MUE can operate under certain constraint and can inform its base station when its threshold is breached. This ensures the MUE operates in an interference free environment. The MUE inference constraint is given as (4-7):

$$I_n^{fm} < \gamma_1 \, p_{noise} \tag{4-7}$$

where γ_1 is the interference coefficient for MUE and is assumed to be known while p_{noise} is the noise power. On the other hand, the FUE also have to be able to operate and have an interference constraint given in equation (4-8):

$$I_n^{ff} + I_n^{mf} < \gamma_2 \, p_{noise} \tag{4-8}$$

where γ_2 is the interference coefficient for FUE, I_n^{ff} is the femto to femto co-layer interference, and I_n^{mf} is the macro to femto cross layer interference with respect to the n^{th} channel.

Two path loss models are used for our propagation analysis. This is categorised as indoor and outdoor path loss models. Equation (4-9) is the indoor path loss (PL_{in}) model recommended by 3GPP while equation (4-10) is a simplistic outdoor to indoor path loss model represented as PL_{out} ([LevEtal2012] and references therein):

$$PL_{in} = 32.8 + 16.9 \log_{10} d + 20 \log_{10} f_c \tag{4-9}$$

where f_c is the centre frequency of the transmission channel and d is the FAP-FUE distance.

$$PL_{out} = 15.3 + 37.6 \log_{10} t + wL_{wall} \tag{4-10}$$

L_{wall} is the loss due to the thickness of the wall, t is the distance between FAP transmitter and MUE receiver (both obeying reciprocity), while w is the number of walls.

4.2.2.1 Assumptions

The algorithm considers a number of assumptions and there it is necessary to summarise them separately. The main assumptions are given as follows.

- The cognitive femtocell is aware of its location and it is also also assumed that it has a geo location capability and can find its location.
- The femtocell is also aware of the location of the MUEs within its range. This information is given to the femtocell by the macrocell BS through the unidirectional broadcast channel.
- It is assumed that whenever a MUE faces interference from nearby femtocells, higher than a predefined threshold, it informs its base station about the higher interference.
- The model is based on path loss only.

4.2.3 Proposed downlink power control algorithm for femtocells

The main purpose of the downlink power control algorithm is to reduce the level of interference a femtocell causes to a nearby MUE in co-channel femtocell deployment and works as follows:

The MUEs in a macrocell containing femtocells would face interference from nearby femtocells. The impact of each femtocell on the MUE depends on the channel between MUE and FAP. When the MUE faces an interference level from the surrounding femtocells that is greater than a pre-defined threshold shown in (4-7), the MUE informs its base station about the increased level of interference. This predefined threshold is the level of interference that an MUE can handle. The MBS similarly receives one bit interference information from any active MUEs under interference. The one bit information only describes if the particular MUE is facing high interference or not. The one bit information is used to make sure there is no communication overhead.

The MSC for this process is shown in Figure 4-17. After receiving this information, the base station uses the unidirectional broadcast channel to inform the femtocells about the MUEs that are facing interference. The femtocell then, first checks if it is one of the “main aggressors” to any of the MUEs that are under interference. If the distance between the FAP and an MUE is less than or equal to the aggressor distance threshold t_{th} the particular FAP is then one of the main aggressors to that MUE. The aggressor distance specifies a circular region around the FAP with radius equal to *t*; any MUE within this area would be counted as a possible victim of interference from the femtocell. If an MUE resides within the aggressor distance, the FAP will then reduce its power; in the case where an MUE is at a distance greater than *t*, the FAP will ignore the message from MBS.

The concept behind the aggressor distance is simple and only those femtocells that are near to the MUE will decrease their power, as the nearby femtocells are the ones causing most of the interference to the MUE. This technique makes sure any unnecessary power reduction in faraway FAPs is avoided without degrading the quality of service to FUEs and also causing unnecessary FUE-FAP processing overhead. The aggressor distance *t* needs to be selected carefully as large values of *t* would cause unnecessary reduction in power (QoS) of femtocells that are faraway thereby increasing processing overhead while a smaller values will result in an underestimation of MUEs and cause high interference to MUEs thereby degrading overall performance. The value of t_{th} for the proposed algorithm has been selected based on best performance via simulation and is explained in the sequel.

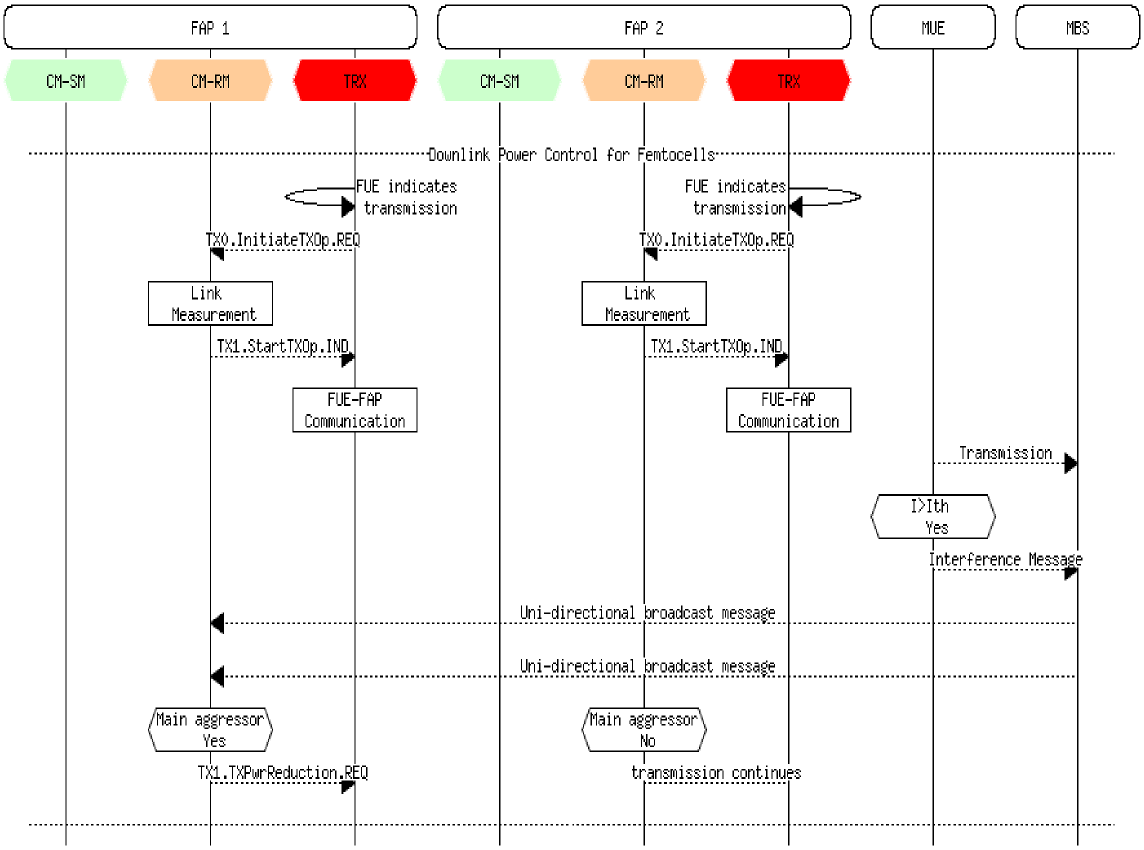


Figure 4-17: Message sequence chart for femtocell downlink power control.

Normally, the FUE sends a transmit operation request to its FAP using a `TX0.InitiateTXop.REQ` command as seen in Figure 4-17. This command is processed at the CM-RM module of the FAP where inferences on the link measurements (channel) and power control configurations are made and passed to the FUE through a `TX1.StartTXOP.IND` indicator. FUE therefore starts transmission based on configuration parameters (power level) received from its FAP. Whenever a FAP have to reduce its power as a result of information from the MBS, it reduces its power by a fixed step size Δ and correspondingly re-calculates configuration parameters which are conveyed to its active FUE. This reduction in power is to make sure the FAP reduces the amount of interference to an MUE shown in equation (4-6). Both the value of aggressor distance *t* and Δ can affect the performance of the algorithm. The algorithm after reduction of power operates normally until it receives another message from the MBS to reduce its power.

4.2.4 Femtocell simulation results and discussion

Simulations are carried out in order to evaluate the performance of the algorithm. The simulation parameters are given in Table 4-4.

Table 4-4: Femtocell simulation parameters.

parameter name	value
number of MUEs	20
number of femtocells	0 - 200
number of channels	20
macro-cell radius	500m
femtocell transmit power	-20dBm
MUE interference threshold	-100dBm
wall losses L_{wall}	15dB
aggressor distant *t*	50m

The simulation is carried out with 20 MUEs in a macrocells with different number of femtocells. The locations of MUEs and femtocells are taken randomly and interference faced by MUEs with increasing number of femtocell is calculated. The results are shown with the total interference in system (I_T), which is the sum of aggregate interference faced by all MUEs and given as (4-11):

$$I_T = \sum_{n=1}^{20} I_n^{fm} \tag{4-11}$$

where I_n^{fm} is the interference faced by each MUE in the n^{th} channel shown in equation (4-6). The result in Figure 4-18 shows that the total interference in the system is reduced with the proposed power control algorithm. Notice that for any given number of femtocells, the algorithm makes sure all the MUEs are interference free. This algorithm gives higher priority to MUEs and therefore, the transmit power is reduced whenever an MUE is under interference due to a nearby femtocell. The number of MUEs within the macrocell also has an impact on the overall performance as shown in Figure 4-19.

Increasing in the number of MUEs increases the total interference in the system as seen in Figure 4-19. It can also be inferred that increasing the number of MUEs would also increase the processing overhead, as the femtocells would then have many MUEs in its main aggressor list.

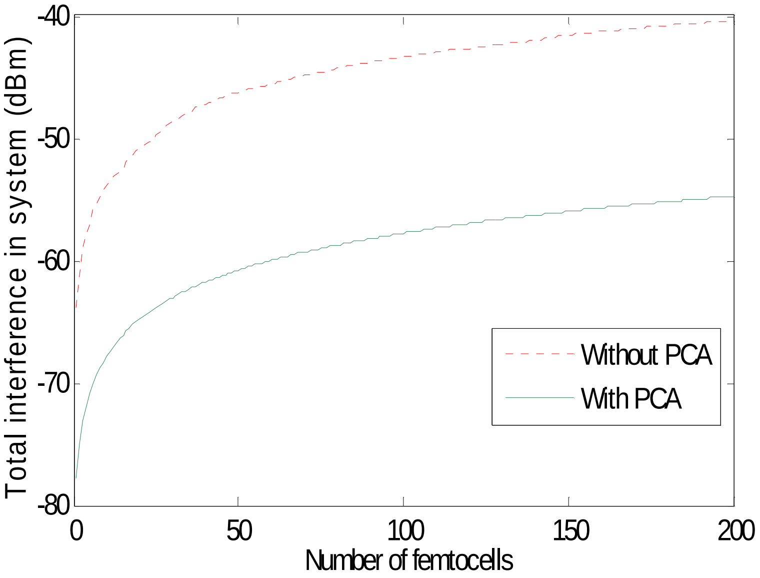


Figure 4-18: Performance of the downlink power control algorithm.

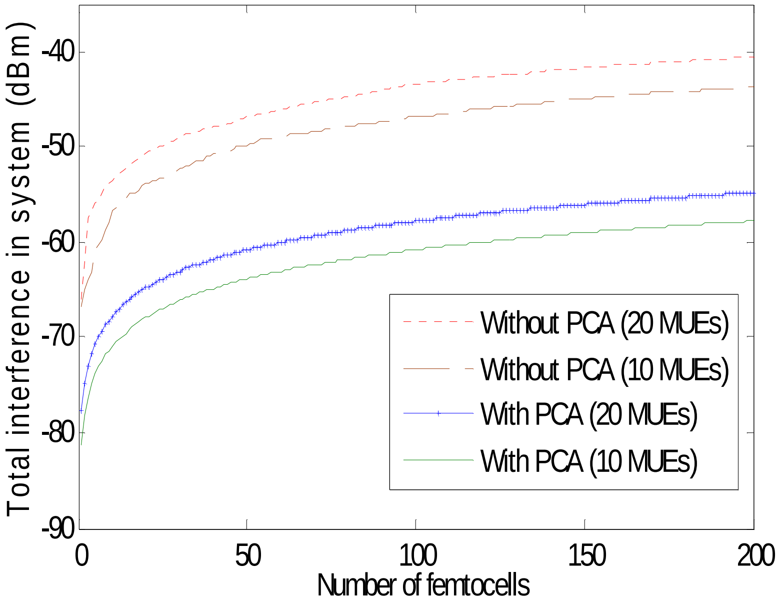


Figure 4-19: Comparison of PCA with different number of MUEs.

The aggressor distance has an impact on the overall performance, which can be seen in Figure 4-20. Decreasing the aggressor distance reduces the overall performance. It is worthwhile to note here that the increase in distance does not increase the performance after a certain value, which in this case is 50 metres. The performance at an aggressor distance 50 metres and 70 metres is almost the same. Therefore, the aggressor distance of 50 metres has been used in this algorithm. The algorithm is a simple one and thus suitable for implementation in low powered FAP devices. It also makes sure the far away FAPs do not need to arbitrarily reduce their power if they are not causing interference to MUEs.

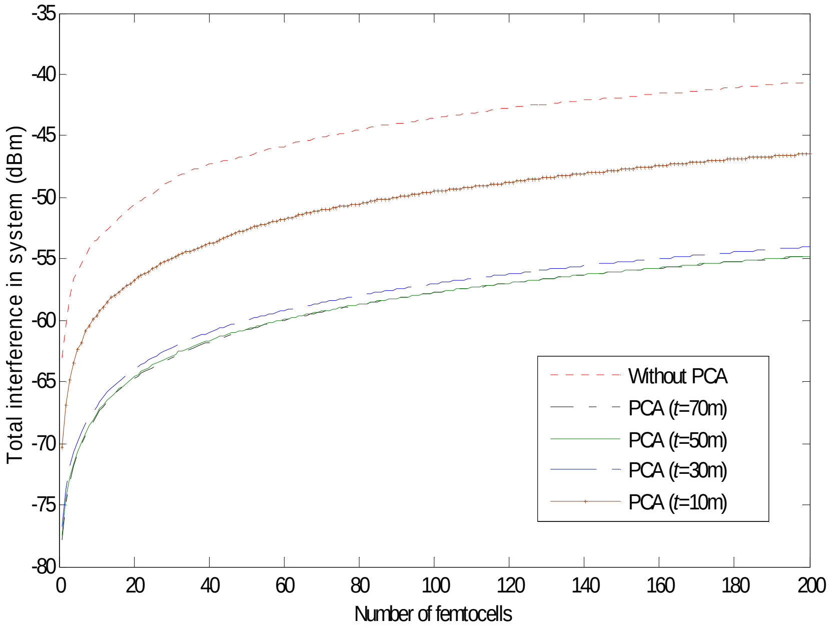


Figure 4-20: Impact of aggressor distance on the overall performance.

4.2.5 Remark

A downlink power control algorithm for co-channel femtocells which allows the co-existence with the underlay macrocell network even with a dense deployment of femtocells was considered. The proposed algorithm in this report is a step towards providing a solution for the problem of strong cross layer interference in such a scenario. The algorithm requires minimal help from the MBS and makes sure no MUE faces unwanted interference from any femtocell. Simulation results show the performance of the algorithm and it is clear that, with the help of the proposed power control algorithm, large number of femtocells can be accommodated within the macrocell, while at the same time protecting all the MUEs from interference.

4.3 Ad hoc networks scenario

Cognitive ad hoc networks can be set up by different types of nodes which may be static, nomadic or mobile depending on the use cases envisaged. Their existence is limited in time (like for emergency or big event) and they allow the operating frequency band to be adapted to the specific needs in bandwidth, range and QoS.

The case of multiple cognitive ad hoc networks sharing the same portfolio of available channels in a given area is considered. Each cluster-based single ad hoc network has a star topology as depicted in Figure 4-21. The nodes/UEs are grouped into one cluster with one node/UE having the additional functionality of cluster head (CH). It is assumed that the management of the resources is centralised and implemented in the CH providing routing, resource allocation and power control functionalities whereby communication flows are exchanged directly between nodes when possible.

In this scenario, the focus is on finding necessary algorithms for sharing efficiently the available channels in the CM-RM portfolio among different clusterized mobile ad hoc networks employing scheduled access protocol and not random access ones.

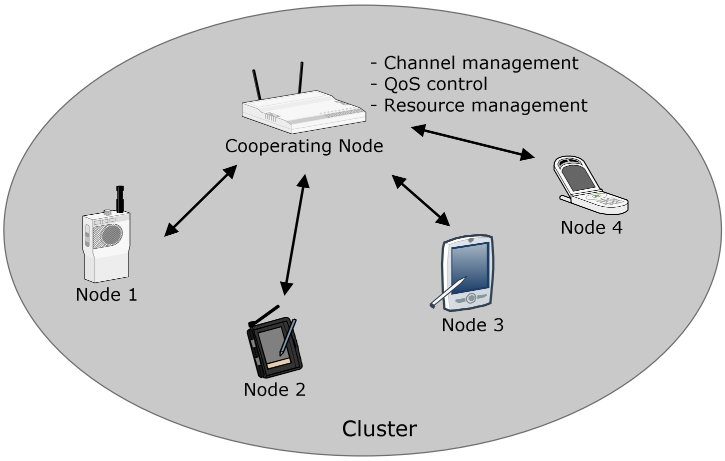


Figure 4-21: Cluster based ad hoc network with star topology.

4.3.1 System description

Time-division multiple access (TDMA) communications in a time-division duplexing (TDD) mode using OFDM transmission technique is considered. Since the nodes inside a cluster share the same band, both for transmitting and receiving, then the nodes must transmit and receive on separate slots. The switch from transmitting state to receiving state can be handled during the guard time which is present at the beginning of any slot (beacon, random access and data). The corresponding TDMA frame structure is illustrated in Figure 4-22 and it is divided into signalling, incumbent detection, and communication phases.

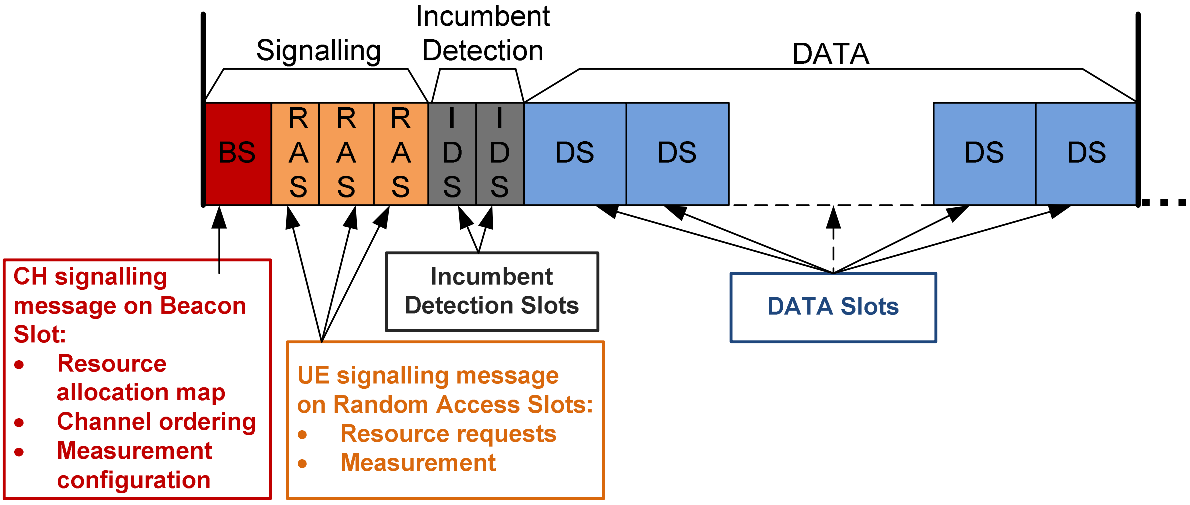

Incumbent Detection
Signalling
DATA
BS
RAS
RAS
RAS
IDS
IDS
DS
DS
DS
DS
CH signalling message on Beacon Slot:
• Resource allocation map
• Channel ordering
• Measurement configuration
Incumbent Detection Slots
DATA Slots
UE signalling message on Random Access Slots:
• Resource requests
• Measurement

Figure 4-22: TDMA frame structure with the signalling, incumbent detection, and the communication phases.

4.3.1.1 Signalling phase

The beacon slot is used by the CHs to send CH signalling message that indicates to the nodes the use map of the data slots in the communication phase, and serves also for time and frequency synchronisation purposes. The map allows the members of the cluster to know on which time/frequency resources the communication links are scheduled.

The beacon indicates also information about available channels in the portfolio and sensing directives.

The random access slots (RAS) have three main purposes:

- the resource allocation requests sent by each node to its cluster head;
- the network synchronisation messages are transmitted and detected during these slots and are used for neighbouring discovery and topology control;
- the sensing measurement reporting by the nodes to their CH.

4.3.1.2 Incumbent detection phase

One or several random access slots are dedicated to incumbent users' detection on the operating channel. It is chosen to rather dedicate for this purpose RAS than data slots as signalling messages are exchanged on each frame, which is not the case of data protocol data units (PDU). The RAS slots being much shorter than the data slots, this choice permits also to preserve the network useful throughput.

4.3.1.3 Communication phase

In this phase, all transmitting users will communicate on their allocated time/frequency resources. Indeed, the multiple access scheme proposed in this section is based upon the orthogonal frequency-division multiple access (OFDMA) which gives a great flexibility since allocations can be managed both in time and frequency. It enables to allocate several transmissions (links) in the same slot compared to only one link with an OFDM scheme. The transmission bandwidth is divided into logical sub channels that correspond to a fixed number of subcarriers. A sub channel is called resource block (RB) and represents the smallest piece of resource than can be allocated for a link.

Figure 4-23 represents one frame of the TDMA/OFDMA structure showing the different resource blocks. Each colour corresponds to a dedicated link.

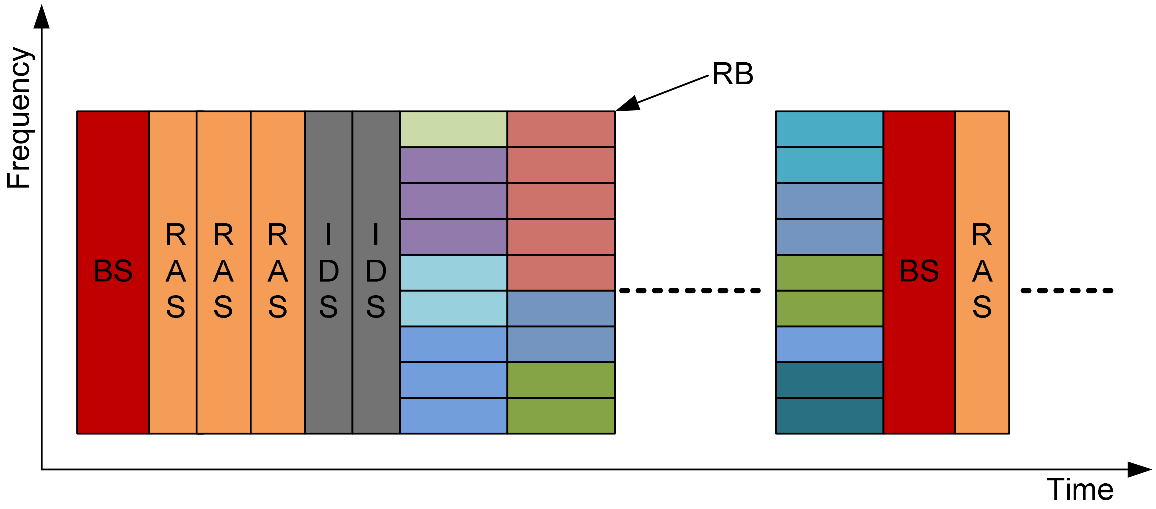


Figure 4-23: Time/frequency representation of a resource block.

4.3.1.4 Resources allocation and QoS support

QoS management is performed at link level for peer-to-peer communications. We consider two QoS classes:

1. best effort (BE) for data (files, images, etc.);
2. real-time (RT) for voice and urgent traffic that is needed in emergency situations.

The resource allocation proposed in the QoSMOS context is split into two stages. The first stage referred to as demand adaptation (DA) consists in pre-processing the brute demand of the queues by taking into account the QoS classes, then a recursive per-slot allocation is performed taking as inputs the list of candidate links elaborated by the DA.

The global allocation process is illustrated in Figure 4-24. A detailed description is given in [D5.2].

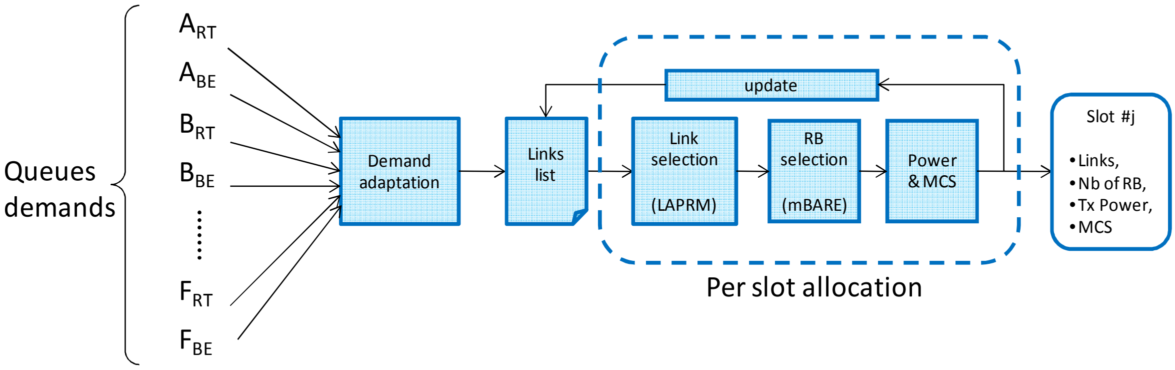


Figure 4-24: Global resource allocation algorithm.

4.3.2 Sensing control

The sensing operation of the cognitive ad hoc network is monitored by the CH. The sensing control directives are part of the beacon message. These directives are given in terms of the list of available channels listed from the best to the worst (channel sorting algorithm is given thereafter), and the target number of measurement to be done on each channel.

Each node performs then the sensing in a distributed and opportunistic manner in order to best fulfil the directives of the CH. In doing so, the need of having the CH sending explicit sensing commands to the nodes is avoided and thus we save signalling bandwidth. Moreover, there is no guaranty that a network is able to sense all available channels unless it dedicates some slots for sensing on every channel.

Table 4-5 depicts the cluster head signalling message on the beacon slot with the sensing directives:

Table 4-5: Cluster head signalling message.

field	description
Ch_id	cluster head MAC address
Op_channel_id	operating channel ID
Res_map_list	resource allocation map (slot, RBs set, source, destinations' list)
Bu_channel_list	list of backup channels sorted from the best to the worst
Nb_target_meas_list	list of the number of target measurement on each channel of the portfolio

4.3.3 Incumbent user protection

Upon the detection of an incumbent user on the operating channel, the nodes stop immediately transmitting data. If a node detects the incumbent presence on a given channel then it informs its CH in its signalling message. The CH marks the channel where incumbent users are detected as not available and excludes it from the channel selection process until it is sensed again as free from incumbent users. If an incumbent is detected on the operating channel, the CH selects another one immediately and informs the nodes in its cluster.

4.3.4 Sensing procedure

The sensing operation is done by each node in the network in a distributed and opportunistic way, driven by sensing directives of the CH.

Indeed, let C={C1,…,Cn} be the list of sorted available channels, T={T1,…,Tn} be the list of the corresponding target number of measurement asked by the CH, and M={M1,…,Mn} be the number of corresponding performed measurements. Then, when a node is not transmitting and has no data to receive on a given slot, it selects to listen to the channel of the portfolio that maximises the cost function:

$$F(C_i) = \frac{T_i - M_i}{i} \qquad (4\text{-}12)$$

The measurement process is of moving average type. The sliding window size corresponds to the frame duration.

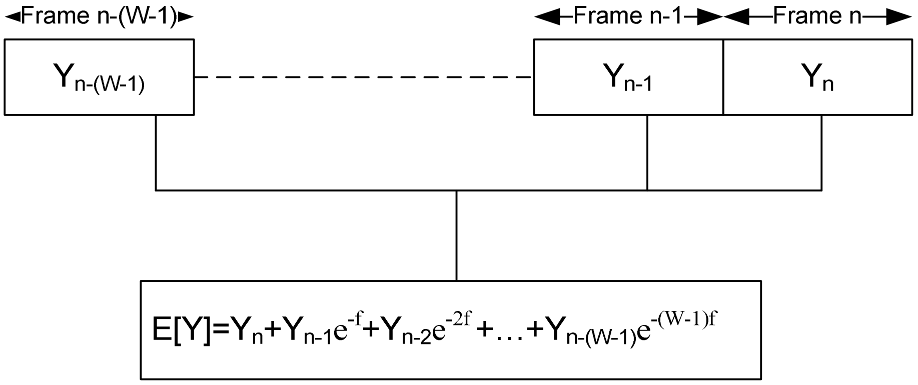

Frame n-(W-1)
Frame n-1
Frame n
$Y_{n-(W-1)}$
Y_{n-1}
Y_n
$E[Y]=Y_n+Y_{n-1}e^{-f}+Y_{n-2}e^{-2f}+\ldots+Y_{n-(W-1)}e^{-(W-1)f}$

Figure 4-25 depicts the measurement process where *Yn* represents the mean value of the measures performed during frame *n*, *W* is the measurement window size, and *f* is a forgetting factor.

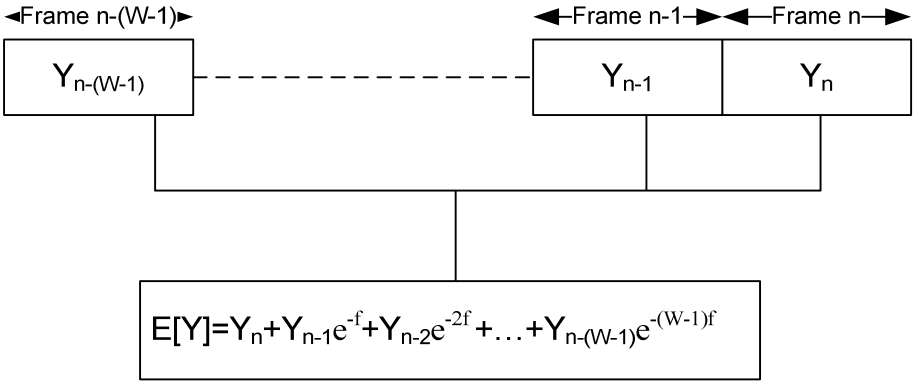

Frame n-(W-1)
Frame n-1
Frame n
$Y_{n-(W-1)}$
Y_{n-1}
Y_n
$E[Y]=Y_n+Y_{n-1}e^{-f}+Y_{n-2}e^{-2f}+\ldots+Y_{n-(W-1)}e^{-(W-1)f}$

Figure 4-25: Measurement process.

The UE reports their measurements in their signalling message once per frame using the RA slots. Table 4-6 depicts the UE signalling message.

Table 4-6: UE signalling message.

field	**description**
Node_id	node MAC address
Ch_id	MAC address of the corresponding CH
Res_req_list	resource requests list (source, destinations' list, queue size, QoS class)
Channels_quality_list	average received power on each channel
Link_quality_list	list of quality descriptor of the nodes' peer links (Dest_id, SINR)

4.3.5 Channel selection algorithm

The role of the channel selection algorithm (CSA) is to provide a list of active channels sorted out from the best (ranked #1) to the worst, the best being taken as the operating channel. The remaining channels are kept to serve as reserve channels in case the operating channel needs to be freed, and are selected in the decreasing quality order. The CSA process is performed at the CHs based on local measurements by the CH itself and reported measurement from the UEs

A metric that reflects the channel quality needs then to be chosen in order to sort out the available channels. It is assumed that at a ggiven time in the ad hoc network all the links that need to be served (called “links in operation” in the sequel) are identified and for each of them also the corresponding doublet (data rate demand (queue size), priority). The metric considered here is then the maximum data rate that the cluster can achieve for the set of links in operation, and for the highest priority. This metric allows fulfilling the QoS since the highest priority will be served first and the maximum data rate ensures the efficient use of the radio resources.

The first approach to select the channel that maximises the above metric is to run the resource allocation algorithm (RAA) for all the available channels. This will guarantee that the channels order will be in accordance with respect to the RAA capabilities. The possible issue with this approach is the computation complexity since the RAA is rather demanding, and this method would need to run the RAA as many times as the number of candidate channels.

To overcome the issue of the direct approach, it is here proposed to use as metric the weighted sum of the link capacity (over the set of operating links) computed using the Shannon capacity as a function of the SINR and is referred to as WSLA (weighted sum link approach). The weight applied to each link in the metrics computation is the rate demand normalised by the total rate demand. This normalisation is for fairness issue between links of the same priority.

The metric is given as follows

$$SR(c,p) = \frac{1}{\sum_{i=1}^{N}\sum_{j=1}^{N} R_{i,j}(p)} \sum_{i=1}^{N}\sum_{j=1}^{N} R_{i,j}(p)\log(1+SINR_{i,j}^{c}) \qquad \text{(4-13)}$$

Where $SINR^c_{i,j}$ is the signal to noise-plus- interference ratio of the link between source i and destination j on channel c of the portfolio, and $R_{i,j}(p)$ is the asked rate of the flow demand on this link having priority P.

This metric requires little computation effort to the expense of eventual performance degradation. In the following, we compute performance results of both approaches to evaluate the degradation of the SLA compared to the RAA.

Indeed, both approaches have been compared through simulations. The simulation settings are: $N = 10, 20, 30$ nodes in the cluster, Rayleigh fading, a fixed transmit power and a set of seven modulation and coding schemes: BPSK 1/2, BPSK 2/3, QPSK 1/2, QPSK 3/4, QPSK 8/9, QAM-16 3/4 and QAM-16 8/9. The multiple access scheme is OFDMA, where the smallest number of subcarriers that can be allocated per users is equal to 16 among a total of 512. The number of candidate channels in the portfolio provided by the CM-SM is 15.

Results are reported in Figure 4-26 which presents the probability that the *k*-th channel identified by WSLA is equal to the operating channel (#1) chosen by the RAA for $N = 10, 20, 30$. It can be seen that the WSLA performs well and the performance is improved as the number of nodes increases. The probabilities that the two methods choose the same operating channel are equal to 0.68, 0.92, 0.99 for $N = 10, 20, 30$ respectively. This can be explained by the average effect since the number of operating links increases with the number of nodes. Thus from the simulations it can be concluded that, when the number of nodes is large enough (e.g. $N \geq 20$), the SLA can be a good alternative to the RAA by achieving comparable performance at a lower computational cost.

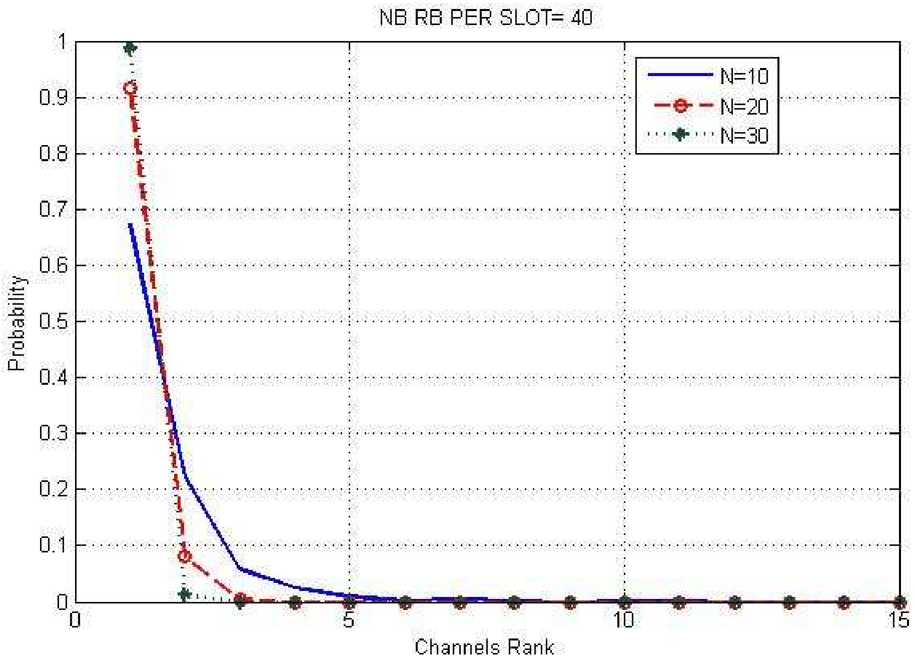


Figure 4-26: Channel selection algorithm performance vs network size.

In Figure 4-27 are compared the results of WSLA to those of systematic sum rate (SSR) maximisation over all possible links of the networks without taking into account difference between active and non-active links or rate demand amplitudes. It is observed that WSLA clearly outperforms SSR. Indeed, SSR is unable to differentiate the channel and choose among them uniformly

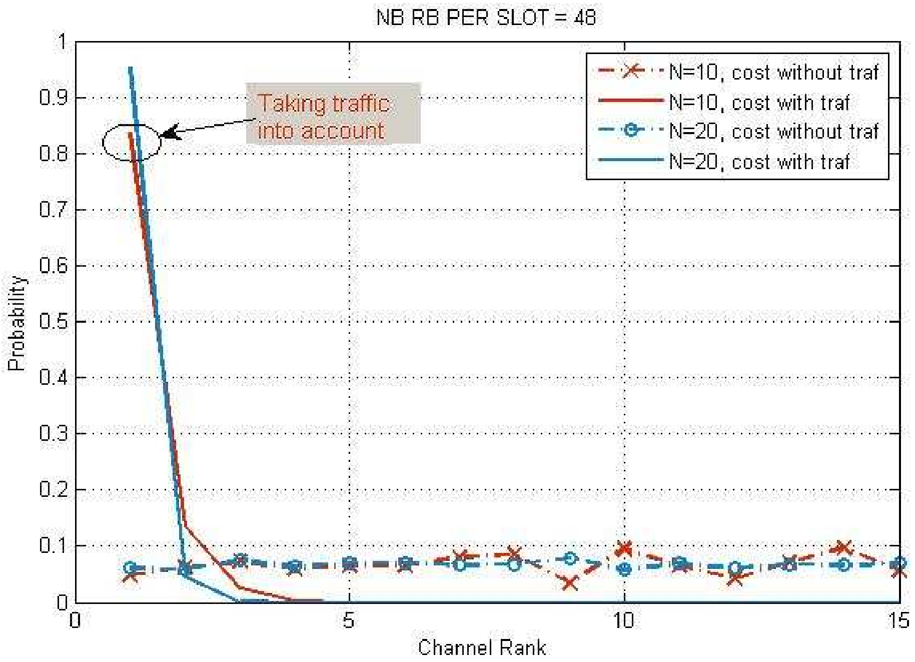


Figure 4-27: Channel selection algorithm performance vs systematic sum rate.

4.3.6 Channel acquisition protocol

Once the CH has identified the best operating channel to use using the channel selection algorithm, it notifies to the nodes in its cluster to switch to the selected channel.

The change of the operating channel may occur due to the detection of incumbent users or due to mutual interference caused by co-localised cognitive ad hoc networks sharing all the same portfolio of available channels. In the two cases, the decision of CHs to change their operating channels will occur somehow synchronously in time and will lead to a conflicting use of the radio channels in the first case, or will not resolve the conflicting use of the radio channels in the second one.

In order to reduce the probability of conflicting use of the operating channel among the cognitive ad hoc networks, the CHs will follow a channel acquisition protocol described in Figure 4-28.

The protocol is of the family of random access protocols and imposes to each CH that decides to change its operating channel to wait a random amount of time before the effective change (local change and notification to the nodes in its cluster).

During this random wait (multiple of frame time), the CH keeps updating its measurement about the available channels (local and remote measurements), and calling the channel selection algorithm on each frame. If the CSA detects that the current operating channel is again the best one then it cancels the channel change procedure and reset the channel change counter, otherwise it decrements the change counter. Once the counter is zero the change is performed and notified to the nodes in the cluster.

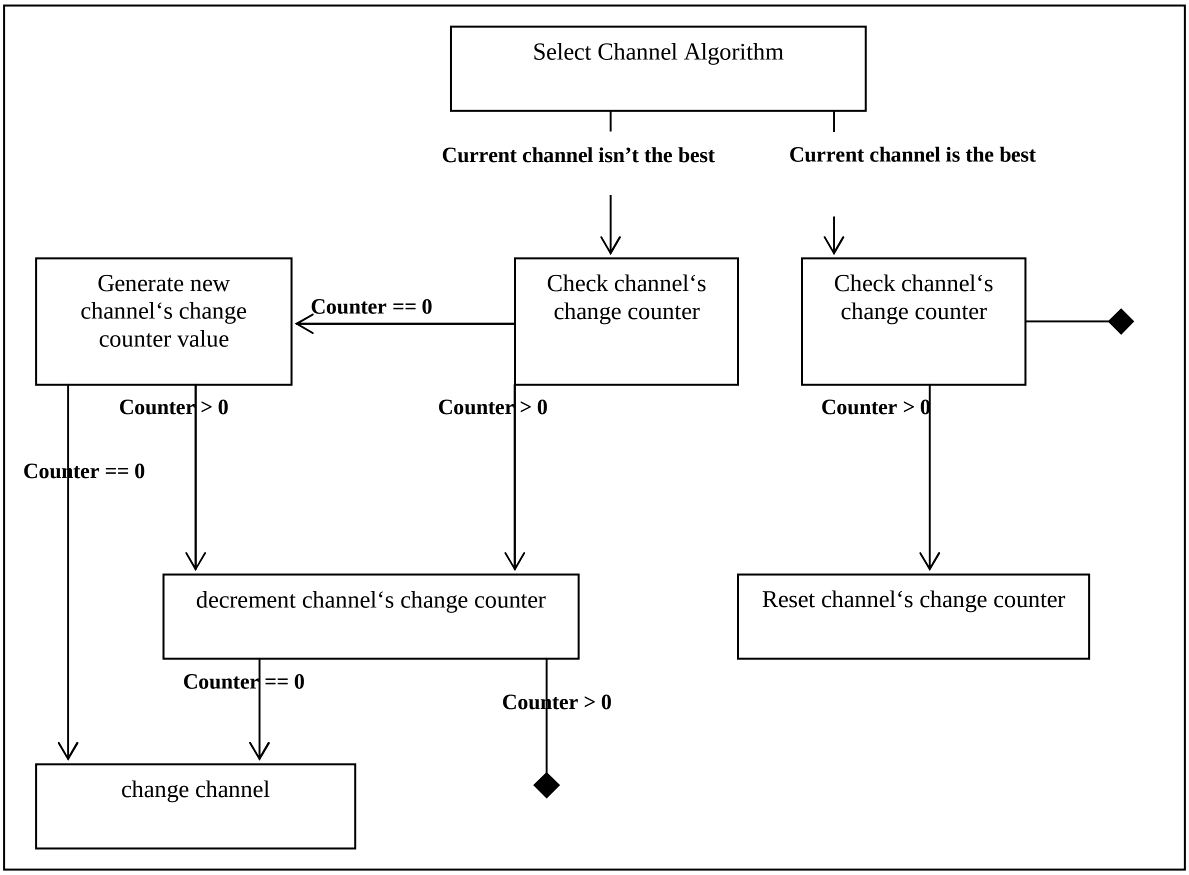

Select Channel Algorithm
Current channel isn't the best
Current channel is the best
Generate new channel's change counter value
Counter == 0
Check channel's change counter
Check channel's change counter
Counter > 0
Counter > 0
Counter > 0
Counter == 0
decrement channel's change counter
Reset channel's change counter
Counter == 0
Counter > 0
change channel

Figure 4-28: Channel acquisition protocol.

Figure 4-29 gives an example of the operation of the channel selection and acquisition protocol (CSAP). We consider two cognitive ad hoc networks (CAN) initially operating on the same channel and not mutually interfering. As soon as the mutual interference is sensed by both networks and reported to the CHs, the channel selection algorithm will detect that the current operating channel is no more the best. Both CHs generate then a random change counter. Each time the CSAP is called and the operating channel is detected as still not the best, the change counter is decreased. In the example considered here, the change counter of CAN2 reaches first zero so that the CH2 decides to change its channel. This change results then in an interference-free operation of both CANs on their respective operating channels. Thus CAN1 detects again that its channel is again the best and cancel its channel change procedure.

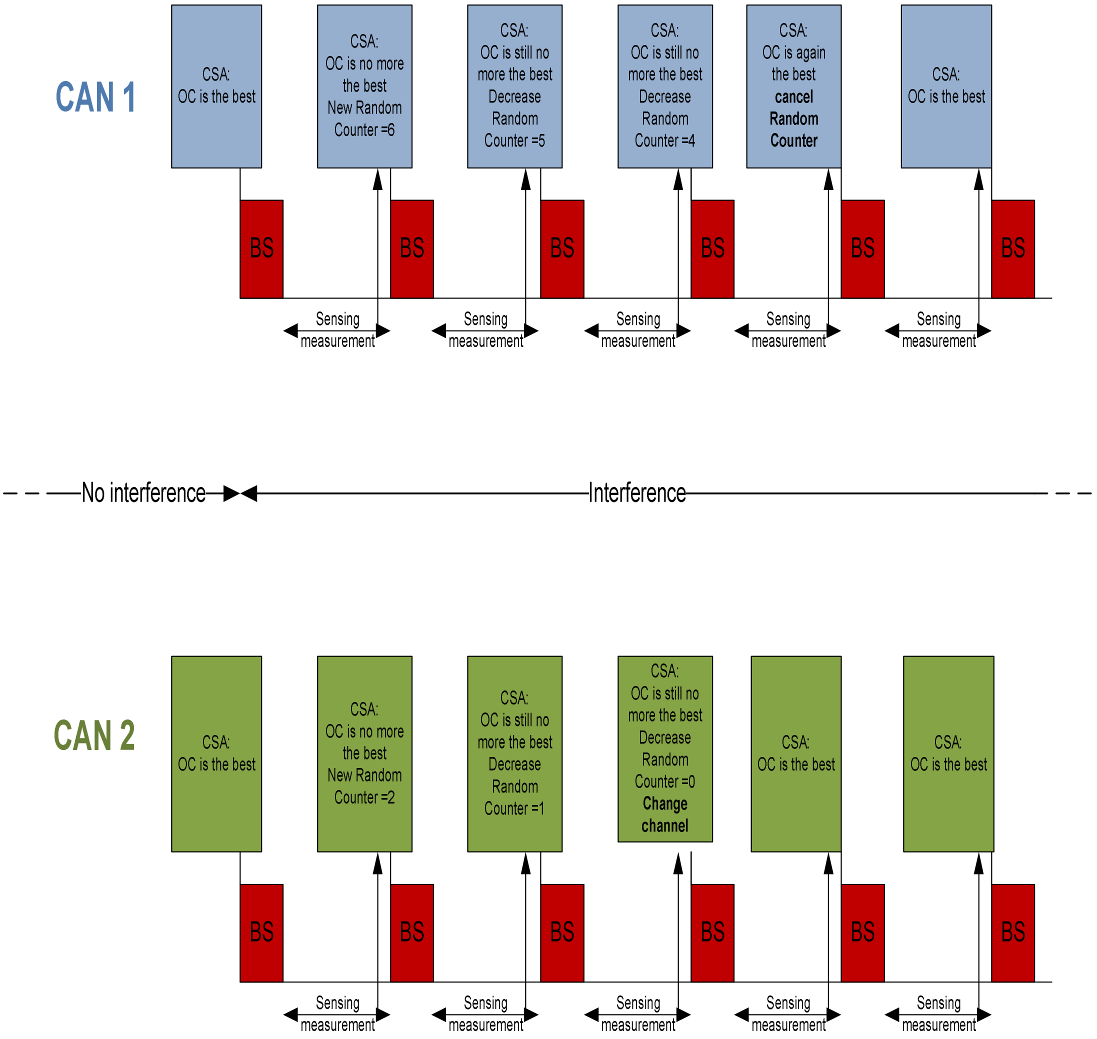

CAN 1
CSA: OC is the best
CSA: OC is no more the best New Random Counter =6
CSA: OC is still no more the best Decrease Random Counter =5
CSA: OC is still no more the best Decrease Random Counter =4
CSA: OC is again the best cancel Random Counter
CSA: OC is the best
BS
Sensing measurement
No interference
Interference
CAN 2
CSA: OC is the best
CSA: OC is no more the best New Random Counter =2
CSA: OC is still no more the best Decrease Random Counter =1
CSA: OC is still no more the best Decrease Random Counter =0 Change channel
CSA: OC is the best
CSA: OC is the best

Figure 4-29: Channel acquisition example.

4.3.7 Simulation results

The simulation framework is based on the event-driven discrete time simulation tool Omnet++ [Var]. The simulation considers complete mobile ad hoc network simulation from protocol layer L1 to L7 [MAS10].

4.3.7.1 Incumbent users protection scenario

In this scenario we are interested in the coexistence behaviour of cognitive ad hoc networks and incumbent networks. one mobile CAN and one fixed incumbent network are considered. At the beginning of the scenario, the CAN uses as operating channel the one of the incumbent network. The CAN then moves to cross the incumbent network area and to finish at the other side of it.

In both networks, every node has one TX/RX flow to/from another node in its network. All the flows have the same throughput and are of best effort class of QoS. Figure 4-30 illustrates the scenario.

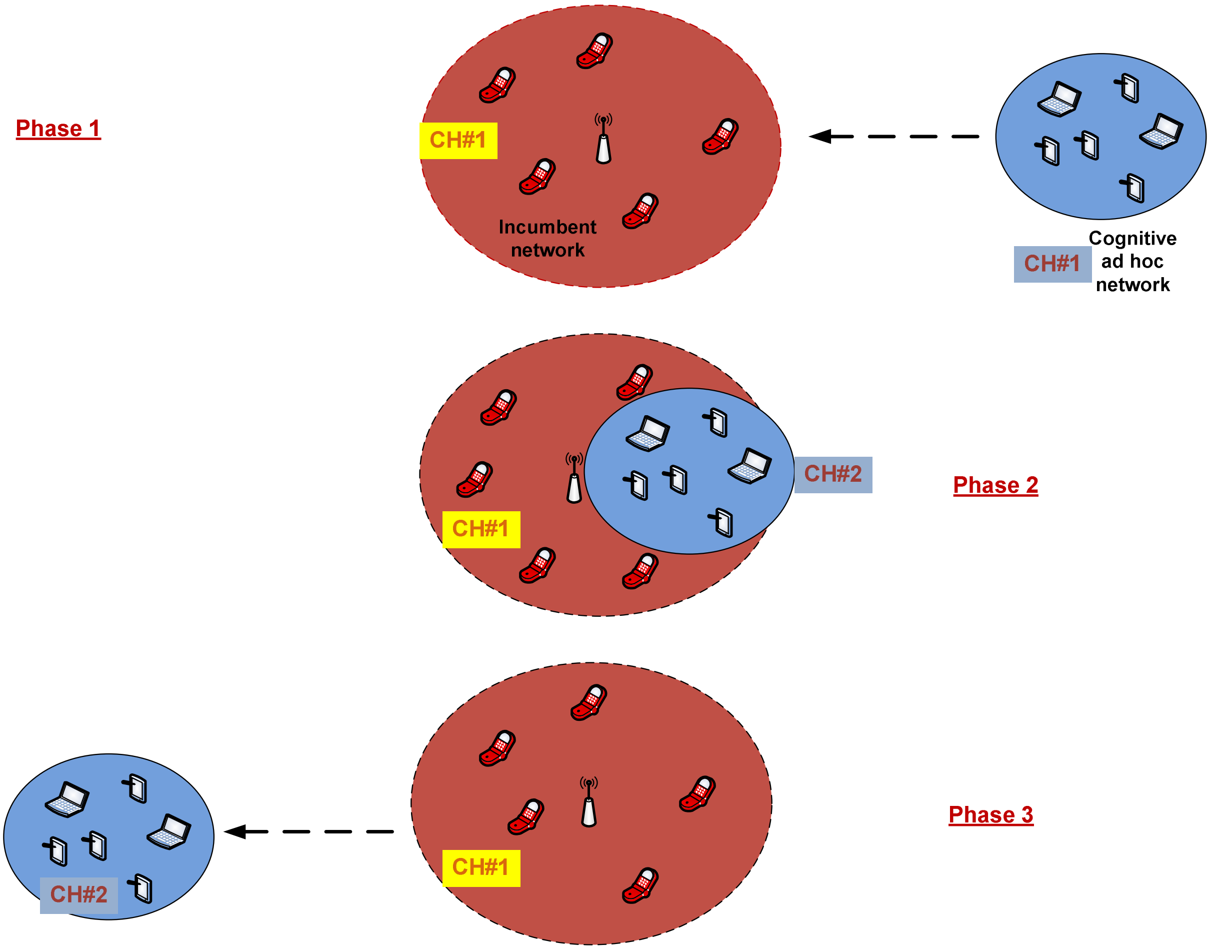


Figure 4-30: Incumbent users protection scenario.

Figure 4-31 shows the achieved sum throughput of each network during the simulation time.

The physical interference between the two networks lasts approximately 40 seconds from t=30s to t=70s.

It can be observed that immediately at the beginning of the mutual interference the throughput of the cognitive networks deeply decreases, while the throughput of the incumbent network remains unchanged. The decrease is due to the fact that as soon as the nodes of the CAN sense the presence of the incumbent users they stop transmitting data to not interfere the incumbent network. The CAN changes then the operating channel and restores the data transfers. The CAN takes approximately three seconds only to change the operating channel and to retrieve the same level of data throughput as before the detection of the incumbent network.

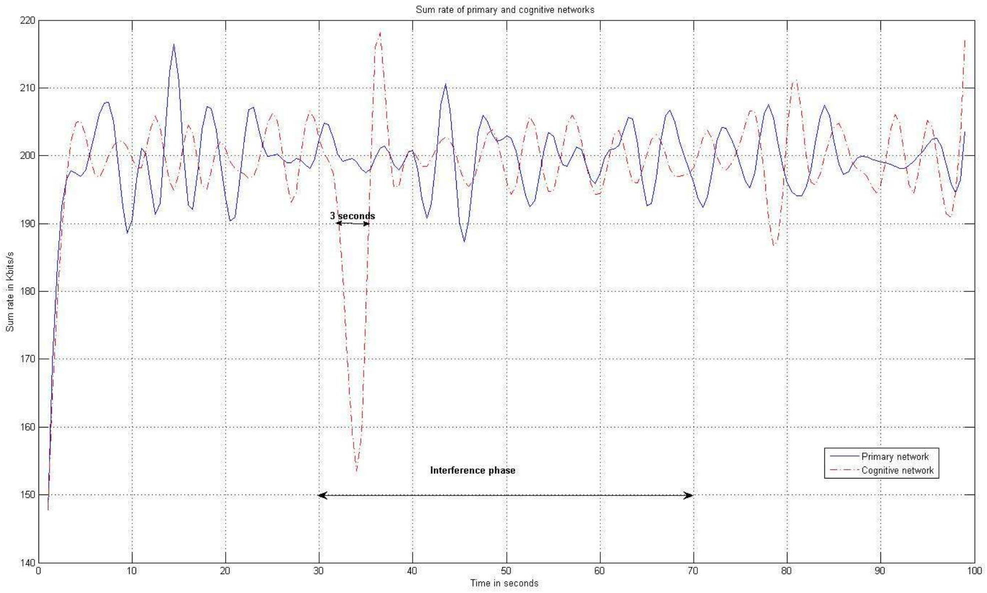


Figure 4-31: Sum rate of incumbent and cognitive networks.

4.3.7.2 Co-localised cognitive ad hoc networks scenario

Let us focus on now the coexistence behaviour of these cognitive ad hoc networks and two mobile CANs are therefore considered. At the beginning of the scenario the CANs use the same operating channel. The CANs then moves toward each other and exchange their initial position at the end of the scenario.

In both networks, every node has one TX/RX flow to/from another node in its network. All the flows have the same throughput and are of best effort class of QoS. Figure 4-32 illustrates the scenario.

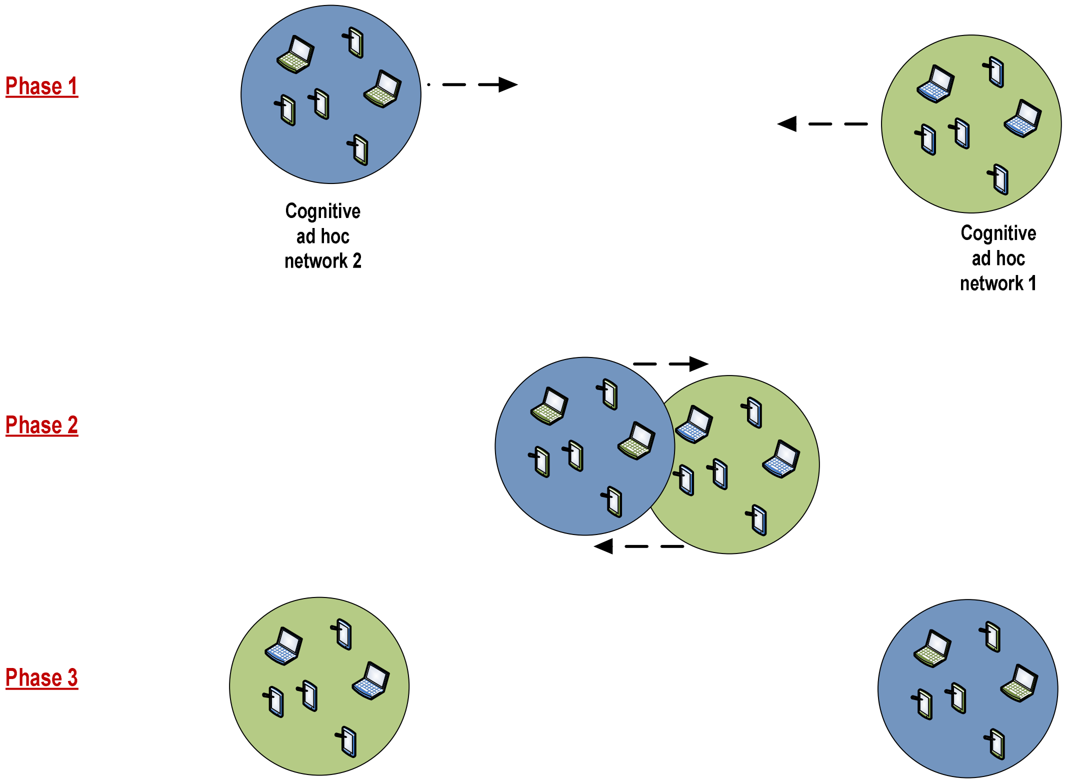


Figure 4-32: Co-localised cognitive ad hoc networks scenario.

Figure 4-33 depicts the achieved sum throughput of each network during the simulation time.

The physical interference between the two networks lasts approximately 40 seconds from t=30s to t=70s.

It can be again observed that immediately at the beginning of the mutual interference, the throughput of both networks decreases, but slightly this time as the CANs do not stop transmitting data. The decrease is due to interference and to operating channel change but last only four seconds approximately.

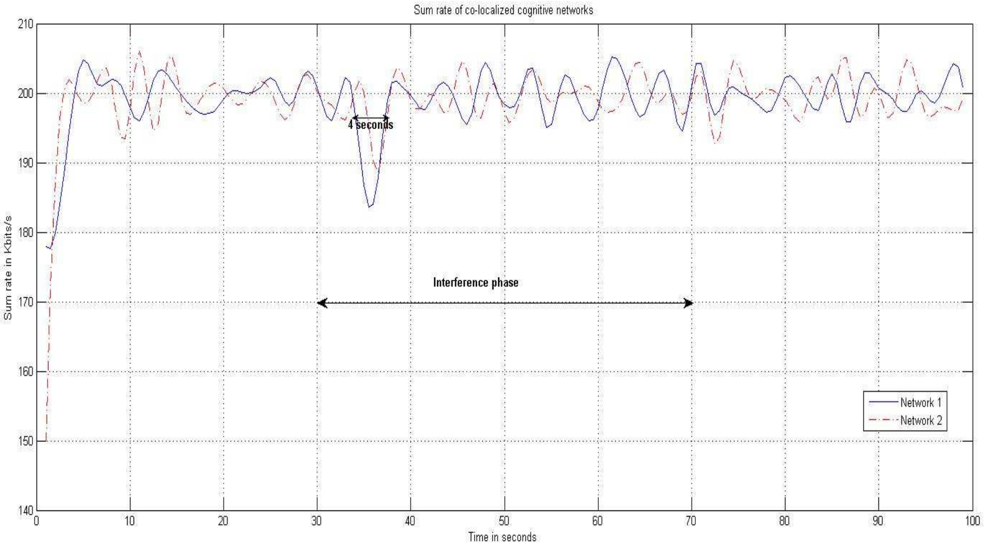


Figure 4-33: Sum rate of co-localised cognitive networks.

4.3.7.3 Channel portfolio size impact scenario

In this scenario, the impact of the channels portfolio size on the performances of cognitive ad hoc networks has been investigated. Now, are considered four CANs in a rectangular area where each CAN exchanges its initial position with the CAN in its opposite diagonal of the rectangle. All CAN uses initially the same operating channel and cross the same region in the centre of the rectangular during their movements.

In the four networks, every node has one TX/RX flow to/from another node in its network. All the flows have the same throughput and are of best effort class of QoS. Figure 4-34 illustrates the scenario.

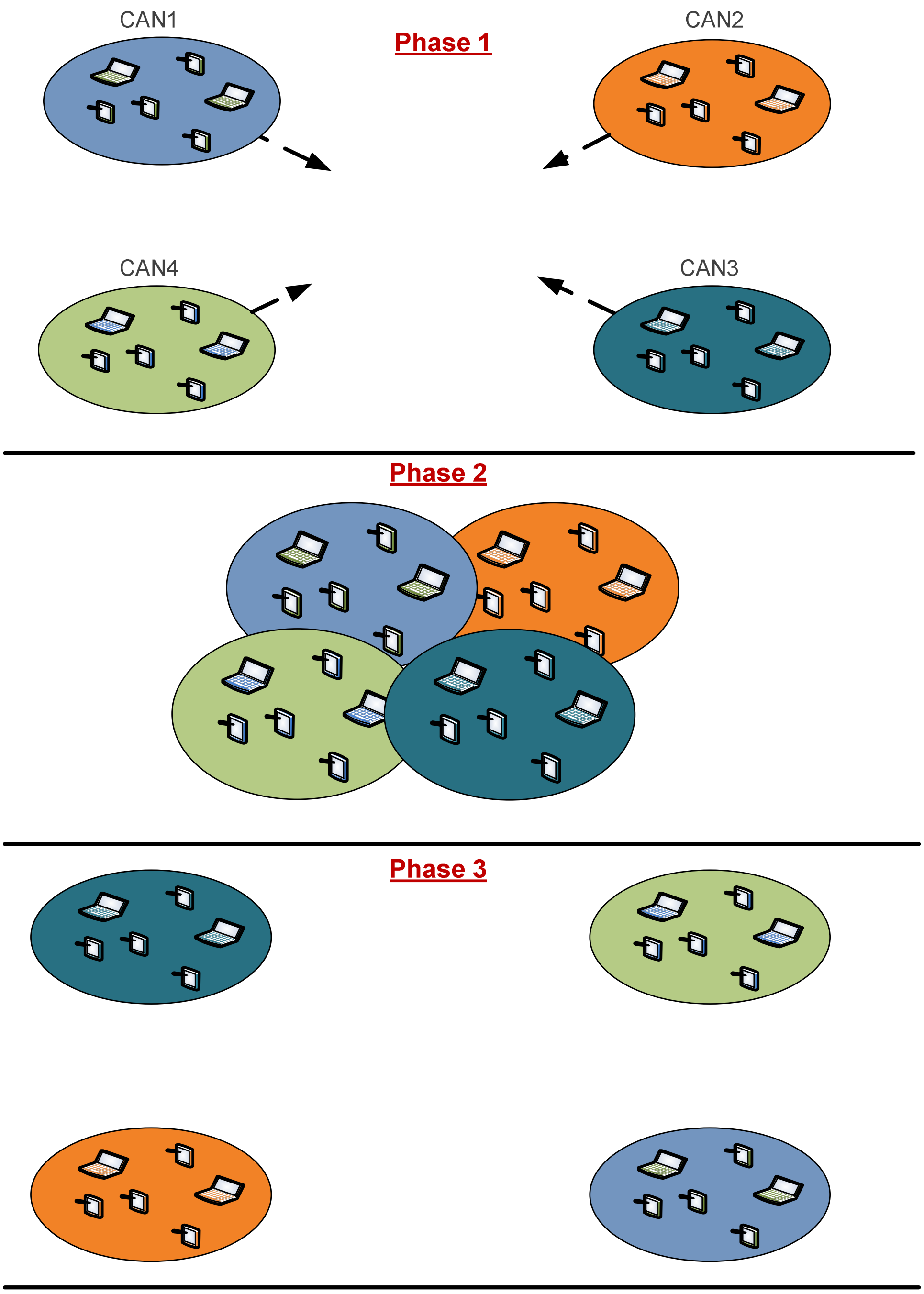


Figure 4-34: Channel portfolio size impact scenario.

Simulations are run considering two, four, and six channels in the portfolio.

Figure 4-35 shows the achieved throughput where only two channels are available for the CANs. It can be then observed that again, all CANs react immediately at the beginning of the interference phase and try to change their operating channels causing a slight decreases of throughput that last few seconds (5 s from t=30 s to t=35 s). The decrease is mainly due to channel change rather than to physical interference. Unfortunately, these changes are not sufficient and the mutual interferences keep increasing, causing this time continuous change of the operating channels by the CANs as no CH is able to find a dedicated (interference free) operating channel. This conflict is not resolved until the end of the interference phase.

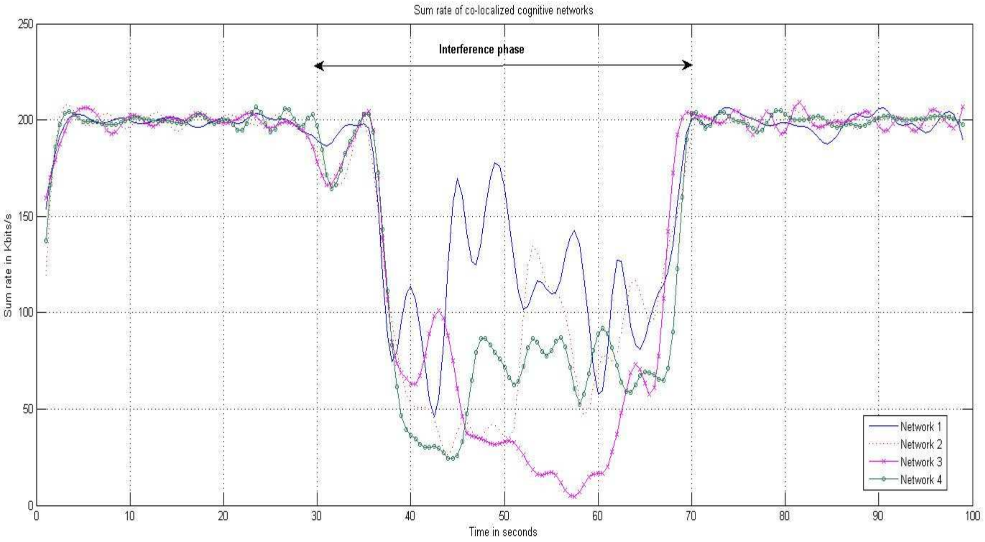


Figure 4-35: Sum rate of co-localised cognitive networks, portfolio of two channels.

Figure 4-36 depicts the achieved throughput where four channels are available for the 4 CANs. Here it is observed that the CANs react rapidly and adapt themselves to the interference situation by selecting each a separate operating channel in only four seconds after the beginning of the interference phase. This result is similar to those of the co-localised 2 CANs scenario of section 4.3.7.2.

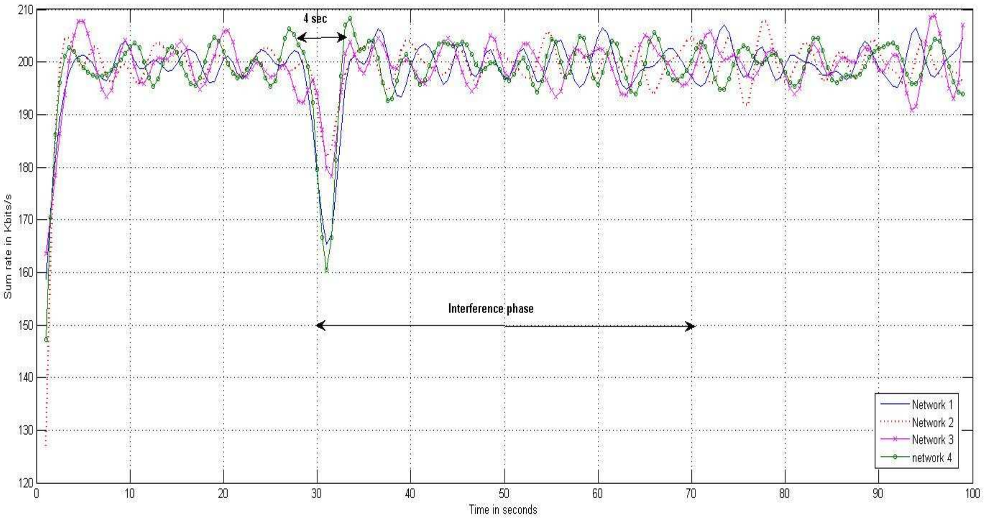


Figure 4-36: Sum rate of co-localised cognitive networks, portfolio of four channels.

Figure 4-37 shows the achieved throughput where six channels are available for the CANs. Surprisingly, the CANs undergo two throughput decrease phases and not only one, each of approximately four seconds, the first starting again immediately at the beginning of the interference phase.

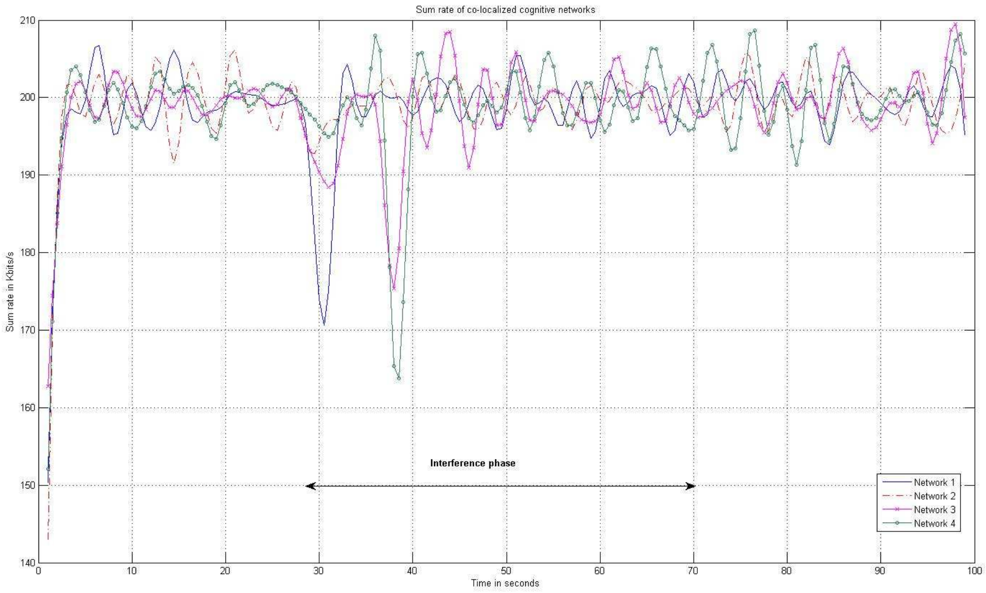


Figure 4-37: Sum rate of co-localised cognitive networks, portfolio of six channels.

To confirm this behaviour, Figure 4-38 is considered, where the performance of CANs sharing eight channels in the portfolio is shown. Here again it is observed that not only the two throughput's decrease regions, but also a slight amplification of these decreases. Thus, the richness of the channel portfolio has a negative impact on the performance of the CANs.

This can be explained as follows: during the channel selection and change procedure, temporary conflicts may occur in the channel acquisition phase as explained in section 4.3.6.

Thus, at the end of the first throughput decrease region, the CANs have all selected a separate operating channel after some eventual temporary collisions. As the portfolio size increases, the channel acquisition conflict probability decreases but also the probability that a given channel is not selected. The no initially selected channels are then better evaluated by the CHs who have by any chance undergo a collision in their past channels acquisition procedure. Nevertheless, the channel change procedure is not free of collision which explains the second throughput decrease region.

Hopefully, not all CAN are concerned by the second decrease region as also the probability for a given CAN to select a channel which is not selected by the others increase as the channel portfolio size increases. This phenomenon is observed in the last two figures (Figure 4-37, Figure 4-38) as only CAN3 and CAN4 are concerned with the second throughput decrease phase.

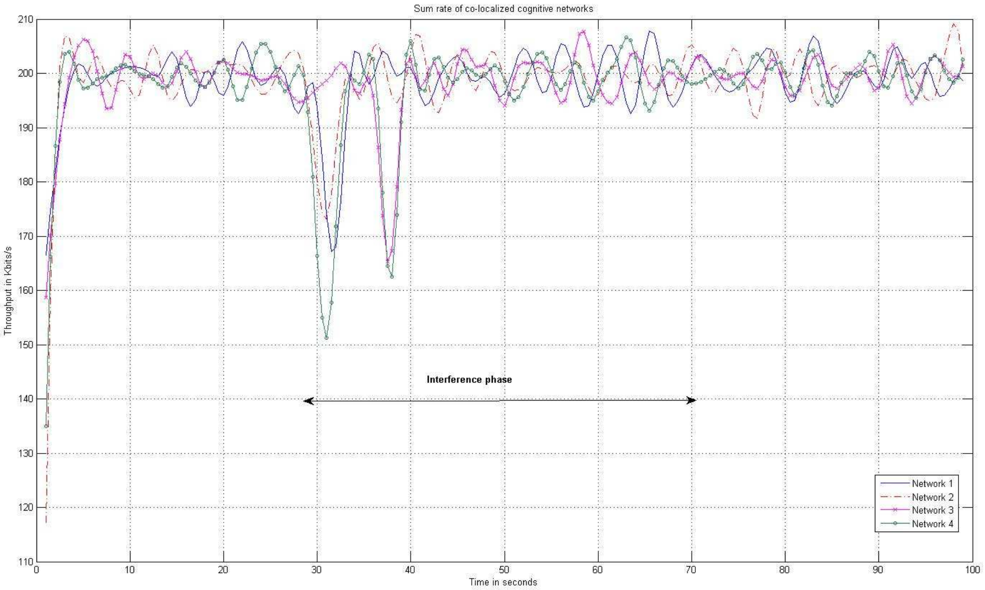


Figure 4-38: Sum rate of co-localised cognitive networks, portfolio of eight channels.

4.3.7.4 QoS support scenario

In this section the focus in on measuring how these cognitive ad hoc networks may support QoS traffic, and also how the presence of CANs does not degrade the QoS of incumbent networks.

Therefore, it is considered the scenario of section 4.3.7.1 where a CAN crosses the area of a fixed incumbent network. We suppose now that each node in each network is the source/destination of one best effort flow and one real-time flow to/from another node.

Figure 4-39 and Figure 4-40 show respectively the achieved sum rate of best effort traffic and real time traffic. As expected, both type of traffic in the incumbent network are not impacted by the presence of the cognitive network.

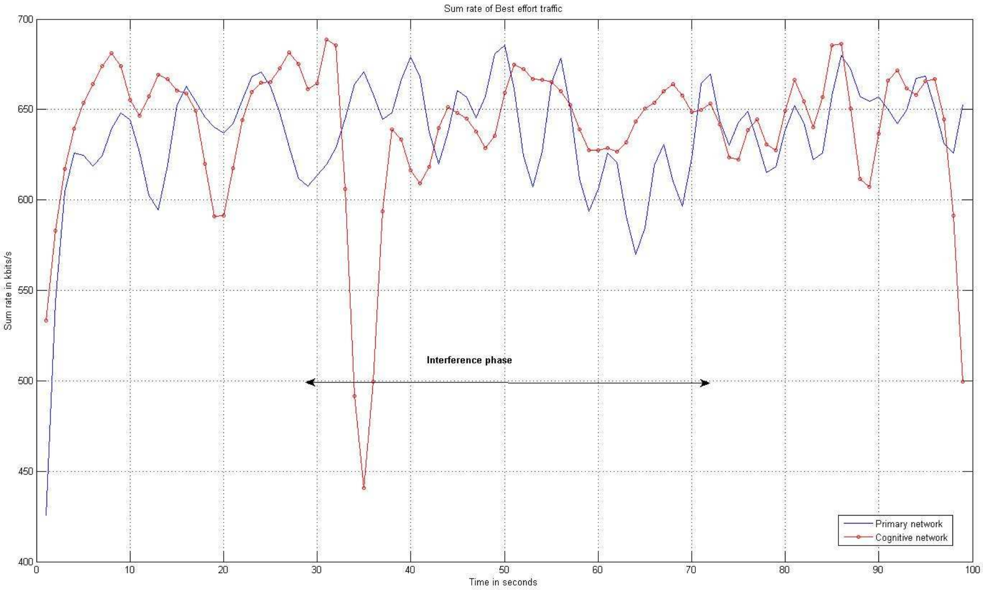


Figure 4-39: Incumbent protection: sum rate of best effort traffic in incumbent and cognitive networks.

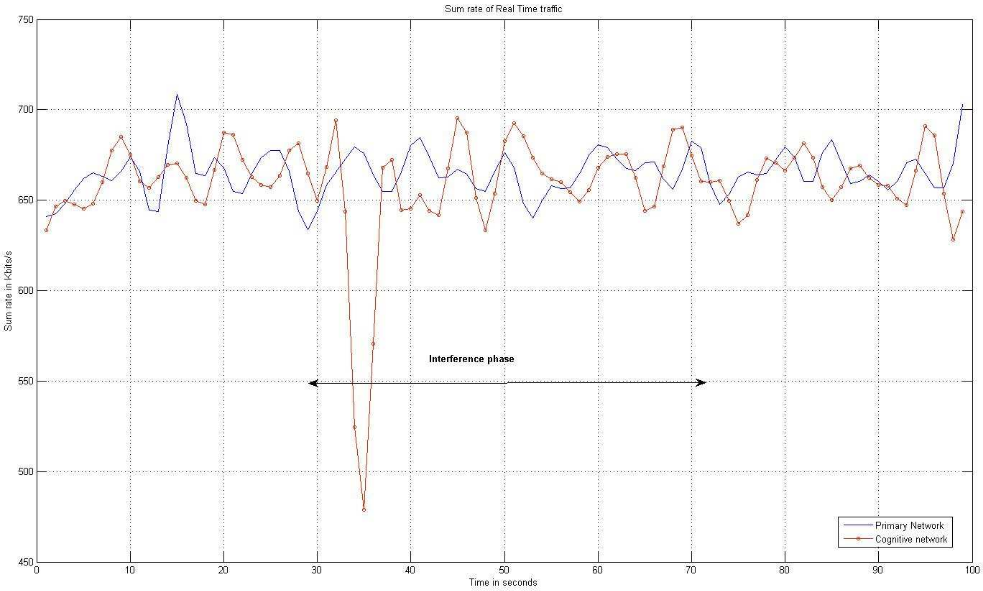


Figure 4-40: Incumbent protection: sum rate of real time traffic in incumbent and cognitive networks.

In Figure 4-41 is plotted the sum rate achieved by the cognitive network for each type of traffic. It is observed that both traffics undergo a deep decrease during channel change as cognitive nodes stop data transfers immediately after the detection of incumbent users on their operating channel. QoS support in this case in the cognitive network is quite difficult due to the presence of incumbent users.

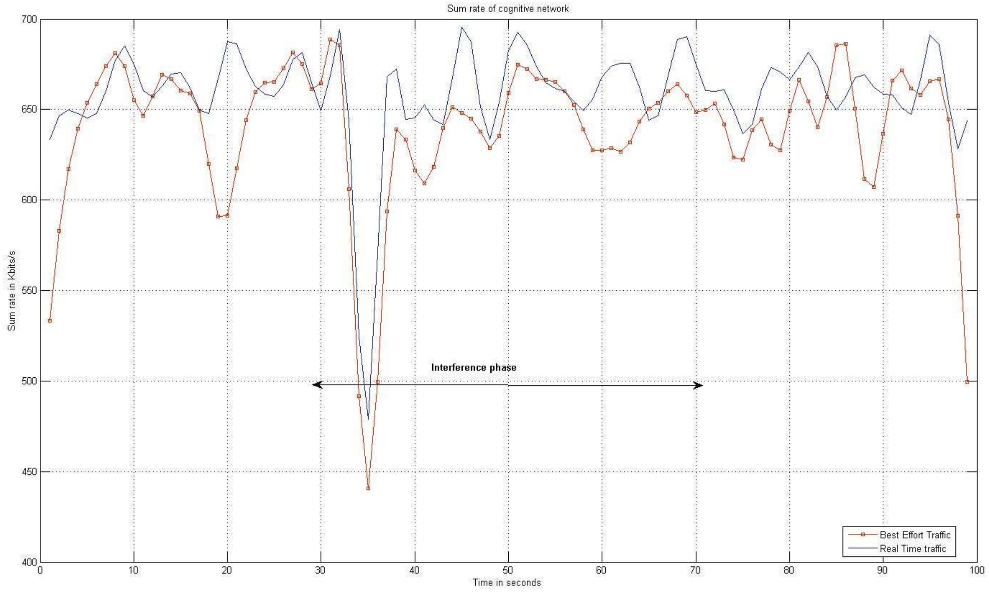


Figure 4-41: Incumbent protection: sum rate cognitive network.

Then the scenario of section 4.3.7.2 of temporary co-localised cognitive ad hoc networks is considered, and the same traffic hypotheses as previously are taken. In Figure 4-42 and Figure 4-43 are plotted respectively the sum rate of each type of traffic in the first and second CANs.

It is again observed that the two networks undergo a throughput decrease phase, but relatively less severe than in the case where incumbent users are presents. Moreover, RT flows seems to be less impacted than BE ones.

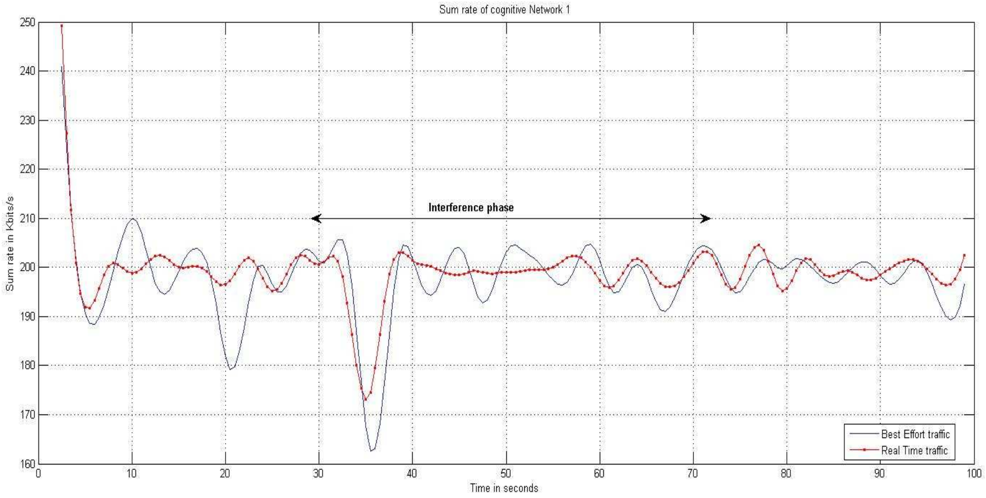


Figure 4-42: Co-localised CANs: sum rate of network #1.

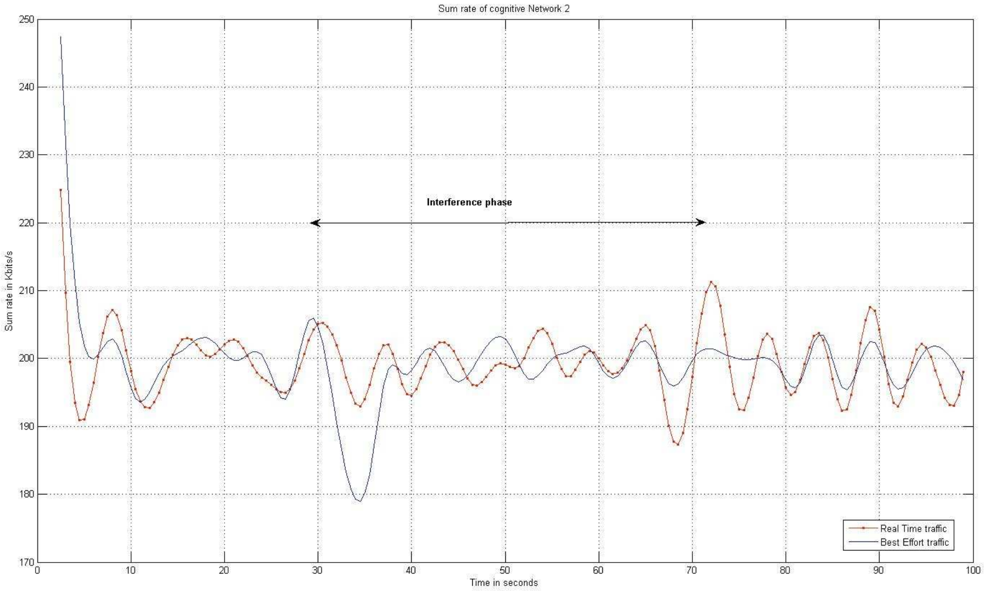


Figure 4-43: Co-localised CANs: sum rate of network #2.

5 Conclusions

The scope of this chapter is two-fold. From one hand, it provides concluding words on the work presented in this report, but putting the material into the overall framework laid down for the research methodology in WP5 [D5.1] and implemented by the WP tasks. On the other hand, it provides an overview of those results not reported in this public report, but presented either in the other restricted deliverables produced by WP5, or published in QoSMOS papers not cited in the present report.

5.1 Framework and cognitive manager for QoS and mobility support

In order to respond to the goals set, the research in WP5 is split into the elements treated in this and the following sections.

5.1.1 Framework and requirements

The first deliverable [D5.1] presents the research methodology set for the work to be done in WP5. The first step is the preparation of a framework to support and facilitate the required functionalities consisting in support of QoS in presence of incumbent appearance (spectrum mobility) and mobile terminals (physical mobility). Correspondingly, the requirements relevant to WP5 identified in [D5.1] and revised in [D5.2] [ManEtal2011] are divided into *QoS optimisation*, *mobility support* and *incumbent protection*. This work also contributed to the corresponding one done at system level and presented in [D1.4]. Stemming from these pre-requisites, the QoS and mobility functions are defined in [D5.1] and include access control, resource allocation and mobility management, as well as cognition and adaptation layer. Those are further discussed in [D5.2] with reference to the six QoSMOS scenarios initially identified[7] in [D1.2] [MacEtal2011]. The deliverable [D5.2] also includes a detailed discussion of the WP5 framework considering the end-to-end QoS with corresponding involved actors.

5.1.2 Cognitive manager for resource management and topological issues

The identified functionalities are rationalised into the internal functional blocks of the cognitive manager for resource management. The high-level functional blocks of the CM-RM are outlined in [D5.1]. Its evolution is tracked by an intermediate version presented in [ManEtal2011] and further detailed in [D5.2]. The architecture comprises, as seen here in [D5.3], a resource control (RC) and a resource use (RU) group, including access control resource allocation, mobility control and network domain cognition at one side and resource control support, resource exploitation and terminating domain internal blocks at the other side. The functional block groups are needed to ensure the flexibility of the system with the aid of the topological domains described in [CelEtal2011].A detailed description of the aforementioned internal functional blocks as well as of the external and internal interfaces in their final version is given here in [D5.3], while a light description of those is published in [LevEtal2012].

Some basic issues concerning context acquisition and decision-making are approached in [D5.1]. The interaction of CM-RM with the CM-SM, the cognitive manager for spectrum management, which is a form of context acquisition and exchange, is addressed first in [D5.2] and further developed here in [D5.3], with some highlights provided with [CelEtal2011].

The QoSMOS scenarios identified by WP1 [D1.2] [MacEtal2011] have been discussed within WP2 to identify the corresponding architecture options [D2.2] [D2.3]. The options for resource control topologies (centralised/distributed), together with architectural views considering topological domains [CelEtal2011], protocol layer, and plane (data/control/management) views are presented in [D5.2]. Topological issues are also discussed in [LevEtal2012]. This is reflected in the approach followed for

[7] The target scenarios have finally stabilised to a set of three, after a business analysis [LehEtal2012].

the mechanism to support mobility and manage QoS for the solutions presented in the present report [D5.3] and throughout the project. An important aid for the operations of the CM-RM as for all the remaining blocks of the QoSMOS system is the adaptation layer (AL) described in [D2.3] [D5.1] [D5.2] and outlined in [CelEtal2011].

5.2 Studies and solutions for the selected scenarios

Solutions to attain the set goals are presented for all the QoSMOS target scenarios. The illustration of the corresponding operations such as context acquisition, network set-up and reconfiguration, service establishment and management, resource allocation, vertical mobility management and eviction, also making use of message sequence charts, started with WP5 in [D5.2] and at project level with [D2.2]. Some examples are published in [ManEtal2011] [CelEtal2011] and [LevEtal2012].

5.2.1 Performance assessment

The metrics for evaluation of the performance of the solutions developed within WP2 are introduced in [D5.2]. Analysis of relevant medium access control (MAC) protocols is provided by use of a Markov chain model in [D5.2] and by simulations in [D5.3] with insights on protocol parameter setting. A Markov chain model is used also to assess the QoS of both incumbent and opportunistic users in presence of inaccurate spectrum sensing. The analysis concerning performance and complexity of the AL is provided here in [D5.3] on a number of aspects with a focus on QoS and mobility management, facilitated by the AL providing seamless connectivity between and enabling the system to react faster to changes in topology, avoiding annoying issues like a service dropping or a notable reduction in the signal.

5.2.2 Cellular extension in the whitespace

Spectrum mobility for the scenario of cellular extension in the whitespace is discussed in [LevEtal2012]. In this [D5.3], are explored access control options for an opportunistic cellular network to prevent network congestion while preserving the QoS of the accepted connections and protecting the incumbent operations. A new algorithm has been proposed, adding cognitive capabilities to the access control of a LTE cellular network in order to enable its operations in the TVWS band by using opportunistic radio resources. Its design has been described as well as the method for assessing its benefits. Then, simulation results have been provided which lead to the following conclusions.

The study has demonstrated that the incumbent apparition has a major impact on the system performance (this has been highlighted with connection blocking rate and connection dropping rate observation), both on the QoS and mobility aspects. It has proved the benefits of the solution by reducing/cancelling the observed impacts. It has highlighted the advantage to predict the incumbent apparition, even if the window size for taking the preventive actions is short: the algorithm is able to adapt to this situation. At last, it has shown the quality of experience for accessing to the service is preserved in comparison to conventional networks.

5.2.3 Cognitive femtocells

A basic overview of femtocells, with their background and history and potentials, is provided by [Zah2012]. Interference management is crucial in a successful deployment, and [Zah2012] shows that cognitive radio technologies can facilitate self-management of femtocell networks. This problem is addressed in more detail in [Zah2011] and in [D5.3] by proposing a downlink power control algorithm that gradually reduces the downlink transmit power to react to the knowledge about interference caused to a macrocell. Resource management for a cognitive femtocell scenario is presented in [LevEtal2012]. [GuoMoe2012] proposes an optimal QoS support policy for joint admission control, eviction control and bandwidth adaptation at a cognitive base station, to reduce the blocking and dropping probability of the cognitive users.

5.2.4 Cognitive ad hoc networks

A distributed power control (DPC) algorithm for cognitive ad hoc networks is initially outlined in [D5.1] and refined in [D5.2] [ManEtal2011] with simulation results. Cognitive DPC algorithms presented in [DurEtal2010] and further analysed in [DurEtal2011b], exploit the radio context information to identify a suitable power control strategy for protecting the incumbent whilst satisfying QoS of opportunistic users under a worst case analysis. An analysis of the effects of a time-variant channel and user mobility on the performance of such transmit power control schemes in terms of QoS of both incumbent and opportunistic users is provided by [DurEtal2011a] [DurEtal2012DurEtal2012] together with algorithms coping with those whilst a joint rate and power optimisation algorithm is presented in [RajEtal2010].Resource allocation algorithms for cognitive ad hoc networks are further discussed in [D5.2],], with an eye to the underlying physical layer and analysed by simulations here in [LevEtal2012] and in [D5.3], which shows that reliable transmission is possible even when radio resources are intermittent due to incumbent pre-emption. The adaptive scheduler and admission control protocols [YuEtal2010] reported in [D3.4] are suitable to distributed MAC and evaluated by simulation for an IEEE802.11e network, while [YuEtal2011] focuses on the reliable exchange of sensing information to serve a distributed admission control.

6 References

QoSMOS deliverables are listed first. The rest of the references, including those generated within QoSMOS, follow in alphabetical order.

[D1.2] R. MacKenzie, P. H. Lehne, U. Celentano, M. Ariyoshi, B. Cendón, A. Medela, “QoSMOS consolidated scenarios”, Deliverable D1.2 (PU), 31 p., 23 Dec 2010.

[D1.4] P. H. Lehne, U. Celentano, J. Lehtomäki, D. Noguet, R. Datta, P. Delahaye, R. Wansch, V. Berg, P. Grønsund, “QoSMOS consolidated system requirements”, Deliverable D1.4 (PU), 64 p., 31 Mar 2011.

[D2.1] M. Ariyoshi, K. Arshad, B. Bochow, U. Celentano, B. Cendón, R. Datta, P. Delahaye, J. Gebert, O. Grøndalen, P. H. Lehne, P. Marchand, K. Moessner, D. Noguet, G. Yuming, “Initial description of system architecture options for the QoSMOS system”, Deliverable D2.1 (RE), 79 p., 7 May 2010.

[D2.2] S. Leveil, C. Le Martret, O. Grøndalen, B. Bochow, C. Rosik, V. Mérat, B. Cendón, J. Herrero, A. Medela, U. Celentano, R. Datta, F. Noack, C. Lange, J. Gebert, K. Moessner, H.Sugahara, K. Muraoka, M. Ariyoshi, “System architecture options for the QoSMOS system”, Deliverable D2.2 (PU), 106 p., 24 Dec 2010.

[D2.3] G. Mange, P. Horváth, V. Berg, B. Bochow, R. Robles, M. Ariyoshi, C. Rosik, O. Grøndalen, P. H. Lehne, J. Rico, A. Medela, R. Datta, U. Celentano, “System specification and evaluation criteria”, Deliverable D2.3 (PU), 95 p., 30 Nov 2011.

[D3.4] U. Celentano, K. Muraoka, M. Ariyoshi, A. Bagayoko, D. Panaitopol, P. Delahaye, X. Yu, P. Navaratnam, K. Moessner, A. Gameiro, F. Kemeth, R. Wansch, D. Noguet, V. Berg, “Reference protocol stack for QoSMOS”, Deliverable D3.4 (PU), 56 p., 2 Apr 2012.

[D5.1] G. Mange, L. Csurgai, V. Mérat, C. Rosik, C. Le Martret, S. Leveil, B. Cendón, J. Herrero, K. Arshad, K. Moessner, U. Celentano, “Initial description of framework for supporting QoS and handling mobility”, Deliverable D5.1 (RE), 60 p., 31 Jul 2010.

[D5.2] G. Mange, P. Horváth, R. MacKenzie, K. Briggs, M. Fitch, C. Rosik, S. Leveil, C. Le Martret, R. Massin, P. Fouillot, A. Sanz, B. Cendón, A. Medela, J. Rico, K. Arshad, T. Guo, U. Celentano, ”Final framework description, preliminary cognitive manager structure and first mechanisms for QoS support”, Deliverable D5.2 (RE), 109 p., 13 Jul 2011.

[3GPP 23.203] Policy and charging control architecture

[ABI2007] ABI Research, Picochip, Airvana, IP access, Gartner, Telefonica Espana, *2nd Intl. Conf. Home Access Points and Femtocells*; available on line at: http://www.avrenevents.com/dallasfemto2007/purchase presentations.htm.

[Arc11] ARCEP, “La qualité des services de voix et de données des réseaux mobiles (2G et 3G) en France métropolitaine”, Nov 2011.

[ArsEtal10] K. Arshad and K. Moessner, “Mobility Driven Energy Detection based Spectrum Sensing Framework of a Cognitive Radio”. UKIWCS 2010, 13-14 Dec 2010, Delhi, India.

[AzaEtal2012] A. Azarfar, J.F. Frigon and B. Sansò, “Service differentiation in cognitive radio networks: A priority queuing model”, *IEEE Communications Surveys and Tutorials, volume 14, issue 2*, May 2012.

[AVC03] “Advanced video coding for generic audio-visual services,” ITU-T Recommendation H.264 and ISO/IEC 14496-10 AVC, 2003

[Bianchi04] G. Bianchi, “Performance analysis of the IEEE 802.11 distributed coordination function”, *IEEE Journal on Selected Areas in Communications*, vol.18, no.3, pp.535-547, Mar 2000.

[CacEtal11] A. S. Cacciapuoti, I. F. Akyildiz and L. Paura, “Primary-user mobility impact on spectrum sensing in cognitive radio networks”, *PIMRC* 2011.

[CelEtal2011] U. Celentano, B. Bochow, C. Lange, F. Noack, J. Herrero, B. Cendón, O. Grøndalen, V. Mérat, C. Rosik, “Flexible architecture for spectrum and resource management in the whitespace”, *Proc. Int. Symp. Wireless Personal Multimedia Commun.* (WPMC 2011), Brest, France, 3-7 Oct 2011.

[Chan2008] V. Chandrasekhar, J. Andrews, “Femtocell networks: A survey”, *IEEE Commun. Mag.*, vol. 46, no .9, pp. 59-67, Sep. 2008.

[Chan2009] V. Chandrasekhar, J.G. Andrews, T. Muharemovic, Zukang Shen, A. Gatherer, “Power control in two-tier femtocell networks”, *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4316-4328, Aug. 2009.

[COR11] Common Object Request Broker Architecture (CORBA/IIOP) v3.2 Nov 2011.

[DatEtal2011] R. Datta, G. Fettweis, Zs. Kollár and P. Horváth, “FBMC and GFDM interference cancellation schemes for flexible digital radio PHY design”, *Proceedings of the 14th EUROMICRO Conference* (Euromicro'11), Oulu, Finland, 2011.

[DufEtal05] K. Duffy, D. Malone, D.J. Leith, “Modeling the 802.11 distributed coordination function in non-saturated conditions”, *IEEE Communications Letters*, vol.9, no.8, pp. 715- 717, Aug 2005.

[DurEtal2010] O. Durowoju, K. Arshad, K. Moessner, “Distributed power control for cognitive radios with primary protection via spectrum sensing”, *IEEE VTC 2010 Fall*, Ottawa, Canada, Sept 2010.

[DurEtal2011a] O. Durowoju, K. Arshad K. Moessner, “Cognitive time variant power control in slow fading mobile channels”, *IEEE VTC 2011-Spring*, Budapest, Hungary.

[DurEtal2011b] O. Durowoju, K. Arshad, K. Moessner, “Distributed power control for cognitive radio networks, based on incumbent outage information”, *IEEE ICC 2011*, Kyoto, Japan.

[DurEtal2012] O. Durowoju, K. Arshad, K. Moessner, “Distributed power control algorithm for cognitive radios with primary protection via spectrum sensing under user mobility”, *Elsevier Journal of Ad Hoc Networks*, Special Issue: CRAHNs, 2012.

[FC08] FCC, “Second report and order”, FCC 08-260, Nov 2008. http://www.fcc.gov/.

[GLPK] GNU linear programming kit: http://www.gnu.org/software/glpk/.

[GuoMoe2012] T. Guo, K. Moessner, “Optimal strategy for QoS provision under spectrum mobility in cognitive radio networks, *VTC2012*, Quebec, Canada.

[Hos05] M. Hossam Ahmed, “Call admission control in wireless networks: A comprehensive survey”, *IEEE Communications Magazine*, Jan 2005.

[Hu11] N. Hu, "Investigations of radio behaviour ans security threats in cognitive radio networks". Stevens Institute of Technology, 2011. http://personal.stevens.edu/~nhu1/papers/dissertation.pdf

[JanEtal12] S. Jana, K. Zeng and P. Mohapatra, "Trusted collaborative spectrum sensing for mobile cognitive radio networks", *Infocom* 2012.

[JiEtal07] Z. Ji and K. J. Ray Liu, 'Dynamic spectrum sharing: A game theoretical overview", *IEEE Communications Magazine*, May 2007.

[Kim09] H. Kim and K. G. Shin, "Optimal admission and eviction control of secondary users at cognitive radio hotspots", *Proc. IEEE SECON'09*, 2009.

[KliEtal2009] A. Kliks, A. Zalonis, I. Dagres, A. Polydoros, H. Bogucka, "PHY abstraction methods for OFDM and NOFDM systems", *Journal of Telecommunications and Information Technology* (JTIT), 3/2009.

[LehEtal2012] P. H. Lehne, R. MacKenzie, O. Grøndalen, P. Grønsund, K. Briggs, "Business opportunities and scenarios for cognitive radio systems", QoSMOS whitepaper, Apr 2012. http://www.ict-qosmos.eu/

[LevEtal2012] S. Leveil, C. Martret, H. Anouar, K. Arshad, T. Zahir, J. Bito, U. Celentano, G. Mange, J. Rico And A. Medela, "Resource management of centrally controlled cognitive radio networks", *Proc. Future Network & Mobile Summit* (FuNeMS 2012), Berlin, Germany, 6-8 Jul 2012.

[MacEtal2011] R. MacKenzie, P. H. Lehne, U. Celentano, "Identifying scenarios with high potential for future cognitive radio networks", *Proc. Future Network & Mobile Summit* (FuNeMS 2011), Warsaw, Poland, 15-17 Jun 2011.

[MacEtal12] R. MacKenzie and K. Briggs, "Comparison of contention-based protocols for secondary access in TV whitespaces", *SDR'12-WInnComm-Europe*, 2012.

[ManEtal2011] G. Mange, C. Rosik, S. Leveil, U. Celentano, O. Durowoju, K. Arshad, "Cognitive resource management for QoS support in mobile opportunistic communications", *Proc. Future Network & Mobile Summit* (FuNeMS 2011), Warsaw, Poland, 15-17 Jun 2011.

[Mas10] R.Massin, C. Lamy-Bergot, C. J. LeMartret, and R. Fracchia, "OMNeT++-based cross-layer simulator for content transmission over wireless ad hoc networks", *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, article ID 502549.

[MinEtal09] A. W. Min and K. G. Shin, "Impact of mobility on spectrum sensing in cognitive radio networks", *CoRoNet* 2009.

[MinEtal11] A. W. Min, K.-H. Kim, J. P. Singh and K. G. Shin, "Opportunistic spectrum access for mobile cognitive radios", *Infocom* 2011.

[MuwEtal09] B. K. Muwonge and H. A. Chan, "Resource management of next generation networks using cognitive radio networks", *Intech* 2009. http://www.intechopen.com/books/cognitive-radio-systems/resource-management-of-next-generation-networks-using-cognitive-radio-networks

[NgoEtal11] D. T. Ngo and T. Le-Ngoc, "Distributed resource allocation for cognitive radio networks with spectrum-sharing constraints", *IEEE Transactions on Vehicular Technology*, 2011.

[Nic88] D. L. Nicholson, *Vulnerability of spread spectrum AJ and LPE communications sytems*, Computer Science Press, The George Washington University, Washington D.C., 1988.

[ParEtal12] S. Parakh and A. K. Jagannatham, “Optimal resource allocation and VCG auction-based pricing for H.264 scalable video quality maximization in 4G wireless systems”. *Advances in Multimedia*, Hindawi, vol. 2012, ID 567217, 13 pages. http://www.hindawi.com/journals/am/2012/567217/

[RajEtal2010] U. Rajeskaran, K. Arshad, K. Moessner, “Joint rate and power optimisation using distributed power control algorithms in cognitive radio networks”, *SDR 2010*, Washington, USA, Dec 2010.

[Red11] Y. B. Reddy, “Spectrum selection through resource management in cognitive environment”, *International Journal on Advances in Systems and Measurements*, vol. 4 no. 1 & 2, year 2011, http://www.iariajournals.org/systems_and_measurements/

[Ros89] K. W. Ross and D. H. K. Tsang, “Optimal circuit access policies in an ISDN environment: A Markov decision approach”, *IEEE Trans. Commun.*, vol. 37, no. 9, pp. 934–939, 1989.

[TanEtal12] L. Tang and J. Wu, “Research and analysis on cognitive radio network security”. *Wireless Sensor Network*, 2012, 4, 120-126, April 2012.

[TemEtal08] H. Tembine, E. Altman, R. ElAzouzi and Y. Hayel, “Stable networking games”. *Forty-Sixth Annual Allerton Conference*, Allerton House, UIUC, Illinois, USA, September 23-26, 2008.

[Var] A. Varga, “OMNeT++ discrete event simulation system”, http://www.omnetpp.org/.

[Var10] J. Vartiainen, Concentrated signal extraction using consecutive mean excision algorithms, PhD thesis, University of Oulu, 2010. http://jultika.oulu.fi/Record/isbn978-951-42-6349-1

[VarEtal12] J. Vartiainen, J. Lehtomäki and R. Vuohtoniemi, “The LAD Methods in WLAN Indoor Multipath Channels”, *Crowncom* 2012.

[WanEtal10] W. Wang, H. Li, Y. Sun and Z. Han, “Securing collaborative spectrum sensing against untrustworthy secondary users in cognitive radio networks”, *EURASIP Journal on Advances in Signal Processing*, vol. 2010.

[WanEtal11] B. Wang and K. J. Ray Liu, “Advances in cognitive radio networks: A survey”, *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, Feb 2011.

[WelEtal08] M. Wellens, A. de Baynast and P. Mähönen, “Exploiting historical spectrum occupancy information for adaptive spectrum sensing”, *WCNC* 2008.

[Xia01] Y. Xiao, P. Chen, and Y. Wang, “Optimal admission control for multiclass of wireless adaptive multimedia services”, *IEICE Trans. Commun.*, vol. E84-B, no. 4, pp. 795–804, 2001.

[XML99] XML-RPC Specification, D. Winer, Jun, 1999.

[YuEtal2010] X. Yu, P. Navaratnam, K. Moessner, “Distributed resource reservation mechanism for IEEE 802.11e-based networks”, *IEEE VTC 2010 Fall*, Ottawa, Canada, Sept. 2010.

[YuEtal2011] X. Yu, P. Navaratnam, K. Moessner , “Distributed resource reservation for real time sessions in multi-hop wireless networks”, *7th International Wireless Communications and Mobile Computing Conference* (IWCMC 2011), Istanbul, Turkey, 5-8 July 2011.

[Zah2011] T. Zahir, K. Arshad, K. Youngwook, K. Moessner, “A downlink power control scheme for interference avoidance in femtocells”, *7th International Wireless Communications and Mobile Computing Conference* (IWCMC), 2011, pp.1222-1226, 4-8 July 2011.

[Zah2012] T. Zahir, K. Arshad, A. Nakata, K.Moessner,”Interference management in femtocells”, *IEEE Communications Surveys & Tutorials*, vol. 99, pp.1-19, 2012.