Université du Québec

Institut national de la recherche scientifique

Énergie, Matériaux et Télécommunications

Modeling, Design and Analysis of Multi-Channel Cognitive Radio Networks

By

Navid Tadayon

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Ph.D.) in Telecommunicationss

External examiner:	Prof. Mustafa K. Mehmet Ali
	Concordia University
External examiner:	Prof. Benoit Champagne
	McGill University
Internal examiner:	Prof. Douglas O'Shaughnessy
	INRS University
Research director:	Prof. Sonia Aïssa
	INRS University

August 26, 2016

Abstract

This dissertation offers a comprehensive treatment of cognitive radio (CR) as an elemental technology capable of improving the spectral inadequacy problem in future generations of wireless networks. As the first milestone in this PhD dissertation, mathematical models are established as the abstract descriptions of CR networks (CRNs). In the realm of wireless communications, such modeling is the skeleton for any ensuing action plan. These models help network planners to comprehend and visualize different aspects of CRNs before deployment. Aside from shedding insight into our understanding, expressions derived from the analytical models can be incorporated at the core of the algorithms and mechanisms for CRNs.

The proposed models in this dissertation are stable and based on well-defined methodologies. The modeling research objectives are sought from three different perspectives. From the first perspective, the efforts are focused on both single-interface and multi-interface multi-channel CRNs (MC-CRNs) with emphasis on generic settings as well as WRANs. From the second perspective, four different issues are worked out, namely, modeling, analysis, synthesis and optimality. Finally, from the last perspective, the MC-CRNs are explored through numerical modeling, closed-form analysis, and simulation validation.

Despite its importance, investigations in the area of designing mechanisms for MC-CRNs have been scarce. In filling this void, the second part of this thesis is devoted to designing smart, efficient, and scalable mechanisms for MC-CRNs. The research work in this part is extensive as it embodies spectrum sensing, decision, mobility, and sharing problems. The first three problems are tackled by proposing an efficient spectrum distributed sensing and fusion mechanism (SFM) for CR based IEEE 802.22 WRANs. The proposed mechanism avoids most of the shortcomings of the SFMs that are featured by the standard, complies with the standard directives, and can be easily implemented into WRANs with minimal changes.

The last chapter deals with the problem of spectrum sharing. Built upon the recently-emerged theory of optimal multi-unit auctions, a spectrum sharing mechanism is introduced. The proposal resolves several deep-rooted issues in MC-CRNs. On the user side, the fairness-optimality trade-off is settled, whereas on the network operator side the total collected revenue is maximized through the imposition of payments. To diminish uncertainties in the decision making process and improve the stability of the resource assignments, the aforementioned payment rule is designed to induce truthfulness among self-centered and rational users. These properties, along with the computational manageability of the formulated problem, as well as the existence and stability of its solution, renders the proposed spectrum sharing mechanism an ideal scheme for exploitation in practice.

Navid Tadayon

Sonia Aïssa

Acknowledgements

I would like to express my most heartfelt gratitude to my PhD supervisor, Prof. *Sonia Aïssa* for her support and mentorship in this research journey. She helped me develop a critical way of thinking about my work and taught me how to aim high. Special recognition goes to the government of Quebec for valuing my work and granting me the prestigious FRQNT merit scholarship. It is also a great privilege for me to have prominent figures like Prof. *Mehmet Ali* from Concordia University, Prof. *Champagne* from McGill University and Prof. *O'Shaughnessy* from INRS as my jury composition.

I will always remain indebted to my *parents* for being a perpetual source of encouragement and positivity to me. I am grateful of my beloved wife, *Ellie*, whose unwavering love, support and patience cannot be honored with words. A special gratitude goes to my two sisters, *Negar* and *Neda*, for helping me become a better person.

The completion of this undertaking would not have been possible without the moral and intellectual support of people from near and far. I am thankful to all of them: First, the friend of my childhood, adolescence and adulthood, *Arash*, whose endless presence in my heart never left me feel the void of a brother I never had. Second, my endearing friends, *Sara*, *Safa*, and *Omid* for being a great source of inspiration and morality to me. I remain thankful of Prof. *Xing* and Prof. *Zarrillo* for their ceaseless guidance and support after I graduated from University of Massachusetts Dartmouth. To my good friend *Georges Kaddoum* for his encouragements and proving to me how ethical and honest a young professor can be. Special thanks goes to my friend *Serge-fabrice* for helping me to prepare the french summary of this dissertation. This list can go on and on and I still won't be able to finish naming all those who played an influential role in my life. I am thankful to each and every one of them.

Once the great Isaac Newton truly said, "*If I have seen a little further, it is by standing on the shoulders of giants*". I wholeheartedly believe in that.

To all those
who taught me how to think

Contents

1 Introduction and Objectives **2**

1.1 Introduction to Cognitive Radio 2

1.2 Motivations and Visions 5

1.3 Research Objectives 11

1.3.1 Modeling and Performance Analysis 11

1.3.2 Mechanism Design 14

1.4 Research Methodology 16

1.4.1 Modeling and Performance Analysis 16

1.4.2 Mechanism Design 17

I Modeling, Analysis and Synthesis **20**

2 Single-Interface CRNs: Priority-Queue Analysis **24**

2.1 Literature Review 25

2.2 Model Development 27

2.2.1 Queue Decoupling 28

2.2.2 CTMC Representation 30

2.3 Network Synthesis 34

2.4 Simulation Development 37

2.4.1 Queue Unfinished Work . . . 37
2.4.2 Queue Length Histogram . . . 37
2.4.3 Validation . . . 39
2.5 Achievable Region . . . 40
2.6 Mixed Strategy in CRNs . . . 43
2.7 Model Refinement . . . 46
2.7.1 Sensing Time . . . 46
2.7.2 Sensing and Channel Imperfections . . . 47
2.8 Summary . . . 48

3 Single-Interface CRNs: Queuing Network's Approach 50
3.1 IEEE 802.22 Centralized Access Mechanism . . . 52
3.1.1 Cell Association . . . 52
3.1.2 Sensing and Incumbent Protection . . . 53
3.1.3 Access Mechanism and Frame Structure . . . 55
3.2 IEEE 802.22 Modeling . . . 59
3.2.1 WRAN Cell Modeling: Network of Tandem Queues . . . 61
3.2.2 Server Characterization and Rate Determination . . . 63
3.2.3 Modeling at the MAC Layer . . . 64
3.3 Closed-Form Probability Density Function . . . 69
3.3.1 Network-Wide PMF . . . 74
3.3.2 Numerical Results . . . 75
3.3.3 How to Choose q_i? . . . 75
3.4 Cell-Level Analysis . . . 78
3.4.1 Cell Stability Condition (SC) . . . 78
3.4.2 Peer-to-Peer Delay . . . 79
3.5 Summary . . . 83

4 Multi-Interface Cognitive Radio Networks: MAC Design and Modeling 84
4.1 Multi-Interface Multi-Channel Cognitive Radio Network 87
4.1.1 Protocol Description and Network Model 87
4.2 Modeling Framework 91
4.2.1 Primary Channel Characterization 91
4.2.2 Decoupling of Secondary Users 92
4.2.3 Continuous-Time Markov Chain 93
4.2.4 Case Study 96
4.2.5 Stability Analysis 97
4.2.6 Energy Sensing and Channel Busyness Probability 100
4.2.7 Rate Determination 104
4.3 Secondary's Queue Length 106
4.3.1 Numerical Analysis 106
4.3.2 Statistical Analysis and Heavy-Regime Characteristics 110
4.3.3 Closed-form Analysis 112
4.4 Simulation Validation 115
4.5 Summary 116

II Mechanism Design 118

5 Spectrum Sensing, Decision and Mobility 120
5.1 Spectrum Management in IEEE 802.22 122
5.1.1 Literature Review 122
5.1.2 Channel Management 126
5.1.3 Normal Mode vs. Coexistence Mode 130
5.1.4 Quiet Period Management 131
5.2 The Proposed Multi-channel Sensing Fusion Mechanism 132

5.2.1 Decision Binarization . . . 133
5.2.2 Decision Combination . . . 138
5.3 Simulation and Validation Results . . . 144
5.3.1 Simulator Platform . . . 144
5.3.2 Performance Measures . . . 145
5.3.3 Comparative Results . . . 148
5.3.4 Adaptive Tuning . . . 152
5.4 Summary . . . 153

6 Spectrum Sharing **154**
6.1 Background . . . 157
6.1.1 Literature Review . . . 157
6.1.2 Fundamentals of Game Theory . . . 158
6.2 Spectrum Pricing and Utility Derivation . . . 162
6.2.1 Demand Profile and Utility Function . . . 164
6.2.2 Problem Formulation . . . 166
6.2.3 Secondary User's Valuation . . . 170
6.2.4 Secondary Operator's Perspective . . . 176
6.2.5 Nonlinear Pricing Rule . . . 177
6.2.6 Pricing Rule: Digest of Findings . . . 179
6.3 Spectrum Assignment . . . 182
6.3.1 Problem Restatement . . . 182
6.3.2 Revenue Maximization . . . 184
6.3.3 Convexity of the Allocation Problem . . . 186
6.3.4 Solution of the Allocation Problem . . . 188
6.3.5 Allocation Mechanism: Digest of Findings . . . 189
6.4 Application Scenarios and Case Study . . . 195

6.4.1 Allocation in Prioritized Access Networks . . . 195
6.4.2 Application in LTE and WRANs . . . 197
6.4.3 Case Study . . . 201
6.5 Summary . . . 204

7 Thesis Conclusion **206**
7.1 Publications . . . 212

References **214**

III Résumé Français **214**

A Introduction et Objectifs **216**
A.1 Introduction à la Radio Cognitive . . . 216
A.2 Motivation et Vision . . . 217
A.3 Objectifs de Recherches . . . 220
A.3.1 Modélisation et Analyse de Performances . . . 220
A.3.2 Conception de mécanisme . . . 221
A.4 Methodologie . . . 222
A.4.1 Modélisation de Performance . . . 222
A.4.2 Conception du Mécanisme . . . 223

B Modelisation, Analyse et Synthèse **226**
B.1 CRNs à Interface-Unique: Analyse Priority-Queue . . . 226
B.1.1 Modèle de Développement . . . 227
B.1.2 Synthèse du Réseau . . . 230

C Conception du Mecanisme **232**
C.1 Mecanisme de Detection . . . 232

C.1.1 IEEE 802.22 Spectrum Sensing Fonction 233
C.1.2 Détection Distribuée Basée sur L'Apprentissage Multi-canal 234
C.2 Mecanisme d'allocation de Ressource 237
C.3 Conclusion 241

List of Figures

1.1 Some 5G applications. 4
1.2 5G challenges and enabling technologies. 5
1.3 Identified challenges in the realm of CR research. 9
1.4 The main CR components. 10
1.5 My research objectives. 12
1.6 The research methodology. 17

2.1 Modeling the interactive behavior of SUs and PUs 29
2.2 Representation of the multi-channel CRN with 2D CTMC. 30
2.3 Joint PMF of the secondary and primary classes' queue lengths. 32
2.4 2D view of the marginal PMF 33
2.5 Numerical, closed-form, and simulation results-average total delays 34
2.6 Amount of unfinished work in secondary's queue vs. simulation time. . . . 37
2.7 Amount of unfinished work in primary queues. 38
2.8 Queue length histograms. 39
2.9 Achievable regions in CRNs with and without constraints. 41
2.10 Cost-function for different values of parameters 45

3.1 An IEEE 802.22 scenario. 51
3.2 IEEE 802.22 *frame* and *superframe* structures. 57

3.3 Queuing model of an IEEE 802.22 cell . . . 59
3.4 2D CTMC representation for the i^{th} tagged CPE. . . . 65
3.5 The marginal CTMC for the number of subchannels. . . . 68
3.6 A demonstration for the accuracy of the PMF . . . 74
3.7 PMF of the number of packets . . . 76
3.8 PMF of the number of packets in a CPE . . . 77
3.9 Peer-to-peer delay vs. CPE subchannel access probability . . . 82

4.1 Proposed network model for MIMC-CRN . . . 88
4.2 Equivalent queue representation . . . 91
4.3 Illustration of the queue decoupling logic in the network. . . . 94
4.4 2D CTMC modeling the dynamics of aN SU in the network. . . . 95
4.5 CDF and PMF as a function of the SU's queue length . . . 108
4.6 Coefficient of variation . . . 109
4.7 Validation results. . . . 113
4.8 The unfinished work in the secondary queue. . . . 116

5.1 Known spectrum sensing methods. . . . 122
5.2 Example of a WRAN . . . 127
5.3 IEEE 802.22 channel transition diagram . . . 130
5.4 Schematic illustration of the updating scheme . . . 132
5.5 Sensor's decision binarization . . . 135
5.6 Spatial footprint of TV incumbents . . . 140
5.7 The MC-LDS block diagram. . . . 142
5.8 Performance comparison-I . . . 149
5.9 Performance comparison-II . . . 150

6.1 Analogy between games and DSA . . . 156

6.2 platform-independent nature of auctions . . . 163
6.3 The axioms of the proposed spectrum allocation mechanism. . . . 164
6.4 Demand profile . . . 165
6.5 Resource allocation block diagram . . . 180
6.6 The axioms of the proposed spectrum allocation mechanism. . . . 183
6.7 Auction-based resource allocation block diagram. . . . 191
6.8 Illustration of the existence of fixed-point in (6.56). . . . 193
6.9 Application of the allocation mechanism to LTE . . . 198
6.10 Modified IEEE 802.22 frame structure . . . 200
6.11 Simulation experiments for $M = 2$ SUs. . . . 202
6.12 Total revenue collected by the SO from two SUs . . . 203

B.1 Modélisation du comportement interactif des SUs et PUs . . . 227
B.2 Représentation du CRN multi-canal avec CTMC 2-D. . . . 229
B.3 Numérique, forme fermée, et les résultats de simulation. . . . 230

C.1 Diagramme de transition de la liste canal. . . . 233
C.2 Un aperçu des 12 WRAN co-existantes. . . . 234
C.3 Allocation de la bande passante aux enchères pour MC-CRNs. . . . 240

Abbreviations

2D: Two Dimensional

AWGN: Additive White Gaussian Noise

BCL: Backup Channel List

BNE: Bayesian Nash Equilibrium

BR: Best Response

BS: Base Station

BW: Bandwidth

CBP: Coexistence Beacon Protocol

CCC: Common Control Channel

CCL: Candidate Channel List

CDF: Cumulative Distribution Function

CE: Consumer Electronics

CM: Computational Manageability

CPE: Customer Premises Equipment

CQI: Channel Quality Indicator

CR: Cognitive Radio

CRN: Cognitive Radio Network

CSI: Channel State Information

CTMC: Continuous Time Markov Chain

CoV: Coefficient of Variation

DB: Database

DS: Downstream

DS-MAP: Downstream Map

DSA: Dynamic Spectrum Allocation

FD: Full-Duplex

FIFO: First-in First-out

HD: Half-Duplex

HDC: Hard Decision Combining

HPQ: High Priority Queue

IAF: Incumbent Alteration Frequency

IAR: Incumbent Activity Ratio

IBS: In-band Sensing

IC: Incentive Compatibility

ICI: Inter-carrier Interference

IR: Individual Rationality

LHS: Left-hand-side

ISI: Inter-Symbol Interference

LoS: Line-of-Sight

LPQ: Low Priority Queue

LTE: Long Terminal Evolution

LTR: Low Traffic Regime

MAC: Medium Access Layer

MC-CRN: Multi-Channel CRN

MC-LDS: Multi-Channel Learning-based Distributed Sensing

MECE: Mutually Exclusive and Collectively Exhaustive

MIMC-CRN: Multi-Interface Multi-Channel CRN

NE: Nash Equilibrium

NWCF: Network Wide Correlation Factor

OBS: Out-of-band Sensing

OCL: Operating Channel List

OFDMA: Orthogonal Frequency Division Multiplexing Access

PCL: Protected Channel List

PDF: Probability Distribution Function

PGF: Probability Generating Function

PHY: Physical Layer

PMF: Probability Mass Function

PMI: Precoding Matrix Indicator

PRB: Physical Resource Block

PU: Primary User

PUSCH: Physical Uplink Shared Channel

QoS: Quality of Service

QP: Quiet Period

RHS: Right-hand-side

RM: Revenue Maximization

RS: Reference Signal

RV: Random Variable

SC: Stability Condition

SC-LDS: Single-Channel Learning-based Distributed Sensing

SCH: Superframe Control Header

SCW: Self-Coexistence Window

SDC: Soft Decision Combining

SER: Symbol Error Rate

SNR: Signal-to-Noise Ratio

SO: Secondary Operator

SSF: Spectrum Sensing Function

TDD: Time Division Duplexing

TDMA: Time Division Multiplexing Access

SU: Secondary User

UCS: Urgent Coexistence Situation

UE: User Equipment

US: Upstream

US-MAP: Upstream Map

WRAN: Wireless Regional Area Network

Nomenclature

Note *Although the following notation system is adopted in this thesis, there are several instances of overlapping usage of same symbols in distinct chapters. Whenever this is the case, a footnote is immediately added to draw the distinction.*

M, m	$\triangleq$	Number of SUs (CPEs)
$\mathcal{M}$	$\triangleq$	Set of SUs (CPEs)
N, n, $\mathfrak{B}$	$\triangleq$	Number of (sub)Channels
T_x	$\triangleq$	Length of Interval x or a Cycle
$\mathcal{S}$	$\triangleq$	Received Power
$\mathcal{S}^{(\mathrm{Tr.})}$	$\triangleq$	Transmit Power
$\mathcal{S}^{\min}$	$\triangleq$	Power Threshold
$\mathrm{S(I)NR}$	$\triangleq$	Received S(I)NR
$\mathrm{S(I)NR}^{\min}$	$\triangleq$	minimum S(I)NR
$\mathcal{S}_N$	$\triangleq$	Noise Power
$\mathtt{W}$	$\triangleq$	Signal Bandwidth
$\mathtt{N}_0$	$\triangleq$	Noise Spectral Density
$\mathcal{G}(z)$	$\triangleq$	PGF
$f_X(x)/\mathcal{F}_X(x)$	$\triangleq$	PDF/CDF of RV X
$\Pr(X=x)=P_x$	$\triangleq$	Probability of Event x
$P_{\mathrm{S,BS}}$	$\triangleq$	Successful Transmission Probability

P_{Out}	$\triangleq$	Outage Probability
P_{FA}	$\triangleq$	False Alarm Probability
P_{D}	$\triangleq$	PU Detection probability
$P^{\mathrm{Sen.}}$	$\triangleq$	Channel Sensing Probability
P_{B}	$\triangleq$	Probability that channel is Perceived Busy
$P_{\mathrm{PU}}^{\mathrm{Active}}$	$\triangleq$	Probability that PU is Active (Channel is Really Busy)
$P_{\mathrm{SU}}^{\mathrm{misSync}}$	$\triangleq$	Probability that SUs are Out of Sync in Sensing
q	$\triangleq$	Subchannel Allocation Probability
$\overline{W}_{\mathrm{PU}}$	$\triangleq$	PU Tolerable Waiting Time
$\overline{W}_{\mathrm{SU}}$	$\triangleq$	SU Tolerable Waiting Time
P_{PL}	$\triangleq$	Packet Loss Probability
$\pi(x, y)$	$\triangleq$	Probability of being in state (x, y) (of a CTMC)
μ	$\triangleq$	Queue Service Rate
λ	$\triangleq$	Queue Arrival Rate (Fresh Packets)
Λ	$\triangleq$	Gross (Net) Arrival Rate (Circulating+Fresh Packets)
ρ	$\triangleq$	Queue Utilization Factor
W	$\triangleq$	Queueing (Waiting) Delay
W^{US}	$\triangleq$	Queueing Delay in Upstream
W^{DS}	$\triangleq$	Queueing Delay in Downstream
D_{access}	$\triangleq$	Access Delay
D	$\triangleq$	Total Delay (Queueing+Service) a.k.a. Response Time
D_{BB}	$\triangleq$	Backbone Delay
D_{PtP}	$\triangleq$	Peer-to-Peer delay
$\mathtt{L}$	$\triangleq$	Queue Length
$\overline{\mathtt{L}}$	$\triangleq$	Average Queue Length
$h(n)$	$\triangleq$	Discretized Fading Process

$\overline{h}$	$\triangleq$	Average Channel Fading Gain
ϵ^{DS}	$\triangleq$	Fraction of Frames Spent on DS
$\epsilon^{\mathrm{sen.}}$	$\triangleq$	Fraction of Frames Spent on Sensing
$\mathcal{B}$	$\triangleq$	Samples Taken from Analog Signal
$\mathcal{H}_0/\mathcal{H}_1$	$\triangleq$	Null/Alternative Primary Channel Hypothesis
$r_{c,\mathrm{BS}}$	$\triangleq$	Signal Coding Rate
$\mathtt{M}$	$\triangleq$	Signal Constellation Size
$\mathfrak{K}$	$\triangleq$	Coefficient (when needed)
$d^{(t)}$	$\triangleq$	Binary CPE's Local Sensing Decision at Time t
$D^{(t)}$	$\triangleq$	Binary WRAN's Global Sensing Decision at Time t
$R^{(t)}$	$\triangleq$	Binary Channel State's DB Reading at Time t
$L^{(t)}$	$\triangleq$	Reward/Penalty Scores at Time t
$X^{(t)}$	$\triangleq$	Continuous Sensing Channel Indicator at Time t
$\mathcal{Z}^{(t)}$	$\triangleq$	Binary Factual Sate of Channel at Time t
$u(\cdot)$	$\triangleq$	Utility of a SU
$\mathcal{U}^{\mathbf{P}}$	$\triangleq$	Total Revenue Collected by the SO
V	$\triangleq$	Valuation Space of a SU
$\Omega(\cdot)$	$\triangleq$	Hazard Rate of Belief
$\mathcal{D}(\cdot,\cdot)$	$\triangleq$	Demand Profile
$\mathcal{R}(\cdot)$	$\triangleq$	Revenue Function of a SU
$\mathcal{K}$	$\triangleq$	Average Spectral Efficiency of a SU
$\mathcal{L}(\cdot)$	$\triangleq$	Likelihood Function
w^*	$\triangleq$	Optimal Allocation of Bandwidth to a SU
C^*	$\triangleq$	Optimal Charge Levied on a SU

Chapter 1

Introduction and Objectives

1.1 Introduction to Cognitive Radio

The proliferation of wireless technologies in all aspects of modern life is a reality. With more than three billions of them in use today, the number of wireless devices is predicted to surpass 25 billions by 2025 [1]. The triggering impetus for this acceleration can be directly attributed to the ever-increasing demand for fast, reliable, machine-centric and boundless communications (Fig. 1.1). Merged with the parallel advancements in new, scalable software architectures and consumer-electronic (CE) fabrication, wireless communications, nowadays, calls for highly capable carrier networks. These networks should have extensive coverage and extremely high speed. More importantly, they should be traffic-agnostic. The latter requirement has important implications. The hundreds of Exabyte of traffic that is expected to inundate wireless networks in the near future originates from a different type of subscribers we have known so far, that is, Machines. These machines can be anything from our garage door controllers, vacuum cleaners, cars to city lamp posts, power meters, traffic lights, environmental sensors, etc. This vision promises an automated life. Nevertheless, there is a big challenge along the way: There seems to be not enough free spectrum

to enable these machines connect to the networks.

Spectrum has always been a scarce resource. But the insufficiency problem has not been as severe in the past as today. Not long ago, the wireless communications was a luxury choice and demand for it was much less. So, it sufficed to allocate spectrum statically, meaning to lease portions of it for exclusive access on a long-term basis. As demand grew from circuit-switched voice-only service to packet-switched IP-based services, so did the need for more bandwidth (BW). However, since the available spectrum could not increase, such insufficiency has been managed by other means (e.g., utilizing high-order modulations, spatial multiplexing (MIMO), spatial reuse, network densification (HetNets), etc.) to improve the spectral efficiency ($\mathrm{bits/s/m^2/Hz}$), instead. Nevertheless, there is so much that these technologies can do before dragging the networks to the point of diminishing return. In fact, the rising concern is that, beyond a point, their huge added complexity may not be worth the marginal improvement they gain in terms of the spectral efficiency. Considering the massive growth of traffic in a decade, there is a consensus among researchers that 5G cannot merely rely on the above-mentioned technologies. In plain language, much more spectrum is needed than what is available today.

One school of thought suggests that the extremely high frequency (EHF) band, that is above 30 GHz (a.k.a. mmW), be opened and used for civilian wireless communications. However, this seems to be the road to a much distant future as the EHF band is quite unknown to us and exhibits unpredictable behavior in terms of attenuation, absorption, scattering, Doppler shift, etc. The alternative school of thought challenges the way spectrum was statically allocated. This paradigm advocates that the command and control philosophy of the static allocation will certainly fail to satisfy the future massive communications needs of modern societies and provides a prescription for it. In supporting this claim, some recent statistical/empirical studies unveil that a significant portion of the spectrum is *temporally* and *spatially* underutilized while small fractions are highly overloaded [2–6]. As an example, [7] reports that the overall spectrum utilization is as low as 6-10% in three

(a) Wearable Technologies (b) Video on-demand (c) Learning

(d) Smart cities (e) Conferencing (f) Virtual group activities

Figure 1.1: Some 5G applications.

urban locations in Europe. This has motivated the research community to reconsider a different paradigm in utilizing the valuable spectrum. This paradigm that is known as dynamic spectrum allocation (DSA), loosely speaking, cognitive radio (CR), is empowered by a new architectural framework called software defined radio (SDR). DSA is believed to be the remedy for BW insufficiency and inefficiency in a more foreseeable future compared to the mmW rival technology. Despite the common misunderstanding, CR and DSA are not precisely the same. More accurately, DSA is a subsidiary of CR, a big one, where radio adaptation to the environment is only restricted to the operating frequency. On the other hand, the realm of adaptation in CR has no bound, including the operating frequency, transmit power, transmit mode and beyond. Having this narrow, yet impactful, distinction in mind, we will use the two terms interchangeably in this thesis.

Since its inception, CR has established itself as a potent solution to remedy the existing spectral inefficiency, as well as to address the challenges identified for 5G networks. These

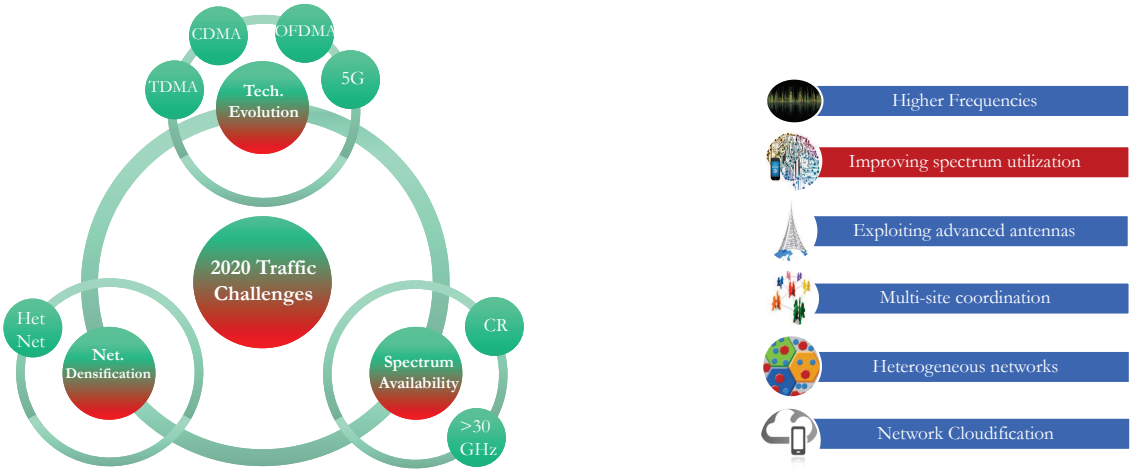


Figure 1.2: 5G challenges and enabling technologies.

challenges are illustrated in Fig. 1.2. By being swift and cognizant, a CR node is able to effectively adapt its parameters, such as power, frequency, data rate, etc., to the changing circumstances in order to maximally exploit the available transmission opportunities in the time, space, and frequency domains. Fig. 1.4 depicts a cognitive cycle from two different perspectives.

1.2 Motivations and Visions

The chief motivation behind this PhD dissertation is to substantiate the DSA as a gateway to a world where wireless communications is highly responsive, reliable and ubiquitous. In the author's view, DSA has the aptitude and competence in expanding the realm of wireless communications into all aspects of life such that humanity can live more comfortably and prosperously. The author believes the fifth generation (5G) of cellular communications will, indeed, host DSA as one of its core components. In spite of the common beliefs seeing DSA as an approach, tool, design, or techniques, the author discerns DSA as a communications concept.

Cognitive radio networks (CRNs) allocate spectrum to users more dynamically, faster, and for shorter intervals. This idea has not emerged suddenly. Instead, DSA is the product of a developmental process that has taken place for over several decades. Having been started with the 2G cellular networks (GSM), where the architecture was connection-oriented with non-adaptive session-based scheduling, and continued to the all-IP packet-switched architecture of 4G networks (LTE), where the resource scheduling is load and channel dependent and is done as frequently every $1\ \mathrm{msec}$, ever since spectrum allocation have become more agile and adaptive. Along the same line of thought, 5G networks with near-zero delay requirement is expected to be more agile by allocating/ de-allocating/ reallocating spectrum in the $\mu\mathrm{s}$ scale.

The orientation towards DSA is also driven by the success story of wireless local area networks (WLANs). Using only $80\ \mathrm{MHz}$ of spectrum in $2.4\ \mathrm{GHz}$ band (a.k.a. ISM band), Wi-Fi is indisputably the most dominant access technology ever known. The reason for such widespread adoption and prosperity is associated to its open spectrum model where devices can opportunistically transmit over the frequency band as long as they conform to certain stipulations. Once the $2.4\ \mathrm{GHz}$ spectrum became overcrowded, other open-access spectrum bands, i.e., $3.6\ \mathrm{GHz}$, $4.9\ \mathrm{GHz}$, $5\ \mathrm{GHz}$ and $5.9\ \mathrm{GHz}$, were introduced. This trend that shows no sign of slowdown promises a future where the entire spectrum will be made open.

With this vision, CR is deemed to have elemental impacts at different levels. First, it offers society more chances for a sustainable development by unleashing communications potentials and opening new doors. Second, it provides individuals with a plethora of new services, who pay less for what they receive as they are subject to a more competitive market. Third, it supplies network operators with more utility as their customers' demands evolve. Finally, it helps governments, directly and indirectly, as they can accrue more taxes by leasing more licenses, and by having new markets and jobs created, respectively. These multifaceted incentives galvanize the emergence of new use cases such as

smart grids, broadband services in cellular backhaul, public safety, femtocells, and wireless medical networks. Some of these applications are just standing in the doorway waiting for a responsive infrastructure that can deliver them to the users.

Despite the stated advantages and use cases, there are challenges involved with CR. Unfortunately after decades of study and focus, obstacles and incertitudes in this realm surpass the determinations and certainties. In a broad sense, what has been accomplished by the research community has been extensive but disoriented along with some deficiencies as stated in the following and portrayed in Fig. 1.3:

1. Choice of assumptions: Network models that are created are mostly based on assumptions that are either too simplistic or unrealistic. Unfortunately, such assumptions have their footprints everywhere, from channel access and channel models to primary network traffic profile, interference constraints, sensing, etc.

2. Validation: Not all models are validated by experimentation or simulation. Validation is an integral task not only to crosscheck the correctness and accuracy of findings but also to assure their superiority over non-DSA, legacy architecture.

3. Single-channel vs. Multi-channel: Existing network models are mainly focused on single-channel CRN (SC-CRN). Considering that the spectrum is becoming a more and more divisible resource, the full potentials of CR are exploitable through a multi-channel CRN (MC-CRN).

4. Synthesis vs. Modeling: An important question that has remained unanswered so far is the synthesis problem. Network synthesis, a.k.a. the reverse analysis problem, addresses whether or not the requirements of a service/application can be satisfied knowing the interference allowed to be inflicted on primary users (PUs). Such a question can barely be answered in a modeling/analysis framework and has not been tackled for CRNs.

5. Distributions vs. Statistics: The uncertainties about the wireless channel, primary network transmission pattern, CRN traffic pattern, random competition, etc. usually leave us with no choice but to use a probabilistic modeling approach, whereby network quantities are characterized by random variables (RVs). Though an RV is fully characterized with its probability density function (PDF), most of the existing studies deal with its primitive statistical moments, such as the *mean* and *variance*. These moments, though insightful, can only provide partial information on the characteristics of random phenomena.

6. High-traffic Regimes and Stability: CRNs do not always operate in low-traffic regimes. Instead, in almost all cases, we need to know how they perform when traffic intensity is high and responsiveness is low. In reality, it is in the latter situation that instability arises. The performance of CRNs in high-traffic regimes has been barely investigated in previous studies.

7. Cross-layer Modeling: In traditional networks, modeling could be confined to a single layer without taking into account the influences that different layers have on each other. In CR architecture, however, the ties between protocols in the protocol stack are stronger and the interfaces between different layers are thicker. Therefore, layers are more interdependent requiring a cross-layer modeling approach.

Similar deficiencies are observed in the area of algorithm design and protocol development. Such disorientation resulted in the slower development of economic models and technical standards for CRNs. All these led to a situation where the industrial sector is still unsure as to whether DSA is an enduring and profitable solution capable of obviating customers' surging demands or not. As a result, after a decade from its emergence, CR has continued to remain a research topic. Nonetheless, the dormant state of CR is no indication of its potentials and capabilities. Like many other research ideas, the time is ripe for CR to materialize into a technology. That highlights the importance of continuing research work

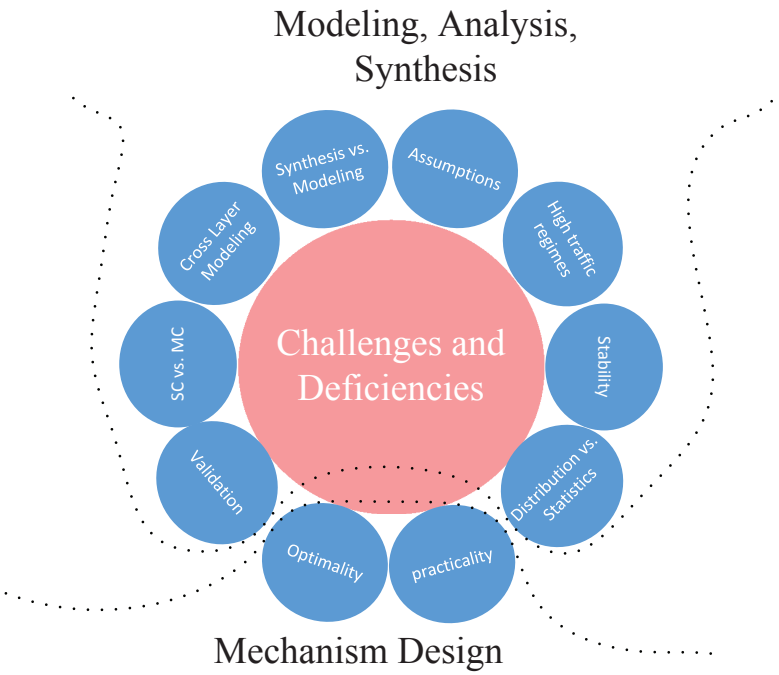
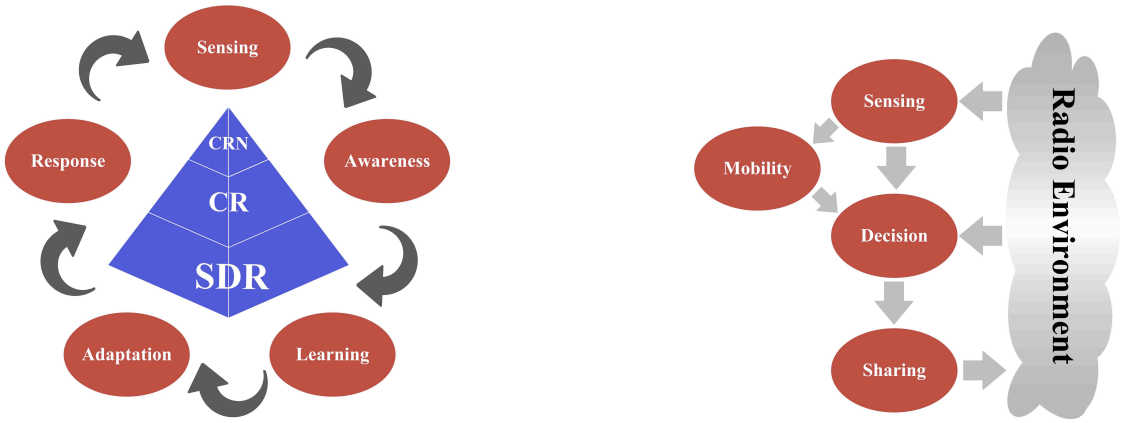


Figure 1.3: Identified challenges in the realm of CR research.

in this domain. Equally important is the standardization step. As a matter of fact, the recently released IEEE 802.22 for wireless regional area networks (WRANs) [8] has been a vigorous starting point along this path.

In line with these facts, the author found it an imperative task to explore the power of multi-channel CRNs by establishing well-defined, well-structured, and more precise models, in the first Part. In particular, when it comes to standardized networks such as WRANs, the exigency is higher as they will more likely be the pioneer DSA systems to be implemented. In fact, likewise the former efforts on the precise mathematical modeling of wireless standards, such as IEEE 802.11, which helped forward the extension of the technology outreach and advanced the succeeding progress, the modeling of IEEE 802.22 WRANs is a must.

In the realm of MC-CRNs, there is another domain parallel to the system modeling and



(a) Cognition cycle.

(b) CR block diagram.

Figure 1.4: The main CR components.

performance analysis that poses a multitude of brand-new, yet important, problems. Put under the heading of mechanism design and protocol development, these four CR components are branched: (i) spectrum management, (ii) spectrum sensing, (iii) spectrum mobility and (iv) spectrum sharing. These components are illustrated in Fig. 1.4b. Despite its importance, the investigations in the area of designing mechanisms for MC-CRNs have been less abundant. The reason is not too difficult to conceive: Compared to the traditional network design methodology, designing mechanisms for DSA should be founded on a completely different line of thinking. In fact, as aforementioned, DSA almost collapses the existing protocol stack erecting a new structure where layers are more intertwined and interfaces are less abstract. As the entities dwelling in these layers, protocols, algorithms, and mechanisms have extended realm of operation. This makes the mechanism design task relatively cumbersome. What is spectrum sharing without spectrum sensing? What is spectrum sensing without spectrum sharing? How can spectrum mobility be triggered if there is no spectrum sensing in place to detect PUs? How can spectrum be effectively managed if application requirements are not known? Questions of this sort make one realize the importance and intricacies of mechanism design in CRNs.

Besides the architectural distinction, the *scale* and *nature* of traffic affect the design problem. CR is one of the 5G enabler technologies where machines will be the dominating network subscribers. The traffic emanating from these machines have characteristics that are far more diverse than the voice/video/web browsing traffic originating from hand-held mobile devices. With the massive number of machines, the other issue would be the scale of the aggregate traffic. Both issues render the algorithms/mechanisms/protocols that are optimized for the existing networks incapable of handling the tasks in 5G networks. In filling this void, the second part of this PhD thesis is devoted to designing smart, efficient, and scalable mechanisms for MC-CRNs. Since the onset of this journey, the goal has been set to profoundly explore each and every CR component as illustrated in Fig. 1.4b. In fact, while the sensing, decision, and mobility components are dealt with in Chapter 5, a separate Chapter (6) is dedicated to the spectrum sharing problem.

1.3 Research Objectives

Having discussed the key motivations behind this dissertation, the goal has been to model and analyze the performance of MC-CRNs and design mechanisms for them while avoiding the common shortcomings identified in the previous section. These objectives are diagrammatically shown in Fig. 1.5 and delineated next.

1.3.1 Modeling and Performance Analysis

Mathematical models are important because they are abstract descriptions of the system (network). In the realm of telecommunications, the modeling is the skeleton for any ensuing action. Futuristic wireless technologies are sophisticated and extensive. Network operators and solution providers need to be sure of the qualities that such networks deliver before they have to make huge investments on them. Models help them to comprehend

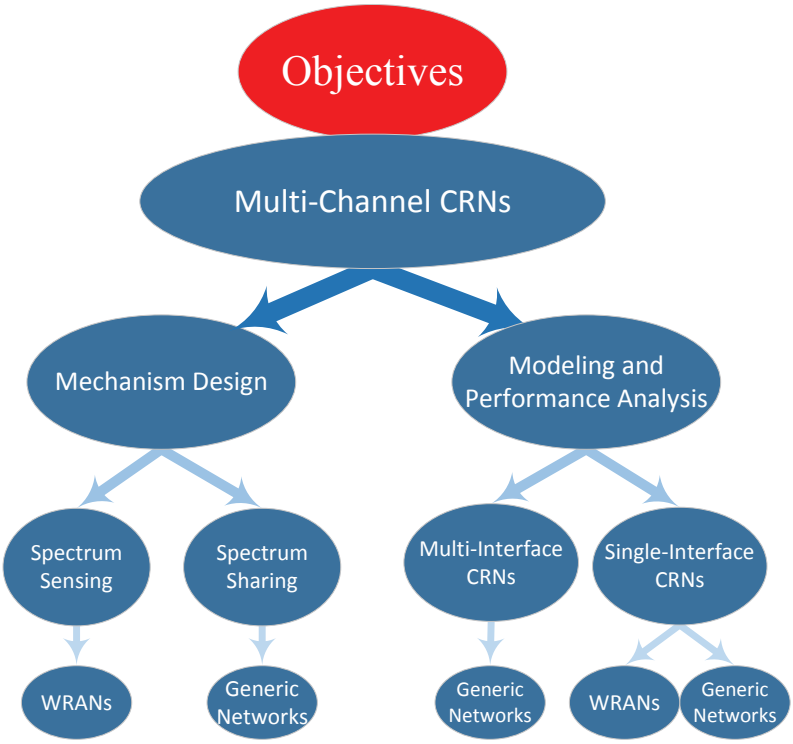


Figure 1.5: My research objectives.

and visualize different aspects of such systems before deployment. They reveal whether an idea is worth the effort and money or not. A good model may not possess all the details but has enough to characterize the behavior of the system and prevent it from collapsing caused by unwanted/unforeseen factors. It also enables researchers to specify the structure and behavior of the system, and provides them with guidance as to how it can be safely and efficiently built, later on. Finally, it can provide an action plan for any subsequent decision for redesigning the network.

Having this vision in mind, the first part of this dissertation proposes solid modeling approaches that could be used to evaluate the performance of multi-channel CRNs. The proposed models are stable and based on well-defined methodologies. In line with this general goal, the modeling research objectives are sought from three different perspectives.

From the first perspective, the efforts are focused on

- Modeling and performance evaluation of single-interface multi-channel (SIMC) CRNs with emphasis on (1) Generic CRNs and (2) WRANs.
- Modeling and performance evaluation of multi-interface multi-channel (MIMC) CRNs.

It should be noted that even though a lot has been said on the modeling and analysis of single-channel CRNs, the area of multi-channel CRNs has been barely subject to investigations, most probably due to the challenges involved.

From the second perspective, this work embodies the following four different aspects:

- Modeling
- Analysis
- Synthesis
- Optimality

While the first two aspects focus on establishing a mathematical framework that can be used for the performance evaluation, the synthesis aspect answers the question that whether or not the requirements of a service/application can be satisfied given the network and operator constraints. The latter aspect becomes a driving force for the optimality discussions. Finally, from the last perspective, the CRNs are explored through

- Numerical modeling,
- Closed-form derivation,
- Simulation validation.

Due to the fact that the proposed models are based on Markov chains, sometimes, solving the models numerically becomes computationally intensive. This can lessen the practical merits and importance of mathematical models as simplicity is the main reason they are sought after. In such cases, closed-form expressions that can estimate the network performance quickly and at no cost were developed, in this thesis. Aside from empowering us to understand CRNs better, closed-form expressions can be incorporated at the core of the algorithms/mechanisms, such as admission and congestion control (as the threshold function) or allocation mechanism (as the utility function). Finally, some effort was expended developing simulations to ascertain about the correctness of the proposed analytical models and to observe the error gaps.

1.3.2 Mechanism Design

For the motives stated in Section 1.2, the second part of this dissertation is devoted to designing mechanisms for CRNs.

Spectrum Sensing/Decision/Mobility

The first chapter in the second part examines the spectrum sensing, decision, and mobility problems. The focus is on IEEE 802.22 WRAN as the most prominent standard that has appeared on CRNs. The reason for such emphasis is clear: At the moment, CR is just a tiny step away from becoming a technology. For this to happen, the best bet would be to direct all the efforts towards the frontrunner candidate, that is IEEE 802.22 WRANs. Since its first release in 2011, this standard has been revisited several times. Nevertheless, there is still much work left to be done given that some WRAN features are explicitly stated as open issues and some functionalities exhibit sub-optimal performance. The sensing and fusion mechanism (SFM) is among those issues for which [8] proposes very basic combining algorithms, namely, AND, OR and VOTING logic rules. Given the importance of the

spectrum sensing process in DSA, these combining rules are not satisfactorily reliable, accurate and stable. In particular, they suffer from unbalanced performance, where a low false-alarm rate comes at the cost of an unbearably high misdetection rate (and vice versa), and exhibit rigid performance, where performance does not adapt to varying conditions. Moreover, due to the lack of performance stabilizing feedback from sensing output to input, operational instability is a possibility, manifesting itself in the form of weak performance in low signal-to-noise ratio (SNR) regimes. Motivated by these facts, in Chapter 5, an efficient distributed SFM is proposed that avoids the above shortcomings, complies with the standard directives, and can be easily embedded in the network.

Spectrum Sharing

The second chapter in the second part of this thesis targets the spectrum sharing problem in CRNs. The resource sharing problem in CRNs is formulated in the form of an auction game. Built upon the recently-emerged theory of optimal multi-unit auctions, the allocation mechanism proposed in Chapter 6 resolves several deep-rooted issues in wireless networks. On the user side, the fairness-optimality trade-off is settled in an elegant way; on the network operator side, the total collected revenue is maximized through the imposition of payments. To diminish uncertainties in the decision making process and improve the stability of the resource assignments, the aforementioned payment rule is delicately designed to induce truthfulness among self-centered and rational users. These properties, along with the computational manageability of the formulated problem, as well as the existence and stability of its solution render the proposed allocation mechanism an ideal scheme for exploitation in CR-based 5G networks.

1.4 Research Methodology

1.4.1 Modeling and Performance Analysis

Among the diverse set of analytical tools, queuing theory has shown to be effective in capturing the dynamics of systems. Queues arise when the short-term demand for service exceeds the capacity. On the other side, the value of queuing theory is in its capability to analyze the capacity and waiting time in the system. These are all important issues in CRNs. In spite of its importance, application of this theoretical modeling tool in analyzing CRNs is new. Work on the imperative class of MC-CRNs has been rarely pursued due to the fact that "the theory of multi-server queues has been barely touched by mathematicians except for very special cases. Motivated by these facts and as depicted in Fig. 1.6, the methodology in this thesis is the exploitation of the following two important disciplines in multi-server queuing theory to fulfill the objectives stated in the Section 1.3:

- Stand-alone Queues
  - FIFO: First-in First-out
  - PQ: Priority Queuing
- Network of Queues

More specifically, the stand-alone FIFO queuing approach is exploited for the modeling of MIMC-CRNs and the PQ approach is used for the analysis of generic SIMC-CRNs. Ultimately, the theory of network of queues is leveraged to model the performance of WRANs. The distinction between what is accomplished in this thesis with any previous studies starts from the approach taken and the extent of progression. More narrowly, our modeling methods are based upon stable and dynamic Markov chains that are easily understandable and logically sound due to the meticulous and factual choice of transitions, states and rates.

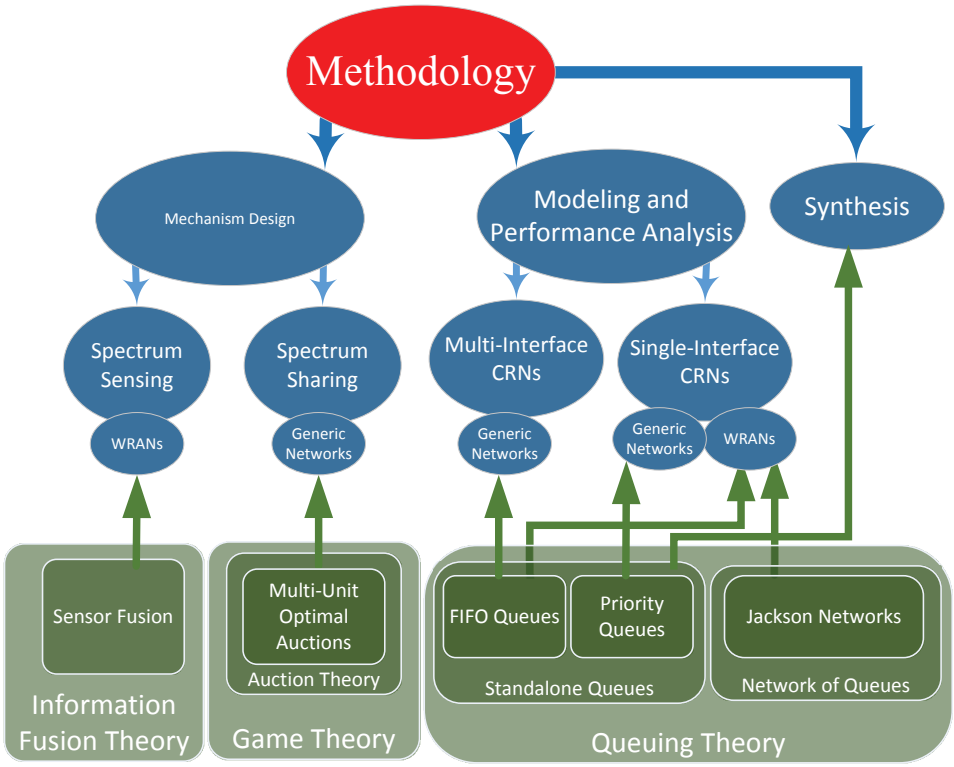


Figure 1.6: The research methodology.

In each case, key performance measures are probed and achievable region as well as optimal solutions are discussed . One last point to mention is about the verification of the accuracy and precision of the proposed models. In this vein, whenever it was possible, the validation task was undertaken. This has been accomplished either by writing meticulous system-level simulations or comparison with existing results in the literature.

1.4.2 Mechanism Design

Spectrum Sensing/Decision/Mobility

From the very beginning of this research journey, the learning theory was seen as an indispensable element of design in CRNs. The reason cannot be clearer; without learning, the

cycle in Fig. 1.4a would become disconnected and CR becomes a lost cause. Therefore, to fulfill the objectives laid out from the onset, it was decided to exploit learning theory for designing the spectrum sensing mechanism. The multi-channel learning-based distributed sensing (MC-LDS) mechanism proposed in Chapter 5 works based on learning concepts and uses differentiation as well as reward penalty procedures to intelligently combine current and past sensing measurements. These two concepts, when combined, are expected to increase the sensing precision and decision reliability since each user not only learns from itself but also from the environment (other users). It will be shown that the individual sensing outcomes are combined in such a way that both accurate and faulty sensors contribute constructively towards an accurate final decision. In order to adapt the sensing mechanism to varying channel/load conditions, differentiation is integrated into the mechanism in two horizons, *temporal* and *spatial*. To validate the performance of MC-LDS, a discrete-event simulator is built in C++ and Maple programming languages. In the simulator, all the relevant management and controlling functions of WRANs are implemented. This chiefly includes the quiet period (QP) and sensing management procedures (including inter-frame and intra-frame sensing, synchronization and reporting) and channel management functions (including channel list updating, and incumbent protection module).

Spectrum Sharing

In this thesis, auction theory was adopted to allocate spectrum among wireless users given the similarities between CRNs and competitive markets and the fast-pacing advancements in the area of artificial intelligence, decision-making, optimization and signal processing. Auctions are well-known for their allocation efficiency and higher revenue generation properties. By structuring the problem of resource allocation as an auction game of incomplete information, the problem is formulated as a Bayesian game where the sequential reasoning of selfish but rational players (nodes) converges to a Nash equilibrium which is stable,

fast-converging, and efficient. In fact, as the Nash equilibrium in games usually suffers from inefficiency, non-uniqueness and stability issues, a pricing rule is developed so that CR users have incentives to reveal their true valuation for the BW resource, thus, leading to improved allocation efficiency. In a systematic approach to deal with this problem, an efficacious allocation is characterized by six axioms, namely incentive compatibility (IC), individual rationality (IR), fairness, efficiency, revenue maximization (RM) and computational manageability (CM). While the first three axioms are incorporated into the allocation mechanism through the non-linear *spectrum pricing*, the last three axioms are integrated into the mechanism through *spectrum assignment*.

Part I

Modeling, Analysis and Synthesis of Multi-Channel Cognitive Radio Networks

As it was pointed out in the previous chapter, the rapid advancements in wireless technologies and the emergence of abundant services in the tiny unlicensed ISM band compelled the international telecommunications union (ITU) and the frequency regulatory bodies to rethink the current spectrum allocation structure that is aided by CR technology. Even though much has been discovered on the analysis of single-channel CR networks (CRNs), the *modeling* and *analysis* [1] of multi-channel CRNs has been barely discovered, whereof [9, 10] can be cited. Motivated by this fact, this chapter fills the existing gap by establishing a modeling framework for multi-channel CRNs using stable and well-defined queuing models, and then, developing a comprehensive analytical procedure for this model in order to obtain key performance measures, draw achievable region, and find optimal solutions. The second important challenge that is dealt with in this chapter is the problem of network *synthesis*. Contrary to the network modeling and performance analysis whereby the approach is bottom-up, the synthesis problem is a top-down methodology. More precisely, network analysis helps us understand what performance can be achieved given the parameters, whereas network synthesis aids us to find the parameters that is needed to reach at our desirable performance. Within the context of CRNs, the synthesis methodology is leveraged to see whether or not the requirements of secondary service/application can be satisfied knowing that the amount of interference to PUs shall remain limited. One should bear in mind that the answer to this question can barely be answered in an analytical framework. Such exploration is not only valuable from the theoretical perspective but it empowers one to assess if application performance criteria are satisfied in the CRN, and if not, what tradeoffs can be made for such criteria to be met.

Finally, a mixed strategy is introduced in section 2.6, where some level of interference on the PUs is allowed to the benefit of better QoS for the SUs. This resembles the case of spectrum underlay CRNs where the maximization of the joint secondary-primary profit

[1] Semantically, analysis is the process of breaking down the complex system into smaller fragments in order to make appropriate deductions and come up with improved understanding of the system.

matters instead of individual segregated utilities. Our observations show that an optimum mixed-strategy always exists, which outputs the least cost for the defined cost function.

Chapter 2

Single-Interface CRNs: Priority-Queue Analysis[1]

In this chapter, we make use of the theory of priority queues to model single-interface multi-channel CRNs. Though this modeling tool has found wide application in other domains in years, its benefit for the analysis of single-channel CRNs was discovered only recently. For instance, [11] introduces an analytical framework based upon preemptive-resume priority queues to characterize the effect of spectrum handoff on performance. In [12–14], the authors use the theory of priority queues to find several network statistics. The value of these studies lies in their simplicity in finding the moments of waiting delay for single-channel [13] and multi-channel CRNs [12, 14]. In [15] the average waiting time of packets is derived for single-channel CRN by leveraging the preemptive priority queues, and several observations on the dependencies between the secondary and primary queues are made. Finally, [16] establishes a Markov transition model to characterize the cumulative handoff delay of SUs with different priorities.

The modeling approach in this chapter is different from previous works, both in the

[1] N. Tadayon and S. Aïssa, "Multi-channel cognitive radio networks: modeling, analysis, and synthesis", *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 11, pp. 2065–2074, Feb. 2014.

methodology that is used and the extent that it proceeds. Indeed, while our modeling delivers the queue length distribution of secondary users (SUs) on the basis of a dynamic model, [9] (or [10]) works on moments of delay (or tail distribution) according to approximate methods. This modeling is based upon a stable and dynamic Markov chain that is easily understandable and logically sound due to the meticulous and factual choice of transitions, states and rates.

2.1 Literature Review

Priority mechanism is used for situations when a service facility has to serve classes with different characteristics and requirements, and therefore one should be treated preferentially compared to others. In this case, a priority scheduling policy can fulfill this aim and analyzing the characteristics of each class is of paramount importance.

The theory of priority queues was initiated with the introduction of the preemptive resume policy by Cobham in [17] and Holley in [18]. The idea was later expanded in [19] to incorporate a higher number of priority levels, more realistic queuing models $(\mathrm{M}/\mathrm{G}/1)$ and higher moments. In a general classification, there are two types of priority policies: *preemptive* and *non-preemptive*. In preemptive policy, all the lower priority queues (LPQs) should instantly vacate the server(s) upon the presence of a packet in a higher priority queue (HPQ), and embark on server(s) once the HPQ is empty. *Preemptive resume* and *preemptive repeat* are two slight variations of this policy [20–23]. In preemptive resume, the interrupted head-of-line packets (due to the presence of HPQ traffic) resume their services from the interruption points, once no HPQ packet is left unserved [22, 23]. In preemptive repeat, the interrupted packet should restart the service from the beginning, no matter how much time was spent in the previous serving period [20, 21]. On the other hand, the non-preemptive policy gives some marginal assurance for LPQs by allowing the under-served LPQ head-of-line packet (and only this packet) to continue its service even if HPQ receives

a fresh packet. Nevertheless, after this packet is serviced, the server is again unconditionally possessed by the HPQs.

In spite of the vast research on the analysis of priority queues with a single-server facility, efforts on the characterization of multi-server queues were not as extensive or fruitful due to their complications. This is essentially unfortunate since the modeling of many telecommunications problems can be suitably placed in the framework of a multi-server system, such as multi-channel CRN, as will be detailed shortly. Perhaps [24–28] are the only studies in this area. In [24], the authors analyze a queuing model in which two classes of customers can share an $\mathrm{M/M/c}$ service facility according to a non-monopolized priority policy. This policy is such that users of class I (II) have preemptive priority over the users of class II (I) for the first $c_1 (c_2 = c - c_1)$ servers. The joint distribution of the packet counts in all the classes is found numerically using a matrix-geometric approach. In [25], the authors apply different numerical approaches known for solving the non-preemptive multi-server priority queuing system. In [26], a priority system with N server facility serving R classes is considered, where classes with lower indices have preemptive priority over classes with higher indices. In the analysis, the average response time (queuing + service delay) of each class is found. In [27], the study in [26] is extended to cover the multi-class case of service with arbitrary distribution. In the latter work, the mean response time of each class in a multi-class $\mathrm{M/G/c}$ queue with preemptive priority scheduling is approximated using an elegant simple method. Study [28], on the other hand, turns the focus to non-preemptive priority policy, for the first time. The approximations in [29] were used for $\mathrm{M/G/c}$ (which has no exact expression for the statistical moments, let alone the probability density function) and applied to the scenario with two traffic classes. Closed-form expressions for the Laplace transform of the waiting time distribution and the mean waiting were derived for both classes.

Among few scheduling policies known in the theory of priority queues (i.e., non-preemptive, preemptive repeat and preemptive resume), the preemptive resume seems to

better emulate the essence of CRNs. Indeed, the lack of awareness of PUs about the existence of SUs, as demanded in drafts and standards [3], renders the non-preemptive policy an unfitting model for CRNs. On the other hand, the preemptive repeat policy neglects the practical aspect of wireless links where data packets are composed of statistically independent subunits, i.e. symbols, that can be decoded independently. Therefore, upon the loss of a packet, the system can resume transmitting the rest of the symbols from the point of interruption. All this said, we recognize the preemptive-resume multi-class multi-server priority scheduling as a well-fitting model to the problem at hand.

At this point, the well-defined goal in this chapter is to be restated, namely, to analyze and synthesize the performance of multi-channel CRNs using a simple and tractable model rather than through complicated numerical and algorithmic approaches. In fact, by considering few widely recognized assumptions,[2] a very simple and insightful model is attainable that not only allows a tractable performance evaluation of the system under study but can also be used as grounds for future explorations in the area, as well as a tool for synthesis studies such as what will be accomplished later on in Section 2.3.

Next, we develop our network model for single-interface multi-channel CRNs.

2.2 Model Development

The analogy between CRNs and priority queues revolves around the fact that, in both cases, resources are shared between traffic classes with different preferential rights. In the latter case, shared resources are tangible service facilities, while in the former case they are the intangible frequency channels. Except for this conceptual difference, both systems are

[2]For example, the packet inter-arrival and service-time distributions on the primary and secondary sides are assumed to be exponentially distributed. These assumptions have been widely used in previous studies in the area, e.g., [9], [11–13, 30, 31], added to the fact that they yield accurate modeling in many situations, in actuality. The latter claim is due to the fact that allocation intervals in many networks are such that the service-time of each user has low variability. Therefore, exponential service-time assumption, indeed, provides a lower (pessimistic) bound on the performance.

naturally similar. For example, in the case of CRNs, PUs have unconditioned privilege to access the channel due to their exclusive right. Inside each traffic class, packets are served according to a first-in first-out (FIFO) policy. Also, the discretized nature of arriving units in both cases enables us to represent customers with data packets.

Consequently, by representing each channel with a server, the problem of modeling multi-channel CRNs can fit within the framework of multi-class multi-server priority queues, as illustrated in Fig. 2.1a. Here, N channels (say servers) represent portions of the spectrum that are authorized for opportunistic access [8] (cf. circles in Fig. 2.1). Focusing on this figure, on the primary side, the i^{th} indexed queue stores data packets arriving with aggregate arrival rate λ_1^i, which is the accumulative traffic of PUs that are authorized to transmit on the i^{th} channel, and serves packets with rate μ_1^i.

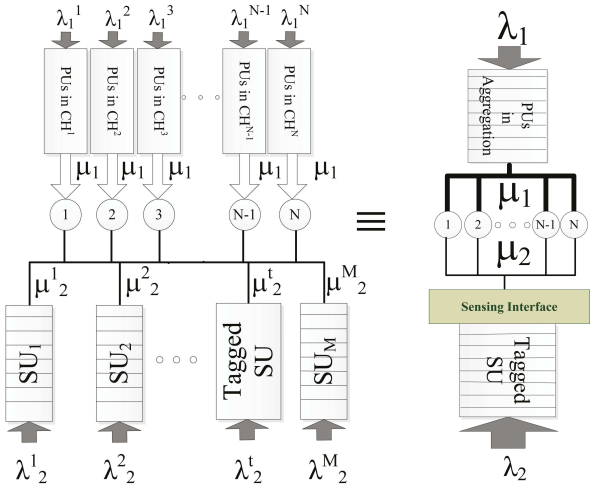
On the secondary side, each queue represents an SU with packet arrival rate λ_2^j and service rate μ_2^j, $j = 1, \cdots, m$. SUs must sense the N channels to discover opportunities for dispatching packets over them once sensed idle. Upon dispatching packets over the empty channels, it is irrelevant to the SU which packet has gone over which server. This is because optimal channel allocation is neither a concern nor our focus in this chapter and will be addressed in full detail in Chapter 6.

2.2.1 Queue Decoupling

At the primary side, the network of N independent single-server queues, with arrival rates λ_1^i, can be approximated with a single N-server queue, as shown in Fig. 2.1b, with equivalent arrival rate $\lambda_1 = \sum_{i=1}^{N} \lambda_1^i$.[3]. We assume the service rates on all servers in the primary side are the same, i.e., $\mu_1^i = \mu_1$.

As for the secondary network, since the performance evaluation of this network with coupled queues, as shown in Fig. 2.1b, is a challenging task, we assume weak coupling

[3]The validity of this equivalency for heavy-traffic regimes has been verified both mathematically and with simulations.



(a) Original model. (b) Decoupled model.

Figure 2.1: Modeling the interactive behavior of SUs and PUs in the framework of preemptive-resume multi-class multi-server priority queues.

among the m SUs, which is accurate for light traffic regimes[4]. Subsequently, one of the SUs (termed tagged SU hereafter) can be detached from the rest of the system as shown in Fig. 2.1b if its service rate is appropriately adjusted to meaningfully reflect the multi-user characteristic of the network (i.e. $\mu_2^t \Rightarrow \mu_2$, with t denoting the tagged SU in Fig. 2.1). Defined as the amount of time it takes for the head-of-line packet to get transmitted successfully, the service time D_{access}, which is a function of almost all network quantities, is related to the service rate μ_2 and the transmission time T_s in the following way,

$$\mu_2 = \frac{1}{D_{\text{access}} + \dfrac{1}{\mu_2^t}}, \tag{2.1}$$

where $T_s = 1/mu_2^t$ depends on the packet length while D_{access} depends on the access mechanism, the number of admitted SUs, etc.[5] A simulator (Section 2.4) written for the

[4]This assumption has been approved and exploited in many studies, e.g., [9, 32, 33]

[5]To calculate μ_2, if SUs access the channel using a 802.11x card, one can use [34] which presents a useful

purpose of validation confirms the precision of the PU aggregation and SU decoupling methods.

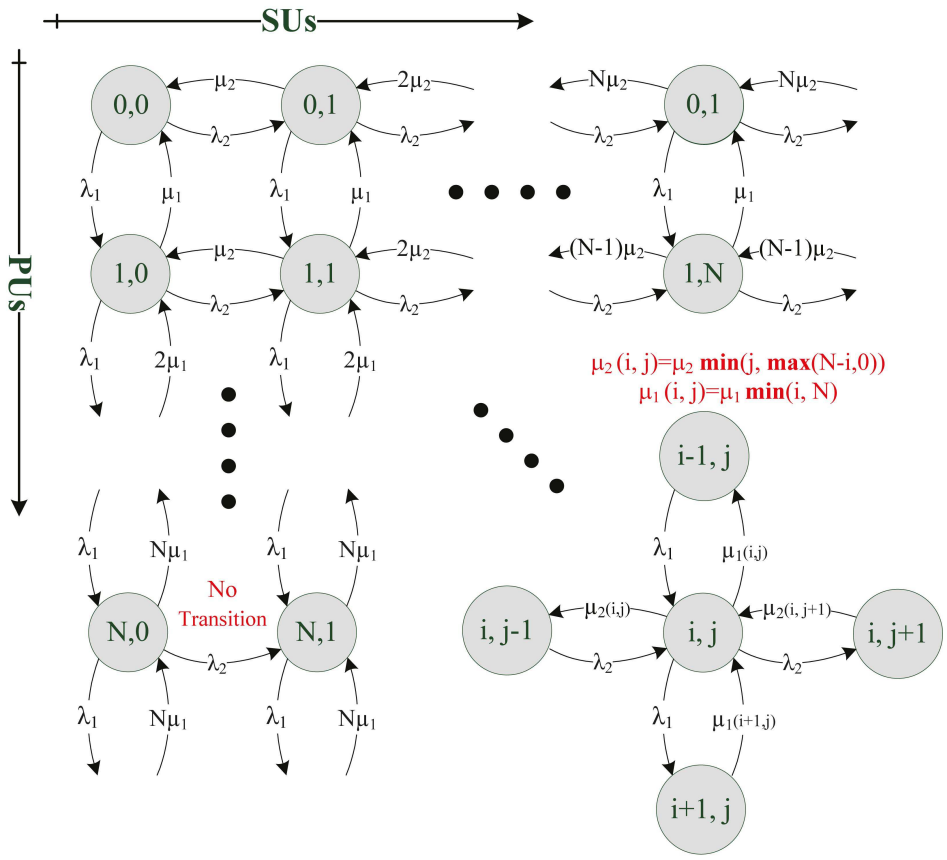


Figure 2.2: Representation of the multi-channel CRN with 2D CTMC.

2.2.2 CTMC Representation

With the decoupling of queues illustrated in Fig. 2.1, we can mathematically model the CRN in Fig. 2.1a as a Markov chain. The quantity of interest is the average number of packets in PU and SU queues of the decoupled model in Fig. 2.1b (i and j, respectively).

non-recursive closed-form expression for delay, or use the formula for the time sharing access delay in the case of TDMA.

An state of the 2D continuous-time Markov chain (CTMC) is represented by the pair (i, j) as shown in Fig. 2.2. To save space while preserving clarity, a generic state of this CTMC is shown in the lower-right corner of Fig. 2.2. Starting from state (i, j), a horizontal (vertical) displacement to the right (down) represents the addition of a packet to the SU (PU) class with rate $\lambda_2(\lambda_1)$, and the horizontal (vertical) displacement to the left (up) represents the departure of a packet from the SU (PU) class with rate $\mu_2(i, j) = \mu_2 \min(j, \max(N - i, 0))$ $(\mu_1(i, j) = \mu_1 \min(i, N))$. The logic underlying this rate determination is pretty obvious, thus, not explained here. According to the above expression, when all channels are busy $(i > N)$ serving PUs, no SU packet is accepted for transmission $(\mu_2(i > N, j) = 0)$.

The infinite state CTMC in Fig. 2.2 must have a stability condition (SC). This SC, as reported in [24–28], is as follows:

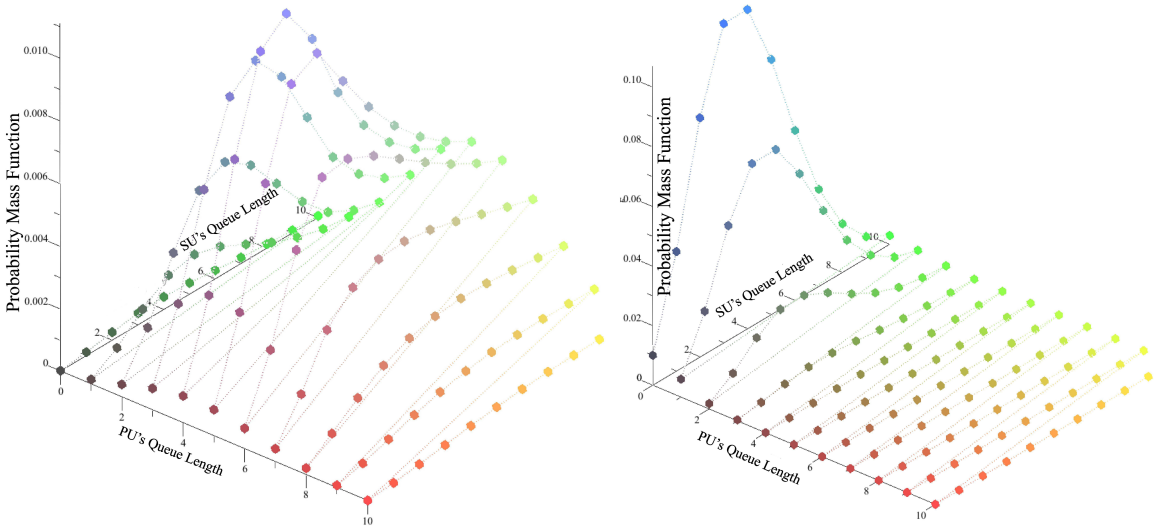
$$0 < \rho = \sum_{i=1}^{r} \frac{\lambda_i}{\mu_i} < N, \tag{2.2}$$

where N is the number of servers (or PU queues), r is the number of priority classes and $\rho_i = \lambda_i/\mu_i$ represents the utilization factor of the i^{th} class.

In the framework of CRNs, two classes of traffic $(r = 2)$ is presumed, where the first class, indexed one $(\rho_1,\ \lambda_1,\ \mu_1)$, represents the primary queue and the second class, indexed two $(\rho_2,\ \lambda_2,\ \mu_2)$, represents the secondary queue. Given this explanation, this CTMC is solved using Z-transform approach to find the steady-state probabilities on each state and, possibly, some moments pertaining to both traffic classes.

Without diverging the focus to the trivial problem of how to solve this CTMC, the corresponding solution is presented here.[6] Fig. 2.3 illustrates the joint probability mass function (PMF) of the number of data packets in the secondary and primary queues in Fig. 2.1b. In both 3-D plots, $N = 10$, $\lambda_2 = 4 \cdot 10^4 \text{pk/s}$, $\mu_2 = 10^4 \text{pk/s}$, $\mu_1 = 0.5 \cdot 10^4 \text{pk/s}$, and they only differ in PUs input rate chosen as $\lambda_1 = 0.3 \cdot 10^4 \text{pk/s}$ in Fig. 2.3b and as

[6] There are many well-established methods for numerically solving Markov chains, see e.g., [35].

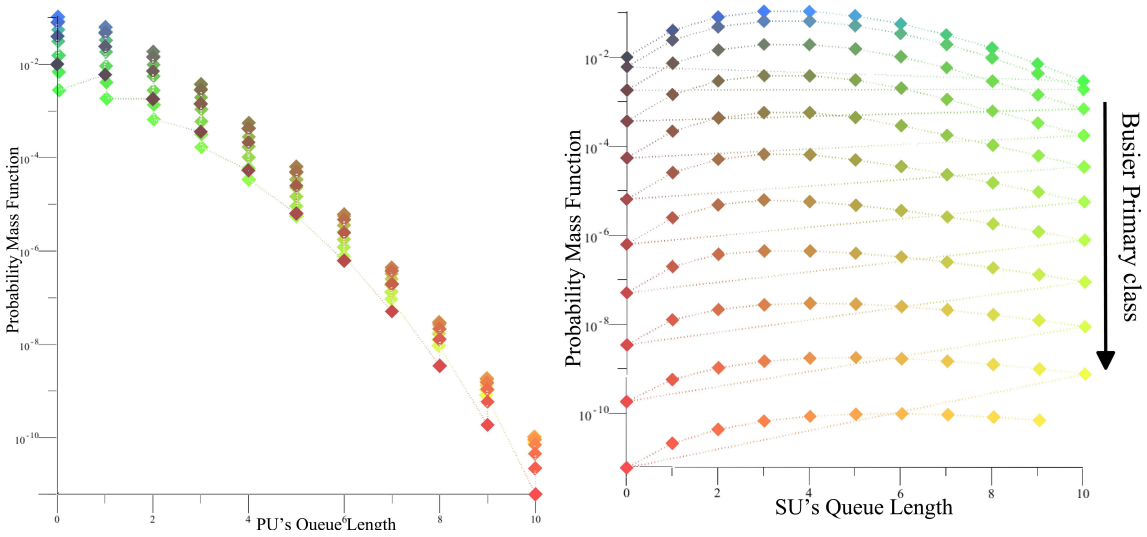


(a) HTR ($\rho_1 = 5.46,\ \rho_2 = 4$) (b) LTR ($\rho_1 = 0.6,\ \rho_2 = 4$)

Figure 2.3: Joint PMF of the secondary and primary classes' queue lengths.

$\lambda_1 = 2.7 \cdot 10^4 \text{pk/s}$ in Fig. 2.3a, resulting in a primary utilization factor of 5.46 ($\rho = 9.46 < 10$) and 0.6 ($\rho = 4.6 < 10$), respectively. Fig. 2.4 shows the 2D views of the joint PMF in Fig. 2.3b with a logarithmic scale. The above ρ values are chosen such that they reflect the extreme cases, i.e. low traffic regime (LTR) and heavy traffic regime (HTR), while not violating the SC of (2.1). Several intuitive observations can be made from these plots. First and foremost, even though the tiny change $\rho_1 \rightarrow \rho_1 + \Delta\rho$ does not shift the primary's marginal PMF in a sensible way, it causes a considerable change for the secondary's, widening it and shifting up its center rapidly. Second, both plots (Fig. 2.4a and 2.4b) demonstrate that the secondary's PMF has a heavy tail characteristic, getting heavier as the PUs activity factor increases (ρ_1 gets larger).

Fig. 2.5 illustrates the average total delay D_i (queuing + service) in both classes for the HTR setting described. Please note that the curves are plotted in double ordinate setting so that the results for PU and SU can be seen together. Moreover, the ordinate pertaining to the SU is plotted logarithmically while the PU's scaling is ordinary. As observed, the increase



(a) Joint PMF vs. primary's queue length

(b) 2D view of the joint PMF vs. secondary's queue length

Figure 2.4: 2D view of the marginal PMF in LTR setting of Fig. 2.3b (logarithmic scale).

of ρ_1 from 0.6 to 5.46 results in the primary's total delay (D_1) to increase only 13% (right ordinate) while shooting up the secondary's total delay (D_2) about 600% (left ordinate). It is fortunate to note that such high value of PU's utilization factor is much higher than what is measure in reality otherwise communications on the secondary's side would have been impossible altogether. In fact, [3] and many other studies prove a low and almost constant utilization in currently occupied primary spectrum.

Since the PUs have exclusive right to access the spectrum and SUs have no control over the traffic and activity of the PUs, all the focus should be concentrated on the secondary side to obtain the best out of what is available. As a result, an individual decision-making approach by SUs will certainly fail in providing the wide-scale QoS that benefits all SUs. Therefore, mechanisms such as admission control and congestion control as well as an efficient, swift and agile resource allocation, play a very important role in multi-channel CRNs. To that end, mathematical expressions that characterize both classes are required. Such expression can be used as the cost (utility) function in the resource allocation problem,

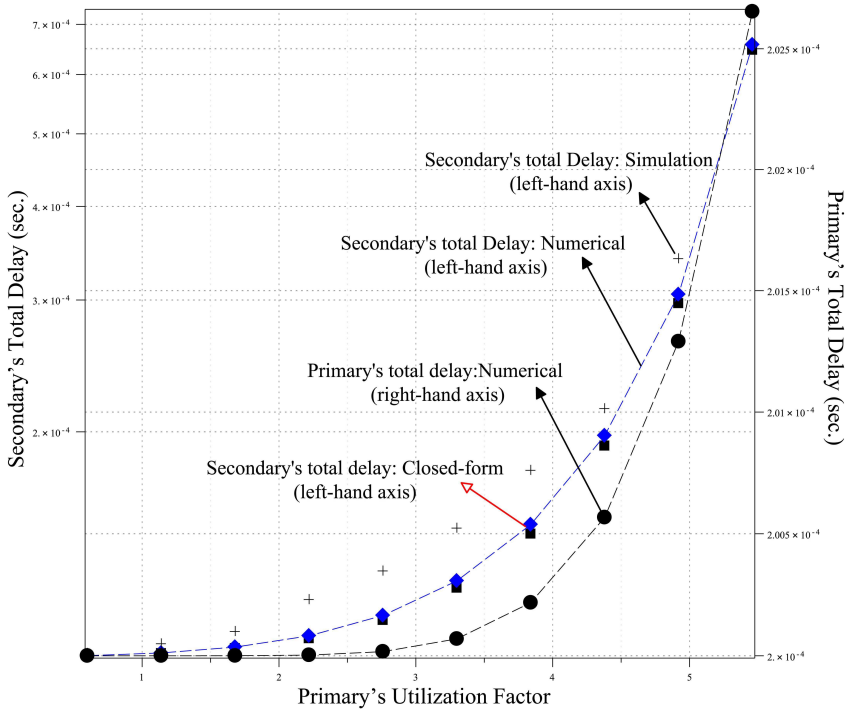


Figure 2.5: Numerical, closed-form, and simulation results obtained for average total delays D_1, D_2 (queuing +service) in 2 classes under HTR regime.

or threshold function in the admission control problem. This motivated us to seek closed-form expressions for the average total delays, D_i, $i = 1, 2$, in the multi-channel CRN.

2.3 Network Synthesis

The question that whether a demanded performance is achievable or not (given the traffic characteristics and queue attributes), and if so under which policy, is important to deal with. In fact, [36] was the first to answer this question in the context of single-server multi-class priority queues, where the synthesis problem was investigated after decades of exclusive work on queue analysis. The point of departure in [36] is Kleinrock's conservation law [37], which states that in stable single-server multi-class (m classes) queuing

systems ($\rho = \sum_{i=1}^{m} \rho_i < 1$), any performance vector $[W_1, \cdots, W_m]$, where W_i denotes the waiting time pertaining to the i^{th} class, must satisfy the condition $\sum_{i=1}^{m} \rho_i W_i < \rho V/(1-\rho)$, where $\rho_i = \lambda_i/\mu_i$ and $V = \sum_{i=1}^{m} \rho_i/\mu_i$. This expression has the following important implication: no class can do better off without another class doing worse off. In other words, the weighted linear sum of the W_i's in a work-conserving single-server multi-class queue is fixed, that is [37],

$$\sum_{i=1}^{l} \rho_i W_i = \frac{\sum_{i=1}^{l} \rho_i \cdot \sum_{i=1}^{l} \frac{\rho_i}{\mu_i}}{1 - \sum_{i=1}^{l} \rho_i}, \qquad l = 1, \ldots, r, \tag{2.3}$$

where $W_i = D_i - \mu_i^{-1}$.

In (2.3), the class indexed one (r) has the highest (lowest) priority. Since the preemptive priority grants the exclusive right of access to the higher priority class (as for PUs in CRN), the lower priority classes are transparent to the higher priority class, thus, W_1 in (2.3) is nothing more than the classic formula for waiting time in the $\mathrm{M}/\mathrm{M}/1$ queue. Hence, the solution to (2.3) can be obtained by reverse plugging. However, this solution is only valid for single-server multi-class priority queues not for the multi-server multi-class priority queue case as needed in the multi-channel CRN of Fig. 2.1 where $r = 2$.

[38] is among the very few studies that sought the existence of conservation law in multi-server priority queues. That work is salient due to the derivation of a semi-conservation law that is valid under identical service time distribution in both classes ($\mu_1 = \mu_2$). Along the process of examining our results, we came up with a closed-form expression for the conservation law in multi-server work-conserving priority queues gen-

eralizing the work in [38] for non-identical service time distributions, as follows:

$$\rho_1 D_1 + \rho_2 D_2 = \frac{1}{1 + \sum_{k=0}^{N-1} \frac{N!\,(N-\rho_1-\rho_2)}{k!N\,(\rho_1+\rho_2)^{N-k}}} \cdot \frac{\frac{\rho_1}{\mu_1} + \frac{\rho_2}{\mu_2}}{N-\rho_1-\rho_2} + \frac{\rho_1}{\mu_1} + \frac{\rho_2}{\mu_2}. \tag{2.4}$$

Even though no proof was attempted for (2.4) in this dissertation,[7] the comparison of the numerically plotted total delay versus the total delay derivable from (2.4) proves the validity of this result, with an accuracy of more than 99.5% (cf. Fig. 2.5). As detailed in the forthcoming section, a discrete-event simulator was written for the network model in Fig. 2.1, to check the preciseness of the analytical treatments and simplifications and the validity of the assumptions made.

To solve (2.4), one should note that the primary's total delay (curve in black with circle marks in Fig. 2.5) is the classic formula for $\mathrm{M/M/N}$ (the reason for this was explained before), that is,

$$D_1 = \frac{1}{\lambda_1}\left(\rho_1 + \frac{1}{N} \cdot \frac{\rho_1^{N+1}}{N!} \cdot \frac{P_0}{\left(1-\frac{\rho_1}{N}\right)^2}\right), \tag{2.5}$$

where [39]

$$P_0 = \left(\sum_{k=0}^{N-1} \frac{\rho_1^k}{k!} + \frac{\rho_1^N}{N!} \cdot \frac{1}{1-\frac{\rho_1}{N}}\right)^{-1}. \tag{2.6}$$

In Fig. 2.5, the negligible difference between D_2 obtained from (2.4) and the one calculated numerically from the CTMC is due to the little imprecision caused by the numerical derivation of the probabilities when solving the CTMC. Thus, we conjecture that the per-

[7] Unfortunately, the shortage of abundant investigations in this area seems to be the main reason for the lack of derivations and proofs on conservation laws. For this general case of multi-server priority queues, unresolved complications were reported by those who tackled the problem. In essence, the *supermodularity* property of the long-run expected amount of work (which has been used in the proof of conservation law for single-server priority systems) cannot be proven to hold for multi-server priority systems with different service rates.

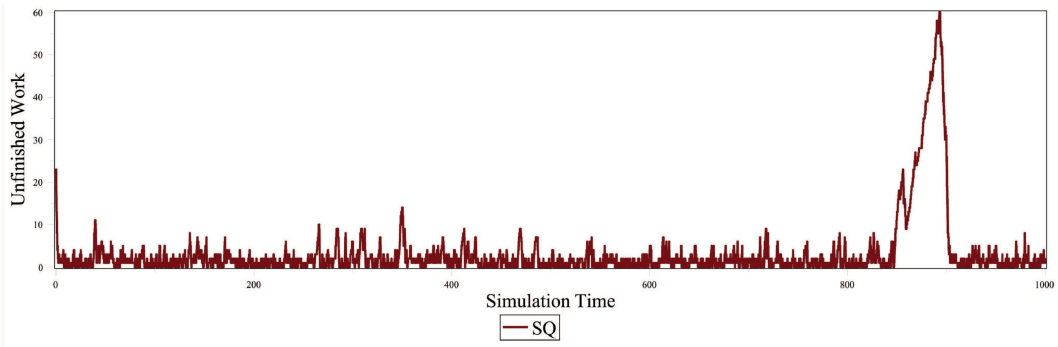


Figure 2.6: Amount of unfinished work in secondary's queue vs. simulation time.

formance vector $[D_1, D_2]$ obtained from (2.4) is exact.

2.4 Simulation Development

2.4.1 Queue Unfinished Work

A discrete-event simulator was built to emulate the behavior of the system under study. This simulator, built in Maple programming language, has been operated for a long time in order to eliminate the transient behavior and enter the steady-state phase (any queuing system has a transient interval whose length depends on many factors). The unfinished work U(t) in the network with $N = 3$ was probed for 1000sec run of the simulator as is plotted in Fig. 2.6, for secondary queue, and Fig. 2.7, for primary queues. As defined by Kleinrock, the unfinished work is a random process that indicates the remaining time required to empty the system of all customers present at time t.

2.4.2 Queue Length Histogram

Fig. 2.8 depicts the histogram of the queue occupancy for the secondary queue and only one primary queue. Since the arrival and departure processes of the queues are chosen to be similar, the histogram of the other queues will look quite the same.

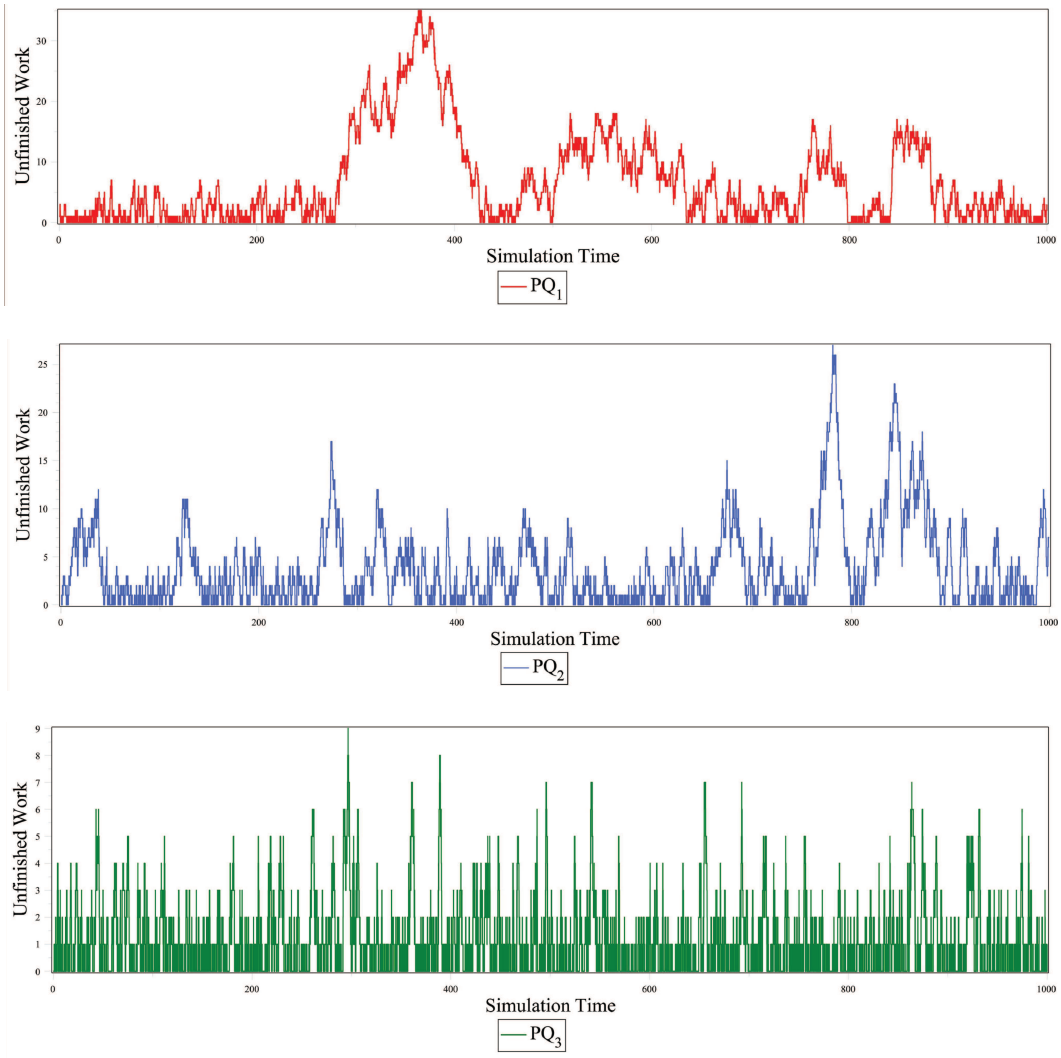
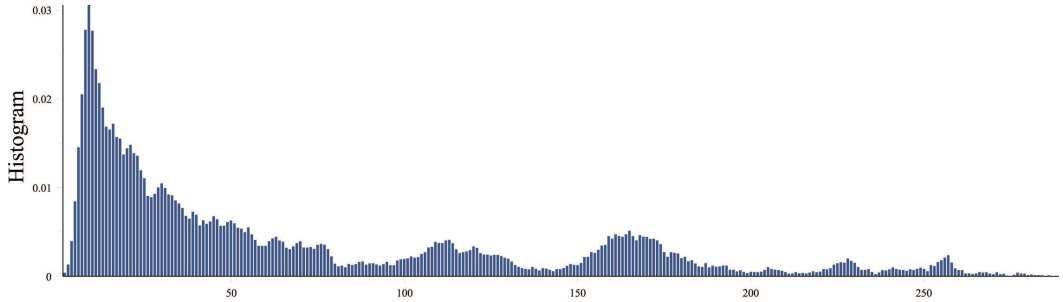
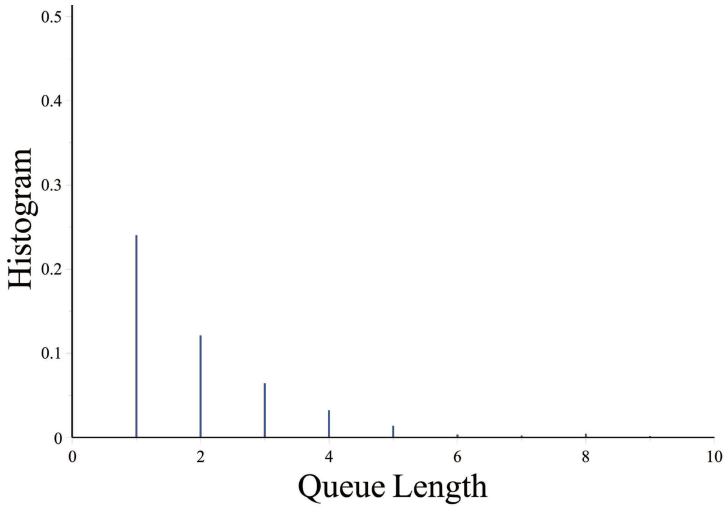


Figure 2.7: Amount of unfinished work in primary queues for the network model in Fig. 2.1 with N=3 servers.



(a) Secondary queue.



(b) Primary queue.

Figure 2.8: Queue length histograms for (a) the secondary queue and (b) one of the primary queues

2.4.3 Validation

For the sake of validation, the simulator was operated for ten different values of ρ_1 (equally-spaced) and the results were illustrated with *crosses* in Fig. 2.5. These results prove that the network model in Fig. 2.1, its associated CTMC in Fig. 2.2 and the closed-form expression in (2.4), all are in tight conformity both in the absolute value and the rate of change. Once again, we draw the reader's attention to the choice of logarithmic (linear) scaling on the left-hand (right-hand) ordinate in Fig. 2.5.

2.5 Achievable Region

The conservation law in (2.4) is not only important in the sense that it provides a simple closed-form expression for the total delay of SUs but is central given that it related total delays in both classes generating an achievable region in the performance space.[8]

This achievability region (convex-hull) for the cognitive scenario with $r = 2$ is illustrated in Fig. 2.9, where each vertex corresponds to one of the classes having unconditional privilege over the other. Furthermore, any other vector inside this region can be achieved by a mixing strategy [36],[9] which assigns privileged access to a fraction of resources to each class ($\alpha_{\rm PU}$, $\alpha_{\rm SU}$), where $\alpha_{\rm SU} + \alpha_{\rm PU} = 1$. For example, for the absolute priority, the upper-left vertex corresponds to $(\alpha_{\rm PU} = 1, \alpha_{\rm SU} = 0)$ and the lower-right vertex corresponds to $(\alpha_{\rm PU} = 0, \alpha_{\rm SU} = 1)$. Taking $\alpha_{\rm PU} = \alpha$, any decrease in α can be interpreted as qualifying the SU to have more exclusive access (proportional to $1 - \alpha$) and enjoy better QoS though it would inflict more interference on the primary.

The next goal is to narrow down the achievable region in Fig. 2.9 given some interference and delay constraints on the primary and the secondary classes. Also, we would like to know whether a performance vector is achievable according to such constraints. This turns the problem at hand into a synthesis problem.

At this point, to better understand the subject, we change our notation for the classes as follows: Classes 1 and 2 (as denoted this far) will be represented by PU and SU, respectively. Then, using the notion of mixed strategy in [37], [36], an arbitrary mixed-strategy

[8] This concept, elaborated in [36, 38, 40], simply states that in a network with r classes and implementing a work-conserving scheduling policy (e.g., preemptive), any achievable performance vector (note that in this section the waiting time W is chosen as performance indicator instead of D) must lie within an r-dimensional *polyhedron* with $r!$ vertices, or equivalently, that any vector lying inside this region is achievable. More importantly, each vertex of this polyhedron corresponds to a different class having absolute priority over the others.

[9] Note that the mixed strategy here is irrelevant to that in Game theory, though conceptually there exist some similarities between them.

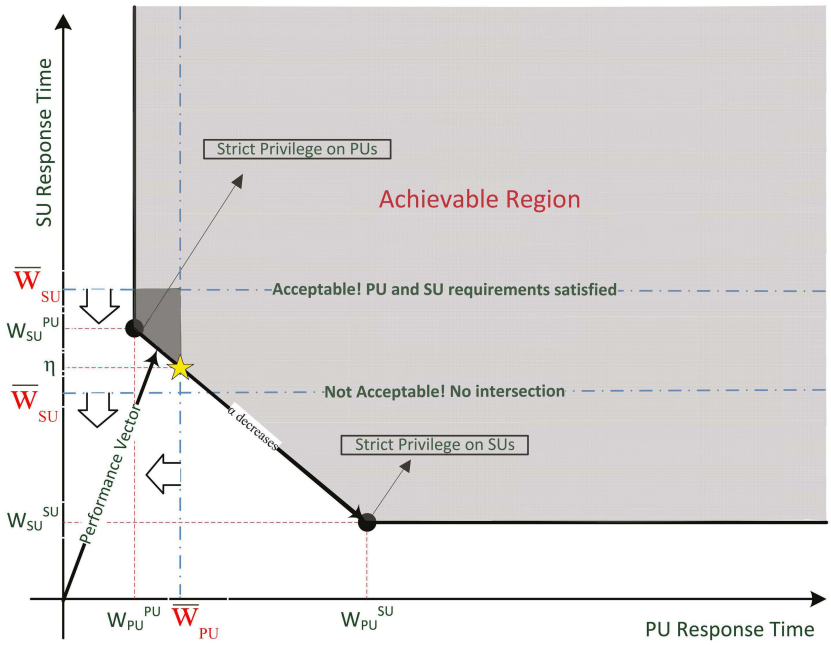


Figure 2.9: Achievable regions in CRNs with and without constraints.

performance vector $W^{\alpha-\text{mix}}$ is obtained using

$$W_i^{\alpha-\text{mix}} = \alpha W_i^{\text{PU}} + (1-\alpha)\, W_i^{\text{SU}}, \quad i \in \{\text{PU}, \text{SU}\}, \tag{2.7}$$

where W_i^{PU} (W_i^{SU}) represents the waiting time on class i $(i \in \{\text{PU}, \text{SU}\})$ when PU (SU) has privileged access to the spectrum. These values are obtained using the following relation between the queue service time and the total delay as

$$\begin{aligned} W_i^{\text{PU}} &= D_i^{\text{PU}} - \frac{1}{\mu_i}, \qquad i \in \{\text{PU}, \text{SU}\} \\ W_i^{\text{SU}} &= D_i^{\text{SU}} - \frac{1}{\mu_i}, \qquad i \in \{\text{PU}, \text{SU}\} \end{aligned}, \tag{2.8}$$

where D_i^{PU} is derived directly from (2.4)-(2.6) and D_i^{SU} is obtained from (2.4)-(2.6) by swapping the priorities such that the secondary has preemptive priority over the primary (a

fictitious case).

A performance vector is achievable if

$$W_i^{\alpha-\text{mix}} < \overline{W}_i, \qquad i \in \{\text{PU}, \text{SU}\}, \tag{2.9}$$

where $\overline{W}_i$ is the waiting delay that class i can tolerate. Such thresholds are imposed by the standards, the QoS requirements, or both. Thus,

$$\begin{cases} W_{\text{PU}}^{\alpha-\text{mix}} = \alpha W_{\text{PU}}^{\text{PU}} + (1-\alpha)\, W_{\text{PU}}^{\text{SU}} < \overline{W}_{\text{PU}} \rightarrow \mathfrak{K}^1(\overline{W}_{\text{PU}}) = \underbrace{\overbrace{W_{\text{PU}}^{\text{SU}} - \overline{W}_{\text{PU}}}^{>0}}_{>0} < \alpha < 1 \\ W_{\text{SU}}^{\alpha-\text{mix}} = \alpha W_{\text{SU}}^{\text{PU}} + (1-\alpha)\, W_{\text{SU}}^{\text{SU}} < \overline{W}_{\text{SU}} \rightarrow 0 < \alpha < \mathfrak{K}^2(\overline{W}_{\text{SU}}) = \underbrace{\overbrace{\overline{W}_{\text{SU}} - W_{\text{SU}}^{\text{SU}}}^{>0}}_{>0} \end{cases}, \tag{2.10}$$

where $\mathfrak{K}^1$ and $\mathfrak{K}^2$ are functions of the thresholds $\overline{W}_{\text{PU}}$ and $\overline{W}_{\text{SU}}$, respectively. In view of (2.10), the constraining pair $(\overline{W}_{\text{PU}}, \overline{W}_{\text{SU}})$ renders the desirable performance vector $[W_{\text{PU}}^{\alpha-\text{mix}}, W_{\text{SU}}^{\alpha-\text{mix}}]$ achievable if and only if the two intervals derived for α in (2.10) overlap, i.e.,

$$0 < \mathfrak{K}^1(\overline{W}_{\text{PU}}) < \alpha < \mathfrak{K}^2(\overline{W}_{\text{SU}}) < 1. \tag{2.11}$$

The expression in (2.11) has an important implication: any target vector $[W_{\text{PU}}^{\alpha-\text{mix}}, W_{\text{SU}}^{\alpha-\text{mix}}]$ corresponds to a unique α and this α should lay within the above interval for this target vector to be achievable. Therefore, either the proper choice of the constraint vector $(\overline{W}_{\text{PU}}, \overline{W}_{\text{SU}})$ or the target vector itself results in an answer. The above problem has no solution if the choices are such that $\mathfrak{K}^1(\overline{W}_{\text{PU}}) > \mathfrak{K}^2(\overline{W}_{\text{SU}})$. For instance, given the

threshold $\overline{W}_{\mathrm{PU}}$ on the primary class (which corresponds to the allowed level of inflicted interference), the lower horizontal dashedfepic line in Fig. 2.9 represents an unrealistic choice for $\overline{W}_{\mathrm{SU}}$ such that $\mathfrak{K}^1(\overline{W}_{\mathrm{PU}}) > \mathfrak{K}^2(\overline{W}_{\mathrm{SU}})$. However, a less constraining while practical choice for $\overline{W}_{\mathrm{SU}}$ can be the upper horizontal line in Fig. 2.9, which forms the small trapezoidal achievable region colored in dark gray. The yellow star corresponds to the performance vector with coordinate $(\overline{W}_{\mathrm{PU}}, \eta)$, where

$$\eta = m' \left(W_{\mathrm{PU}}^{\mathrm{SU}} - \overline{W}_{\mathrm{PU}}\right) + W_{\mathrm{SU}}^{\mathrm{SU}}, \tag{2.12}$$

with

$$m' = \frac{W_{\mathrm{SU}}^{\mathrm{PU}} - W_{\mathrm{SU}}^{\mathrm{SU}}}{W_{\mathrm{PU}}^{\mathrm{SU}} - W_{\mathrm{PU}}^{\mathrm{PU}}}, \tag{2.13}$$

which is feasible, provides the best QoS for the SU while being at the achievability border. If $\overline{W}_{\mathrm{SU}} < \eta$, then the only way out of this dilemma is to settle for larger $\overline{W}_{\mathrm{PU}}$ that would turn around this inequality. If $\eta < \overline{W}_{\mathrm{SU}} < W_{\mathrm{SU}}^{\mathrm{PU}}$, then the primary class suffers from observing some amount of interference equivalent to the excess delay $\Delta_{\mathrm{PU}} = (W_{\mathrm{SU}}^{\mathrm{PU}} - \overline{W}_{\mathrm{SU}})/m'$ on the primary side compared to the ideal case. Finally, if $\overline{W}_{\mathrm{SU}} > W_{\mathrm{SU}}^{\mathrm{PU}}$, then no interference is inflicted by the secondary at the cost of the larger excess delay it accumulates. It should be noted that the excess delay Δ_{PU} that the primary class experiences is equivalent to the amount of inflicted interference through the relationship between time, power and energy.

2.6 Mixed Strategy in CRNs

In this section a cost function is defined that relates the primary's interference cost and the secondary's QoS metric in an intuitive manner. With the insights shed in the last section, the length of the performance vector in Fig. 2.9, which starts from the origin and ends at a point lying in the achievable region, seems a natural choice for the cost function given

that it relates the inflicted interference and experienced delay in a methodologically correct way. Thus, the said function is given by

$$F(\alpha) = \sqrt{\underbrace{\left(\alpha W_{\mathrm{PU}}^{\mathrm{PU}} + (1-\alpha) W_{\mathrm{PU}}^{\mathrm{SU}}\right)^2}_{\text{interference factor}} + \underbrace{\left(\alpha W_{\mathrm{SU}}^{\mathrm{PU}} + (1-\alpha) W_{\mathrm{SU}}^{\mathrm{SU}}\right)^2}_{\text{SU's QoS}}}, \tag{2.14}$$

where α can only take values imposed by the constraint in (2.11). For simplicity, let us introduce the notation $W_{\mathrm{PU}}^{\mathrm{PU}} = A_1$, $W_{\mathrm{PU}}^{\mathrm{SU}} = A_2$, $W_{\mathrm{SU}}^{\mathrm{PU}} = A_3$ and $W_{\mathrm{SU}}^{\mathrm{SU}} = A_4$. Then after simplification, the quadratic cost-function in (2.14) becomes

$$F(\alpha) = \sqrt{C_1\alpha^2 - C_2\alpha + C_3}, \tag{2.15}$$

where

$$\begin{aligned} C_1 &= (A_2 - A_1)^2 + (A_3 - A_4)^2 > 0, \\ C_2 &= 2\left(A_2\left(A_2 - A_1\right) - A_4\left(A_3 - A_4\right)\right), \\ C_3 &= A_2^2 + A_4^2. \end{aligned}$$

The coefficient C_1 in (2.15) is always positive. This means that the cost function may have a local minimum at

$$\alpha^* = \frac{C_2}{2C_1} = \frac{A_2\left(A_2 - A_1\right) - A_4\left(A_3 - A_4\right)}{\left(A_2 - A_1\right)^2 + \left(A_3 - A_4\right)^2}. \tag{2.16}$$

Since $A_1 < A_3$, $A_1 < A_2$, $A_4 < A_2$ and $A_4 < A_3$, α^* can be negative, positive, or zero, depending on the sign of the coefficient C_2 in (2.15). Since it was shown in (2.11) that imposing constraints would limit the range of values for the factor α, a local minimum exists only when $0 < \mathfrak{K}^1(\overline{W}_{\mathrm{PU}}) < \alpha^* < \mathfrak{K}^2(\overline{W}_{\mathrm{SU}}) < 1$. Three cases may arise for the value of α that minimizes the cost function (say $\alpha^{\min}$) depending on where within the above

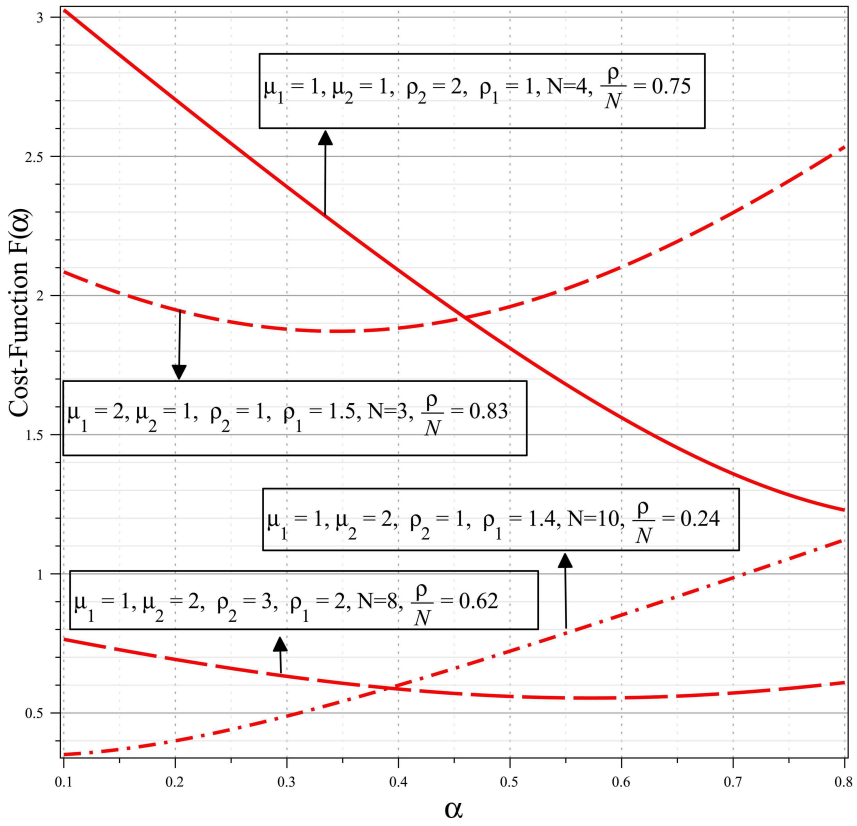


Figure 2.10: Cost-function $F(\alpha)$ for different values of parameters and $\mathfrak{K}^1(\overline{W}_{\mathrm{PU}}) < \alpha < \mathfrak{K}^2(\overline{W}_{\mathrm{SU}})$.

interval $\alpha^{\min}$ falls:

- if $\alpha^* < \mathfrak{K}^1(\overline{W}_{\mathrm{PU}})$, then $\alpha^{\min} = \mathfrak{K}^1(\overline{W}_{\mathrm{PU}})$.
- if $\alpha^* > \mathfrak{K}^2(\overline{W}_{\mathrm{SU}})$, then $\alpha^{\min} = \mathfrak{K}^2(\overline{W}_{\mathrm{SU}})$.
- if $\mathfrak{K}^1(\overline{W}_{\mathrm{PU}}) < \alpha^* < \mathfrak{K}^2(\overline{W}_{\mathrm{SU}})$, then $\alpha^{\min} = \alpha^*$.

This is better illustrated in Fig. 2.10, which shows how the choice of parameters may lead to a cost-function that does not have a local minimum but have minimum value on the borders. Since the primary and secondary constraints only allow α to take values in

$[\mathfrak{K}^1(\overline{W}_{\rm PU}), \mathfrak{K}^2(\overline{W}_{\rm SU})]$ (as shown in (2.10) and (2.11)), our choice was for the threshold values $\overline{W}_{\rm PU} = 0.9A_2 + 0.1A_1$ and $\overline{W}_{\rm SU} = 0.8A_3 + 0.2A_4$. However, in a practical scenario, $\overline{W}_{\rm PU}$ and $\overline{W}_{\rm SU}$ might be independent parameters and should be chosen with regard to service requirements. In comparing these curves, it is remarkable that the lower curves are not necessarily better as they are related to settings with lower overall utilization (smaller ρ/N).

The importance of this result in practical realization of CRNs is related to the proper choice of the target vector in the performance space of Fig. 2.9. More precisely, if the target vector and the thresholds $\overline{W}_{\rm PU}$ and $\overline{W}_{\rm SU}$ are chosen such that the value of α obtained from (2.7) is equal to $\alpha^{\min}$, then the cost of inflicting interference on the primary would be minimum while the maximum quality is offered to the secondary.

2.7 Model Refinement

In the modeling and analysis developed this far, we intentionally dropped some details to avoid making the model unnecessarily complex. In this section, we add to the model complexity by taking into account the channel and sensing imperfections.

2.7.1 Sensing Time

Considering that the periodic sensing task in half-duplex (HD) radio requires the interruption of transmission, the previous model overestimates the performance of CRNs as it does not take into account the resource wasted on not transmitting data payloads when the channels are not occupied by PUs. As a matter of fact, the proposed model does not need any refinement for the case where the cognitive nodes are equipped with full-duplex (FD) radio, as FD allows SUs to perform sensing and transmitting, at the same time. Thus, any discussion in the rest of this section is only limited to the HD radio.

The new adjustment is equivalent to having two equally-ranked processes (instead of one) privileged over the secondary process, i.e., the primary process and the sensing process. Note that these latter processes are not privileged upon each other and considered independent, thus, the existence of one does not hinder the occurrence of the other. The idea is to reflect the impact of the sensing process on the secondary. Analogous to Theorem I in Section 4.2.5, where the independence assumption for channels will lead to the representation of the number of busy channels with a Binomial RV, here the number of sensed channels during a time period T would be a Binomial RV as well. In other words, when each channel (server) in Fig. 2.1 is sensed for ΔT sec every T seconds, the probability that a station senses a channel ($P^{\text{Sen.}}$) (rather than transmit) would be $P^{\text{Sen.}} = \Delta T/T$, and thus, on the long term, the fraction of channels used for payload transmission would be a RV with distribution $\text{Binomial}\left(N,\, 1 - P^{\text{Sen.}} = (T - \Delta T)/T\right)$. Therefore, on average, $N\left(1 - P^{\text{Sen.}}\right)$ channels would be available to the SUs and (2.2) is refined after replacing N with $N(1 - P^{\text{Sen.}})$, i.e.,

$$\frac{\rho}{1 - P^{\text{Sen.}}} = \rho' < N. \tag{2.17}$$

This equation implies that a busier system with larger utilization factor $\rho' > \rho$ is experienced on the secondary side. With the changes in (2.17), all the equations derived before remain valid by leaving the primary's quantities intact while changing ρ_2 to $\rho_2/(1 - P^{\text{Sen.}})$ and μ_2 to $\mu_2(1 - P^{\text{Sen.}})$.

2.7.2 Sensing and Channel Imperfections

The inaccuracies in sensing is manifested by false alarm (P_{FA}) and misdetection ($1 - P_{\text{MD}}$) events. Misdetection is a serious problem as it lets undesirable interference be inflicted on the PUs. The other destructive effect is due to the channel impairments, such as fading and noise. Assuming that the occurrence of misdetection leads to packet loss (no possible recovery), then a packet is successfully detected if neither misdetection occurs nor the

channel is in deep fading. Thus, the probability of packet loss, P_{PL}, is given by

$$P_{\mathrm{PL}} = 1 - P_{\mathrm{D}}\left(1 - \mathrm{PER}\right), \tag{2.18}$$

where PER represents the packet error rate. Since these two effects translate into a reduction in the effective transmission rates for both the SUs and the PUs, then, by assuming that the successive transmissions of collided packets are probabilistically independent, the number of times a data packet gets retransmitted is geometrically distributed. Therefore, the model would be properly refined by performing the changes $\mu_2 \rightarrow \mu_2 P_{\mathrm{D}}\left(1 - \mathrm{PER}_2\right)$ and $\mu_1 \rightarrow \mu_1 P_{\mathrm{D}}\left(1 - \mathrm{PER}_1\right)$.

2.8 Summary

The theory of priority queues has been applied to model and analyze the performance of single-interface multi-channel CRNs. This modelling effort differs from previous studies in this area in the methodology and the extent of investigations. More specifically, the queue length distribution of SUs were derived and the stability condition of the secondary network was discussed. Aside from the modeling and performance analysis problems, the network synthesis was conducted and the achievable region for a 2-class MC-CRN was derived. Finally, the model was refined through consideration of sensing time as well as sensing and reporting channels' imperfections.

Chapter 3

Single-Interface CRNs: Queuing Network's Approach[1]

As it was explained earlier, the inadequacy of available spectrum, when accompanied with the existing inefficient rigid allocation of spectrum, has exacerbated the situation and resulted in a poor utilization of this valuable resource. This has motivated the research community to take the initiative and introduce efficient methods for spectral utilization. All these efforts recently appeared in the form of a CR-based IEEE standard, namely the IEEE 802.22 [8], which targets the temporally and spatially underutilized VHF/UHF frequency band [3], [5] for opportunistic exploitation.

From a practical viewpoint, the huge cost reduction that CR brings (gained by widening the reach of the network and lessening the amount of required equipment) is even more pronounced for highly remote areas (targeted by IEEE 802.22). The rural and remote locations are where the demand for broadband access exists but the population sparsity does not justify to deploy optical fibers.

Likewise the former efforts on the mathematical modeling of wireless standards, such as

[1] N. Tadayon and S. Aïssa, "Modeling and analysis of cognitive radio based IEEE 802.22 wireless regional area networks", *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4363–4375, Sep. 2013.

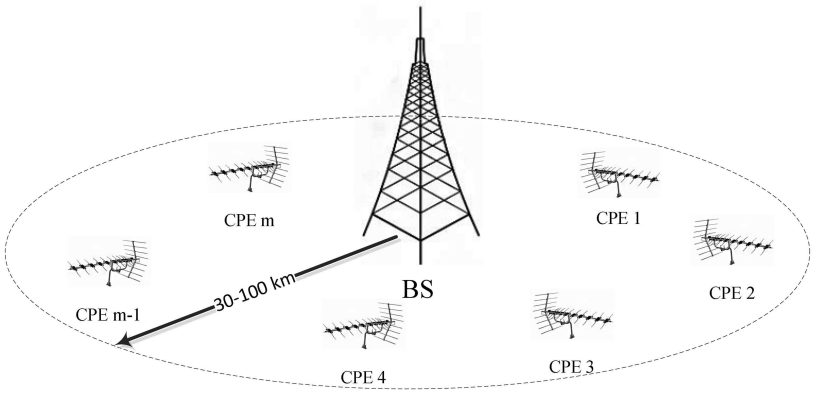


Figure 3.1: An IEEE 802.22 scenario.

IEEE 802.11, which helped forward the extension of the technology outreach and advanced the succeeding progresses, the modeling of IEEE 802.22 WRANs is an essential task. As a matter of fact, the importance of such modeling is even more significant due to concerns about the incumbent protection. An analytical model delivers an easy and cost-effective tool for the pre-deployment evaluations. This constitutes the main focus of this dissertation in this chapter. Contrasting this thesis's contribution from earlier studies, in this chapter the modeling and performance evaluation of IEEE 802.22 WRAN is pursued. In so doing, it will be shown that a WRAN cell that works in a centralized manner can be modeled as a *Jackson* network. In analyzing this model, the dynamics of each queue is represented with a two-dimensional CTMC (whose details are discussed in Chapter 4). Through numerical analysis, the PMF and the cumulative distribution function (CDF) of the queue length of the customer premises equipment (CPE) are found as the fundamental quantities from which any higher moment can be obtained. Moreover, a closed-form expression for the PMF of the CPE's queue length will be derived and the accuracy of precision compared. At last, an analysis of the end-to-end delay in a WRAN cell will be conducted. For ease, the main

Table 3.1: Main notations in this chapter.

Symb.	Definition	Symb.	Definition	Symb.	Definition
D_{BB}	Backbone delay	n_i	# of subch. of CPE$_i$	λ_i	Traffic arrival rate
N	Total # of vacant subch.	P_{B}	Subch. busyness prob. (Perceived)	m	Number of active CPEs.
$P_{\mathrm{S},i}$	Succ. trans. prob	$P_{\mathrm{S,BS}}$	Succ. trans. prob. of BS	μ_i	Subch. service rate
μ_{BS}	Subch. service rate	$\mathtt{M}_i$	Constellation size	$r_{c,i}$	Coding rate
T_{F}	Frame length	t	DS/Frame ratio	f	Quiet period/frame ratio
η_i	Subch. addition rate	ζ_i	Subch. elimination rate	$\pi_i(n_i,\ell)$	Prob. of having n_i subch. and ℓ packets
$\pi_i(n_i)$	Prob. of having n_i subch.	q_i	Access prob.	$\rho_{\mathrm{BS}}^{\mathrm{US}}$	Subch. utilization factor
$\rho_{\mathrm{BS}}^{\mathrm{DS}}$	Subch. utilization factor	ρ_i	Utilization factor	$\Lambda_{\mathrm{BS}}^{\mathrm{US}}$	Gross arrival rate
$\Lambda_{\mathrm{BS}}^{\mathrm{DS}}$	Gross arrival rate	μ_{BB}	BS service rate	$P_{\mathrm{D},i}$	Detection prob.
$P_{\mathrm{Out},i}$	Outage prob.	$\mathrm{SNR}^{\min}$	Receiver threshold	D_{PtP}	Peer-to-Peer delay
W_i	Queuing delay	$W_{\mathrm{BS}}^{\mathrm{US}}$	Queuing delay in US	$W_{\mathrm{BS}}^{\mathrm{DS}}$	Queuing delay in DS

notations used in this section are summarized in Table 3.1.[2]

3.1 IEEE 802.22 Centralized Access Mechanism

In this subsection, the mechanics behind the physical (PHY) and medium access control (MAC) layers in the IEEE 802.22 standard are explained, which are required for the understanding of the modeling that will follow in the subsequent sections.[3]

3.1.1 Cell Association

As detailed in [8], the MAC layer defined in IEEE 802.22 is connection-oriented similar to WiMax, GSM, etc. Such choice was made to have more control on the QoS of cognitive nodes while controlling the interference inflicted on incumbent base stations (BSs). To accomplish this task, the access to the spectrum resources is granted by the secondary (WRAN) BS that is aware of the channel activities of incumbent BSs and its associated cognitive nodes (CEPs). Before a CPE joins the network, a series of tasks shall be compiled;

[2] Subscripts $_i$ and $_{\mathrm{BS}}$ in Table 3.1 indicate that a parameter pertains to the i^{th} CPE or base station (BS), respectively. The same logic is used for superscripts $^{\mathrm{DS}}$ and $^{\mathrm{US}}$ to indicate the association of a parameter to downstream and upstream paths, respectively.

[3] To the best of the author's knowledge, so far, only [41] and [42] overview the IEEE 802.22 functionality in a brief manner.

that CPE should first sense the spectrum and choose an operating WRAN cell (say k), and then initiate the second sensing stage by sensing the channels that the k^{th} BS declared as operating (by overhearing the superframe control header-SCH) to figure out which channel is vacant and which one is not. Once the incumbent BS is detected, the CPE shall notify the WRAN BS of such situation. Then, this CPE obtains its location information using the geolocation system and synchronizes itself with the WRAN BS using downstream MAP (DS-MAP) message. The initial ranging is then initiated to estimate the delay and radiated power (EIRP).

Based on the periodic updates that the WRAN BS receives from the incumbent database (DB) about the inhibitive and safe locations for cognitive operation, and given that the WRAN BS is aware of the location of CPEs at the network association phase, it decides whether the soliciting CPE is located in an interfering area or not and, thus, if it can join the network. Because the non-interfering operation of an IEEE 802.22 network is of utmost importance, it is required that a local sensing be performed by both the BS and CPEs.

3.1.2 Sensing and Incumbent Protection

The standard defines two types of measurements during which all CPEs operating on the same channel shall keep silent by ceasing their transmissions: *automated* (carried out by CPEs autonomously), and *hierarchical* (initiated by management messages). The MAC is able to fully manage the periodic sensing through the quiet period (QP) advertisements coming in SCH message at the beginning of each superframe.[4] During QPs, CPEs/BS should stop their transmissions by keeping silent while listening to channel activities for a sufficient time span. Even though some recommendations were made in [8] for the preferred sensing methods in WRANs, the ultimate choice, the sensing rate and the sensing length, are left open to manufacturer as long as the requirements on detection probabil-

[4]Note that the network-wide quiet period management is only mandatory for the in-band channels (operating channels plus their two immediate adjacent channels) not for the out-of-band channel measurements.

ity, allowed interference to incumbent operator and channel detection time, are met. This requires that CPEs/BS be synchronized in starting their QPs in order to have a precise measure of whether the primary operator has returned or the channel is still vacant.

Different sensing methods perform differently in terms of the false alarm and misdetection metrics, complexity, and spectrum efficiency. For instance, the implementation ease [43], [44] and cost effectiveness associated with the energy detection sensing method comes at the price of its lower bandwidth efficiency [45] (due to a larger false alarm probability) when CPEs are unsynchronized. Indeed, this sensing method is unable to recognize whether a transmission is from a CPE or a licensed user. On the other hand, sensing methods such as spectral correlation, time domain cyclostationarity, eigenvalue detection, etc. [46], [47], [48] are more precise but more complex as well.

The sensing results from different CPEs are to be reported back to the corresponding BS. This is possible either (i) immediately during urgent coexistence situation (UCS) intervals (transmission by CDMA (code division multiple access)-based or contention-based mechanisms), (ii) during allocated upstream (US) intervals, (iii) or, non-periodically, using BLM-REQ/RSP messages in the following frames. Once the BS collects the CPEs' sensing results, it combines them along with its own sensing outcome in order to decide whether a frequency channel has been occupied by a licensed operator or is still empty. This cooperative method reduces the possibility of the hidden terminal problem and many other uncertainties due to fading and shadowing [44], [49], [50].

In case the BS deduces that the incumbent has returned to the channel or that the quality of channel has degraded noticeably compared to those available in backup channels, that TV channel is removed from the channel list of operating channels, and the affected CPEs hop to a channel in the backup list within a predefined interval so as to avoid interference with licensed users (or to accomplish the communications over a channel with better quality).

There are three different types of quiet periods during which the sensing task by all

CPEs is performed:

- *DS-length quiet period* (<10 msec): sensing lasts for the length of one DS (downstream) subframe (Fig. 3.2).
- *Intra-frame quiet period* (10 msec): sensing lasts for the duration of a frame.
- *Inter-frame quiet period* (160 msec): sensing lasts for the length of a *superframe*.

Depending on the requirement imposed by the BS, any of these quiet periods might be adopted during the network operation. While the first two types yield a medium level of accuracy, the third type provides a more precise result on the presence/absence of licensed user. Per standard, the detection threshold for digital TV signals is -116 dBm and -94 dBm for analog TV signals.[5]

3.1.3 Access Mechanism and Frame Structure

The MAC mechanism of IEEE 802.22 is a combination of orthogonal frequency division multiplexing (OFDM) and time division multiple access (TDMA), in order to take advantage of both in terms of scheduling granularity, flexibility, and centralized access while being able to tackle the channel impairments. Transmissions from CPEs and BS are jointly structured in subsequent intervals called *Superframes* with equal length of 160 msec. Each *superframe* is split up to 16 *frames*, each of length 10ms. Moreover, each *frame* is comprised of two *subframes* with non-equal lengths (DS subframe for data transmissions from BS to CPEs and US subframe for transmissions from CPEs to BS) and possibly a third part called a SCW window during which coexistence task is performed by transmitting the coexistence beacon protocol (CBP) packets. We neglect this latter part as we assume that WRAN operates in normal mode. Even though the boundary between downstream

[5] In this thesis, we only consider the TV incumbent services as these occupy the major portion of UHF/VHF RF band compared to wireless microphones.

(DS) and upstream (US) subframes is non-rigid, their sum must not exceed the 10ms *frame* length constraint. The US/DS intervals alternation is according to time division duplexing (TDD). It should be noted that it is impossible to consider all the details of the standard into the modeling; for example, the self-coexistence window, the CBP bursts, and the short separating intervals TTG and RTG are neglected in this analysis as they are either irrelevant or could unnecessarily add to the complexity of the model.

This being said, the access mechanism for data transmission in IEEE 802.22 belongs to the category of non-contending channelized MAC protocols.[6] The OFDM mechanism brings up a great deal of flexibility by letting each node transmit at a different rate and with different EIRP while exploiting different modulation/coding combinations. Furthermore, OFDM can neutralize the destruction caused by the frequency selectivity/time dispersion of the channel in long-reach communications as well as the Doppler spread (negligible in a WRAN scenario). The impact of the frequency selectivity is further alleviated by adding a temporal cyclic prefix (of ratio 1/4, 1/8, 1/16 and 1/32) to the end of each OFDM symbol depending on the BS-CPE distance. Fig. 3.2 shows the *superframe* structure in IEEE 802.22. The number of OFDM symbols that can be transmitted in an IEEE 802.22 *frame* (DS+US) is restricted to the total of 26, with the minimum of seven symbols that each CPE should be assigned during DS.[7] As such, it is impossible to schedule more than two CPEs to transmit in a *frame* over the same frequency subchannel.

As shown in Fig. 3.2, two different preambles are defined in the IEEE 802.22 *superframe* standard:

- **Superframe preamble:** It is embedded at the beginning of a *superframe* and entails time and frequency synchronization information (e.g., for initial network association).

[6]The contention-based access is also used for other controlling and management purposes in situations such as BW request, co-existence discovery, urgent incumbent presence notification, ranging, etc.

[7]That is because pilot carriers are spread over seven symbols.

- **Frame preamble:** It is embedded at the beginning of each *frame* and entails channel estimation information, as well as data for tuning the frequency offset.

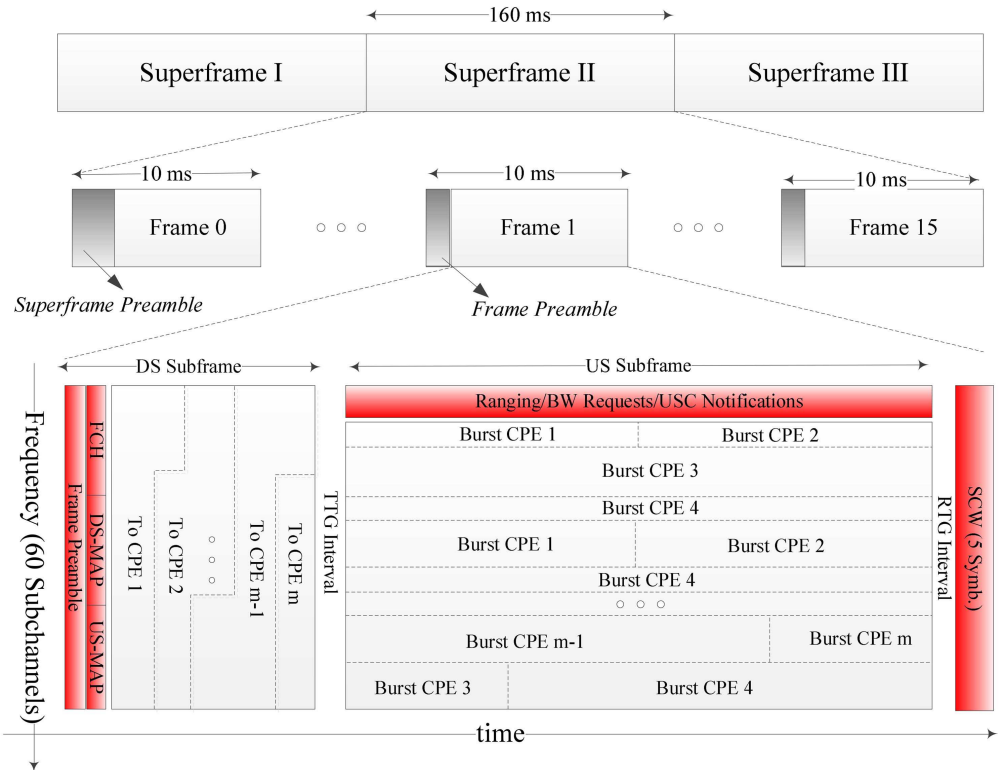


Figure 3.2: IEEE 802.22 *frame* and *superframe* structures.

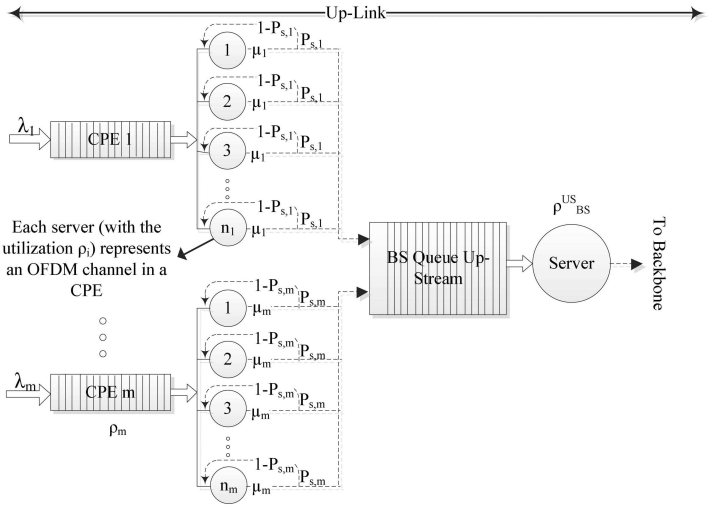
Similar to the IEEE 802.16, in orthogonal frequency division multiple access (OFDMA)-based systems, the total bandwidth is divided into a number of orthogonal subcarriers. However, despite IEEE 802.16, the number of subcarriers is fixed to 2048 all given equal widths. These subcarriers are bundled into 60 groups (subchannels), each having 28 carriers. The total of 368 null and and guard subcarriers is needed to decrease the chance of inter-carrier interference (ICI) and cross-talk. Since the frequency bands are the traditional analog/digital TV channels with bandwidth $\mathtt{W} = 6$ MHz (7 and 8 MHz in some other countries), each carrier would be of width 2.9 KHz. From 28 subcarriers in each subchannel, only 24 are assigned for data transmission and the other 4 are used as the pilot subcarriers

for the purpose of channel estimation. These reference pilot signals are spread among resource elements and are essential for time/frequency domain synchronization, determining MIMO weights, and estimating channels' gains. The remaining 1440 carriers (equivalent to 5.6 MHz) are scheduled for different CPEs according to their traffic demand and some network considerations (distance and proximity to BS, number of associated CPEs, etc.) in such a way that no two CPEs are assigned the same subchannel at the same time. This way, the orthogonality among the subcarriers is always sustained and the ICI is prevented. Assigning subcarriers to subchannels are done such that neighboring subcarriers fall in different subchannels in order to increase the frequency diversity and decrease the undesirable frequency correlations. Moreover, the CPEs might be assigned different numbers of subchannels over time, according to a cyclic pattern in order to enhance the fairness and access level.

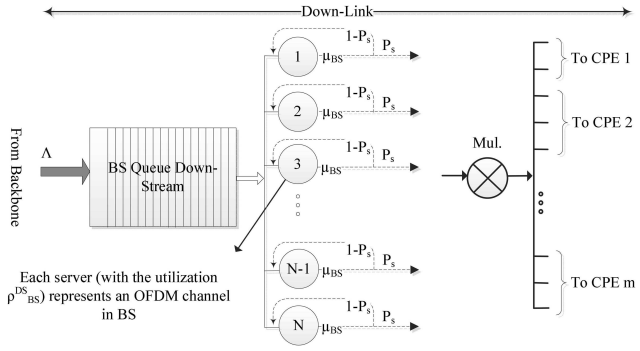
In terms of allocation, the BS distributes the available subchannels vertically (horizontally) during DS (US) in the form of $bursts$ (header and payload). Resource schedules are then embedded inside DS/US-MAPs, which are broadcast in the first OFDM symbol (DS) of every frame. Management messages and MAC protocol data units (PDUs) can then be transmitted by each CPE within its allocated $bursts$ with the possibility of fragmentation/concatenation if a burst is not sufficient to transmit a whole PDU. The details of management plane, service scheduling, ARQ, etc. are not discussed in this thesis due to lack of space and versatility of these functions. For further details, interested readers are referred to [8]. Meanwhile, we assume that the WRAN cell operates in $normal$ mode (as permitted in [8]) where no-coexistence situation can arise, as opposed to the $coexistence$ mode, which allows several overlapping WRAN cells to share the available spectrum in a synchronized manner.

Given that a CPE is assigned a set of subcarriers, a total of 14 PHY modes involving different modulation/coding schemes is defined in IEEE 802.22. This gives the CPE enough degree of freedom to choose a suitable transmission mode. Particularly, the combination of

BPSK, QPSK, 16QAM and 64QAM modulations along with coding rates 1/2, 2/3, 3/4 and 5/6 open up opportunities to establish a reasonable trade-off between robustness, data-rate and power consumption.



(a) Upstream (US) direction



(b) Downstream (DS) direction

Figure 3.3: Modeling of an IEEE 802.22 cell with a network of interconnected queues.

3.2 IEEE 802.22 Modeling

Denote the total number of subchannels in the network by n. As pointed out before, for a single 6 MHz-based TV channel $n = 60$, and for a higher number of TV channels (say N_c),

$n = 60N_c$. If a subchannel in the i^{th} channel is perceived as occupied, with probability P_{B}, then that channel as well as the $(i-1)^{\text{th}}$ and the $(i+1)^{\text{th}}$ channels will be declared occupied in the subsequent notification interval. On the other hand, a channel is declared vacant, if all its subchannels are found vacant, each occurring with probability $1 - P_{\text{B}}$. As explained before, we do not deal with the fact that P_{B} is affected by many factors such as all local sensing results and the incumbent database. Thus, we handle P_{B} as a lump quantity. That granted, the average number of available subchannels that can possibly be assigned to a tagged CPE would be equal to $N = (1 - P_{\text{B}})\, n$.

The available resources have to be shared among m CPEs, each receiving n_i subchannels. As previously pointed out, the 386 null/guard control subcarriers and 240 pilot control subcarriers in each TV channel can never be used for the transmission of payloads. We assume that such provisioning of carriers for control/management purposes works perfectly and fixes any frequency drift and synchronization problem while estimating the channel condition with no errors. The other valid assumption in this section is related to the statistical independence of frequency subchannels. This assumption is valid because (i) for the proper non-interfering operation, the OFDMA carriers allocated to different CPEs are distinct, (ii) the orthogonality principle in OFDMA guarantees that the transmission on different subcarriers do not interfere with each other, (iii) and, more importantly, subcarriers are deliberately shuffled inside a subchannel to gain frequency diversity. Therefore, any set of arbitrary chosen subcarriers can be considered independent. Since the modeling approach taken in this section is based on the theory of queues, the above-mentioned assumption builds the footstone by affirming that a single subchannel dedicated to a CPE can be modeled as a single *queuing server* in a queue element.

3.2.1 WRAN Cell Modeling: Network of Tandem Queues

The task starts by modeling the WRAN cell in Fig. 3.1 with a tandem network of queues as shown in Fig. 3.3 where m CPEs and the BS are modeled with $m+1$ distinct elementary queues. A complete characterization of this model requires determining the input/output queues' processes and the queuing discipline. Depending on transmission histories of CPEs and the remoteness/closeness of them to the BS, they might be assigned different numbers of subchannels by the scheduler. Two distinct subchannels scheduled for two different CPEs do not necessarily provide the same rate due to the fact that they may transmit using different modulation/coding modes with different spectral efficiency (because they experience different channel gains). However, as per the standard, the same mode is to be used for transmission over the subchannels assigned to a given CPE. Therefore, subchannels assigned to that CPE can be modeled with servers of the same transmission rate. The input process to a CPE is a stream of data with average rate λ (Poisson process). The queuing policy is FIFO. Even though the OFDMA access mechanism well tackles the impairments caused by channel fading, our model takes into account an accumulative inaccuracy factor $1-P_{\mathrm{S},i}$ for each CPE.

The interactions among queues is such that a data packet from CPE (BS) already embarked on a server (or subchannel) in the US (DS) interval gets transferred to BS (CPE) with probability $P_{\mathrm{S},i}$ ($P_{\mathrm{S,BS}}$) and is retransmitted in the next allocated transmission opportunity (a.k.a. burst) with probability $1-P_{\mathrm{S},i}$ ($1-P_{\mathrm{S,BS}}$). CPEs experience different $P_{\mathrm{S},i}$ depending on their locations, channel conditions and transmission modes. Once a transmission from a server (subchannel) of the i^{th} CPE is completed, the next head-of-line chunk of data embarks on that server regardless of the status of other servers.

The channel is assumed to be slowly varying (slow fading) with mild time-dispersive behavior (flat fading),[8] thus, $P_{\mathrm{S},i}$ will be constant over time.

[8]Both these assumptions are justified by the fact that WRANs are meant for household deployment in rural areas where the propagation environment changes slowly and the cyclic prefix is chosen long enough to

On the BS side, during the US, the data packets from different CPEs are detected simultaneously and piled up in the BS's queue (Fig. 3.3a) for delivery to the network core. On the DS side, data packets from the network core are delivered with rate Λ to the BS. These packets are served by the BS during the DS (Fig. 3.3b) in each *frame* using the N servers (OFDM subchannels). The multiplier operator in Fig. 3.3b is a simplistic representation of an OFDMA receiver. The network of queues in Fig. 3.3 fully describes the end-to-end behavior of the IEEE 802.22 cell.

The modeling approach in this section consists of applying the theory of Jackson networks [51] to the model illustrated in Fig. 3.3, which requires the following three conditions to be met:

- **Independent CPEs (queues):** The reasoning given before on the statistical independence of distinct subchannels, along with the fact that the transmission lengths are fixed, reinforce the temporal and spectral independence of CPEs from each other and, thereby, confirms that the node independence condition is satisfied.

- $\mathrm{M/M/c}$ **queue:** At the MAC layer, each CPE is modeled with an $\mathrm{M/M/c}$ queue.

- **Stability:** The traffic rate that is destined to a queue from upper layers shall be less than the capacity that such connection can sustain, otherwise, traffic overflow takes place resulting in an unstable queue model. For the queue to be stable, the proper choices of the input rate (by congestion/flow control, e.g., in the transport layer) and PHY parameters (output rate) are necessary. That being the case, the queue has steady state PMF (bounded queue), which is a direct consequence of the ergodicity and irreducibility of its corresponding CTMC.

nullify the inter-symbol interference (ISI) caused by time-dispersion (multi-path).

3.2.2 Server Characterization and Rate Determination

As known, the OFDM symbol rate can never exceed the corresponding allocated bandwidth. Since for the 6 MHz TV channels in IEEE 802.22 with 2048 subcarriers only 1440 are used for pure payload transmission (equivalent to 5.62 MHz), the maximum achievable symbol rate would be 5.62 MSymbol/s. Due to the fact that this symbol rate is to be transmitted in a *frame* of length 10ms, the maximum number of OFDM symbols that can be transmitted would be 26. That is why a number of 26 symbols is mentioned in [8].

The objective is to find the service rates μ_i and μ_{BS} bit/(s $\cdot$ subch.) in the network of Fig. 3.3a for the i^{th} CPE, that is, the achievable data rate per subchannel. Assuming that a CPE is utilizing QAM modulation with constellation size $\mathtt{M}_i$ and FEC coding of rate $r_{c,i}$, the bit rate of the i^{th} CPE and the BS would be equal to

$$\begin{aligned}\mu_i =& P_{\text{S},i}\frac{\overbrace{\frac{\#\text{symbols}}{\text{subcarrier}}}^{26(1-\epsilon^{\text{DS}})}\overbrace{\frac{\#\text{subcarriers}}{\text{subchannel}}}^{24}\overbrace{\frac{\text{bit}}{\text{symbol}}}^{\log_2(\mathtt{M}_i)} r_{c,i}(1-\epsilon^{\text{sen.}})}{T_{\text{F}}} = \\ & 62.4\cdot\left(1-\epsilon^{\text{DS}}\right)\cdot r_{c,i}\cdot\log_2\left(\mathtt{M}_i\right)\cdot(1-\epsilon^{\text{sen.}})\,P_{\text{S},i}\,\text{kbit/s},\end{aligned} \tag{3.1}$$

$$\begin{aligned}\mu_{\text{BS}} =& P_{\text{S},i}\frac{\overbrace{\frac{\#\text{symbols}}{\text{subcarrier}}}^{26\epsilon^{\text{DS}}}\overbrace{\frac{\#\text{subcarriers}}{\text{subchannel}}}^{24}\overbrace{\frac{\text{bit}}{\text{symbol}}}^{\log_2(\mathtt{M}_{\text{BS}})} r_{c,\text{BS}}(1-\epsilon^{\text{sen.}})}{T_{\text{F}}} = \\ & 62.4\cdot\epsilon^{\text{DS}}\cdot r_{c,\text{BS}}\cdot\log_2\left(\mathtt{M}_{\text{BS}}\right)\cdot(1-\epsilon^{\text{sen.}})\,P_{\text{S,BS}}\,\text{kbit/s}.\end{aligned}$$

respectively, where $T_{\text{F}} = 10$ msec is the length of a frame and $26\left(1-\epsilon^{\text{DS}}\right)$ represents the number of OFDM symbols that is allocated for US transmission. The factor $\epsilon^{\text{sen.}}$ in (3.1) is the fraction of time spent for periodic sensing (QP). For instance, for an always vacant TV channel, when only $\epsilon^{\text{DS}} = 1/3$ portion of a *frame* is allocated for DS, QPSK modulation with coding rate 2/3 yields a service rate of 55.4kbit/s per subchannel while 64-QAM with coding rate 5/6 delivers 208 kbit/s per subchannel. The combination of available coding

and modulation schemes results in 14 different transmission modes at the PHY layer.

Using the above analysis, the service time on each server in Fig. 3.3 is modeled with an exponential RV of mean μ_i. In fact, the latter is a requirement for the Jackson theory. Modeling the service times with an exponential RV has strong supporting logc. Concretely, the equally-sized timely-bounded *frames* make the behavior of each CPE more like a $\mathrm{M/D/c}$ queue rather than a $M/M/c$ queue as denoted above. However, the performance achievable by $\mathrm{M/M/c}$ queue is a lower bound (worse performance) to that of $M/D/m$ due to variation in service time that packets experience in the former model. To improve the precision, we can modify the model by decreasing the mean service rate of each server with a factor of $\sqrt{2}$. This can be proven by equating the mean queue length in an $\mathrm{M/M/c}$ queue (with service rate μ as given in (3.1)) with that of an $M/D/m$ queue to find the service rate μ' of the latter, thus resulting in $\mu' = \mu/\sqrt{2}$ [52].

The difficulty related to this modeling task using classic queuing approaches lies in the fact that the number of servers assigned to a CPE could be changing every 160 msec (*superframe*) because that is how often the scheduler changes the allocation. Therefore, direct application of the Jackson theory to a network of IEEE 802.22 CPEs, as depicted in Fig. 3.3, is ill-formed. Next, we tackle this problem.

3.2.3 Modeling at the MAC Layer

In order to deal with the abovementioned analytical intricacy, we resort to the theory of queues in a random environment. In simple language, a queue is said to operate in a random environment if some of its related characteristics change during its operation period. This general definition is broad enough to entail diverse cases such as phase dependent input/output processes, where the input/output rates or the number of servers in a queue can vary. The modeling efforts by mathematicians in the last 30 years were mostly motivated by the practical importance of such queues in modeling real-life scenarios [19, 21, 22, 53–55].

Despite the applicability of findings to other fields, they have rarely been applied to the study of communications networks.

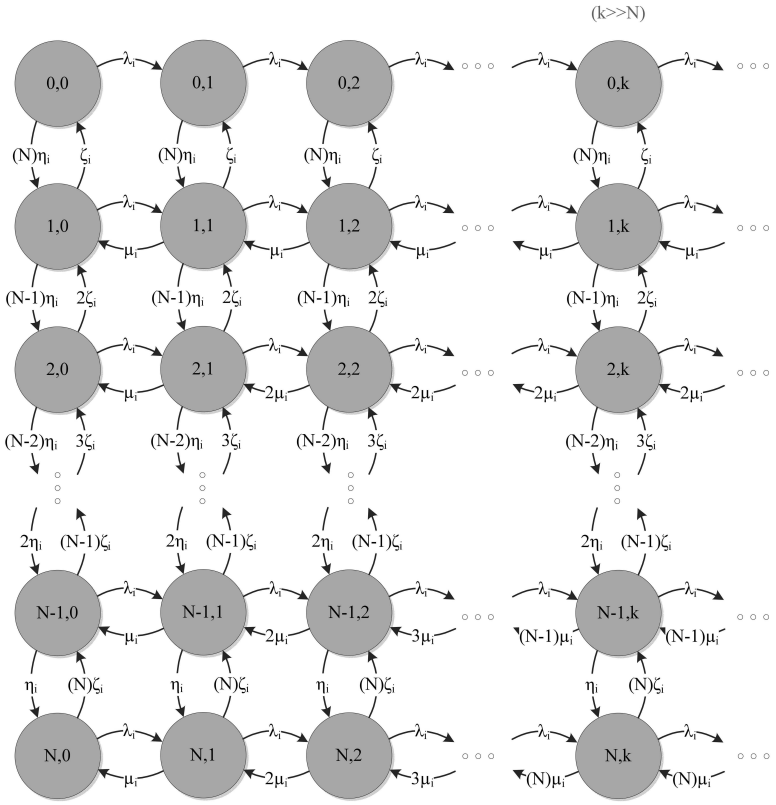


Figure 3.4: 2D CTMC representation for the i^{th} tagged CPE.

Let's assume there is only one CPE (tagged as i^{th}, hereafter) in the WRAN cell of Fig. 3.3. That being the case, one can model this CPE with the 2D CTMC $\{(E(t), X(t)) : t > 0\}$ as shown in Fig. 3.4, where $E(t)$ denotes the parent process and $X(t)$ denotes the queue process. This CTMC evolves in two directions: The horizontal transitions with rates λ_i and μ_i (of the queue process $X(t)$) stand for the arrival and departure rates of packets, respectively. The vertical transitions with rates η_i and ζ_i (of the parent process $E(t)$) represent the rate with which a channel is assigned to and revoked from the i^{th} CPE, respectively. Therefore, the states (n_i, ℓ) of this CTMC are samples of the state space $E * X$ whereby n_i represents the total number of allocated subchannels to CPE$_i$ and ℓ

represents the number of data packets stacked in its CPE$_i$'s queue. With this state definition, it should be clear why the CTMC in Fig. 3.4 is an equivalent representation for the protocol description in the previous section. For example, the announcement of radio spectrum resources in US/DS MAPs, which results in deallocation, reallocation of subchannels, is reflected into the model through the vertical transitions.

The explicit assumption in Fig. 3.4 is that vertical transitions are only possible to the neighbouring states of each state. This assumption turns out to be quite accurate due to the fact that the interval between two sequential US/DS MAP messages (10 msec) is much shorter than the average interval between the time a present PU becomes absent or vice versa (in the scale of a minute or more). Knowing that subchannels are shuffled by the scheduler before allocation to CPEs, it is extremely unlikely for two or more wireless subchannels to get evacuated/occupied by incumbent during a 10 msec allocation interval. Therefore, only the vertical transitions to neighboring states have significant rates (η_i and ζ_i).

The first row of the CTMC in Fig. 3.4 corresponds to the scenario where no subchannel is available. In this case, packets get piled up in the queue without any chance of departure. Rate λ_i in this model is the sum of exogenous arrival rate (fresh packets) and the circulating traffic rate (caused by retransmission). Departure rate $\mu_i = \sqrt{2}(62.4\left(1-\epsilon^{\mathrm{DS}}\right) r_{c,i} \log_2\left(\mathtt{M}_i\right) \cdot (1-\epsilon^{\mathrm{sen.}}))P_{\mathrm{S},i}$ derived in (3.1) accounts for the channel impairment and synchronization drift impacts. Having defined rates λ_i, μ_i, ζ_i, η_i, the SC for the CTMC in Fig. 3.4 is given by

$$\frac{\lambda_i}{\mu_i} < N\left(\frac{\eta_i}{\zeta_i+\eta_i}\right), \tag{3.2}$$

where $N = (1-P_{\mathrm{B}})\, n$, as explained at the beginning of Section 3.2.

Representing the probability of being in state (n_i, ℓ_i) with $\pi_i(n = n_i, l = \ell_i)$, the probability generating function (PGF) $\mathcal{G}_k(z)$ of each row of the CTMC in Fig. 3.4 would

be equal to

$$\mathcal{G}_{n_i}(z) = \sum_{\ell_i=0}^{\infty} z^{\ell_i} \pi_i(n = n_i, \ell = \ell_i). \tag{3.3}$$

In order to obtain the probability of having n_i servers (subchannels) allocated, i.e. $\pi_i(n_i)$, (3.3) is written for two successive rows of Fig. 3.4. Relating these latter and plugging $z = 1$ into the resulting equation gives rise to the following recursive formula:

$$\left.\begin{aligned} \mathcal{G}_{n_i+1}(1) &= \mathcal{G}_{n_i}(1) \left[\frac{(N - n_i)\,\eta_i}{(n_i + 1)\,\zeta_i}\right] \\ \mathcal{G}_{n_i}(1) &= \pi_i(n_i) = \sum_{\ell_i=0}^{\infty} \pi_i(n = n_i, \ell = \ell_i) \end{aligned}\right\}, \qquad n_i = 0, \ldots, N. \tag{3.4}$$

The equation above, when combined with the normalization condition $\sum_{n_i=0}^{N} \mathcal{G}_{n_i}(1) = 1$, yields the absolute value of $\pi_i(n_i)$, or $\mathcal{G}_{n_i}(1)$, as follows:

$$\mathcal{G}_{n_i}(1) = \pi_i(n_i) = \binom{N}{n_i} \left(\frac{\eta_i}{\eta_i + \zeta_i}\right)^{n_i} \left(1 - \frac{\eta_i}{\eta_i + \zeta_i}\right)^{N - n_i}. \tag{3.5}$$

Hence, the PMF of the i^{th} CPE having n_i vacant subchannels is $\text{Binomial}(N\eta_i/(\eta_i + \zeta_i))$ with $\eta_i/(\eta_i + \zeta_i)$ (already represented as $1 - P_{\text{B}}$) being the probability that a specific subchannel is vacant.

The other intriguing observation in (3.5) is related to its independence from the exogenous arrival rate λ_i and service rate μ_i seemingly decomposing vertical and horizontal progressions. With a precise look at (3.5), we can find out that this equation is the steady state probability distribution for one column of the CTMC in Fig. 3.4 when it is detached from its left and right hand-side states, as shown in Fig. 3.5.

This interesting result on the decomposition of 2D CTMC is due to the ergodicity and will later be used for the derivation of closed-form expressions for performance metrics of

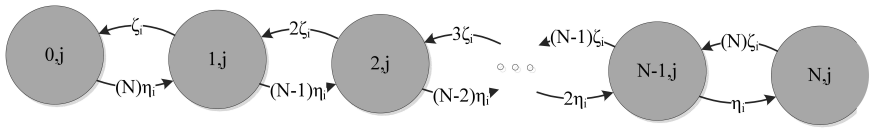


Figure 3.5: The marginal CTMC for the number of subchannels.

interest.

In order to extend the result in (3.5) to the general case with m CPEs (as illustrated in Fig. 3.3), the total number of N subchannels should be appropriately dispensed among m CPEs rather than to a single one. According to the WRAN standard, this allocation should be a comprehensive and non-overlapping partitioning of subchannels in such a way that the CPE closer to a given BS (or with better transmission history) gets less OFDM subcarriers than the farther ones (for fairness). Inspired by (3.5), and assuming that N subchannels are to be allocated to m associated CPEs with distributing probabilities $q_1, \ldots, q_m$, the probability that the partitioning $n_1, \ldots, n_m$ is chosen among all possibilities would be equal to:

$$\Pr(X_1 = n_1, \ldots, X_m = n_m) = \begin{cases} \dfrac{n!}{n_1! n_2! \ldots n_m!} q_1^{n_1} q_2^{n_2} \ldots q_m^{n_m}, & \sum_{i=1}^{m} n_m = N \\ 0, & \text{otherwise} \end{cases} \tag{3.6}$$

where $\sum_{i=1}^{m} q_i = 1$. This is a multinomial distribution. The interesting property of such distribution is that its i^{th} marginal distribution (the number of servers allocated to the i^{th} CPE) would be binomially distributed: $\text{Binomial}(n_i, q_i)$, $i = 1, \ldots, m$, agreeing with (3.5). Also, this property is valid regardless of the fact that $n_1, \ldots, n_m$ are dependent variables. Therefore, (3.5) and the CTMC in Fig. 3.4 are still applicable and each CPE in a network of CPEs (Fig. 3.4) can be modeled with a separate 2D CTMC similar to that of Fig. 3.4. Consequently, $m+1$ 2D CTMC will appropriately model the network of Fig. 3.3. Now, since the vertical transitions in Fig. 3.4 might change at the end of every *frame*, the

suitable choice for η_i and ζ_i such that $\eta_i/(\eta_i+\zeta_i)=q_i$ gives the following results:

$$\eta_i=\frac{q_i}{T_{\rm F}} \quad \text{and} \quad \zeta_i=\frac{1-q_i}{T_{\rm F}} \qquad ,i=1,\ldots,m, \tag{3.7}$$

where $T_{\rm F}$ is the *frame* length of 10 msec. The determination of the subcarrier distribution probabilities among CPEs, i.e., q_i's, will be done later.

3.3 Closed-Form Probability Density Function

In the previous section, we were able to obtain the PMF of the number of subchannels (servers), namely $E(t)$, that a CPE receives from the serving BS. This PMF was shown to be of Binomial type in an IEEE 802.22 cell (cf. (3.5)). It remains to find the PMF of the 2D process $\{(E(t),X(t))\colon t>0\}$and, particularly, the marginal process $X(t)$, which represents the total number of packets (in kbits as denoted in (3.1)) in the queue of a CPE.

A desirable closed-form expression cannot be obtained through writing the balance equations for the CTMC in Fig. 3.4, as reported in [54, 55]. Moreover, as we will see later, the numerical complexity for solving this CTMC is prohibitively huge for a large number of subchannels. In fact, any inference problem is NP-hard [56]. Lastly, such a numerical PMF, even though insightful, does not provide us with a handy tool for the performance evaluation of the network. Here, we are motivated to find a closed-form expression for the aforementioned PMFs using some recent mathematical methods. For the most part, recent advances in the field of queuing networks proved that product-form expressions exist for multidimensional CTMCs when certain requirements are met. We will discuss this matter next.

Assume that $\{(E(t),X(t)):t>0\}$ is a 2D irreducible and positive-recurrent Markov chain with state space $E*X$. Further, let the marginal environmental process $\{E(t)\}$ be ergodic with transition matrix $G_E=(g_E(n,n'):n,n'\in E)$. Also, in between any

two environmental transmissions (change of n to n'), the queuing process $\{X(t)\}$ has the transition matrix $G_{X|E} = (g_{X|E}(\ell, \ell'|e) : \ell, \ell' \in X)$. Therefore, the transition rates for the 2D CTMC are defined by

$$g\left((n,\ell),(n',\ell')\right) = \begin{cases} g_{X|E}(\ell,\ell'|n), & \text{if } n = n',\ \ell \neq \ell' \\ g_E\left(n,n'\right), & \text{if } n \neq n',\ \ell = \ell' \end{cases} \tag{3.8}$$

Representing $\overline{\Pi} = (\pi(n,\ell)\,, n \in E,\ \ell \in X)$ as the steady state solution for the 2D CTMC and $\overline{\Pi}_\ell = (\pi_X(\ell)\,,\ \ell \in X)$, $\overline{\Pi}_n = (\pi_E(n)\,,\ n \in E)$ as the marginal distributions, then the theorem below establishes a straightforward relationship between these PMFs.

Theorem 1. *If the process $\{X(t)\}$ is an ergodic Markov chain and has stationary distribution, then the* 2D *CTMC as defined above can be decomposed into two independent 1-D Markov chains, and a product-form solution for $\pi(n,\ell)$ exists such that:*

$$\pi\left(n,\ell\right) = \pi_E\left(n\right)\pi_{X|E}\left(\ell|n\right) \quad , \quad n \in E,\ \ell \in X. \tag{3.9}$$

Proof. The proof is given in [57]. □

After comparing the requirements of this theorem with our modeling facts, we find the theorem applicable to the problem at hand where the queue process $\pi_{X|E}\left(\ell|n\right)$ represents the PMF of the number of packets in an $M/M/n_i$ queue and the environmental process $\pi_E\left(n\right)$ represents the PMF of the number of servers n_i. Therefore, using this theorem and plugging (3.5) and (3.7) into (3.9), we obtain the secondary's joint PMF of the queue length and available channels as follows:

$$\begin{aligned} &\pi_i(n = n_i, \mathtt{L} = \ell_i) = \tbinom{N}{n_i}\left(q_i\right)^{n_i}\left(1-q_i\right)^{N-n_i}\pi_{X|E}\left(\mathtt{L} = \ell_i|n = n_i\right), \qquad i = 1,\ldots,m, \\ &\pi_{\mathrm{BS}}^{\mathrm{DS}}(n = N, \mathtt{L} = \ell_i) = \pi_{X|E}\left(\mathtt{L} = \ell_i|n = N, \rho = \rho_{\mathrm{BS}}^{\mathrm{DS}}\right), \\ &\pi_{\mathrm{BS}}^{\mathrm{US}}(n = 1, \mathtt{L} = \ell_i) = \pi_{X|E}\left(\mathtt{L} = \ell_i|n = 1, \rho = \rho_{\mathrm{BS}}^{\mathrm{US}}\right), \end{aligned}$$

where

$$\pi_{X|E}\left(\mathtt{L}=\ell_i|n=n_i\right)=\begin{cases}\dfrac{P_{0,i}\left(\rho_i\right)^{\ell_i}}{\ell_i!}, & \ell_i<n_i\\ \dfrac{P_{0,i}\left(\rho_i\right)^{\ell_i}}{n_i!n_i^{\ell_i-n_i}}, & \ell_i\geq n_i\end{cases}$$
$$\pi_i(\mathtt{L}=0|n=n_i)=P_{0,i}=\left(\sum_{j=0}^{n_i-1}\frac{\left(\rho_i\right)^j}{j!}+\frac{\left(\rho_i\right)^{n_i}}{n_i!}\frac{1}{1-\dfrac{\rho_i}{n_i}}\right)^{-1}, \tag{3.10}$$

is the well-known queue length PMF of $\mathrm{M/M/c}$ queue. Here, for the US direction (Fig. 3.3a) $\rho_i=\lambda/\mu_i$ in each CPE and $\Lambda_{\mathrm{BS}}^{\mathrm{US}}=\sum_{i=0}^{m}\lambda_i$ (hence $\rho_{\mathrm{BS}}^{\mathrm{US}}=\Lambda_{\mathrm{BS}}^{\mathrm{US}}/\mu_{\mathrm{BB}}$), whereas for the DS direction (Fig. 3.3b) $\Lambda_{\mathrm{BS}}^{\mathrm{DS}}=\Lambda$ (hence, $\rho_{\mathrm{BS}}^{\mathrm{DS}}=\Lambda_{\mathrm{BS}}^{\mathrm{DS}}/\mu_{\mathrm{BS}}$). Here, we recall that the arrival rates $\Lambda_{\mathrm{BS}}^{\mathrm{US}}$ and $\Lambda_{\mathrm{BS}}^{\mathrm{US}}$ are obtained after writing the traffic equations [51], and that μ_{BB} represents the BS service rate in the US direction, which is related to the rate at which data are transferred from the BS to the backbone of the network. For situations where IEEE 802.22 cells are connected by fibers or wires, μ_{BB} is very large ($\rho_{\mathrm{BS}}^{\mathrm{US}}\approx 0$) and, hence, the queuing delay caused by BS in the US direction (Fig. 3.3a) is negligible. Later on, we will exploit this simple product-form solution in (3.10) to obtain the m-dimensional joint PMF for the number of packets in the whole network. Then from this PMF, any statistic can be derived.

Another essential measure is the PMF for the marginal process $\{X(t)\}$. This PMF is obtained by summing the $\pi\left(n=n_i,\mathtt{L}=\ell_i\right)$ in (3.10) over the first variable, i.e.,

$$\pi_X\left(\mathtt{L}=\ell_i\right)=\sum_{n_i=0}^{N}\pi\left(n=n_i,\mathtt{L}=\ell_i\right), \tag{3.11}$$

which yields (3.12)

$$\pi_X\left(\mathtt{L}=\ell_i\right)=$$
$$\begin{cases} \rho_i^{\ell_i}\left(\sum\limits_{n_i=\rho_i}^{\ell_i} P_{0,i}\dfrac{\pi_i(n_i)}{n_i!n_i^{\ell_i-n_i}}+\sum\limits_{n_i=\ell_i+1}^{N} P_{0,i}\dfrac{\pi_i(n_i)}{\ell_i!}\right)+\mathfrak{K}_i\cdot\left(\pi_i(n_i=0)+\pi_i(n_i=1)\right), & \ell_i<N, \\ \rho_i^{\ell_i}\sum\limits_{n_i=\rho_i}^{N} P_{0,i}\dfrac{\pi_i(n_i)}{n_i!n_i^{\ell_i-n_i}}+\mathfrak{K}_i\cdot\left(\pi_i(n_i=0)+\pi_i(n_i=1)\right), & \ell_i\geq N, \end{cases} \tag{3.12}$$

knowing that

$$\pi_i(n_i)=\binom{N}{n_i}\left(q_i\right)^{n_i}\left(1-q_i\right)^{N-n_i},$$

$$\mathfrak{K}_i=(1-\rho_i^{'})(\rho_i^{'})^{\ell_i},$$

$$\rho_i^{'}=\frac{\rho_i+\lambda_i q_i^2\left(1-q_i\right)}{Nq_i+\lambda_i q_i^2\left(1-q_i\right)}.$$

Here, $\pi_i(n_i=0)$ and $\pi_i(n_i=1)$ in (3.12) correspond to the case when zero or one server is allocated to the i^{th} CPE, respectively. We remark that the factor $\mathfrak{K}_i$ in (3.12) is added in order to enhance the precision of the closed-form based on the following logic: since $\pi_i\left(\ell_i|n_i=0\right)$ (zero server case) in (3.10) results in an unstable queue for which no stationary PMF exists, the cases with one and zero subchannels ($n=0,1$) are extracted out from the summation in (3.11) and are jointly represented by an $\mathrm{M/M/1}$ queue subject to *breakdown*. Adopting the approach in [58], where an expression (Eq. 24) for the mean number of packets in an $\mathrm{M/G/1}$ queue subject to server breakdown was derived, the mean number of packets in an $\mathrm{M/M/1}$ queue subject to server breakdown is obtained as:

$$\overline{\mathtt{L}}_i^{(n_i=0,1)}=\frac{1}{q_i-\rho_i}\left(\rho_i+\lambda_i q_i^2\left(1-q_i\right)\right). \tag{3.13}$$

It should be noted that the SC for such queue is $\rho_i < q_i$ (note that $q_i < 1$) which is more restricted compared to the SC $\rho_i < 1$ for $\mathrm{M}/\mathrm{M}/1$ regular queue. Knowing that the PMF of the number of packets in an $\mathrm{M}/\mathrm{M}/1$ queue is geometrically distributed, the PMF of $\mathrm{M}/\mathrm{M}/1$ queue with breakdown is approximated with the geometric distribution $(1-\rho_i')\,\rho_i'^{\ell_i}$ whose utilization factor ρ_i' is adjusted to account for the breakdown effect. This ρ_i' can be obtained by equating the mean number of packets in $\mathrm{M}/\mathrm{M}/1$ queue (that is $\rho_i'/(1-\rho_i')$) to (3.13). This operation yields

$$\rho_i^{'} = \frac{\rho_i + \lambda_i q_i^2 (1-q_i)}{q_i + \lambda_i q_i^2 (1-q_i)}, \tag{3.14}$$

giving rise to the result in (3.12).

Fig. 3.6 shows the accuracy of the closed-form queue length distribution in (3.12) which is compared against the exact result obtained by numerically solving the CTMC in Fig. 3.4 for $N = 15$ subchannels. The reason for not choosing the nominal value $N = 60$ is the computational complexity of solving the CTMC for such high volume of N. Note that for the CTMC in Fig. 3.4 this complexity increases exponentially as N grows. This is because the inference problem of such kind are well-known to be NP-hard, in nature. Nevertheless, Fig. 3.6 substantiates that the precision of (3.12) increases for larger number of subchannels.[9] The ordinate scale in Fig. 3.6 is chosen logarithmically to magnify the tiny differences. In should be clarified that the differences noted when the number of packets increases are due to the approximation we had to make for the zero-server situation as explained above.

[9]For example, in another setting where $N = 20$ and $P_{\mathrm{B}} = 0.2$, the comparison for $\ell \in [0, 500]$ (12 times wider range compared to Fig. 3.6) illustrated that for a queue length of 450 packets the difference between the exact and approximate results was as low as 8%. For $N > 20$, the numerical computations involved with the derivation of queue length PDF could not be contained by the PC, resulted in frequent crashes in the middle of computations. This computational intensiveness, justifies the motive behind the obtainment of a closed-form expression for the statistics of CPE queue length.

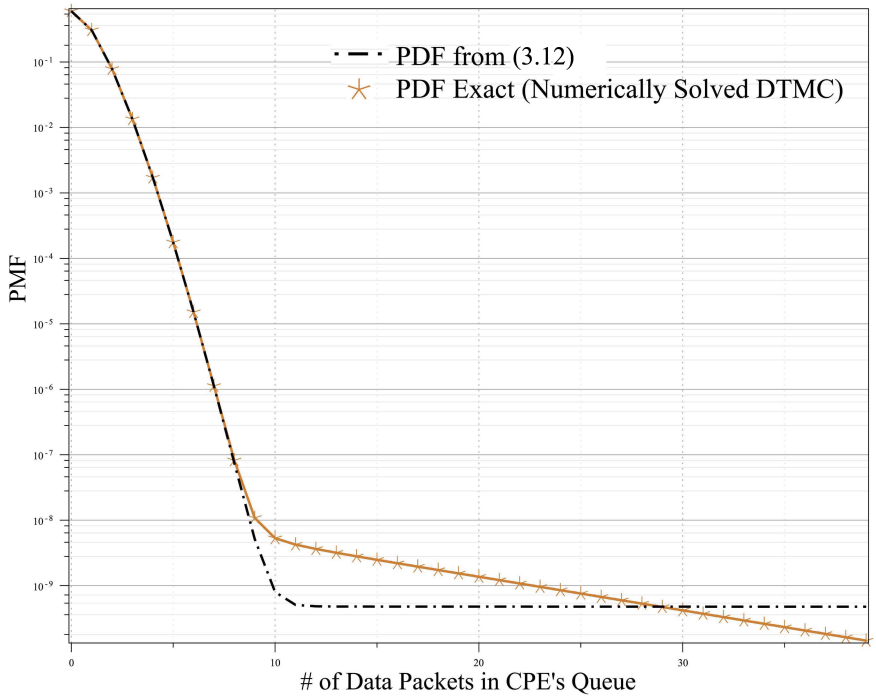


Figure 3.6: A demonstration for the accuracy of the PMF obtained in (3.12) compared to that obtained by numerically solving (exact) the CTMC.

3.3.1 Network-Wide PMF

The steady state solution for the network of Fig. 3.3 with $\ell_1, \ldots, \ell_m$ packets in CPEs $1, \ldots, m$ and with $\ell_{\text{BS}}^{\text{US}}$ and $\ell_{\text{BS}}^{\text{DS}}$ packets in BS (respectively for the US and DS directions) exists if, according to (3.2), $\forall i \in \{1, \cdots, m\}$we have (i) $\rho_i < Nq_i$, (ii) $\rho_{\text{BS}}^{\text{US}} < 1$, (iii) and $\rho_{\text{BS}}^{\text{DS}} < N$. If these conditions are satisfied, then the network-wide solution is obtained using the theory of Jackson networks as follows:

$$\begin{aligned} &\pi\left(\mathrm{L}_1 = \ell_1, \ldots, \mathrm{L}_m = \ell_m, \mathrm{L}_{\text{BS}}^{\text{US}} = \ell_{m+1}^{\text{US}}, \mathrm{L}_{\text{BS}}^{\text{DS}} = \ell_{m+1}^{\text{DS}}\right) = \\ &\pi_{\text{BS}}\left(\mathrm{L}_{\text{BS}}^{\text{US}} = \ell_{m+1}^{\text{US}}\right) \pi_{\text{BS}}\left(\mathrm{L}_{\text{BS}}^{\text{DS}} = \ell_{m+1}^{\text{DS}}\right) \prod_{i=1}^{m} \pi_{\ell_i}\left(\mathrm{L}_i = \ell_i\right), \end{aligned} \tag{3.15}$$

where $\pi_{\mathrm{BS}}\left(\mathtt{L}_{\mathrm{BS}}^{\mathrm{US}} = \ell_{m+1}^{\mathrm{US}}\right)$ and $\pi_{\mathrm{BS}}\left(\mathtt{L}_{\mathrm{BS}}^{\mathrm{DS}} = \ell_{m+1}^{\mathrm{DS}}\right)$ are given by (3.10).

3.3.2 Numerical Results

Using the closed-form expressions for the PMF of the CPE's queue length, the response of the network to the changing parameters such as modulation/coding, allocation probability q, channel busyness probability P_{B}, etc. can be easily observed, which was costly to do by numerically solving the CTMC of Fig. 3.4. Fig. 3.7 shows plots of PMF of queue length (on a logarithmic scale) for four different modulations and two different probabilities that channel being estimated as busy ($P_{\mathrm{B}} = 0.1$ in *black*, $P_{\mathrm{B}} = 0.4$ in *red*). The network parameters used in this setting are $r_c = 5/6,\ q = 0.3,\ t = 1/3,\ \lambda = 15\,\mathrm{pk/s}$. As expected, a active incumbent service results in heavily occupied CPE queues which becomes less piled-up as higher-order modulations are employed. However, as it will be discussed shortly, the data-rate/robustness trade-off that is inherent to any modulation technique does not last longer. For example, in this situation, QAM constellations with levels beyond 128 sacrifice both resiliency and data-rate at no advantage.

Fig. 3.8 depicts the PMF of queue length $\mathtt{L}$ when the OFDM subchannel allocation probability q varies (for fixed P_{B}) and when P_{B} varies (for fixed q). Once more, the logarithmic scale was chosen in order to magnify the differences. As observed, a larger OFDMA subchannel allocation probability shifts the benefit towards having less piled-up queues. The same trend is observed for varying q.

3.3.3 How to Choose q_i?

In this part, we make several suggestions on specifying the unknown quantities $P_{\mathrm{S},i}$ and q_i for $i = 1, \ldots, m$. Parameter q_i, which vary from CPE to CPE, shows how the IEEE 802.22 cell differentiates between CPEs in allocating the resources. For more detailed discussion about optimal allocation of resources to nodes, readers are referred to Chapter

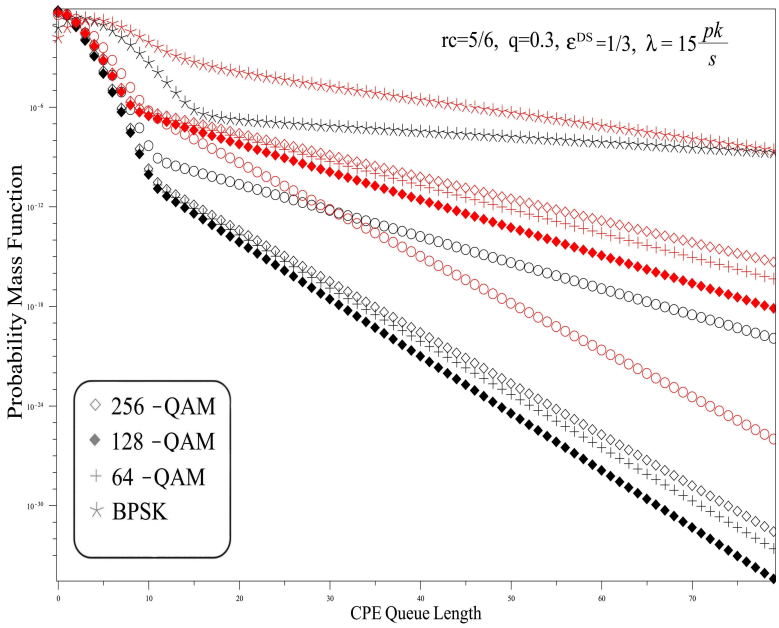


Figure 3.7: PMF of the number of packets in a CPE for different modulation schemes. The black (red) curves are related to scenarios with less (more) active incumbent BSs, that is, $P_{\mathrm{B}} = 0.1$ ($P_{\mathrm{B}} = 0.4$).

6. In reality, such differentiation is made by proportionately relating q_i with the successful transmission rate $P_{\mathrm{S},i} = 1 - \mathrm{SER}_i$ as SER_i entails the impact of all impairments such as frequency drift, mis-synchronization, channel effects, etc. For instance, in the case of square M-QAM constellation under an AWGN propagation channel, which is the recommended carrier modulation in IEEE 802.22, we have

$$P_{\mathrm{S},i} = \left(1 - \left(2\left(1 - \frac{1}{\sqrt{\mathtt{M}_i}}\right)\mathbb{Q}\left(\sqrt{\frac{3\log_2(\mathtt{M}_i)}{\mathtt{M}_i - 1}\frac{E_{b,i}}{\mathtt{N}_0}}\right)\right)\right)^2, \tag{3.16}$$

for $i = 1, \ldots, m$. For Rayleigh fading channel, the formula for the special case of square QAM constellation over single-input single-output Nakagami channels with shape parameter unity [59] can be directly applied to (3.1) to calculate μ_i and μ_{BS}. Here, $E_{b,i}/\mathtt{N}_0$ represents the bit energy per noise, $\mathtt{M}_i$ is the number of symbols in an M-QAM constellation and

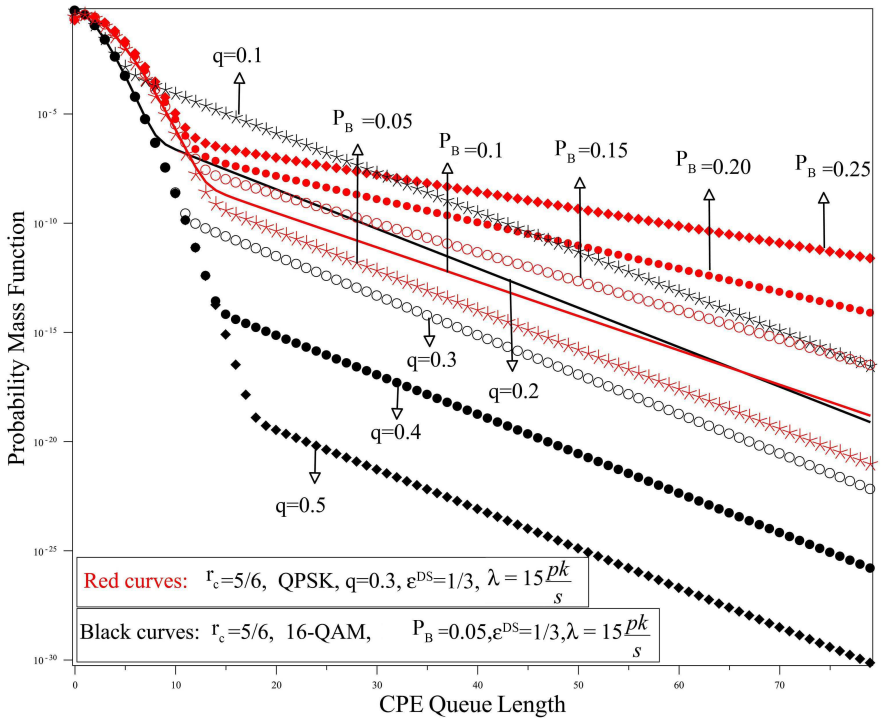


Figure 3.8: PMF of the number of packets in a CPE when q and P_{B} vary.

$\mathbb{Q}(x)$ is the Q-function. Several approaches are possible:

- **Detection-based Linear Weighting:** Since the BS receives the local sensing results from CPEs, it can estimate their detection probabilities $P_{\mathrm{D},i}$ according to their past histories and their latest reported results. With that information, the BS allocates more OFDM subchannels to the CPE that has higher $P_{\mathrm{D},i}$ (thus is more reliable in terms of sensing outcome) according to

$$q_i = \frac{P_{\mathrm{D},i}}{\sum_{j=1}^{m} P_{\mathrm{D},j}}. \tag{3.17}$$

- **Outage-based Linear Weighting:** As in the previous method, the linear weighting approach is taken. However, in this case, more allocation is given to the CPE that

is closer to the BS and, as such, produces less interference to the primary receiver. By definition, the outage probability is given by $P_{\mathrm{Out},i} = \Pr\left(\mathrm{SNR}_i < \mathrm{SNR}^{\min}\right)$. Therefore,

$$q_i = \frac{1 - P_{\mathrm{Out},i}}{\sum_{j=1}^{m} (1 - P_{\mathrm{Out},j})}. \tag{3.18}$$

- SER**-based Linear Weighting:** In this case, more allocation is given to the CEP with the most successful transmission history. Hence,

$$q_i = \frac{1 - P_{\mathrm{S},i}}{\sum_{j=1}^{m} (1 - P_{\mathrm{S},i})}. \tag{3.19}$$

3.4 Cell-Level Analysis

The results obtained till now empower us to conduct a “cell-level analysis" and analyze cell SC and access delay (referred to as peer-to-peer delay).

3.4.1 Cell Stability Condition (SC)

Having found the PMFs of the numbers of packets in the $m+1$ nodes (m CPEs and one BS) of an IEEE 802.22 cell (cf. Section 3.3), the SC for the non-saturated operation of the cell is given by

$$\begin{aligned} \rho_i = \frac{\lambda_i}{\mu_i} < N q_i, \qquad & i = 1, \ldots, m, \\ \rho_{\mathrm{BS}}^{\mathrm{US}} = \frac{\Lambda_{\mathrm{BS}}^{\mathrm{US}} = \sum_{i=1}^{m} \lambda_i}{\mu_{\mathrm{BB}}} < 1, \quad & \rho_{\mathrm{BS}}^{\mathrm{DS}} = \frac{\Lambda_{\mathrm{BS}}^{\mathrm{DS}} = \Lambda}{\mu_{\mathrm{BS}}} < N, \end{aligned} \tag{3.20}$$

where μ_{BS} and μ_i as obtained from (3.1). These conditions, if met simultaneously, should guarantee the seamless functioning of the network as well as existence of steady state joint

PMF of $\mathtt{L}_1, \cdots, \mathtt{L}_m$. However, as long as the stable operation of a single CPE (say the j^{th}) on the US is concerned, the satisfaction of $\rho_i < Nq_j$ and $\rho_{\text{BS}}^{\text{US}} < N$ would be sufficient for the existence of the steady state solution. A necessary condition that obviates reliance on q_i $(i = 1, \ldots, m)$ is obtained by summing up $m + 1$ equations in (3.20) yielding

$$\rho_{\text{BS}}^{\text{DS}} + \rho_{\text{BS}}^{\text{US}} + \sum_{i=1}^{m} \rho_i < 2N + 1. \tag{3.21}$$

This insightful equation indicates that for stable operation, the sum of utilization factors should be smaller than twice the number of subchannels. The first consequence would be on the choice of parameters such as coding rate, modulation level, etc., which cannot take arbitrary values. In addition, the selection of probabilities q_i should be done precisely not to violate (3.20) while keeping the necessary differentiation between CPEs.

Table 3.2 shows the restriction on the input packet rate to a CPE for different quantities. The first column is derived for varying symbol length, $M = 2, 4, 16, 64, 128, 256, 512$. In the second column, q varies between 0.05 and 0.8, and in the third column the coding rate r_c takes value in $[1/2,\ 2/3,\ 3/4,\ 5/6]$. In this table, the trend in the first column is noteworthy since it proves that the exploitation of modulation with larger constellation is not always advantageous. Finally, it should be noted that the conditions in (3.20) might not guarantee the QoS which requires tighter restrictions on $\rho_i,\ i = 1, \ldots, m$.

3.4.2 Peer-to-Peer Delay (D_{PtP})

One metric of interest that reflects on the performance of CRNs is the round-trip delay that a CPE experiences. This quantity, termed peer-to-peer delay, is the total time elapsed from the instant a packet enters a CPE's queue for transmission to a destination until a transport acknowledgement (sent by the destination) is received by that CPE to imply the correct reception of the packet. The significance of this metric is in the fact that not all services

Table 3.2: Varying limits on a secondary's input rate for different network parameters.

$q = 0.1,\ r_c = 5/6$	$M = 16,\ r_c = 5/6$	$M = 16,\ q = 0.1$
$\lambda < 29.80$	$\lambda < 58.20$	$\lambda < 69.8$
$\lambda < 58.20$	$\lambda < 174.5$	$\lambda < 93.0$
$\lambda < 116.3$	$\lambda < 290.8$	$\lambda < 104.7$
$\lambda < 169.4$	$\lambda < 470.1$	$\lambda < 116.3$
$\lambda < 178.8$	$\lambda < 523.5$	
$\lambda < 164.4$	$\lambda < 639.8$	
$\lambda < 131.7$	$\lambda < 756.1$	
	$\lambda < 872.5$	

are delivered unilaterally (from BS to CPEs). Instead, many applications are interactive requiring the flow of data on both directions. Hence, an end-to-end reliable operation of the network in both directions is essential. Given this definition, the peer-to-peer delay D_{PtP} is given by

$$D_{\mathrm{PtP},i} = D_i + D_{\mathrm{BS}}^{\mathrm{US}} + D_{\mathrm{BB}} + D_{\mathrm{BS}}^{\mathrm{DS}}, \tag{3.22}$$

where D_i and $D_{\mathrm{BS}}^{\mathrm{US}}$ are the total delays[10] of the CPE$_i$ and BS in the US direction, and $D_{\mathrm{BS}}^{\mathrm{DS}}$ is the BS total delay in the DS direction. Finally, D_{BB} represents the amount of time a packet spend in the core of the network.

Using the *Little theorem*, which links the average queue length $\left(\overline{\mathrm{L}}_i, \overline{\mathrm{L}}_{m+1}^{\mathrm{US}}, \overline{\mathrm{L}}_{m+1}^{\mathrm{DS}}\right)$ to the queuing delay, we obtain

$$D_{\mathrm{PtP},i} = \frac{\overline{\mathrm{L}}_i}{\lambda_i} + \frac{\overline{\mathrm{L}}_{\mathrm{BS}}^{\mathrm{US}}}{\sum_{i=1}^{m} \lambda_i} + D_{\mathrm{BB}} + \frac{\overline{\mathrm{L}}_{\mathrm{BS}}^{\mathrm{DS}}}{\Lambda_{\mathrm{BS}}^{\mathrm{DS}}}, \tag{3.23}$$

where $\overline{\mathrm{L}}_{\mathrm{BS}}^{\mathrm{US}}$ and $\overline{\mathrm{L}}_{\mathrm{BS}}^{\mathrm{DS}}$ are obtained from (3.24) when $(n = 1,\ \rho = \rho_{\mathrm{BS}}^{\mathrm{US}})$ and $(n = N,\ \rho =$

[10] Total delay is a.k.a. sojourn time or system time.

$\rho_{\mathrm{BS}}^{\mathrm{DS}}$), respectively,

$$\begin{aligned}\overline{\mathtt{L}}_{\mathrm{BS}}^{\mathrm{US}} &= \rho_{\mathrm{BS}}^{\mathrm{US}} + \frac{{\rho_{\mathrm{BS}}^{\mathrm{US}}}^{n+1}}{n!}\frac{P_0}{\left(1-\frac{\rho_{\mathrm{BS}}^{\mathrm{US}}}{n}\right)^2},\\ \overline{\mathtt{L}}_{\mathrm{BS}}^{\mathrm{DS}} &= \rho_{\mathrm{BS}}^{\mathrm{DS}} + \frac{{\rho_{\mathrm{BS}}^{\mathrm{DS}}}^{n+1}}{n!}\frac{P_0}{\left(1-\frac{\rho_{\mathrm{BS}}^{\mathrm{DS}}}{n}\right)^2},\end{aligned} \tag{3.24}$$

with P_0 as given in (3.10). Equation (3.24) is the popular expression for the average number of packets in an $\mathrm{M/M/n}$ queue since all the OFDM subchannels (fixed number) are always available to the BS. To obtain $\overline{\mathtt{L}}_i$, we use (3.25) since the number of subchannels available to CPE$_i$ is random. Thus,

$$\overline{\mathtt{L}}_i = \sum_{\ell=0}^{\infty} \ell \pi_i \left(\mathtt{L}_i = \ell\right), \tag{3.25}$$

where $\pi_i \left(\mathtt{L}_i = \ell\right)$ is given in (3.12). Finally, the value of D_{BB} will be assumed zero as the swiftness of equipment (servers, switches, etc.) in the backbone of the network is irrelevant to the analysis of the cognitive network under study.

The numerical results demonstrate the trend reversal in D_{PtP} (decrease and then increase) as the modulation degree increases like that observed in Table 3.2. Therefore, beyond a point, the downside of the low resistibility of the higher-order QAM scheme dominates the benefit obtained by higher data rate they offer rendering this choice unsuitable.

The question that how to partition 26 OFDM symbols in a frame between DS and US links (that is, factor ϵ^{DS} in (3.1)) can be answered at this point. Concretely, the results show that the fair splitting of *frame* symbols between US and DS of a CPE yields the least D_{PtP}. For instance, for different choices $t = 1/4\,, 1/3,\ 1/2,\ 2/3,\ 3/4$, the values $D_{\mathrm{PtP}} = 0.144,\ 0.11, 0.093, 0.13, 0.53$ were observed and, regardless of the choice of network parameters, the trend of D_{PtP} is always decreasing and then increasing. This

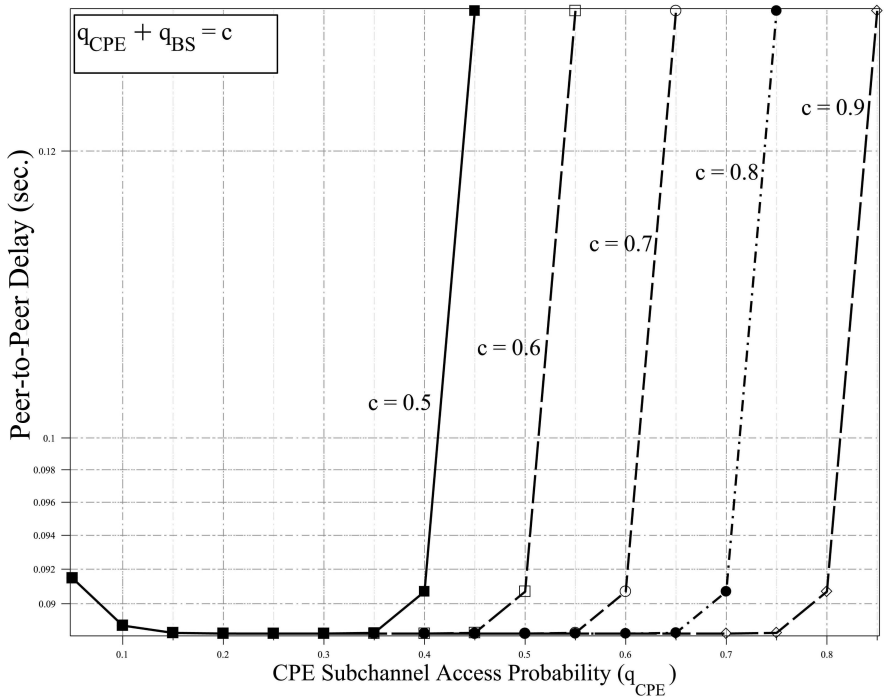


Figure 3.9: Peer-to-peer delay vs. CPE subchannel access probability q_{CPE}. Each curve represents D_{PtP} for fixed resource sum $q_{\mathrm{BS}} + q_{\mathrm{CPE}} = c$, $c = 0.5, \cdots, 0.9$.

demonstrates that $1/3 < t < 2/3$ is the best choice and confirms that fair split gives the least D_{PtP}.

In a separate experiment, we assessed the situation when the BS allocates a fixed amount of resource to the path rather than a CPE. In other words, the total two-way allocated resource in DS and US remains constant as $q_{\mathrm{BS}} + q_{\mathrm{CPE}} = c < 2$. By varying the transmission rate on the DS, the rate on the US link changes reversely and vice versa. The reason for selecting this special setting is related to its implementation practicality. As seen in Fig. 3.9, a larger two-way allocated resource yields smaller delay. The interesting result here is that even for a constant c, depending on how the resource is split between the DS and US, the performance would be different. Obvious from Fig. 3.9, the non-radical split (with $q_{\mathrm{CPE}}/q_{\mathrm{BS}}$ not so large or so small) causes the least D_{PtP} to be experienced. This figure is plotted for the symbol split ratio $t = 1/3$ which gives the larger weight to the US

link than to the DS one. For $t = 2/3$, where the *frame* length is more dedicated to the DS, the same trend as in Fig. 3.9 is observed but with the left wings being higher in value compared to the right wings.

This being said, once the BS allocates a share of resource, it is always better (delay-wise) to fairly (not necessarily evenly) split this amount of resource between the DS and US directions not to cause a bottleneck in the path.

3.5 Summary

In this chapter, WRAN was considered as per the specifications of the IEEE 802.22 and a novel and simple analytical model was presented. The proposed model is based on a tractable node-based two-dimensional continuous time discrete state Markov chain (2-D CTMC) but with the advantage that it takes features of the PHY and MAC layers at the same time. Using the proposed model, the network performance was evaluated by finding the numerical and closed-form expressions for the marginal distribution of the CPE's queue length as well as the joint distribution of all the data packets in a WRAN cell. As a guideline to future investigations one can think of solving the generalized CTMC with non-neighboring transitions. Also, relaxing the Poisson assumption for input traffic and using more complicated models (such as fractional Brownian motion-fBm, which takes fractality and long-range dependency characteristics of input traffic into account) can bring about more precision to the model and adapt it better to practical scenarios. There is also a lot to say and investigate on the sharing of resources among CPEs in a WRAN cell in order to obtain the best mutual benefit for all CPEs. For instance, the derivations for the peer-to-peer delay of a CPE showed that the fair manner of sharing the available resources between the upstream and downstream achieves the best results. The investigation of this matter is of extreme importance to have higher spectral efficiency in the network.

Chapter 4

Multi-Interface CRNs: MAC Design and Modeling[1]

In the previous chapter, modeling frameworks were proposed for the performance analysis of single-interface multi-channel CRNs where payload could only be transmitted over one radio interface at a time. In this chapter, we aim for a different scenario where each node is equipped with multiple radios (interfaces).

CR technology has undergone a fair progress in different aspects, e.g., signal sensing, transmission techniques, antenna design, software defined radio (SDR) design, etc. In the latter two domains, efforts have reached a point where it is now possible to implement multiple radios on a generic-purpose hardware using SDR technology and attach multiple antennas (with reasonable sizes) such that each antenna is able to radiate at different frequencies concurrently with others and radio parameters can be adaptively tuned. This software implementation of radios provides a great deal of flexibility and is seen as an important spurring factor for the widespread adoption of CR networks in 5G networks. Termed multi-interface multi-channel cognitive radio networks (MIMC-CRN) in this dis-

[1] N. Tadayon and S. Aïssa, "Modeling and analysis framework for multi-interface multi-channel cognitive radio networks", *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 935–947, Oct. 2014.

sertation, each node in MIMC-CRN can opportunistically access the spectrum and transmit over multiple channels in a simultaneous manner, which allows much higher data rates and better seamless communications while decreasing the possibility of interference to licensed users. Due to the tight interdependence between MAC and PHY layers in CRNs that is spawned by CR-specific functionalities such as sensing, probing, etc., any change in one layer directly propagates to the other. For that reason, the impact of having multiple interfaces in CRNs should be properly reflected in the MAC layer. This can be done either by tweaking the protocols and mechanisms in the MAC layer or devising a new MAC protocol. The latter is the approach that is taken in this chapter.

CR MAC protocols are fundamentally different from conventional MAC protocols [60]. Several contemporary research efforts resulted in the emergence of CR MAC protocols of both centralized and distributed nature [61–69]. In brief, these protocols are different in aspects such as the number of transceivers, the existence of a dedicated common control channel (CCC), network synchronization, the ability to combat the hidden terminal problem, etc. [70], [71], [72]. Nevertheless, the distributed MAC architecture still has a long path ahead in order to be considered as a reliable solution for an interference-free cognitive communications holding to its strict requirements. On the other hand, the centralized architecture seems to be the prudent choice for the current needs [8, 73, 74], which, on the contrary, suffers from the lack of scalability. Therefore, a hybrid centralized/decentralized approach can seize the best of both worlds; thus, its performance modeling is important.

In this chapter, the goal is on protocol design, modeling, and analysis of MIMC-CRNs where cognitive nodes operating in half-duplex mode are able to transmit or receive over several channels. The model remains valid for full-duplex CRNs as well, where the simultaneous transmission and reception over each channel is possible [75, 76]. The modeling framework is aimed to be simple, easy to understand and use, and tractable and the proposed MAC protocol is aimed to be generic, intuitive, and based on practical considerations.

In comparing this contribution with the settings considered in previous works, the proposed protocol and its subsequent modeling are different from similar studies from an operational point-of-view. More importantly, the focus here is on the queue length and the queuing delay of the SUs. Indeed, in spite of its importance, the delay analysis has received little attention, possibly due to the complexity involved. Maybe the closest works to the research in this chapter are due to recent studies in [9], [10]. While in [10] the stationary queue tail distribution of a single SU is analyzed using a large deviation approach, in [9] the first two moments of delay are derived using fluid approximation and constraints on the stability of SUs' queues are introduced. Exploring the same problem as in [9], the first contribution in this chapter is a generalization for an arbitrary number of interfaces while taking channel and sensing errors into account. To accomplish the modeling task, the classical queuing theory and advanced Markov theory are leveraged. Second, different from the said work, which only derives the first two moments of SU's queuing delay using approximation-based methods, this work derives the queue-length PMF as well as its first-moment. The contributions in this chapter and the one in Chapter 2 are different in terms of the problem scope and the methodology used. Indeed, while in Chapter 2 the PMF and CDF of the queue length in single-interface MC-CRNs are numerically worked out and the achievability region and mixed-strategies discussed, in the present chapter the general multi-interface setting is considered. In particular, the PMF and CDF of the SUs' queue length for MIMC-CRNs are numerically obtained, closed-form expressions for the delay and queue length performance metrics are derived, and the distribution tail properties of the SUs' queue occupancy are investigated through statistical analysis.

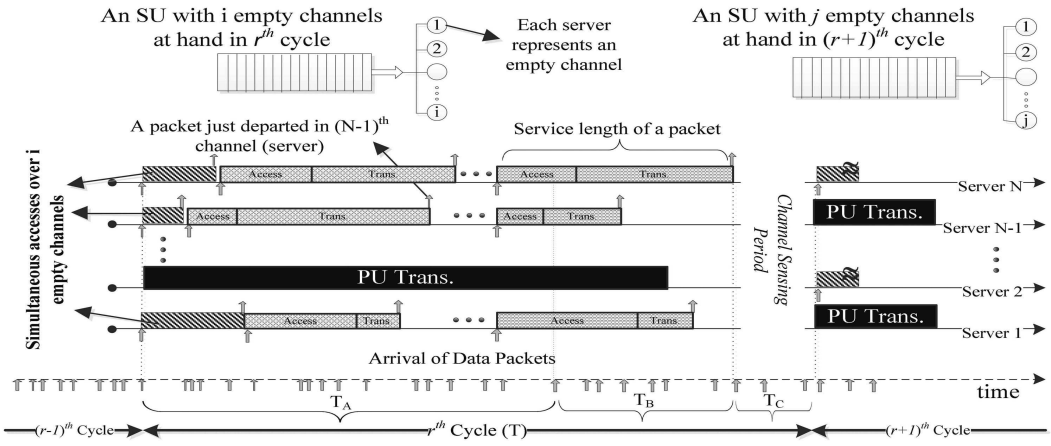
4.1 Multi-Interface Multi-Channel Cognitive Radio Network

MIMC-CRN is a network wherein the cognitive nodes (SUs) are able to transmit over a number of channels simultaneously, rather than sequentially, using multiple PHY interfaces (radios). In essence, the uplink access mechanism introduced in this section works as follows: transmissions from the SUs have to be completed within cycles with synchronized boundaries, which are closely observed and directly controlled by the BS.[2] However, the access right to each available channel (the list of which is to be periodically broadcast to the SUs by the BS as will be explained later) in the uplink can be obtained opportunistically by the SUs through contention. Knowing that the SUs are equipped with multiple interfaces, they are able to contend for channel access on all the available channels, concurrently.

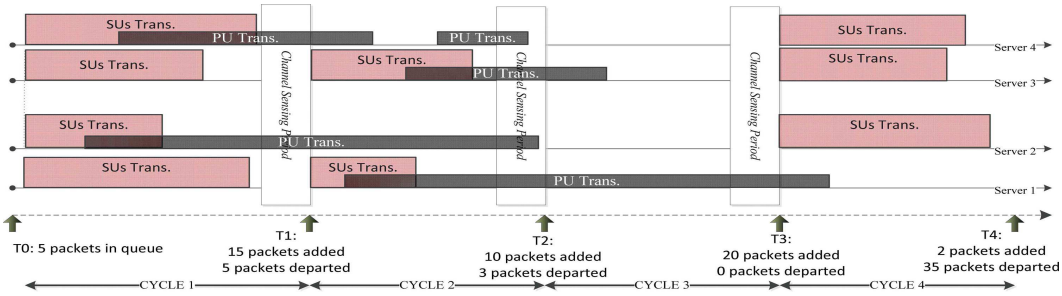
4.1.1 Protocol Description and Network Model

Assume the portion of spectrum that is of interest contains N channels of equal width. A SU is allowed to transmit its data packets on distinct channels simultaneously and monitor (sense) to detect the presence of PUs. This cyclic behaviour of CR intrinsically divides the time axis into periodic intervals, called *cycles*. Fig. 4.1a shows the cyclic functionality of a single SU from its own vantage point, which should not be confused with a multi-user access depiction. As shown in this figure, a transmission cycle of length T is called *cognitive cycle*, which is comprised of three non-overlapping sub-intervals: (i) channel sensing interval (T_{C}), (ii) transmission start-up interval (T_{A}), (iii) and transmission

[2] It should be noted that cyclicity is a trait of interweave CRNs, and not an option. On a different note, cycles on different channels are better to be synchronized for the purpose of interference free operation (in case of inter-channel leakage) and better efficacy [77]. Most importantly is the mis-synchronization among SUs sensing a specific channel, the existence of which can severely degrade the network performance through increasing the possibility of false alarm and fallaciously putting off the usage of a free channel whilst primary network is dormant [78].



(a) Cyclic functionality of an SU in multi-interface multi-channel CRN.



(b) Four consecutive cycles of an SU in multi-interface multi-channel CRN.

Figure 4.1: Proposed network model for multi-interface multi-channel cognitive radio network (MIMC-CRN).

wrap-up interval (T_{B}). Assume that the total of N frequency channels has to be scanned by each SU during T_{C}. As demonstrated in [50], T_{C} depends on the width and number of the channels and total number of samples taken from the received signals pertaining to PUs. At the end of the sensing interval T_{C}, a decision is made, one or more channel IDs are removed from the list of available channels (PU's presence) or are kept in list of available channels (PU's absence).

As illustrated in Fig. 4.1a, chunks of data packets (small arrows) flow down the protocol stack entering the SU's transmission buffer. Based on the previous sensing interval outcomes, this SU accesses the available channels in the transmission start-up phase, which

lasts for a constant period of time T_{A}. This interval is followed by the variable length wrap-up phase (T_{B}), whose aim is to complete the transmission bursts carried over from the former interval, and during which no new packet is admitted for transmission (until T_{A} in the next cycle fires). Transmissions from the primary network start independently of the SUs (interweave CR) when presence of primary transmission is detected during T_{C}, whereupon, the sensed channel is deemed occupied for as long as another sensing interval along the path discovers the absence of a primary transmission. It should be noted that in this protocol the cumulative phase, $T_{\mathrm{A}} + T_{\mathrm{B}}$, represents an interval with generalized length and that the decomposition into T_{A} and T_{B} is just for the purpose of modeling. Then, and as explained before, all SUs perform the sensing task during T_{C}, though the BS is the sole entity to make the final decision on whether a channel is occupied or not based on the sensing results to be cyclically received from individual SUs. The spectrum sensing and reporting mechanism introduced in Chapter 5 can be into this protocol. The entire sensing and reporting procedure in our scheme works as follows:

- SUs sense the channels during the sensing intervals T_{C}.
- Similar to IEEE 802.22 WRAN (section 3.1.2) the SUs report the local sensing results to the BS over their US frequency channels.
- Upon receipt of the sensing results from the SUs, the BS fuses the information and constructs a channel availability map. To accomplish this, the OR, AND and VOTING rules can be applied, whereby different levels of confidence can be achieved.
- The BS broadcasts the sensing results back to the SUs, in the following cycle, so that they know the specific channels over which they are allowed to transmit.

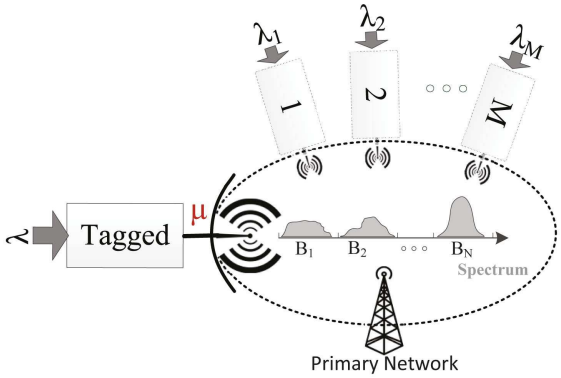
It is worth noting that for the SUs to be able to use channels originally licensed for PUs, several requirements need to be satisfied. The most important of these requirements is related to the interference detection threshold ($-116\ \mathrm{dBm}$) and the minimum channel

Table 4.1: Main notations in this chapter.

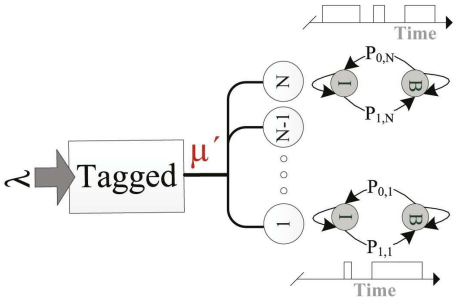
Symbol	Definition	Symbol	Definition	Symbol	Definition
N	Total # of Channels	D_{access}	Access Delay	m	Mean # of SUs per Channel
T	Transmission Cycle	T_s	Transmission Time	$\mathcal{B}$	# of Signal Samples
T_{C}	Sensing Interval	λ	SU Input Rate	P_{B}	Perceived Ch. Busyness Prob.
T_{A}	Start-up Interval	η	Channel Addition Rate	$\mathcal{S}^{\min}$	Power Threshold
T_{B}	Wrap-up Interval	ζ	Channel Elimination Rate	$\overline{D}$	Average Queuing Delay
B	Busy State	$\text{SNR}^{\min}$	Threshold SNR	$\overline{h}$	Fading CH. Gain
I	Idle State	$\overline{\text{L}}$	SU Mean Queue Length	$\mathtt{W}$	Signal Effective BW.
$P_{\text{PU}}^{\text{Active}}$	PU Busyness Prob.	P_{D}	Detection Prob.	$\mathcal{S}$	Received Power
μ'	Decoupling Service Rate	P_{FA}	False Alarm Prob.	$\mathcal{S}_N$	Noise Power
μ	SU Service Rate	P_{Out}	Outage Prob.	$P_{\text{SU}}^{\text{misSync}}$	SU's Mis-Synchronization Prob.

detection time (2 sec) as mandated in [8]. Given this, the former constraint imposes a lower limit on the number of samples taken by the detector ($\mathcal{B}$) and, hence, on the sensing time T_{C}, while the latter puts an upper limit on T_{C} and $\mathcal{B}$ as $\overline{T}/2 < 2\,s$.

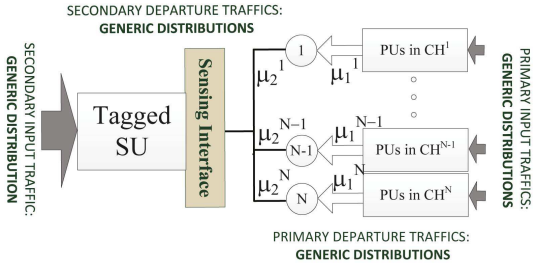
Now, let us assume that k channels were found accessible in the last cycle's sensing interval. Then, SUs attempt to transmit their data packets over these k channels simultaneously, with each packet assigned to a single channel. According to the FIFO policy, upon the departure of a data packet, another one is served immediately. This process continues until the interval T_{A} expires and T_{B} fires, whereupon no new packets get transmitted and those already in service are transmitted until they all depart. Then, the sensing phase of the current cycle (say r) ignites and the next cycle ($r+1$) follows thereafter. Note that the service length from each server (channel) is the sum of access time and packet transmission time. The former time accounts for the fact that SUs have to contend with each other to gain transmission opportunities over a channel. Such contention is separate from the fact that primary transmission should be absent for the channel to be available. For full clarity, the network model and the modeling are delineated with an example in Section 4.2.4. Table 4.1 summarizes the main notations used in this chapter.



(a) Original queuing model.



(b) Equivalent decoupled queuing model.



(c) The simulated queuing model.

Figure 4.2: Equivalent queue representation of the model in Fig. 4.1b.

4.2 Modeling Framework

4.2.1 Primary Channel Characterization

On the primary side, it is assumed that PUs transmitting on distinct channels are statistically independent. The rationale backing this assumption is the fact that these channels are non-overlapping portions of spectrum and that the PUs holding the right of access to them are distinct entities. Given the independence assumption, and as shown in Fig. 4.2b, the ON/OFF model is used where (due to the presence/absence of primary transmission) the i^{th} channel transitions between the busy (B) and idle (I) states with probabilities $P_{0,i}$ and $P_{1,i}$, respectively. Therefore, the i^{th} channel is occupied with probability $P_{\text{PU}}^{\text{Active}} = P_{1,i}/(P_{0,i} +$

$P_{1,i}$) and is idle with probability $1 - P_{\text{PU}}^{\text{Active}}$.

4.2.2 Decoupling of Secondary Users

According to the above description, the interactions among the SUs in the MIMC-CRN can be captured using a number of interacting multi-server queues, where each server represents an available channel. Packets embark on servers according to the FIFO policy. The fundamental difference between this queueing system and the classical multi-server queueing systems is that the former is a multi-access queuing system where the number of servers (channels) is constantly changing in time due to the presence/absence of PUs.[3] Since the performance evaluation of this network with coupled queues, as in Fig. 4.2a, is very challenging, we adopt a different approach called decoupling,[4] whose validity has been approved and exploited in many studies, for instance [9], [32] and [33].

To elucidate the logic behind queue decoupling, an exemplary single-channel network is shown in Fig. 4.3, wherein $M + 1$ users are competing to access the channel. From the vantage point of this tagged SU, the channel is comprised of alternating intervals, namely, transmissions originating from this user (μ) and the access delay (D_{access}), which is the impact of the rest of the network (i.e., m other nodes) on the channel. The latter quantity is generic and, depending on the type of access mechanism, number of competing users, etc., exhibits different probabilistic behaviors. This characteristic is illustrated in Fig. 4.3a. Along this vision, it would be valid to claim that Fig. 4.3b is an equivalent representation for Fig. 4.3a, where transmission times are virtually prolonged (or, equivalently, transmission rate is reduced) for an entire cycle (including μ and D_{access}). With this shift of viewpoint, during the modified transmission intervals (μ'), first the real packets are transmitted, which is followed by the transmission of *dummy* packets that are only for the

[3] As pointed out in Section 3.2.3, the notion of queues with random number of servers, also known as queuing networks in a random environment, is well established in [55], [58].

[4] Here, we emphasize again that the decoupling approach is not a simplifying assumption but a modeling abstraction.

modeling purpose and their departures do not reduce the size of the queue. This is shown in Fig. 4.3b. since the virtual transmissions occur in sequence and the input process and the policy of the tagged user's queue has not changed, the network models in Fig. 4.3a and 4.3b are equivalent in all aspects but the service rates, which are related according to:

$$\mu' \cong \frac{1}{D_{\text{access}} + \frac{1}{\mu_2^t}}, \tag{4.1}$$

The access delay D_{access} depends on the choice of access mechanism, number of contending SUs, traffic, etc.[5]

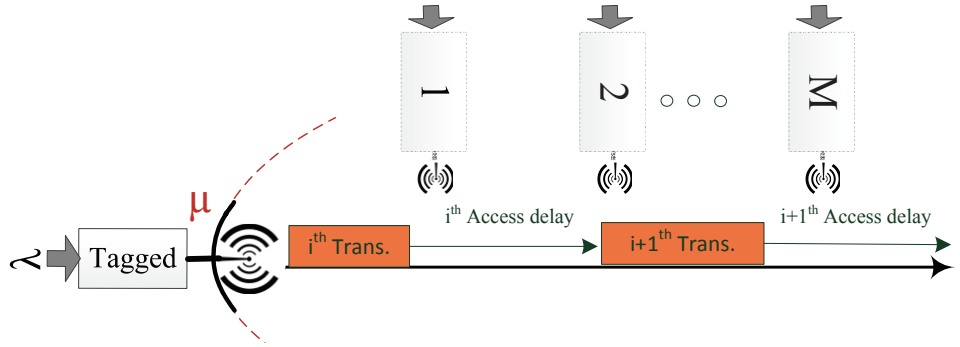
Remarkably, no simplifying assumption has been made in the derivation of this equation. Therefore, as it was shown, queues are still coupled; for example, adding one more node to the rest of the network will supposedly increase the access delay which, subsequently, decreases μ'. Indeed, the only difference between the coupled and decoupled settings is that the impact of the rest of the network is abstracted as a quantity, and integrated into the heart of the simplified network model through modification of the departure process. Obviously, in (4.1), the only thing that is needed is an expression for the access delay.

Fig. 4.2c shows an equivalent representation for the queue model in Fig. 4.2b, which is simulated and used for validation purposes in Section 4.3.

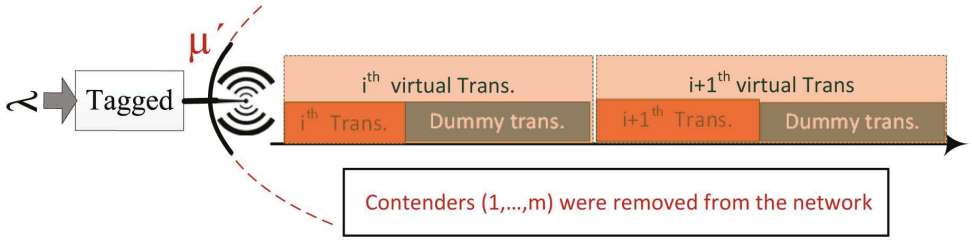
4.2.3 Continuous-Time Markov Chain

The decoupled queue representation in Fig. 4.2b is modeled with the 2D CTMC shown in Fig. 4.4a. The states in this Markov chain are represented by pairs (n_i, ℓ), whereby n_i represents the total number of free channels (out of N) and ℓ represents the number of data packets stacked inside decoupled *tagged* SU. The horizontal transition rates, λ and

[5]For instance, [34] presents a non-recursive closed-form expression for D_{access} in IEEE 802.11 networks.



(a) A coupled network where the tagged user, with service rate μ, contends with m users.

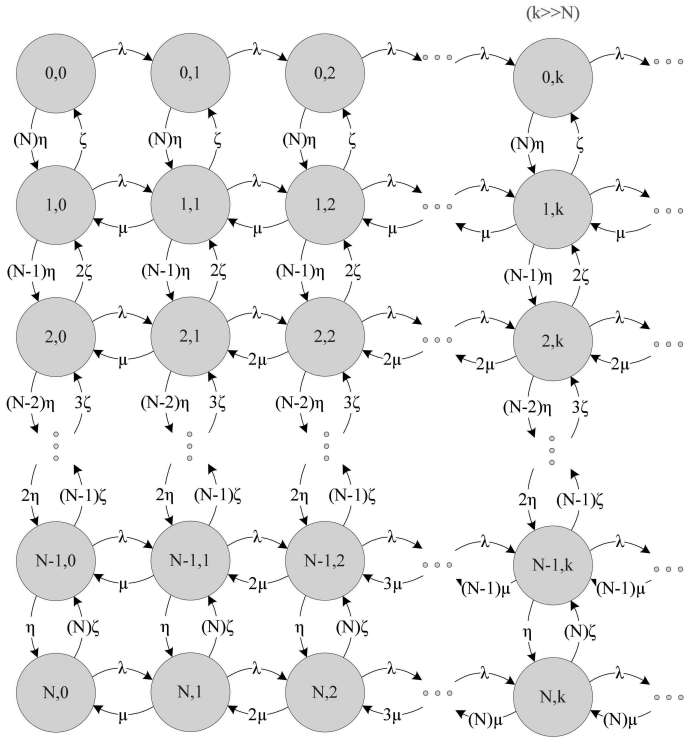


(b) An equivalent decoupled network where all contenders are eliminated and the tagged user transmits with service rate μ'.

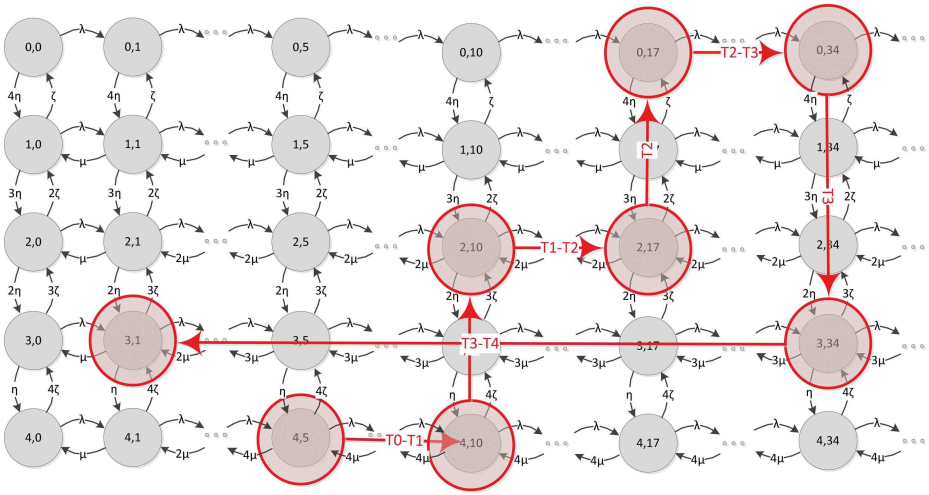
Figure 4.3: Illustration of the queue decoupling logic in the network.

μ' (hereinafter, represented by μ for simplicity of notation), correspond to the arrival and departure of packets, respectively. The vertical transition rates, η and ζ, denote the addition/elimination of a specific channel to/from the list of channels available to the decoupled SU, respectively. Given this, it should not be difficult to understand why the model in Fig. 4.4a represents the SU's dynamics if one starts from a state and follows the transitions. The packet arrival process is Poisson distributed (similar to [9, 11–13, 30]), which is rigorous in many situations, especially when the elements that contribute to generating the input traffic at a node (SU) come in large numbers and are statistically independent [79], e.g., resultant traffic due to aggregation of many independent sessions and applications.

The other assumption this model entails is related to the fact that the vertical transitions to the farther states with the same queue length are neglected. This assumption, which was



(a) Generic model.



(b) Exemplification with $N = 4$. The red identifiers (circles and arrows) are meant to aid readers track the system behavior in four cycles of Fig. 4.1b.

Figure 4.4: 2D CTMC modeling the dynamics of aN SU in the network.

tested by simulation and theory, proved to be satisfactory, sustained by the fact that within T cycles of msec length it is unlikely to have long vertical transitions among states. Later on, we will find these transition rates based on PHY layer quantities such as PU activity factor ($P_{\mathrm{PU}}^{\mathrm{Active}}$), SUs' mis-synchronization ($P_{\mathrm{SU}}^{\mathrm{misSync}}$), packet length, signal-to-noise-ratio threshold $\mathrm{SNR}^{\mathrm{min}}$, etc. Most notably, the first row of the CTMC in Fig. 4.4a is different since it corresponds to the case of no available channel where packets get piled up in the queue upon arrival.

4.2.4 Case Study

For the purpose of better elucidation, we follow a chain of four cycles in Fig. 4.1b and the corresponding Markov chain in Fig. 4.4b. Four channels ($N = 4$) are available to the SU (not necessarily vacant). Transitions on this Markov chain happen in either of these cases: (i) Packet departure from any channel; (ii) packet arrival to the queue; (iii) channel is occupied by the primary; (iv) channel is released by the primary.

Let us focus on the last two cases. Four cycles, with lengths $T_0, \cdots, T_4$ are considered during which server transitions occur (see Fig. 4.1b). The PU may appear on the channel at any moment due to its unconditional right of access. However, this situation may not be instantly perceived as the MIMC-CRN cannot afford asking SUs to sense the channels continuously. Back to Fig. 4.4b, and starting from state $(4, 5)$ (instant T_0), the system is transmitting on channels $\mathrm{CH}_{1,2,3,4}$. The addition of 5 packets to the queue drags the system to state $(4, 10)$. At the end of the first sensing interval, it is recognized that two channels are occupied, thus taking the system to state $(2, 10)$ to protect $\mathrm{CH}_{2,4}$. Then, during the second cycle, 7 more packets are added and the new system state becomes $(2, 17)$. At the end of the second sensing interval, even $\mathrm{CH}_{2,4}$ are sensed busy, leaving the system with no channel to transmit over, i.e., state $(0, 17)$. This goes on, and that is how our modeling is justified.

$$\left(\ell\frac{\mu}{M}+n\zeta+\lambda+(N-n)\,\eta\right)\pi(n,\ell)=(N-n+1)\,\pi(n-1,\ell)+(n+1)\,\zeta\pi(n+1,\ell)+\lambda\pi(n,\ell-1)+\frac{(\ell+1)\,\mu}{M}\pi(n,\ell+1), \qquad \begin{matrix} n=0,\cdots,N \\ \ell<n \end{matrix}$$

$$\left(n\frac{\mu}{M}+n\zeta+\lambda+(N-n)\,\eta\right)\pi(n,\ell)=(N-n+1)\,\pi(n-1,\ell)+(n+1)\,\zeta\pi(n+1,\ell)+\lambda\pi(n,\ell-1)+n\frac{\mu}{M}\pi(n,\ell+1), \qquad \begin{matrix} n=0,\cdots,N \\ \ell\geq n \end{matrix} \tag{4.3}$$

4.2.5 Stability Analysis

Like all infinite-buffer queuing systems, there is a SC for this MAC protocol and its corresponding CTMC beyond which the network is considered unstable. Finding SC for multi-channel CRNs was first attempted in [9]. Our derivation in this part is a generalization of [9, Eqs. (1), (16)] to account for sensing faults, mis-synchronization and channel conditions in the MIMC-CRN under study. The result is presented in the following theorem.

Theorem 2. *In a MIMC-CRN with a total of N frequency channels, where on average m SUs attempt to access each channel, the necessary and sufficient condition for stability in each SU is given by*

$$\frac{\lambda}{\mu}<\frac{N}{M}\left(\frac{\eta}{\zeta+\eta}\right). \tag{4.2}$$

Proof. Writing the balance equation of the CTMC, we obtain (4.3), where $\pi(n,\ell)$ represents the probability of being in state (n,ℓ) of the chain. By multiplying both sides of (4.3) with z^{ℓ} and then summing up over the variable $0\leq\ell<\infty$, we get

$$\begin{aligned}&\left(\lambda z(1-z)-n\frac{\mu}{M}(1-z)+n\zeta z+\eta(N-n)z\right)\mathcal{G}_n(z)\\&=(n+1)\zeta z\mathcal{G}_{n+1}(z)+(N-n+1)\eta z\mathcal{G}_{n-1}(z)\\&-(1-z)\sum_{\ell=0}^{n-1}(n-\ell)\frac{\mu}{M}z^{\ell}\pi(n,\ell), \qquad n=0,\ldots,N,\end{aligned} \tag{4.4}$$

where $\mathcal{G}_n(z)$ is the probability generating function (PGF) of the CTMC in each row, ob-

tained as

$$\mathcal{G}_n\left(z\right) = \sum_{\ell=0}^{\infty} z^{\ell}\pi(n,\ell) \ . \tag{4.5}$$

In the second step, we take the first derivative of (4.4) *w.r.t.* z and let $z = 1$:

$$\begin{aligned} &\mathcal{G}_n'(1)(\eta(N-n)+n\zeta) + \mathcal{G}_n(1)\left(\eta(N-n) + n\frac{\mu}{M} - \lambda + n\zeta\right) \\ &= (n+1)\zeta\mathcal{G}_{n+1}(1) + (N-(n-1))\,\eta\left(\mathcal{G}_{n-1}(1) + \mathcal{G}_{n-1}'(1)\right) \\ &+ (n+1)\zeta\mathcal{G}_{n+1}'(1) + \sum_{\ell=0}^{n-1}(n-\ell)\frac{\mu}{M}\pi(n,\ell), \qquad n = 0,\ldots,N. \end{aligned} \tag{4.6}$$

Then, we sum (4.6) over the variable $0 \leq n \leq N$, which after simplifications and taking into account that $\sum_{n=0}^{N}\mathcal{G}_n\left(1\right) = 1$, yields

$$-N\eta\mathcal{G}_N'(1) + N\eta - \lambda + \frac{\mu}{M}\sum_{n=0}^{N} n\mathcal{G}_n(1) = \frac{\mu}{M}\sum_{n=1}^{N}\sum_{\ell=0}^{n-1}(n-\ell)\pi(n,\ell) + N\eta\sum_{n=0}^{N-1} n\mathcal{G}_n'(1). \tag{4.7}$$

Since, due to (4.7), $\mathcal{G}'\left(z=1\right) = \sum_{n=0}^{N}\mathcal{G}_n'\left(z=1\right) = \overline{\mathrm{L}}$ is the SU's mean queue length, the following identity holds true:

$$\sum_{n=1}^{N}\sum_{\ell=0}^{n-1}(n-\ell)\pi(n,\ell) > 0. \tag{4.8}$$

Finally, by putting together the first term on the left-hand-side (LHS) of (4.7) with the last term on the right-hand-side (RHS) of (4.7), and using the inequality in (4.8), the necessary and sufficient condition for the stability of the SU's queue is obtained, thus completing the proof. □

We affirm that the SC in (4.2) is generic and valid for non-Poisson arrival and service processes at both the secondary and the primary sides. This claim is predicated on the fact

that SCs are always independent from any higher moments of a queue's input and output processes, a fact that is a principle in classical queuing theory. Moreover, (4.2) is identical to the SC derived in [9] for ideal sensing (detection probability P_{D} and false alarm probability $P_{\mathrm{FA}} = 0$) and fully synchronized operation ($P_{\mathrm{SU}}^{\mathrm{misSync}} = 0$). Therefore, by induction, the generalized SC for practical MIMC-CRN with arbitrary traffic and service model, faulty sensing ($P_{\mathrm{D}} \neq 1$ and $P_{\mathrm{FA}} \neq 0$), imperfect synchronization under heterogeneous and erroneous channels (i.e., $P_{\mathrm{PU}_i}^{\mathrm{Active}} \neq P_{\mathrm{PU}_j}^{\mathrm{Active}}, i = 1, \ldots, \#$ of SUs, and $P_{\mathrm{Out},i} \neq 0$) is given by:

$$\lambda_i < \frac{N(1 - P_{\mathrm{Out},i})}{M} \mu_i \left(1 - \frac{1}{N} \sum_{j=1}^{N} \frac{\zeta_j}{\eta_j + \zeta_j} \right), \quad i = 1, \cdots, M, \tag{4.9}$$

where $P_{\mathrm{Out},i}$ is the outage probability that the i^{th} SU experiences, and η_j and ζ_j (μ_i and λ_i) are vertical (horizontal) transition rates for the j^{th} channel (i^{th} SU) in the CTMC of Fig. 4.4a, respectively. For ideal channel and sensing conditions, (4.9) reduces to [9, Eq. (16)].

An intuitive interpretation of the SC in (4.2) is that the expected amount of service demand (the LHS) must be smaller than the fraction of available servers to each SU (the RHS). As a special case, the probability of having n servers (free channels) in the system, denoted by $\pi(n)$, is obtained by plugging $z = 1$ into (4.5), which yields

$$\mathcal{G}_n(1) = \pi(n) = \sum_{\ell=0}^{\infty} \pi(n, \ell) \ . \tag{4.10}$$

Then, given the total number of channels N, the PGF of the CTMC is given by

$$\mathcal{G}(z) = \sum_{n=0}^{N} \mathcal{G}_n(z) \ . \tag{4.11}$$

Without going through the derivation details, it is possible to show that the recursive formula in (4.12) can be obtained from (4.3), (4.5) and (4.10), relating the probabilities of

having n and $n+1$ idle channels (severs) in the network.

$$\mathcal{G}_{n+1}(1) = \mathcal{G}_n(1)\left[\frac{(N-n)\,\eta}{(n+1)\,\zeta}\right] \qquad n = 0, \ldots, N. \tag{4.12}$$

This equation, when combined with the normalization condition $\sum_{n=0}^{N} \mathcal{G}_n(1) = 1$, yields the absolute value of $\pi(n)$ (or $\mathcal{G}_n(1)$) as follows:

$$\mathcal{G}_n(1) = \pi(n) = \binom{N}{n}\left(\frac{\eta}{\eta+\zeta}\right)^n\left(1-\frac{\eta}{\eta+\zeta}\right)^{N-n}. \tag{4.13}$$

This is a similar argument as the one made in Section 3.2.3 where it is seen that the PMF of having n frequency channels available to the SUs is of the Binomial type. Therefore, the parameter $\eta/(\eta+\zeta)$ would be equal to the probability of finding a specific channel busy. Analogously, (4.13) is independent of the exogenous arrival rate λ and the service rate μ. Later on, this result will be used to explain some important properties of the network and find an interesting approximation to the 2D CTMC illustrated in Fig. 4.4a. Such approximation comes handy provided that no closed-form expression exists for the joint PMF $\pi(n,\ell)$ of the CTMC in Fig. 4.4a or, even, for 2D birth-death processes of much simpler nature.

Having the model defined as above, next we need to relate the transition quantities to the factual network quantities, such as packet length, data rates, sensing precision, fading rate, etc.

4.2.6 Energy Sensing and Channel Busyness Probability

As stated earlier, energy detection is used as the sensing technique to be performed during T_{C} sub-intervals (cf. Fig. 4.1a), as is the most simple and cost effective sensing technique when no information about the transmitted signal is a-priori known to SUs. To have a reliable sensing output in this method, the condition $T_{\mathrm{C}} > \mathcal{B}/2\mathtt{W}$ is to be satisfied, where

$\mathcal{B}$ is the number of samples taken from the received signal of bandwidth $\mathtt{W}$ [50], [80]. Apparently, a higher number of samples yields higher resolution through higher detection probability and fewer false alarm events. However, this comes at the price of a larger sensing time, which, at some point, can compromise the channel utilization by the SUs.

In this subsection, a closed-form expression is found for the perceived channel busyness probability P_{B}, which is needed for the characterization of the vertical rates η and ζ (cf. Fig. 4.4a). To this end, we first need to characterize the sensing procedure based on its fundamental metrics, namely, the detection probability (P_{D}) and the false alarm probability (P_{FA}). Unfortunately, the expressions of P_{D} and P_{FA} in a Rayleigh fading environment are very complicated to get insight from and to calculate. Therefore, accurate enough approximations for P_{D} and P_{FA} are introduced in what follows, which not only simplify our analytical treatment in solving the CTMC in Fig. 4.4a but also can be used as a baseline for future works in this context.

Following the notations in [50] and denoting $\mathcal{H}_0$ and $\mathcal{H}_1$ as the events of signal absence and signal presence at the receiver, respectively, P_{D} is defined as

$$P_{\mathrm{D}} = \Pr\left(\mathcal{S} = \sum_{n=1}^{\mathcal{B}} |y\left(n\right)|^2 > \mathcal{S}^{\min} | \mathcal{H}_1\right) = \Pr\left(\sum_{n=1}^{\mathcal{B}} |h\left(n\right) x\left(n\right) + z(n)|^2 > \mathcal{S}^{\min} | \mathcal{H}_1\right), \tag{4.14}$$

where $h\left(n\right)$, $x\left(n\right)$ and $z\left(n\right)$ represent the samples of instantaneous channel fading, transmitted symbol amplitude, and noise with zero-mean Gaussian PDF $\mathcal{N}\left(0, \mathcal{S}_N\right)$, respectively, and where $\mathcal{S}^{\min}$ denotes the power detection threshold. Knowing that the processes $h\left(n\right)$ and $z\left(n\right)$ are statistically independent, and as long as the contiguous received signal samples, $y\left(n\right)$, are statistically independent (which requires the inter-sampling intervals to be larger than the channel coherence time such that taken samples experience independent realizations of channel), then the central limit theorem is applicable to approximate the PDF of the energy detector's output ($\mathcal{S} = \sum_{n=1}^{\mathcal{B}} |y\left(n\right)|^2$ in (4.14)). Finally, for the sake

of high precision, the number of samples $\mathcal{B}$ should be large enough (which is almost always the case for typical values of T_{C}). Accordingly, the PDF of the detector's output is approximated by $f_s\left(x\right) \sim \mathcal{N}\left(a, b\right)$, where

$$a = \mathcal{B} \cdot \left(\mathcal{S}_N + \overline{h}\mathcal{S}\right) \quad \text{and} \quad b = \mathcal{B} \cdot \left((\overline{h}\mathcal{S})^2 + 4\overline{h}\mathcal{S}\mathcal{S}_N + 2\mathcal{S}_N^2\right) . \tag{4.15}$$

Here, $\mathcal{S}$ is the received power, $\mathcal{S}_N$ is the noise power, and $\mathcal{B} = 2T_{\mathrm{C}}\mathtt{W}$ represents the number of samples taken. By applying the above result to (4.14), the expression for P_{D} in Rayleigh fading environment is obtained as

$$P_{\mathrm{D}} = \mathbb{Q}\left(\frac{\mathrm{SNR}^{\min} - \left(1 + \overline{h}\mathrm{SNR}\right)\mathcal{B}}{\sqrt{\mathcal{B}\left((\overline{h}\mathrm{SNR})^2 + 4\overline{h}\mathrm{SNR} + 2\right)}}\right) , \tag{4.16}$$

where $\mathrm{SNR}^{\min} = \mathcal{S}^{\min}/\mathcal{S}_N$ and SNR represent the minimum and instantaneous SNR values, respectively, and $\mathbb{Q}\left(\cdot\right)$ is the Gaussian Q-function. As for the probability of false alarm, since no signal is actually transmitted in this case, the expression for P_{FA} is independent of the fading environment and is only dependent on noise. Thus, the corresponding expression is as given in [50], and can further be approximated by the following equation.

$$P_{\mathrm{FA}} = \frac{1}{\Gamma\left(\dfrac{\mathcal{B}}{2}\right)}\Gamma\left(\frac{\mathcal{B}}{2}, \frac{\mathrm{SNR}^{\min}}{2}\right) \approx \mathbb{Q}\left(\frac{\mathrm{SNR}^{\min} - \mathcal{B}}{\sqrt{2\mathcal{B}}}\right) , \tag{4.17}$$

which is equal to $P_{\mathrm{D}}(\mathrm{SNR} = 0)$. Here, $\Gamma\left(\cdot\,,\cdot\right)$ and $\Gamma(\cdot)$ denote the upper incomplete Gamma function and the Gamma function, respectively. Assuming that the application requirements enforce $P_{\mathrm{D}} > P_{\mathrm{D}}^{\min}$ and $P_{\mathrm{FA}} < P_{\mathrm{FA}}^{\max}$, the following inequality is to be satisfied

through the proper choice of sampling count $\mathcal{B}$:

$$\begin{aligned}\mathbb{Q}^{-1}\left(P_{\mathrm{FA}}^{\max}\right)\sqrt{2\mathcal{B}}+\mathcal{B}<\mathrm{SNR}^{\min}<\mathbb{Q}^{-1}\left(P_{\mathrm{D}}^{\min}\right)\times\\ \sqrt{\mathcal{B}\left((\overline{h}\mathrm{SNR})^2+4\overline{h}\mathrm{SNR}+2\right)}+\left(1+\overline{h}\mathrm{SNR}\right)\mathcal{B},\end{aligned} \tag{4.18}$$

where $\mathbb{Q}^{-1}(\cdot)$ is the inverse of the Gaussian Q-function. For the sake of confidence, we checked the accuracy of (4.16) and (4.17) using the *Kullback–Leibler* divergence method and we found that the precision error is very negligible for $\mathcal{B} \gg 5$, thus validating the proposed expressions.

Once the BS receives the local sensing results from all sensors, it fuses them using a data combining method to make the relevant decisions on the status of the corresponding channels. Among few available choices, in this section, the OR-rule is adopted with the following detection and false alarm probabilities:[6]

$$P_{\mathrm{FA},i}^{\mathrm{OR}}=1-\prod_{j=1}^{m_i}(1-P_{\mathrm{FA},j}),\qquad P_{\mathrm{D},i}^{\mathrm{OR}}=1-\prod_{j=1}^{m_i}(1-P_{\mathrm{D},j}), \tag{4.19}$$

where the superscript ^{j} indicates that the sensing result belongs to the j^{th} SU ($1 \leq j \leq M$) and m_i stands for the number of sensors that are to sense the i^{th} channel, provided that $\sum_{i=1}^{N} m_i = M$. Since each sensor j ($0 \leq j \leq m_i$) senses the i^{th} channel with a different number of samples ($\mathcal{B}_j$) and experiences different channel fading gain $\overline{h}_j$, $P_{\mathrm{FA},j}$ and $P_{\mathrm{D},j}$ will be different among sensor nodes sensing the same channel. For the simpler case that all channels exhibit a similar sensing profile, the probability that a specific channel is perceived

[6]Our choice is based upon the logic that the OR-rule results in the lowest mis-detection rate compared to its rivals and for that reason mandated in [8] for the U.S. regulatory domain. In the next chapter of this dissertation, it will be shown that such superiority in terms of mis-detection rate come at an expensive price of uneven performance w.r.t spectral efficiency (false-alarm rate) caused by the rigid and non-adaptive behavior of this fusion rule rendering it inefficient in dealing with variety of situation. As such, a more complicated and efficient spectrum sensing mechanism is proposed for IEEE 802.22 WRANs in the corresponding section.

busy (P_{B}) during T_{C} would be the same across channels and equal to

$$P_{\mathrm{B}} = \sum_{r=0}^{1} \Pr\left(\mathcal{S} > \mathcal{S}^{\min} | \mathcal{H}_r\right) \Pr\left(\mathcal{H}_r\right) = P_{\mathrm{D}}^{\mathrm{OR}} \Pr\left(\mathcal{H}_1\right) + P_{\mathrm{FA}}^{\mathrm{OR}} \Pr\left(\mathcal{H}_0\right) \ . \tag{4.20}$$

In (4.20), the only unknown quantity is $\Pr\left(\mathcal{H}_1\right)$, knowing that $\Pr\left(\mathcal{H}_0\right) = 1 - \Pr\left(\mathcal{H}_1\right)$. Since a transmission during T_{C} either originates from PU (with probability $P_{\mathrm{PU}}^{\mathrm{Active}}$ as discussed in subsection 4.2.1) or is due to unwanted asynchronism among SUs (for some SUs may continue transmitting during T_{C} with probability $P_{\mathrm{SU}}^{\mathrm{misSync}}$), then $\Pr\left(\mathcal{H}_1\right)$ would be equal to

$$\begin{aligned} \Pr\left(\mathcal{H}_1\right) &= \Pr\left(\text{PU on}\right) + \Pr\left(\text{SUs on} | \text{PU off}\right) \times \Pr\left(\text{PU off}\right) \\ &= P_{\mathrm{PU}}^{\mathrm{Active}} + P_{\mathrm{SU}}^{\mathrm{misSync}} \left(1 - P_{\mathrm{PU}}^{\mathrm{Active}}\right), \end{aligned} \tag{4.21}$$

where on/off is used to refer to a node being active/silent during a sensing interval. The question on how different factors come together to result in $P_{\mathrm{SU}}^{\mathrm{misSync}}$ is not our concern in this section and is only taken into account as a lump quantity. This completes the determination of P_{B} in terms of $P_{\mathrm{D}}^{\mathrm{OR}}$ and $P_{\mathrm{FA}}^{\mathrm{OR}}$ and $\Pr\left(\mathcal{H}_1\right)$ found in (4.19) and (4.21), respectively.

4.2.7 Rate Determination

Having found P_{B}, it is easy to see that the vertical rates of the CTMC are related to the perceived channel busyness probability according to the following set of equations:

$$\begin{aligned} \eta &= \frac{1 - P_{\mathrm{B}}}{\overline{T}}, \\ \zeta &= \frac{P_{\mathrm{B}}}{\overline{\overline{T}}}, \end{aligned} \tag{4.22}$$

where $\overline{T}$ represents the average cognitive cycle duration. As defined in Section 4.1.1, $\overline{T}$ is composed of three intervals (cf. Fig. 4.1a). That is,

$$\overline{T} = T_{\mathrm{A}} + \overline{T}_{\mathrm{B}} + \overline{T}_{\mathrm{C}}, \tag{4.23}$$

where, as aforementioned, $\overline{T}_{\mathrm{A}}$, $\overline{T}_{\mathrm{B}}$ and $\overline{T}_{\mathrm{C}}$ represent the average of the start-up, wrap-up and sensing phases, respectively.

At this point, we just need to find appropriate expressions for T_{B} since T_{A} is set constant and $T_{\mathrm{C}} > \mathcal{B}/2\mathtt{W}$. Assuming that the service periods (each transmission in Fig. 4.1a) are exponentially distributed with average $1/\mu$ (and μ derived from (4.1)), then the transmission wrap-up period T_{B} is a RV that is the maximum of n such exponentially distributed RVs, as follows:

$$\begin{aligned} T_{\mathrm{B}} &= \max\left(y_1,\, y_2, \ldots, y_n\right) \ , \\ n &\sim \mathrm{Binomial}\left(N, \eta/(\eta+\zeta)\right) \ , \\ y_j &\sim \exp\left(\frac{1}{\mu}\right) \ . \end{aligned} \tag{4.24}$$

Therefore, the PDF of the wrap-up period is given by

$$\begin{aligned} f_{T_{\mathrm{B}}}\left(y|n\right) &= \frac{\partial \Pr\left(T_{\mathrm{B}} < y\right)}{\partial y} = \frac{\partial \Pr\left(\max\left(y_1,\, y_2, \ldots, y_n\right) < y\right)}{\partial y} \\ &= \frac{\partial \left(\Pr\left(y_1 < y\right) \ldots \Pr\left(y_n < y\right)\right)}{\partial y} = \frac{\partial \Pr\left(y_j < y\right)^n}{\partial y} \\ &= n\mu e^{-\mu y}\left(1 - e^{-\mu y}\right)^{n-1} \quad n = 1, \ldots, N \ , \end{aligned} \tag{4.25}$$

Since we need to find $\overline{T}_{\mathrm{B}}$, we need to compute the mean of (4.25) and then remove the

condition on n as follows:

$$\begin{aligned}\overline{T}_{\mathrm{B}|n} &= \frac{\mathsf{c}n + n\psi\left(n\right) + 1}{n\mu}, \\ \overline{T}_{\mathrm{B}} &= \sum_{n=0}^{N} \overline{T}_{\mathrm{B}|n}\pi(n),\end{aligned} \tag{4.26}$$

where c and $\psi\left(n\right)$ are the *Euler-mascheroni* constant ($\mathsf{c} \approx 0.577$) and *digamma* function, respectively, and where $\pi(n)$ is the probability of having n channels at hand. Also, it is possible to find the PDF of T_{B} by making use of (4.13) to eliminate the condition in (4.25), which yields

$$f_{T_{\mathrm{B}}}\left(y\right) = \sum_{n=0}^{N} f_{T_{\mathrm{B}}}\left(y|n\right)\pi(n) = N\mu\left(\frac{\eta}{\eta+\zeta}e^{-\mu y}\right)\left(1-\frac{\eta}{\eta+\zeta}e^{-\mu y}\right)^{N-1}. \tag{4.27}$$

After plugging T_{A}, T_{B} and T_{C} into (4.23), we become able to determine the vertical rates, η and ζ, from (4.22) in terms of the fundamental network quantities. Lastly, the horizontal rates, λ and μ, are determined from the SU's input traffic intensity and the service delay length (access+transmission), respectively.

4.3 Secondary's Queue Length

4.3.1 Numerical Analysis

Obtaining a closed-form expression for probabilities $\pi(n,\ell)$ directly from the 2D CTMC under study is seemingly impossible. Nevertheless, finding the marginal PGFs accomplishes the same task from a numerical point-of-view, keeping in mind that even the latter is an NP-hard problem.

In (4.13), we were able to find the marginal PMF for the number of servers. Plugging

(4.5) into (4.11), we obtain the marginal PGF for the number of packets in SU's queue:

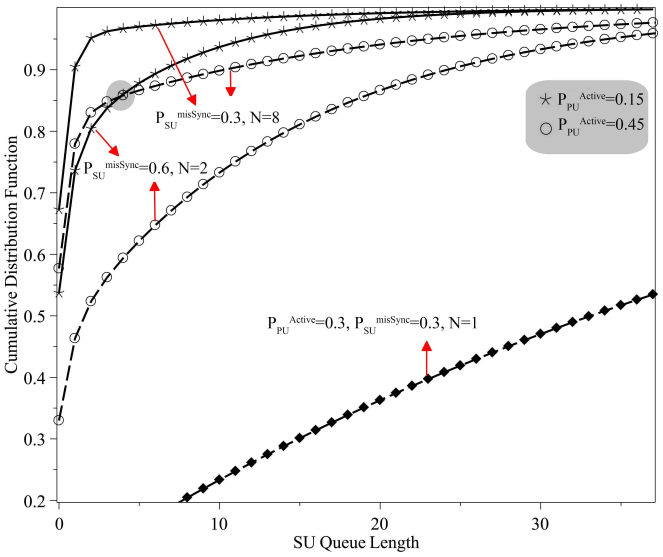
$$\mathcal{G}(z) = \sum_{n=0}^{N} \sum_{\ell=0}^{\infty} z^{\ell} \pi(n, \ell) = \sum_{\ell=0}^{\infty} z^{\ell} \pi_{\ell}, \tag{4.28}$$

where

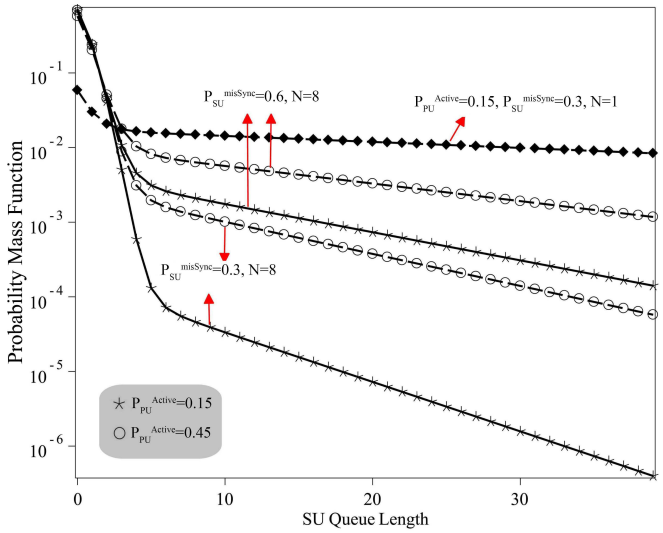
$$\pi_{\ell} = \sum_{n=0}^{N} \pi(n, \ell) = \frac{\mathcal{G}^{(\ell)}(z=1)}{\ell!} \ . \tag{4.29}$$

Here, $\mathcal{G}^{(\ell)}(z)$ represents the ℓ^{th} derivative of $\mathcal{G}(z)$. Even though equation (4.29) does not provide us with a closed-form expression, it allows obtaining the PMF and CDF of the number of data packets in SU's queue by solving the CTMC in Fig. 4.4a using the z-transform numerical method from (4.4) and (4.29) (rather than directly striving to solve an unlimited set of equations with unlimited number of variables in (4.3)). Figs. 4.5 illustrate the mentioned CDF and PMF when PUs exhibit different levels of activity ($P_{\text{PU}}^{\text{Active}}$) and the SUs are not perfectly synchronized ($P_{\text{SU}}^{a} \neq 0$). Furthermore, in Fig. 4.5b, the maximum number of channels that can be exploited by the SUs is variable. The vertical scale in Fig. 4.5b is logarithmic in order to magnify the differences. As per intuition, a larger number of channels (N) leaves less data packets waiting in the queues. Also, according to these plots, a higher activity of PUs gives less chance to the SUs to transmit and results in more piled-up SU queues. The most important inference from Fig. 4.5b is related to the tail of the distribution, which makes it clear that any change in N, $P_{\text{PU}}^{\text{Active}}$ and $P_{\text{SU}}^{\text{misSync}}$ directly reflects in a change in the PMF slope. This, in fact, signifies that if a closed-form expression existed for the PMF of queue length, the above-mentioned quantities would have most likely appeared as exponents rather than coefficients.

We also observed that SUs that collect more sensing samples, with the hope of lessening the interference to PUs, may severely face the diminishing return of sacrificing their QoS, since this increases the delay with a rate higher than exponential. For instance, when $P_{\text{PU}}^{\text{Active}} = 0.42$, the 20% increase in the number of samples, increases the delay 10 times



(a) CDF vs. Number of data packets in SU's queue



(b) PMF vs. number of data packets in the SU's queue

Figure 4.5: CDF and PMF as a function of the SU's queue length, where the probability of PU being active ($P_{\mathrm{PU}}^{\mathrm{Active}}$), the mis-synchronization level ($P_{\mathrm{SU}}^{\mathrm{misSync}}$) and the number of channels (N) are taken as variables.

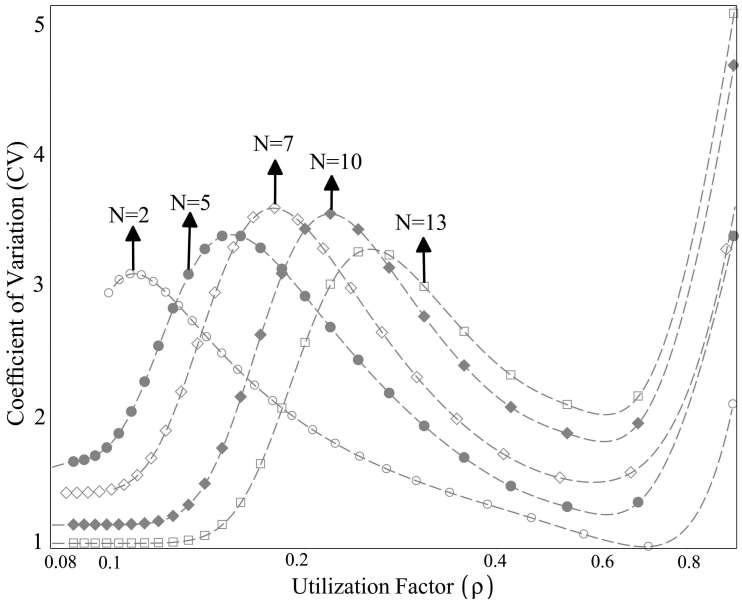


Figure 4.6: Coefficient of variation versus ρ and N.

more (results not presented as figures here). The situation exacerbates for larger $P_{\mathrm{PU}}^{\mathrm{Active}}$. In other words, the choice of the sensing method and its associated parameters seems to be the turning point beyond which the SU's QoS is sacrificed for the PU's protection or vice versa.

Finally, we draw attention to Fig. 4.5a, where the gray circle is drawn to illustrate that two curves, each from a different parametric category, intersect. As observed, in lower traffic regimes of this gray circle, the PMF due to $(P_{\mathrm{SU}}^{\mathrm{misSync}} = 0.3, N = 8, P_{\mathrm{PU}}^{\mathrm{Active}} = 0.45)$ corresponds to a network that outperforms the one due to $(P_{\mathrm{SU}}^{\mathrm{misSync}} = 0.6, N = 2, P_{\mathrm{PU}}^{\mathrm{Active}} = 0.15)$. Passing through this critical point, the latter takes turn and suppresses the former. There is no need to emphasize that an upper PMF curve always indicates a less piled-up queue compared to a lower one. This behavior advocates an existing trade-off between the three quantities, $P_{\mathrm{PU}}^{\mathrm{Active}}$, $P_{\mathrm{SU}}^{\mathrm{misSync}}$ and N, which represent the PU's activity in a channel, the SU's mis-synchronization, and the total number of channels, respectively.

4.3.2 Statistical Analysis and Heavy-Regime Characteristics

As shown in Fig. 4.5b, the variability in the PMF curves triggered by any change in the operating environment are really diverse. The statistical analyses show that the impact of such changes is not just restricted to the SU's average queue length but it influences all the higher moments and, worse, the tail of the distribution. In this direction, we conducted extensive curve-fitting experiments on the secondary's numerical PMF obtained from the CTMC in Fig. 4.4a. We found that different levels of channel occupancy result in the PDFs decaying faster than an exponential PDF (underutilized), or slower than a power-law PDF (over-utilized), such as *Pareto* with shape parameter $\alpha < 1$. For instance, for $N = 5$, $P_{\text{PU}}^{\text{Active}} = 0.05$ and $M = 10$, we changed the SU's input traffic rate λ to be able to catch a wide range of channel utilization, i.e., $\rho \in [0.04, 0.99]$, where $\rho = (\lambda M/\mu N)\left((\eta + \zeta)/\eta\right)$ (as defined in (4.2)), and estimated the tail of the distribution with *Pareto* and *Exponential* PDF models. We matched the curve-fitted PDF cx^{-d} with the standard definition for a *Pareto* distribution, i.e.,

$$f_x(x) = \begin{cases} \dfrac{\alpha x_0^{\alpha}}{x^{\alpha+1}}, & x \geq x_0 \\ 0, & x < x_0. \end{cases} \tag{4.30}$$

The equivalent standardized shape (α) and scale (x_0) parameters are shown in the rightmost two columns of Table 4.2.

As observed in low-utilization regimes, the exponential distribution with moderate values of b is a reasonable tight fit. But as ρ grows (either due to larger number of SUs, larger $P_{\text{PU}}^{\text{Active}}$, etc.), the exponent gets smaller rendering the finite-average finite-variance *Pareto*$(\alpha \in (2, \infty))$ with its heavy-tail properties a more matching model. Drastically, for $\rho > 0.6$ (*Pareto* fit) the variance of queue length obtained through the model is infinite $(\alpha \in (1, 2])$. Once the utilization factor is $\rho \approx 1$, the queue becomes unstable and the average queue length grows boundlessly, which agrees with the immediate declination of

Table 4.2: *Pareto* and *Exponential* fits for the tail of SU's queue length PMF for different values of the utilization factor ρ.

ρ	Exponential: $f(x) = ae^{-bx}$		Power-law: $f(x) = cx^{-d}$		
	a	b	c	d	*Pareto* estimation: (x_0, α)
0.99	0.0004	0.0012	0.0006	0.14	n/a
0.8	0.004	0.0182	6.617	2.03	$(6.13, 1.03)$
0.5	0.0004	0.0365	1102.1	4.06	$(6.84, 3.06)$
0.2	0.00002	0.088	4.54 E10	9.76	$(12.85, 8.76)$
0.04	0.00003	0.39	1.57 E63	43	$(29.24, 42)$

α to below 1 as shown in Table 4.2. These observations demonstrate that the CTMC model presented in this section can capture long-range dependency and heavy-tailedness behavior as well as short-range dependency and light-tailedness of secondary's queue length.

In another set of experiments, we evaluated the coefficient of variation (CoV) of the secondary's queue length vs the utilization factor ρ and the number of channels N. Concisely, the CoV is defined as the ratio of the standard deviation of the SU's queue length to its mean. Each curve in Fig. 4.6 corresponds to a CRN of $M = 10$ SUs contending to access a number of frequency channels with utilization factor ρ continually raising between the two extremes (triggered by magnifying the arrival rates or the PU's activity level). The results are striking. Indeed, regardless of the choice of parameters, the CoV profiles are alike as they first rise and then decline. Nonetheless, the final trends are always a skyrocketing increase, revealing how fast or slow the standard deviation and average queue length change when the network utilization increases. Obviously, SUs working in lower (heavier) traffic regimes, larger (smaller) N and(or) smaller (larger) ρ, have $\mathrm{CoV} \cong 1$ $(\mathrm{CoV} >> 1)$, a fact that confirms our former deduction on exponential (power-law) curve-fitting. On the other side, even though for smaller values of ρ less variability is obtained at larger N, it is not clear why the trend is reversed for higher values of ρ, which could be another observation

to previous conjectures on the ineptness of the CoV as a reliable indicative of a large-scale variability, burstiness, etc.

4.3.3 Closed-form Analysis

Unfortunately, and as explained before, the 2D CTMC in Fig. 4.4a does not provide exact analytical expressions even for the lower degree statistical moments of the SU's queue length. On the contrary, the secondary's average queuing delay and queue length are the quantities that ascertain how good the MIMC-CRN performs averagely and, as such, it is of principal importance to have estimates or bounds defined for them. For this reason, and in order to make the proposed model usable for more practical cases, approximations for the above-mentioned quantities are found in exact closed-form in the rest of this subsection.

The point of departure is (4.13), whereby the marginal PMF of the number of empty channels for an SU was derived. The procedure is as follows. First, let us suppose that the number of servers (empty channels) available to the SUs is fixed. In this situation, each SU can be modelled as a $\mathrm{M/M/n}$ queue, where the average queue length ($\overline{\mathrm{L}}$) is found from (4.31) as derived in [39],

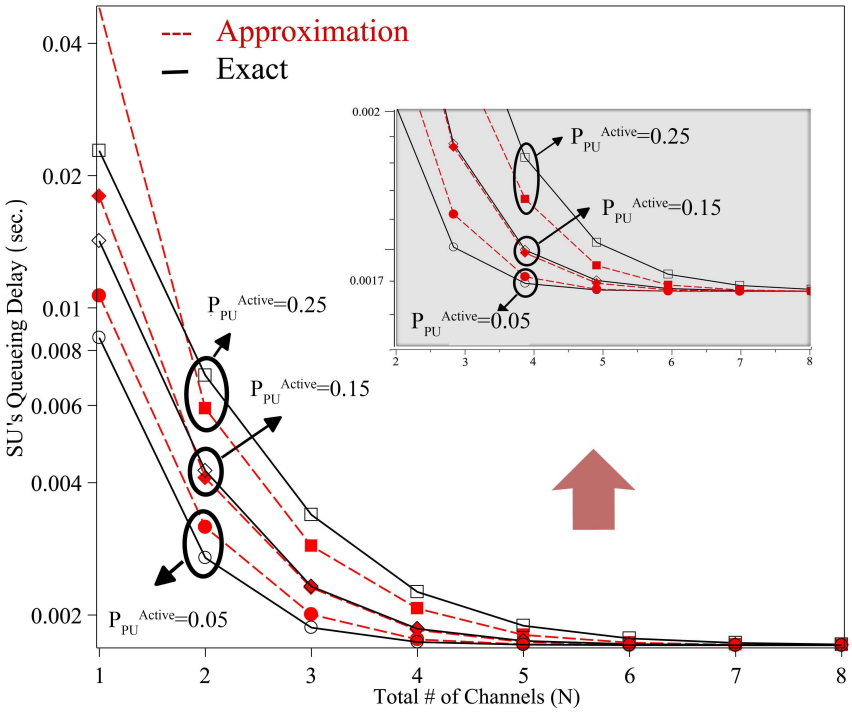
$$\overline{\mathrm{L}}_{|n} = n\rho + \frac{\rho\,(n\rho)^n}{n!}\frac{P_0}{(1-\rho)^2}, \tag{4.31}$$

where

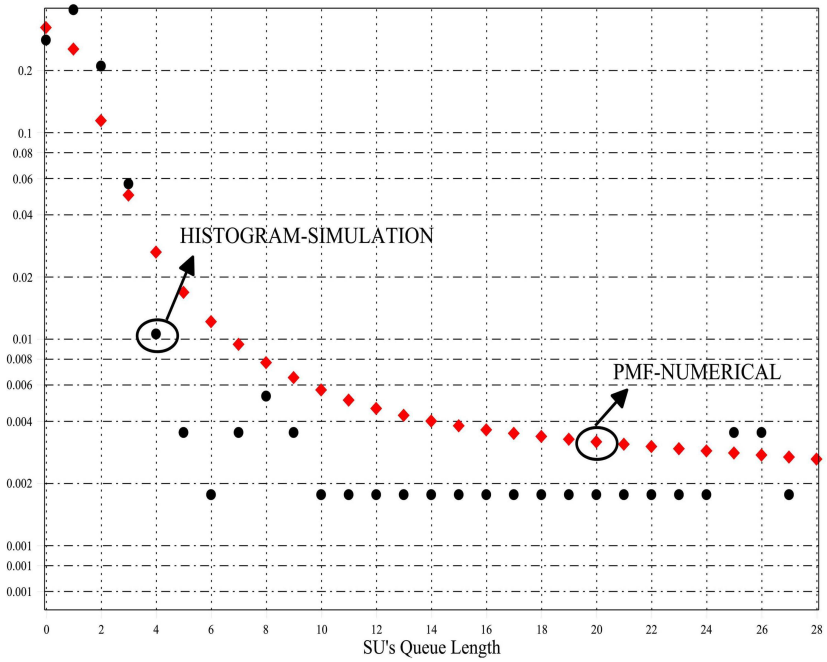
$$P_0 = \left(\sum_{j=0}^{n-1}\frac{(n\rho)^j}{j!} + \frac{(n\rho)^n}{n!}\frac{1}{1-\rho}\right)^{-1} \tag{4.32}$$

and $\rho = \lambda/n\mu$. Next, we use $n \sim \mathrm{Binomial}\,(N, 1-P_{\mathrm{B}})$ obtained in (4.13) to probabilistically remove the aforementioned condition on k,

$$\overline{\mathrm{L}} = \sum_{n=0}^{N}\overline{\mathrm{L}}_{|n}\cdot\mathrm{Binomial}\,(N, 1-P_{\mathrm{B}}). \tag{4.33}$$



(a) Average SU's queuing delay vs. number of channels: comparison between numerical and closed-form results.



(b) PMF of secondary queue occupancy: comparison between simulation and numerical results

Figure 4.7: Validation results.

Subsequently, we carry out curve-fitting and some mathematical manipulation to improve the precision of this estimation. Hence, we obtain

$$\overline{\mathrm{L}} = \frac{\lambda}{\mu} + 1.5\frac{\lambda}{\mu} \cdot \frac{1}{a-b} \cdot \frac{\left(\frac{\lambda}{\mu}\right)^{a-b}}{(a-b)!} \cdot \frac{P_0}{\left(1-\frac{\lambda}{\mu a}\right)^2}, \tag{4.34}$$

where P_{B} is given by (4.20) and P_0, a and b are given by

$$\begin{aligned} P_0 &= \left(\sum_{j=0}^{a-1} \frac{\left(\frac{\lambda}{\mu}\right)^j}{j!} + \frac{\left(\frac{\lambda}{\mu}\right)^{a-(1-P_{\mathrm{B}})}}{(a-(1-P_{\mathrm{B}}))!} \cdot \frac{1}{\left|1-\frac{\lambda}{\mu a}\right|} \right)^{-1}, \\ a &= (1-P_{\mathrm{B}})N, \\ b &= \sqrt{aP_{\mathrm{B}}}. \end{aligned} \tag{4.35}$$

Finally, using *Little*'s theorem, the total delay $\overline{D}$ that SU experiences is obtained through

$$\overline{D} = \frac{\overline{\mathrm{L}}}{\lambda}. \tag{4.36}$$

Due to space limitation, we only apply this idea to derive the first moment but the same idea can be leveraged to derive approximations for π_ℓ in (4.29) and higher moments. Fig. 4.7a shows the accuracy of such approximation for the queuing delay considering different total numbers of channels (N) and $P_{\mathrm{PU}}^{\mathrm{Active}}$. As is evident from this figure, approximate curves are in good match with the exact curves in value, slope and convexity. Once more, the scale is chosen logarithmic to reflect the precision with higher resolution.

It should never be neglected that satisfying the SC in (4.2) is obligatory for all SUs in order for the given analysis to be effective. From a practical point of view, since the value of P_{B} cannot be altered by SUs, then the number of admitted SUs and their corresponding

traffic intensities should never be chosen in such a way that the SC values $\lambda_i < N(1 - P_{\mathrm{B}})\mu_i/M, \forall i = 1 \cdots M$ are violated.

4.4 Simulation Validation

For the purpose of validating our analytical derivations, a discrete event simulator (DES) was built for the network model shown in Fig. 4.1a and its corresponding queue representation in Fig. 4.2c, which is a more realistic representation of Fig. 4.2a.[7] Fig. 4.2c shows an SU's queue with N servers where each server is shared with a primary network queue. These primary queues act independently and have pre-emptive resume access privilege over their corresponding secondary queues, thus emulating a multi-channel cognitive scenario.

The values chosen in the simulations are as follows: $N = 8$, $\lambda_p = 200\ \mathrm{pk/s}$, $\lambda_s = 1200\ \mathrm{pk/s}$, $\mu_s = 600\ \mathrm{pk/s}$ and $\mu_p = 1000\ \mathrm{pk/s}$. In this simulator, the arrival and departure trends can be drawn from any generic stochastic process. Nonetheless, in order to keep the conformity between the simulator and developed model, the *Poisson* input/output processes are adopted in the simulator.

As an imperative and comprehensive quantity, the unfinished work in a queue is probed, which is a step function of time representing the instantaneous occupancy (length) of a queue. Given this definition, the unfinished work for the secondary queue is plotted in Fig. 4.8. Moreover, the validation results comparing the histogram of the SU's queue occupancy with the relevant PMF obtained numerically by solving the CTMC are presented in Fig. 4.4a. As shown in Fig. 4.7b, a reasonably tight match between the simulation and numerical results exists. It is important to note that when the PMFs match, then, unarguably, other moments do as well, which obviates the need to go over the examination

[7] By equating the secondary queue's input/output rates (λ, μ) in Fig. 4.2b to those of Fig. 4.2c (μ_1, λ_1), associating the busyness probability $P_{\mathrm{PU}}^{\mathrm{Active}}$ to the primary input/output rates to have $P_{\mathrm{PU}}^{\mathrm{Active}} = \mu_1/\lambda_1$ on all primary queues, and leaving all the other settings alike, the models will be equivalent.

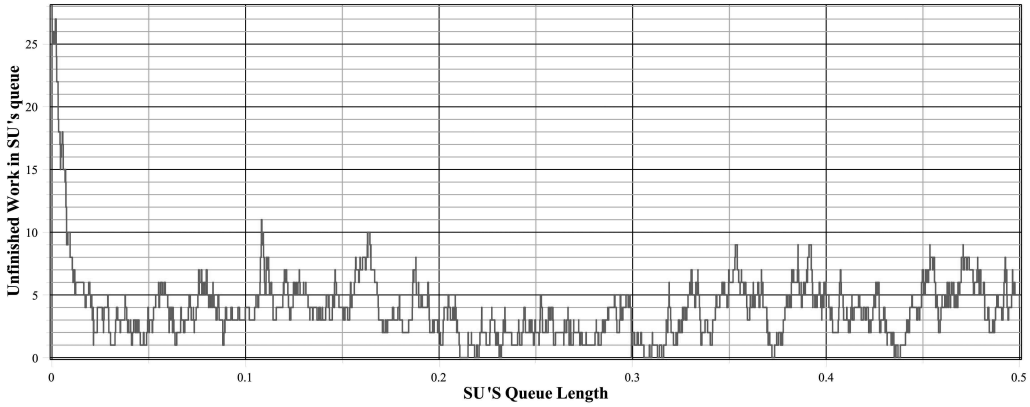


Figure 4.8: The unfinished work in the secondary queue.

of single moments one by one.

This summarizes the modeling and analytical study of multi-channel CRN. The next chapter looks at the problem at hand from a new perspective. In fact, while the network modeling and analysis is an endeavor to quantify the qualities of a CRN with predefined features, characteristics, and procedures, the process is reversed in the realm of mechanism design (the subject of this thesis in Part II) where features and characteristics are to be determined to meet certain network qualities.

4.5 Summary

In this chapter, a detailed model was proposed to analyze the performance of multi-interface multi-channel cognitive radio networks. The queue representation was established as a set of interacting queues with respect to the SUs, where the amount of available resources is subject to variability due to the presence or absence of PUs. Considering that channels (N in total) of non-overlapping widths are statistically independent, SUs were modeled as a set of $\mathrm{M}/\mathrm{M}/\mathrm{N}$ interacting queues, with variable number of servers (N), where each server represents an empty channel. By properly tuning the service rates, the coupling among SUs was removed, which led to the representation of each SU with a 2D continuous state

Markov chain. The direct result of this representation was the derivation of a generalized stability condition for each SU, which can directly be used to decide on the maximum allowed input traffic rate, maximum number of admitted SUs, etc. Moreover, in spite of many previous works, the model presented in this chapter takes into account realistic PHY facts such as mis-synchronization, sensing faultiness, channel fading, etc. Among the other findings of the proposed modeling is the joint and marginal probability density functions of the said chain from which any higher moment can be directly extracted, as well as a statistical analysis of the tail behaviour and variability of the secondary's queue length. Finally, closed-form expressions for the average service delay and queue length were obtained.

Part II

Mechanism Design for Multi-Channel Cognitive Radio Networks

In the previous chapter, the approach adopted was bottom-up. Having pronounced the gravity of the modeling task in Section 1.3.1, we, established well-defined, dynamic, and precise models for MC-CRNs with *pre-designed* components. Our aim was to comprehend the behavior of the MC-CRNs in terms of quantities such as delay, queue length, and throughput (or utilization factor). The erected models unveiled different facets of MC-CRNs, such as the degree of influence of spectrum sensing imprecision on the aforementioned metrics, the QoS trade-off between primary and secondary classes, the heavy-tailedness of the secondary's queue length distribution, existence of a boundary SC and achievability region and so forth.

Up until now, nothing has been said to address questions like "*what if the observed performance does not meet the service requirements?*". Among many causes that might lead to this situation, the source of the problem is oftentimes a bad, inefficient, non-optimal, or incompatible design of one or more components somewhere in the network. Such component may be a protocol, mechanism, or algorithm in any of the architectural layers of the network protocol stack, with more severe degradation when the problem stems from lower layers. When this is the case, modeling and performance analysis can do nothing but reveal the flaws, if we are lucky. At this point, architects and engineers hustle to root the problem and redesign the network spending so much time, effort, and assets, a circumstance that could have been avoided, had there been deployed the proper mechanisms in the first place. This constitutes the research theme of the thesis in this chapter, meaning a top-bottom approach that takes us from service requirements to defining the bolts and nuts of new components. Given the title *Mechanism design*, and keeping in mind the substantial and far-reaching role of spectrum in the context of CRNs, the focus of the second part of the thesis will be on designing efficient spectrum management mechanisms. As the main three identified spectrum management functionalities, spectrum sensing and decision, and spectrum sharing will be at the center of attention in this part.

Chapter 5

Spectrum Sensing, Decision and Mobility[1]

As explained in Section 3.1, IEEE 802.22 is the outcome of a decade of painstaking efforts in the research community, which aimed at diminishing the rural-urban divide by extending the reach of technology to remote areas. At this time, CR is literally at the doorstep of the market, lingering for several causes; first and foremost, CR is more than a technical concept. It is an entanglement of technical matters and legal issues (due to the involvement of spectrum license owners). The involvement of such a dual ownership model raises the stake much higher for the tech sector forcing them to ponder this solution more cautiously. In other words, the tech sector is uncertain whether the reliability that the spectrum sensing interface brings into play would be sufficient to protect them against unknown consequences. Second, standardization in the area of CR is new and untested; IEEE 802.22 is yet at the inception point with some ambiguities and clauses that are explicitly left as open

[1]N. Tadayon and S. Aïssa, "A multi-channel spectrum sensing fusion mechanism for cognitive radio networks: Design and application to IEEE 802.22 WRANs", pp. 359–371, vol. 1, no. 4, *IEEE Transactions on Cognitive Communications and Networking*, Mar. 2016.

[1]N. Tadayon and S. Aïssa, "A learning-based distributed spectrum sensing mechanism for IEEE 802.22 wireless regional area networks", IEEE GLOBECOM, San Diego, CA, Dec. 2015, pp. 1–6.

issues [8]. Surprisingly, much of these open issues are in the area of spectrum sensing. This is an indication of one key point: spectrum sensing and decision operations are to be improved if CR is hoped to be adopted as an enabling 5G technology.

Given that the choice of a ubiquitous distributed sensing mechanism is among these open issues, limited effort has been dedicated whereof [81–83] can be mentioned.[2] The standard proposals for the sensing fusion mechanism are basically AND, OR and VOTING logic rules. Given the significance of the spectrum sensing process for networks operating in VHF/UHF bands, w.r.t. reliability and accuracy, these combining rules are plainly incompetent. In particular, they suffer from unbalanced performance, where a low false-alarm rate comes at the cost of an unbearably high mis-detection rate (and vice versa), and exhibit rigid performances, i.e., are incapable of adjusting performances to varying situations. Moreover, due to the lack of feedback from sensing output back to the input to stabilize the performance, operational instability can become a major issue, manifesting itself in the form of weak performance in low SNR regimes. Therefore, the need for an efficient distributed sensing fusion mechanism that evades the above shortcomings, complies with the standard directives, and can be easily embedded in the system, is entirely felt.

Motivated by these facts, in this section, a sensing fusion mechanism, termed *MC-LDS*, is proposed for its dependence on learning through observation fostered by a reward-penalty rationality. For the latter attributes, MC-LDS showcases a stable and self-trained behavior as well as robustness in detecting the faulty sensing reports. The other salient characteristics of MC-LDS are improved performance in all traffic regimes, fairness (reduced false-alarm/mis-detection gap), adjustability (operability with several degrees of freedom) and bandwidth efficiency (increased transmission opportunities for more CPEs). Simulation results and comparisons unanimously prove that MC-LDS surpasses the fusion mechanisms proposed in the IEEE 802.22 standard. Furthermore, MC-LDS can be integrated

[2]Hereafter, terms "sensing-data fusion", "decision combining" and "distributed sensing" are used interchangeably as they represent the same idea.

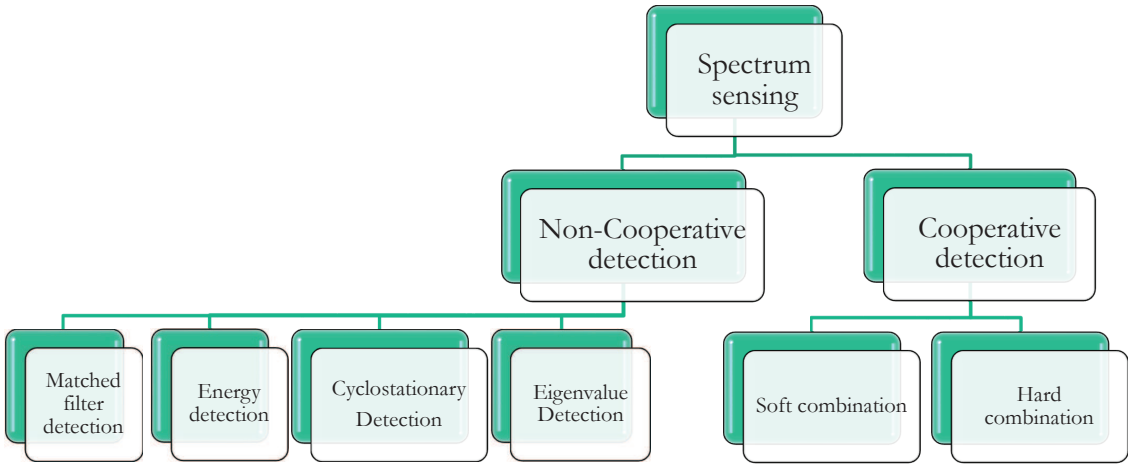


Figure 5.1: Known spectrum sensing methods.

to boost the sensing performance of other promising technologies and standards such as *White-Fi* (IEEE 802.11af WLANs), wireless personal area networks (WPANs), ZigBee (IEEE 802.15), cognitive WiMax (IEEE 802.16h) and the IEEE 1900.6b standard emerged to support spectrum databases using spectrum sensing information.

5.1 Spectrum Management in IEEE 802.22

In this section, some of the recent investigations on spectrum sensing for WRANs are discussed, followed by the most relevant spectrum management and scheduling functionalities whose knowledge is essential for understanding the proposed MC-LDS technique elaborated in Section 5.2.

5.1.1 Literature Review

Spectrum aimed for the operation of WRANs [8] hugely overlaps with UHF/VHF TV bands in the range of $54 - 862$ MHz [3]. Despite the fact that these bands are often empty and temporally underutilized, in time and space, they have been conceded to incumbent

operators for interference-free operation, which legally inhibited others from using them. With the new insights about what has provoked the spectrum inefficiency, now WRAN operators are able to transmit on these bands as far as they do not cause harmful interference to the incumbent service. This requires reliable channel sensing prior to transmitting on a channel. To enhance the reliability and accuracy of the a sensing function, collaborative spectrum sensing has been proposed and shown to alleviate the destructive effects of the hidden and exposed-terminal problems [46] and the channel fading [84]. In fact, these adversaries can sometimes be so degrading that they render the communications over a channel practically futile [85].

Collaborative sensing can be realized in a centralized or distributed fashion [86]. A different classification is based on the nature of the sensing information being exchanged, resulting in a soft/hard decision combining (SDC/HDC). In fact, while the sensing decision in HDC is binary (with 0/1 indicating the absence/presence of incumbent transmission), in SDC any test statistics may be outputted. While static approaches such as AND, OR, VOTING, and linear quadratic combination [87] are among HDC methods, an example of SDC can be found in [85]. It has been shown in [88, 89] that SDC methods often output more accurate sensing decisions than HDC methods. Fig. 5.1 summarizes the existing spectrum sensing methods.

Another way to create a different decision combining methods is to choose a different test statistic. In particular, while opting for descriptive statistics, as in OR, AND, VOTING rules, is one approach, one could define a test metric based on inferential statistics. In this latter case, the choice of the following likelihood function has been of interest in recent years

$$L = \prod_{i=1}^{m} \frac{p(d_i|\mathcal{H}_1)}{p(d_i|\mathcal{H}_0)}, \tag{5.1}$$

where d_i is the sensing result of the i^{th} reporting sensor, m is the total number of sensors

and $p(d_i|\mathcal{H}_0)$ and $p(d_i|\mathcal{H}_1)$ are a sensor's associated posterior probabilities conditioned on the null-hypothesis that the channel is empty ($\mathcal{H}_0$) and the alternative hypothesis that the channel is occupied ($\mathcal{H}_1$) by the primary network, respectively. Indeed, classes of decision combining methods emerged, such as Bayesian detection, Neyman-Pearson detection, sequential probability methods, etc., that uses the likelihood function in (5.1) at their cores. Though the performance of the mechanisms in this class are comparably better than their rivals, they still suffer from some important flaws due to the following reasons: First, the choice of threshold value to which this likelihood function is to be compared with is often arbitrary. Second, the likelihood function in (5.1) requires the a-priori knowledge of the probabilities $p(d_i|\mathcal{H}_0)$ and $p(d_i|\mathcal{H}_1)$, which does not fully abide by practical considerations. Third, these algorithms seems to be slowly responding and less sensitive to changes, mainly due to their reliance on non-instantaneously reproducible probabilities $p(d_i|\mathcal{H}_0)$ and $p(d_i|\mathcal{H}_1)$.

The last class of decision combining mechanism is based on the reputation evaluation models. In these mechanisms, malicious sensors that consistently output wrong sensing information are detected and eliminated from the decision process. Procedures that are used in this regard are usually pre-filtration, data fusing, reputation vector updating, etc. The investigations in this area have been less extensive and are more recent, with good examples such as [81, 90, 91]. Analogous to the previous class, the arbitrarily chosen threshold is a source of the problem in some of the reputation-based methods.

Since the release of the IEEE 802.22 WRAN standard, limited effort has been dedicated to one of the most fundamental open issues pointed out in [8], namely, sensing-data fusion (see e.g., [81–83]). In fact, while the two WRAN sensing fusion mechanisms proposed in [81] operate based on the idea of exploiting the sensors' confidence metrics, [82] presents a sensing clustering mechanism for in-band sensing (IBS). The latter work also discusses the issue of "how often to sense the channel?" in WRANs and proposes a sensing scheduling mechanism for such. Despite its implementation complexity, the approach

in [82] was concluded to alleviate some of the shortcomings associated with the "OR" sensing fusion mechanism. Finally, the core clustering idea in [83] is the same as [82], except that the former puts emphasis on security aspects of the mechanism by using attack-tolerant collaborative sensing.

As delineated in the previous chapter, the collaborative sensing fusion mechanism proposed in the standard [8] requires CPEs (generally, referred to as CR nodes or SUs) to first sense their operating channels individually and, then report the results to the BS, intermittently. The sensing tasks should be carried out during synchronized quiet periods (QPs), whereby all CPEs shall go silent to reduce false-alarm events. The network-wide and periodic coordination of QPs is one of the responsibilities of the WRAN BS. In this regard, different QP schedules may be required depending on the number of *operating channels* being used by CPEs within a cell. Also, QPs can be overlapping or non-overlapping. Once the sensing outcomes of each CPE on the status of an operating channel is reported to the BS,[3] it is the responsibility of this BS to make the decision upon the status of the sensed channels and to administer any respective action. To enhance the reporting reliability, [8] devises different notification and reporting mechanisms including contention-based (or CDMA-based) UCS, bulk measurement messaging (BLM), as well as reporting through CPEs' allocated US bursts.

The rest of this section details the relevant WRAN functionalities, whose knowledge is needed for understanding the sensing mechanism proposed in this chapter. Before further progression, the following notation is used for the rest of this chapter: subscripts $_i$, $_j$, $_k$ represent the CPE index, WRAN cell index (in a network with $\mathfrak{M}$ cells) and the operating channel index (in a network with $\mathfrak{B}$[4] channels), respectively. Table 5.1 summarizes the acronyms and notations used, in this chapter.

[3] Hereinafter, sensor and CPE may be used interchangeably anywhere the context is channel sensing.

[4] Note that symbol N has been used in the previous chapter to represent the number of channels as well.

Table 5.1: Main acronyms and symbols in this chapter.

Acronym	Expanded Form	Acronym	Expanded Form
CH	channel	DCL	disallowed CL
CL	CH list	LPS	local priority set
QP	quiet period	RB	resource block
OCL	operating CL	CBP	coexistence beacon protocol
PCL	protected CL	SCH	superframe control header
BCL	backup CL	CCL	candidate CL
IAR	incumbent activity rate	IAF	incumbent alteration freq.

Symbol	Definition	Symbol	Definition
i, j, k	CPE, WRAN, CH indices	$P_{\mathrm{MD},i,j}^{(t),k}$	misdetection probability
$\mathfrak{M}$	# cells	$P_{\mathrm{FA},i,j}^{(t),k}$	false-alarm probability
$\mathfrak{B}$	# CHs	$d_{i,j,k}^{(t)}$	local sensing decision
Θ_i	logistic estimator parameter	${\chi_{j,k}^{(n)}}^2$	Pearson's test statistic
$^{(t)}$	QP index	$m_{j,k}$	# sensors in WRAN$_j$
$\mathcal{H}_{1,k}^{(t)}$ ($\mathcal{H}_{0,k}^{(t)}$)	alternative (null) hypotheses	$\mathcal{N}_i$	# training samples
$\mathcal{S}_{i,j,k}^{(t)}$	received power	$\overline{h}_{i,j,k}^{(t)^{\mathrm{Rep}}}$	reporting CH gain
$\mathcal{S}_{j,k}^{\min}$	power threshold	$\mathcal{B}_i$	# signal samples
$\mathcal{S}_N$	noise power	$R_{j,k}^{(t)}$	database reading
$L_{i,j,k}^{(t)}$	reward-penalty score	$D_{j,k}^{(t)}$	central decision
$\alpha_{j,k}$	temporal discount	$N_{j,k}$	historic count
$w_{i,j,k}^{(n)}$	confidence metric	$X_{i,j,k}^{(n)}$	CH indicator
$\mathcal{Z}_j^{(t)}$	factual CH status		

5.1.2 Channel Management

Channel Switching

Once the BS in a WRAN$_j$ cell concludes that one of its operating channels, say CH$_b$, sensed by a subset of its associated CPEs is busy, CH$_b$ and its two adjacent channels, CH$_{b\pm1}$, shall be vacated, removed from the *operating channel list* (OCL) and added to the *protected channel list* (PCL), which is then followed by the high priority channel switching mechanism [8, Section 7.22]. For this switching to occur smoothly, the BS assigns a timer with

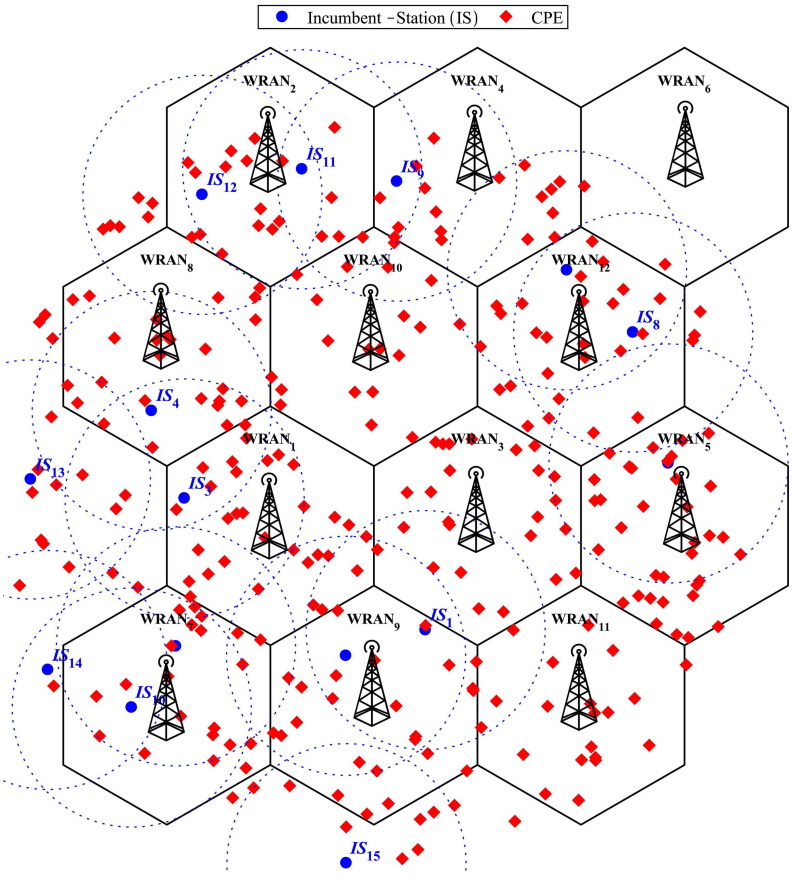


Figure 5.2: An example of a WRAN with $\mathfrak{M} = 12$ coexisting cells (WRAN_j) and CPEs uniformly distributed in the network area. A total of 15 incumbent (primary) stations are shown.

long enough duration to $\text{CPE}_{i,j}$ in WRAN_j to ascertain that they are all prepared for the aforementioned switching operation, provided that such duration shall be no longer than the maximum allowed channel moving time. Subsequently, a channel from the *backup channel list* (BCL) of WRAN_j is selected and the switching procedure gets initiated. Because an operating channel can become obsolete at any time during the operation of the network, it is necessary to maintain sufficient number of channels as backup. Both operations (that is, monitoring the status of operating channel and non-operating channels) are mandatory for smooth and continued network functioning and are possible through sensing. While the mechanism for sensing operating channel(s) is called in-band sensing (IBS) [8, clause

10.3.3], out-of-band sensing (OBS) [8, clause 10.3.1] is performed on protected, candidate, and backup (all non-operating) channels [8, clause 10.2.3].

The importance of OBS is due to the vital need to always maintain sufficient number of channels in the OCL, through a closed-loop mechanism and by giving fair opportunities to all channels to return to the OCL as the incumbent transmission vanishes. That requires that some CPEs be assigned to perform OBS on channels in BCL, PCL and the *candidate channel list* (CCL), which does not need to be necessarily done during network-wide QPs.

Notwithstanding the similarity between OBS and IBS reporting and decision making procedures, these two mechanisms differ in the way the corresponding channel lists are updated. More specifically, a channel in PCL that is sensed idle during OBS (say CH_b) is removed from this list and added to the CCL. This channel may be directly added from PCL to BCL only if it has remained incumbent-free for no less than 30 sec, with inter-sensing intervals not exceeding 6 sec each. On the contrary, CH_b remains in PCL as long as it is sensed busy. In the former case, and after having been added to the BCL, CH_b either remains still in BCL for as long as the OBS flag is not raised (otherwise it shall return to the PCL) or enters the OCL upon the channel switching event (which is triggered when an operating channel from the OCL is added to the PCL).

Finally, *disallowed channel list* (DCL) bears the indices of channels that are precluded from use by incumbents and so can be freely used by the network. Table 5.2 illustrates a state of the simulated WRAN shown in Fig. 5.2 where 12 coexisting cells operate on 10 channels. The channel transitions between lists are implemented according to the above rules, which can be compactly represented in the form of the transition diagram in Fig. 5.3.

List Updating

When operated in coexistence mode (see Section II.B), neighboring WRAN cells may share similar channel pools. In such a situation, changes in BCL and CCL of WRAN_j may be

Table 5.2: A complete dynamic channel list (CL) for a network of twelve coexisting WRANs extracted at a random time from the simulator. $\{\}$ indicates an empty list.

	Operating CL	Disallowed CL	Backup CL	Protected CL	Candidate CL
WRAN_1	$\{\text{CH}_4\}$	$\{\}$	$\{\}$	$\{\text{CH}_2, \text{CH}_{10}\}$	$\{\}$
WRAN_2	$\{\text{CH}_8\}$	$\{\text{CH}_{10}\}$	$\{\text{CH}_4, \text{CH}_6\}$	$\{\}$	$\{\}$
WRAN_3	$\{\text{CH}_4\}$	$\{\}$	$\{\}$	$\{\text{CH}_2\}$	$\{\text{CH}_5\}$
WRAN_4	$\{\text{CH}_{10}\}$	$\{\text{CH}_{10}\}$	$\{\}$	$\{\}$	$\{\}$
WRAN_5	$\{\text{CH}_2\}$	$\{\text{CH}_5\}$	$\{\}$	$\{\}$	$\{\}$
WRAN_6	$\{\}$	$\{\text{CH}_1 \cdots \text{CH}_{10}\}$	$\{\}$	$\{\}$	$\{\}$
WRAN_7	$\{\text{CH}_7\}$	$\{\}$	$\{\}$	$\{\}$	$\{\}$
WRAN_8	$\{\text{CH}_1\}$	$\{\text{CH}_1\}$	$\{\}$	$\{\}$	$\{\}$
WRAN_9	$\{\text{CH}_9\}$	$\{\}$	$\{\text{CH}_6\}$	$\{\text{CH}_2, \text{CH}_{10}\}$	$\{\}$
WRAN_{10}	$\{\text{CH}_3\}$	$\{\}$	$\{\text{CH}_7\}$	$\{\}$	$\{\text{CH}_8\}$
WRAN_{11}	$\{\text{CH}_7\}$	$\{\}$	$\{\text{CH}_5\}$	$\{\text{CH}_{10}\}$	$\{\text{CH}_6\}$
WRAN_{12}	$\{\text{CH}_5\}$	$\{\text{CH}_5\}$	$\{\}$	$\{\}$	$\{\}$

triggered for two reasons: (i) by a change in the status of the channels sensed in WRAN_j (as explained above), (ii) or by any change in the channel lists of WRAN_j's neighboring cells [8, subclause 10.2.3]. This is due to the restrictions imposed by [8] on the choice of the operating and backup channels in neighboring cells. For instance, neighboring cells are not initially allowed to use similar operating and backup channels at the same time. Therefore, in case of channel swapping in the lists of WRAN_j, affected cells may have to update their PCL and CCL lists accordingly. For the sake of updating lists, WRAN_j uses the following *local priority sets* (LPS),

$$\begin{aligned} \text{LPS}_j{}^{\text{I}} &= \{\text{BCL}_j \cup \text{CCL}_j\} \setminus \bigcup_{l \in (-j)} \{\text{CCL}_l \cup \text{BCL}_l\} \\ \text{LPS}_j{}^{\text{II}} &= \{\text{BCL}_j \cup \text{CCL}_j\} \setminus \bigcup_{l \in (-j)} \{\text{OCL}_l\} \\ \text{LPS}_j{}^{\text{III}} &= \bigcup_{l \in (-j)} \{\text{OCL}_l\} \end{aligned} \quad , \tag{5.2}$$

where the subscript $_l$ is to indicate the association with WRAN_l. The operators $\setminus$ and $\cup$ represent the set difference (exclusion) and union operators, respectively, whereas $(-j)$ sig-

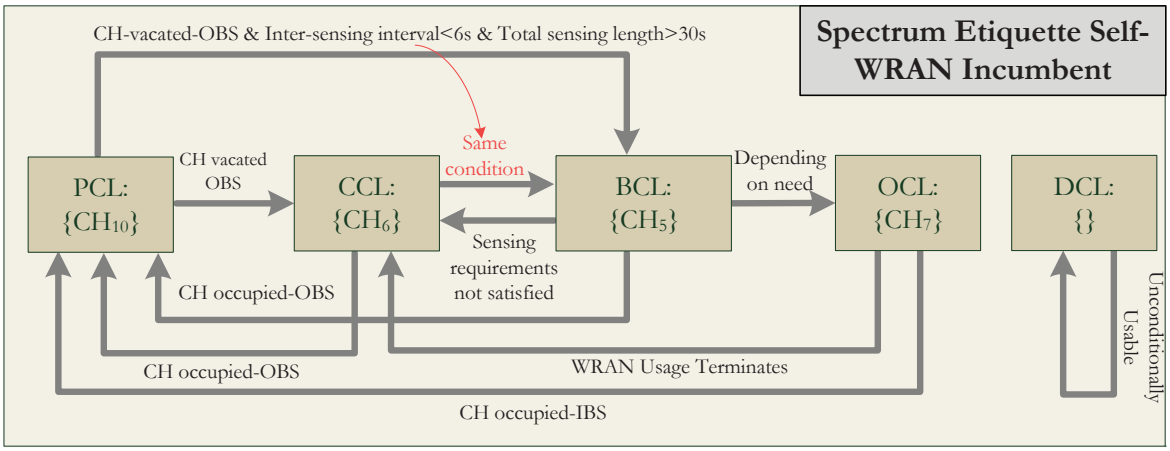


Figure 5.3: IEEE 802.22 channel transition diagram for WRAN_{11} in the simulation scenario of Fig. 5.2.

nifies that the union operation is to be carried out over all the neighboring cells of WRAN_j. Since the updating decisions are all to be made centrally, the BS utilizes the above sets by their priorities, meaning ${\text{LPS}_j}^{\text{III}}$ is used for list updating only when ${\text{LPS}_j}^{\text{I}}$ and ${\text{LPS}_j}^{\text{II}}$ are both empty, whereas ${\text{LPS}_j}^{\text{II}}$ is used only if ${\text{LPS}_j}^{\text{I}}$ is empty (otherwise, ${\text{LPS}_j}^{\text{I}}$ will be utilized). This updating mechanism is illustrated in Fig. 5.4. In the end, the BS uses the SCH (transmitted every 16 msec) to update the local OCLs of CPEs, and uses the *downstream channel descriptor* (DCD) to update the local PCLs and CCLs of CPEs [8].

5.1.3 Normal Mode vs. Coexistence Mode

The IEEE 802.22 standard allows two modes of operation, namely, *normal* and *coexistence*. Unlike the normal mode of operation in which only a single WRAN transmits in all the resource blocks or neighboring cells use totally isolated frequency bands, in the coexistence mode [8, section 7.20] one or more channels may be shared among neighboring WRAN cells. Without delving into details, the CBP is responsible to achieve synchronization among operating cells by transmitting synchronizing control packets during self-coexistence windows (SCW).

The interference-free requirements in IEEE 802.22 WRANs not only necessitate that a channel discovered as busy and its two adjacent channels get immediately vacated to avoid inter-carrier interference but also impose constraints on the choice of the operating channels in neighboring WRAN cells (in the coexistence mode). More narrowly, the CBP transmissions are utilized during the SCW and act as a coordinator by prohibiting neighboring cells from using a similar operating channel or by settling a compromise.

5.1.4 Quiet Period Management

One of the prime challenges of IBS is its uncompromising requirement to silence all CPEs for a span of time (i.e., QPs). Such an obligation can result in QoS degradation as no data payload is transmitted in those interims. In the particular situation when inter-frame sensing is obligated, which lasts for an entire superframe duration, the QoS damage can sometimes be even more severe [8, clause 7.21.1]. On the other hand, the OBS execution does not require network-wide QP scheduling as the target channels are not used for communications, which ultimately burdens less cost on the network. To deal with this inherent shortcoming and meet channel detection time requirements [8, section 7.21], IBS in WRAN is carried out in two different scales: *intra-frame* IBS and *inter-frame* IBS. This two-stage QP management enables the network to adjust the sensing length and repetition rate, as per the needs. The sensing procedure always starts with the more frequent, faster (lasting for a frame length of $\cong$ 10ms) and easier intra-frame sensing. However, if the detection requirements are not satisfied through inter-frame sensing, the more precise, but longer inter-frame sensing (lasting for up to a superframe length $\cong 160$ msec) that utilizes more complicated algorithms is followed. When WRAN is operated in coexistence mode, both QP types shall be synchronized between neighboring cells (meaning they start and end at the same time) using SCH or channel quiet request/response messages.

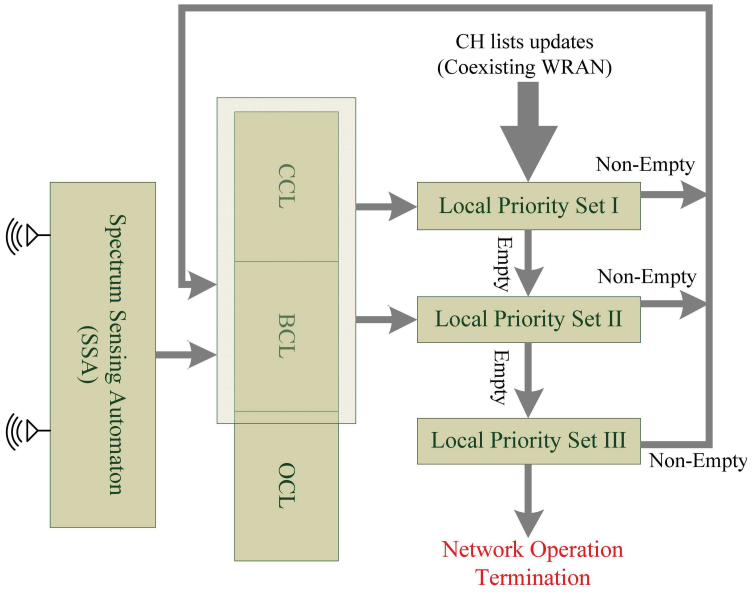


Figure 5.4: Schematic illustration of the updating scheme of the backup (BCL) and candidate (CCL) channel lists.

5.2 The Proposed Multi-channel Sensing Fusion Mechanism

As pointed out earlier, the standard left the choice of a ubiquitous sensing fusion mechanism an open issue for most of the regulatory domains [8, subclause 8.6.3.1]. Seemingly, this is due to the uncertainties that the relevant working groups bore about the performance optimality, efficiency and practicality of the existing sensing solutions and their adaptability to the standard framework. Therefore, [8] proposed very basic sensing fusion mechanisms with deterministic performances, such as AND, OR and VOTING rules, as an interim solution until more fitting solutions appear. Moreover, [8] does not even mandate the aforementioned proposed sensing fusion mechanisms for most of the regulatory domains, except for the United States, which certainly is an unduly cautious choice whose ramification is the waste of resources caused by the manifestation of high false-alarm probability. All these became our motivation to work out an efficient, standard compatible, and easily

implementable, multi-channel learning-based distributed sensing fusion mechanism: the proposed MC-LDS. Seeing what is proposed in this section as a truly generic data fusion scheme, we discern a broader range of scenarios (beside WRANs) wherein the proposed mechanism can be integrated to boost the performance.

Let us focus on Fig. 5.2, which depicts a network layout consisting of twelve coexisting cells where a number of CPEs are uniformly distributed in the coverage area. The incumbent stations (IS) are represented with overlapping circles. A number of channels conceded to the IS are chosen to be opportunistically accessed by WRAN cells. The proposed MC-LDS sensing fusion mechanism, reinforced from underneath by a reward-penalty rationality as well as two levels of differentiation, is detailed next.

5.2.1 Decision Binarization

As detailed in [8], sensors should individually sense the channels within synchronized QPs. No matter the type of sensing method harnessed, the ensuing sensing outcomes should be binary, representing the presence or absence of primary transmission. Herein, energy detection is adopted for it is analytically tractable and prevalent [46]. Spectrum sensing by energy detection requires the least processing, which is basically done by extracting $\mathcal{B}$ discrete samples from the received signal y and forming the power-sum $\mathcal{S} = \sum_{j=1}^{\mathcal{B}} |y[j]|^2$. The major drawback of sensing by energy detection, which is related to its vulnerability to noise uncertainties, can be tackled through cooperation.

Irrespective of the type, the performance of any spectrum sensing mechanism is measured by the following four rates: (1) true-positive, (2) true-negative, (3) false-positive (false-alarm rate, P_{FA}), and (4) false-negative (mis-detection rate, P_{MD}). It is to be noted that the consolidation of (1) and (2) into a single quantity is also known as successful discovering probability, P_{SD}. The false-alarm probability, P_{FA}, quantifies the sensor's misperception degree to detect the primary signal while it actually does not exist. On the other

hand, the mis-detection rate, P_{MD}, denotes the sensor's inability to detect the primary signal when it actually exists.

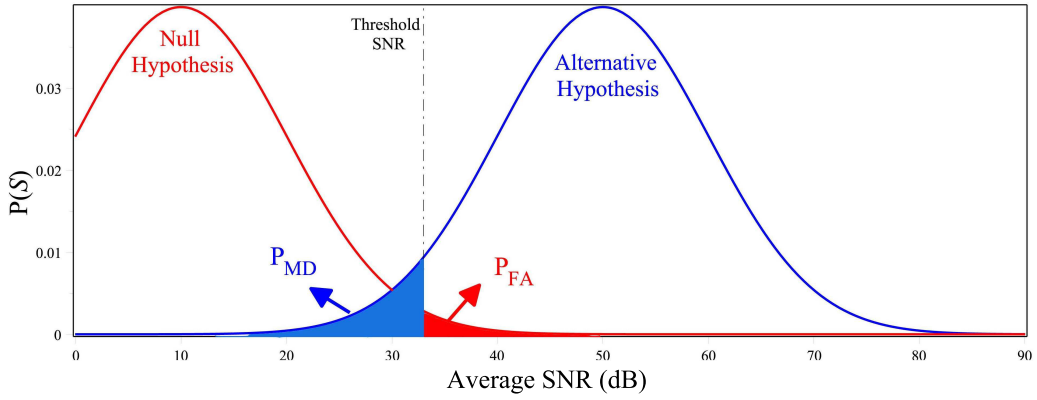
Hereinafter, the superscript $^{(t)}$ over symbols denotes the QP index, which may be signified by $\mathrm{QP}^{(t)}$ as well. Adhering to this notation system, and recalling that indices i, j and k denote the sensor, WRAN cell and operating channel, respectively, the aforementioned quantities are mathematically expressed as:

$$\begin{array}{l} P_{\mathrm{MD},i,j}^{(t),k} = \Pr\left(\mathcal{E}_1\right) \\ P_{\mathrm{FA},i,j}^{(t),k} = \Pr\left(\mathcal{E}_0\right) \\ P_{\mathrm{D},i,j}^{(t),k} = \Pr\left(\neg\mathcal{E}_1 \vee \neg\mathcal{E}_0\right) \end{array}, \quad \begin{array}{l} \mathcal{E}_1 = \{\mathcal{S}_{i,j,k}^{(t)} < \mathcal{S}_{j,k}^{\min} \,|\mathcal{H}_{1,k}^{(t)}\} \\ \mathcal{E}_0 = \{\mathcal{S}_{i,j,k}^{(t)} > \mathcal{S}_{j,k}^{\min} \,|\mathcal{H}_{0,k}^{(t)}\} \end{array}, \tag{5.3}$$

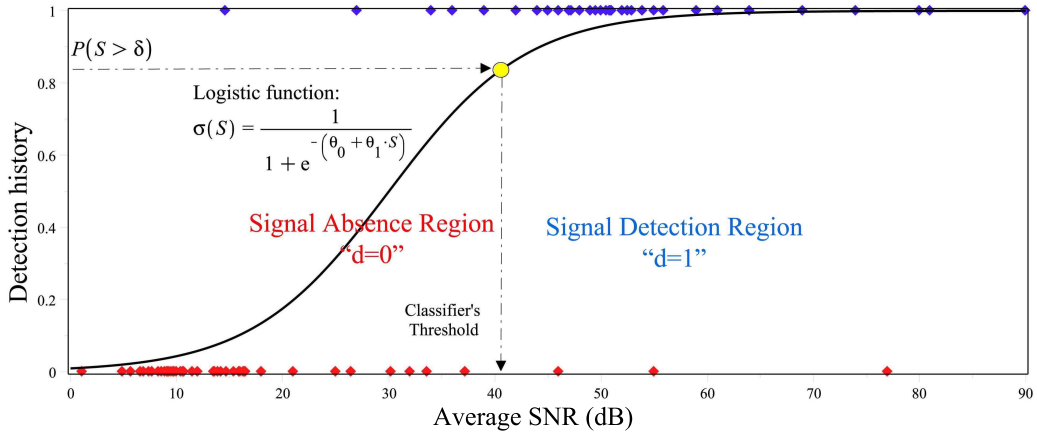
where $\mathcal{H}_{1,k}^{(t)}$ ($\mathcal{H}_{0,k}^{(t)}$), $\mathcal{S}_{i,j,k}^{(t)}$ and $\mathcal{S}_{j,k}^{\min}$ are the alternative (null) hypothesis that the k^{th} channel is taken (not taken) by the IS, the collected power at the sensor's receiver, and its corresponding SNR threshold, respectively. $P_{\mathrm{MD},i,j}^{(t),k}$ ($P_{\mathrm{FA},i,j}^{(t),k}$) can be geometrically represented as the area confined under the received power $\mathcal{S}_{i,j,k}^{(t)}$'s PDF and the left (right) hand-side of the SNR threshold $\mathcal{S}_{j,k}^{\min}$, as shown in Fig. 5.5.

The MC-LDS mechanism works with binary sensing decisions (denoted $d_{i,j,k}^{(t)}$, hereafter), whereby sensors should repeatedly solve their respective binary classification problems after each QP. Among many classification supervised learning models,[5] the logistic regression model chosen in this research for its outcome is a best fitting predictor that not only can be utilized to forecast the operating channel state for future moments but also the value of this function is a meaningful continuum signifying the occurrence likelihood of the event [92]. Also, it is well-known that the logistic estimator is *consistent* (chooses the right parameter values in the limit of a large number of training examples) and *efficient* (no other estimator can converge to the correct parameter values faster in the mean squared sense).

[5]Example models are Bayesian network, support vector machines, neural networks, *Probit* regression, logistic regression, Naive-Bayes, and so forth.



(a) Mis-detection and false-alarm tradeoff.



(b) Logisitc classifier. Solid diamonds are the binary training samples of the sensor that are used to find the logistic regressor $\sigma(\mathcal{S})$ (solid curve).

Figure 5.5: Sensor's decision binarization to discover the status (d) of its operating channel.

The traits identified above make the logistic regression the appropriate classification approach in problems where the nature of events is stochastic. Termed logistic (*Sigmoid*) function and defined as

$$\sigma\left(\mathcal{S}=\mathcal{S}_{i,j,k}^{(t)}\right)=\left.\frac{1}{1+e^{-(\theta_0+\theta_1\cdot\mathcal{S})}}\right|_{\mathcal{S}=\mathcal{S}_{i,j,k}^{(t)}}, \tag{5.4}$$

$\sigma(\mathcal{S})$ is exploited by the i^{th} sensor in WRAN$_j$ (hereinafter, represented as Sen$_{i,j}$) in the following maximum likelihood estimation problem, to find its corresponding regression

parameters $\Theta_i = [\theta_{0,i}, \theta_{1,i}]$:

$$\underset{\Theta_i}{\text{maximize}} \quad \log \prod_{t \in \mathcal{N}_i} \mathcal{L}\left(\sigma\left(\mathcal{S}_{i,j,k}^{(t)}\right) | \Theta_i\right) \quad \begin{array}{l} 0 \leq i \leq m_{j,k} \\ , \ j \in \mathbb{M} = \{1 \cdots \mathfrak{M}\} \ , \\ k \in \mathbb{B} = \{1 \cdots \mathfrak{B}\} \end{array} \tag{5.5}$$

where $m_{j,k}$ is the total number of CPEs in WRAN$_j$ using CH$_k$, $\mathcal{N}_i$ is the set of training samples needed to train the classifier to obtain parameters (Θ_i), $\mathcal{S}_{i,j,k}^{(t)}$ is the measured power of the t^{th} sample in $\mathcal{N}_i$, and $\mathcal{L}(\cdot)$ is the likelihood function to be evaluated for all training samples in $\mathcal{N}_i$. Provided the binary nature of the sensing outcomes $d_{i,j,k}^{(t)} \in \{-1, 1\}$, the likelihood function is a Bernoulli function given by

$$\mathcal{L}\left(\sigma\left(\mathcal{S}_{i,j,k}^{(t)}\right) | \Theta_i\right) = \left(\sigma\left(\mathcal{S}_{i,j,k}^{(t)}\right)\right)^{d_{i,j,k}^{(t)}} \left(1 - \sigma\left(\mathcal{S}_{i,j,k}^{(t)}\right)\right)^{1 - d_{i,j,k}^{(t)}}. \tag{5.6}$$

The maximization problem in (5.5) is convex, hence, optimal coefficient $\hat{\Theta}_i$ (for i^{th} sensor) can be found globally and within a polynomial time. Once obtained, the optimal $\hat{\Theta}_i$ is plugged into (5.4) to characterize the predictor $\sigma\left(\mathcal{S}_{i,j,k}^{(t)}\right)$, which is used to foresee the channel busyness state by calculating its corresponding probability $\sigma\left(\mathcal{S}_{i,j,k}^{(t)} = x\right) = \Pr\left(d_{i,j,k}^{(t)} = 1 \middle| \mathcal{S}_{i,j,k}^{(t)} = x\right)$. The remaining issue to be addressed lies in the fact that the predictor $\sigma\left(\mathcal{S}_{i,j,k}^{(t)} = x\right)$ is a continuous function in $[0, 1]$, $\forall x \in [-\infty, \infty]$, as depicted in Fig. 5.5b, whereas we are in need of a discriminant able to decisively determine whether the channel is busy or idle. In other words,

$$d_{i,j,k}^{(t)} = \mathbb{1}_{\mathcal{A}_i}\left(\mathcal{S}_{i,j,k}^{(t)} = x\right) = \begin{cases} 1 & \text{if } x \in \mathcal{A}_i \\ 0 & \text{if } x \notin \mathcal{A}_i \end{cases}. \tag{5.7}$$

To duly characterize this discriminant function, exclusively disjoint sub-intervals $\mathcal{A}_i$ are to be identified. It turns out that in case of logistic regression, this task is not difficult

as the predictor functions $\sigma(\mathcal{S}_{i,j,k}^{(t)})$ are monotonically increasing and, hence, $\mathcal{A}_i$ is a convex lower-bounded interval. Therefore, as shown in Fig. 5.5b, if one can manage to know the probability $\Pr(\mathcal{S}_{i,j,k}^{(t)} > \mathcal{S}_{j,k}^{\min})$, then the corresponding power obtainable from the inverse of the predictor function acts as the interval boundary, such that $\mathcal{A}_i = [\sigma^{-1}(\Pr(\mathcal{S}_{i,j,k}^{(t)} > \mathcal{S}_{j,k}^{\min})), \infty]$. This narrows down the problem at hand to finding the probabilities $\Pr(\mathcal{S}_{i,j,k}^{(t)} > \mathcal{S}_{j,k}^{\min})$ for $\text{Sen}_{i,j}$ in CH_k at $\text{QP}^{(t)}$ which can be obtained from the following equation

$$\begin{aligned}\Pr\left(\mathcal{S}_{i,j,k}^{(t)} > \mathcal{S}_{j,k}^{\min}\right) =& \sum_{l=0}^{1} \Pr\left(\mathcal{S}_{i,j,k}^{(t)} > \mathcal{S}_{j,k}^{\min} | \mathcal{H}_{l,k}^{(t)}\right) \Pr\left(\mathcal{H}_{l,k}^{(t)}\right) \\ =& 1 - P_{\text{MD},i,j}^{(t),k} + \Pr\left(\mathcal{H}_{0,k}^{(t)}\right)\left(P_{\text{FA},i,j}^{(t),k} + P_{\text{MD},i,j}^{(t),k} - 1\right)\end{aligned}, \tag{5.8}$$

which depends on the mis-detection and false-alarm rates that $\text{Sen}_{i,j}$ experiences by sensing CH_k. Two approaches are conceivable to know the latter quantities. The first approach, which has higher precision, uses labelled training set $\left\{(\mathcal{S}_{i,j,k}^{(t)}, d_{i,j,k}^{(t)})\right\}_{k=1}^{\mathcal{N}_i}$ in order to compute antecedent empirical rates after classification; otherwise, the less precise approximate expressions derived in Section 4.2.6 (i.e., (4.16),(4.17)) can be applied, yielding

$$\begin{aligned}P_{\text{MD},i,j}^{(t),k} &\cong 1 - \mathbb{Q}\left(\frac{\text{SNR}_{j,k}^{\min} - \dfrac{1 + \text{SNR}_{i,j,k}^{(t)}}{\overline{h}_{i,j,k}^{(t)^{\text{Rep}}}}\mathcal{B}_i}{\sqrt{\mathcal{B}_i\left(\left(\dfrac{\text{SNR}_{i,j,k}^{(t)}}{\overline{h}_{i,j,k}^{(t)}} + 2\right)^2 - 2\right)}}\right), \\ P_{\text{FA},i,j}^{(k)} &\cong \mathbb{Q}\left(\frac{\text{SNR}_{j,k}^{\min} - \mathcal{B}_i}{\sqrt{2\mathcal{B}_i}}\right),\end{aligned} \tag{5.9}$$

where $\text{SNR}_{j,k}^{\min}$ and $\text{SNR}_{i,j,k}^{(t)}$ denote $\text{Sen}_{i,j}$'s minimum and instantaneous SNR, respectively, and $\mathbb{Q}(\cdot)$ is the Normal Q-function. Moreover, $\overline{h}_{i,j,k}^{(t)}$ and $\mathcal{B}_i$ represent the channel

loss/fading gain and the number of samples $\text{Sen}_{i,j}$ takes from the received signal, respectively. One should note that $\text{SNR}_{j,k}^{\min} = \mathcal{S}_{j,k}^{\min}/\mathcal{S}_N$ where $\mathcal{S}_N$ is the noise power and $\mathcal{S}_{j,k}^{\min}$ is the threshold power level as in (5.3).

In the next stage, and after having the binarized individual sensing outcomes $d_{i,j,k}^{(t)}$ readily available, MC-LDS combines these values to find an optimum global decision. Before tackling the decision combining problem, it should be noted that the optimization in (5.5), needed to rework the predictor in (5.4), does not have to be solved at every QP. Instead, it is worked out after enough time is elapsed since the last estimation, or upon the availability of new strong evidences that indicate the predictor is performing erroneously.

5.2.2 Decision Combination

The aptness of the sensing fusion mechanism in this research hinges on two internalized levels of differentiation: spatial and temporal. The spatial differentiation among CPEs is sustained by dedicating larger masses on the opinions of sensors with more dependable sensing output, as will be explained later on in this section.

Before describing the temporal differentiation, we need to introduce the reward-penalty logic that is adopted by MC-LDS. The idea here is to initially affix different scores (positive or negative) to sensors depending on situations that can come up, using the following rules: at $\text{QP}^{(t)}$ and given the set $\{d_{i,j,k}^{(t)}\}_{i=0}^{m_{j,k}}$ of the most recent sensing measurements from $m_{j,k}+1$ sensors participating in sensing CH_k in WRAN_j,[6] the latest (penultimate) sensing decision $D_{j,k}^{(t-1)}$ made by the BS[7], and the channel state estimate $R_{j,k}^{(t)}$ obtainable from the database service, the BS assigns rewards and penalty scores, $L_{i,j,k}^{(t)}$ in (5.10) to $m_{j,k}+1$ sensors,

[6] Index $i = 0$ is used to denote the BS of WRAN_j sensing CH_k since, according to [8], all WRAN BSs are self-appointed to carry out separate sensing tasks alongside the CPEs. Thus, $m_{j,k}+1$ sensors sense the channel at any time.

[7] We draw reader's attention to the fact that symbol D has been used in Part I of this thesis to represent the "total delay" that a SU experiences.

$\forall j \in \mathbb{M}, k \in \mathbb{B}$, where $0 < \gamma_k < \zeta_k, \forall k \in \mathbb{B}$.

$$
\begin{aligned}
L_{i,j,k}^{(t)^{\text{rew}}} &= \begin{cases} \gamma_k & \text{if } d_{i,j,k}^{(t)} = R_{j,k}^{(t)} \quad \wedge\, d_{i,j,k}^{(t)} = D_{j,k}^{(t-1)} \\ \zeta_k & \text{if } d_{i,j,k}^{(t)} = R_{j,k}^{(t)} \quad \wedge\, d_{i,j,k}^{(t)} \neq D_{j,k}^{(t-1)} \end{cases} \\
L_{i,j,k}^{(t)^{\text{pen}}} &= \begin{cases} -\zeta_k & \text{if } d_{i,j,k}^{(t)} \neq R_{j,k}^{(t)} \quad \wedge\, d_{i,j,k}^{(t)} = D_{j,k}^{(t-1)} \\ -\gamma_k & \text{if } d_{i,j,k}^{(t)} \neq R_{j,k}^{(t)} \quad \wedge\, d_{i,j,k}^{(t)} \neq D_{j,k}^{(t-1)} \end{cases}
\end{aligned}
\quad . \tag{5.10}
$$

Participation of database reading $R_{j,k}^{(t)}$ as done in this thesis is not arbitrary. Instead, this information shall be utilized towards enhancing the sensing precision and reliability as this mandate has been repeatedly emphasized in the standard, see, e.g., [8, subclause 10.2.2]. Accordingly, WRAN BSs shall retrieve the available channel list as well as the maximum allowed EIRP[8] by sending queries to an incumbent database. Operated by third parties (c.f. FCC R&O 08-260 [93]), the service provided by such database is only a crude estimate of whether the queried channel is occupied at a given location and time or not. Concretely, factors such as passiveness of primary receivers (hidden terminal effect), non-real time availability of the aforesaid database, and the unpredictable and varying nature of communications channels make exclusive reliance on database service extremely inefficient. Figure 5.6 is a spatial snapshot (state) of a database wherein disks with different colors denote the keepout footprint of UHF/VHF channels $k = 5, 6, 7$, ($R_k^{(t)} = 1$). Note that the time index t is frozen and WRAN subscript j is omitted since the latter is irrelevant in that context. Readers are referred to [1, subclause 10.2.2] for a more detailed treatment of the subject. It should also not come as a surprise that a cell can possess multiple values for $R_{j,k}^{(t)}$ and $D_{j,k}^{(t-1)}$ at any moment as there can be more than one operating channel in a WRAN cell.

Returning to (5.10), the designation of the reward-penalty scores gives the highest reward (i.e., ζ_k) to the sensor that has presumably made the right decision (i.e. $d_{i,j,k}^{(t)} = R_{j,k}^{(t)}$)

[8] Effective Isotropic Radiated Power.

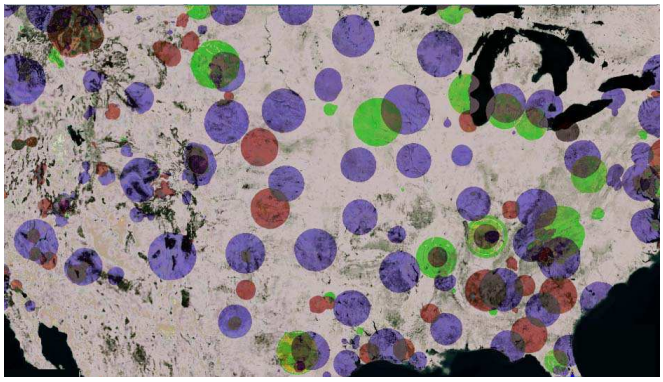


Figure 5.6: Spatial footprint of incumbents occupying TV channels $k = 5$ (green), $k = 6$ (blue), and $k = 7$ (red) in USA, collected from the *White Space Plus* database. It should be noted that this map does not exhibit the temporal footprint of incumbents (source: www.spectrumbridge.com).

although its corresponding BS was deceived in the previous sensing interval (i.e., $D_{j,k}^{(t-1)} \neq R_{j,k}^{(t)}$). On the other hand, the highest penalty (i.e., $-\zeta_k$) is imposed when the local decision neither agrees with the database nor the latest central decision. Table 5.3 is an equivalent representation of (5.10), which reduces to the following mathematical representation.

$$L_{i,j,k}^{(t)} = \gamma_k \left(d_{i,j,k}^{(t)} \odot R_{j,k}^{(t)} - d_{i,j,k}^{(t)} \oplus R_{j,k}^{(t)} \right) - (\zeta_k - \gamma_k) \left(\overline{d_{i,j,k}^{(t)}} - d_{i,j,k}^{(t)} \right) \left(\overline{D_{j,k}^{(t-1)}} R_{j,k}^{(t)} - D_{j,k}^{(t-1)} \overline{R_{j,k}^{(t)}} \right), \quad j \in \mathbb{M}, k \in \mathbb{B}, 0 \leq i \leq m_{j,k} \tag{5.11}$$

where $\oplus$ and $\odot$ represent the XOR and XNOR logic operators, respectively. Note that the total number of sensors in cell WRAN$_j$ is $m_j = \sum_{k \in \mathbb{B}} m_{j,k}$ and that $m_{j,k} \neq 0, \ \forall k \in \text{OCL}_j$ (operating channel list).

Table 5.3: Reward-penalty scores in the MC-LDS mechanism.

$d_{i,j,k}^{(t)}$	$D_{j,k}^{(t-1)}$	$R_{j,k}^{(t)}$	$L_{i,j,k}^{(t)^{\text{rew}}}$	$d_{i,j,k}^{(t)}$	$D_{j,k}^{(t-1)}$	$R_{j,k}^{(t)}$	$L_{i,j,k}^{(t)^{\text{pen}}}$
0	0	0	γ_k	1	0	0	$-\gamma_k$
1	0	1	ζ_k	0	0	1	$-\zeta_k$
0	1	0	ζ_k	1	1	0	$-\zeta_k$
1	1	1	γ_k	0	1	1	$-\gamma_k$

Continuing on, the first level of differentiation in MC-LDS is *temporal*. The key idea is to allow both recent and historical sensing results of sensors to participate in the decision making process. To account for the fact that recent measurements are more dependable compared to older ones,[9] the temporal discount tensor $[\alpha_{j,k}]$ is introduced, which may be chosen either adaptively or statically.

By merging the historical reward-penalty scores of $\text{Sen}_{i,j}$ at the present sensing interval $\text{QP}^{(t=n)}$, the confidence metric[10] $w_{i,j,k}^{(n)}$ is derived as

$$w_{i,j,k}^{(n)} = \sum_{t=n-N_{j,k}}^{n-1} \alpha_{j,k}^{n-t} \cdot L_{i,j,k}^{(t)} \quad , \quad \begin{matrix} j \in \mathbb{M} \\ k \in \mathbb{B} \\ 0 \leq i \leq m_{j,k} \end{matrix} \quad , \tag{5.12}$$

where the superscript $^{n-t}$ is the mathematical exponentiation, as opposed to the indexing superscript $^{(t)}$. The historic count tensor $[N_{j,k}]$[11] represents the effective temporal length.[12] An implication of the above equation is that a sensor which unceasingly inputs wrong information to its BS possesses a smaller (more negative) confidence metric and vice versa.

In the next stage, a discrete-to-continuous sensing channel indicator $X_{i,j,k}^{(n)}$ is established for each sensor at the present sensing interval using the corresponding confidence metric

[9] The existing coherence in the intensity of channel fading as well as the recurring transmission pattern of PUs are two underpinning reasons to sustain temporal differentiation.

[10] Mandated in [8, subclause 10.4.1.2] as one of the spectrum sensing function (SSF) outputs.

[11] We draw the distinction between the usage of symbol N in the first part of the thesis (to represent number of (sub)channels) and here.

[12] In the simulator we built, both static and dynamic approaches were examined for $[N_{j,k}]$ and $[\alpha_{j,k}]$. The observations proved that the adaptive approach results in better performance, in particular, when they are properly related to the instantaneous false-alarm and mis-detection rates, meaning $N_{j,k} = f_1\left(P_{\text{MD,T},j}^{(t),k}, P_{\text{FA,T},j}^{(t),k}\right)$, $\alpha_{j,k} = f_2\left(P_{\text{MD,T},j}^{(t),k}, P_{\text{FA,T},j}^{(t),k}\right)$. It was observed that dynamic adaptation is swifter but computationally heavier to calculate and keep track of compared to the static case when the averaged rates (as will be introduced later on) are utilized.

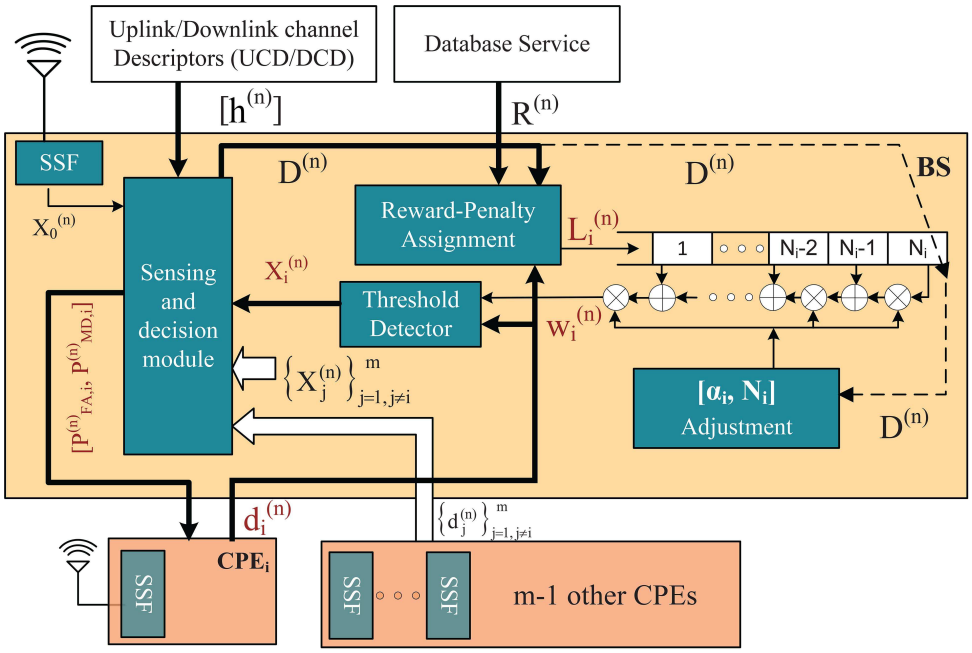


Figure 5.7: The MC-LDS block diagram (cell/channel indices of symbols are removed for better clarity).

$w_{i,j,k}^{(n)}$ and the sensing result $d_{i,j,k}^{(n)}$, as follows:

$$X_{i,j,k}^{(n)} = \begin{cases} w_{i,j,k}^{(n)} & \text{if} \quad d_{i,j,k}^{(n)} = 1 \\ -w_{i,j,k}^{(n)} & \text{if} \quad d_{i,j,k}^{(n)} = 0 \end{cases} , \quad \begin{matrix} j \in \mathbb{M} \\ k \in \mathbb{B} \\ 0 \leq i \leq m_{j,k} \end{matrix} . \tag{5.13}$$

So far, the mechanism has dealt with the sensing outputs from sensors in solidarity. According to (5.12), sensors exhibiting more truthfulness behavior are given larger weights before fusing sensing results. To make an intelligent decision and improve the sensing reliability, the participation of all the sensors is important. In fact, after the vectors $\left[d_{i,j,k}^{(n)}, w_{i,j,k}^{(n)}, X_{i,j,k}^{(n)}\right]$ are reported to the BS of WRAN$_j$, it is the responsibility of WRAN$_j$'s BS to make a global decision about the status of operating channel CH$_k$.

When combining outcomes, there is another source of error, which may degrade the

sensing performance. More narrowly, while the said sensing reliability of each sensor depends on the strength of its sensing channel, the quality of the CPE-to-BS link, known as reporting channel, is as important, and should be taken into account. The idea is that a sensor with strong sensing channel whose reporting channel is weak may not be able to deliver its reliable sensing result to the WRAN BS, thus, is no different from an unreliable sensor from BS's vantage point. Thereby, the second level of differentiation, known as *spatial*, is applied by weighting the indicators $X_{i,j,k}^{(n)}$ with reporting channel gains $\overline{h}_{i,j,k}^{(n)^{\mathrm{Rep}}}$ as follows,

$$D_{j,k}^{(n)} = \begin{cases} 1, & X_{0,j,k}^{(n)} + \sum_{i=1}^{m_{j,k}} \overline{h}_{i,j,k}^{(n)^{\mathrm{Rep}}} X_{i,j,k}^{(n)} > 0 \\ 0, & X_{0,j,k}^{(n)} + \sum_{i=1}^{m_{j,k}} \overline{h}_{i,j,k}^{(n)^{\mathrm{Rep}}} X_{i,j,k}^{(n)} \leq 0 \end{cases}, \quad \begin{matrix} j \in \mathbb{M} \\ k \in \mathbb{B} \end{matrix}, \tag{5.14}$$

where $m_{j,k}$ is the number of sensors sensing CH_k in WRAN_j, with the property $m_{j,k} \neq 0,\ \forall k \in \mathrm{OCL}_j \neq \varnothing$. Once again, the confidence metric $X_{0,j,k}^{(n)}$ is to account for the involvement of the BS in the sensing process as an independent sensing entity.

According to (5.14), a more accurate sensor (larger $X_{i,j,k}^{(n)}$) with a more reliable reporting channel (larger $\overline{h}_{i,j,k}^{(n)^{\mathrm{Rep}}}$) contributes constructively in the final fusion process compared to a sensor with less sensing reliability or one with an erroneous reporting channel. There is another important implication about the way (5.13) is defined in conjunction with (5.14). More precisely, a positive $X_{i,j,k}^{(n)}$, which strengthens the possibility of declaring a busy channel ($D_{i,k}^{(n)} \rightarrow 1$), is attained when one of these two cases occurs in (5.13): (i) $\left[w_{i,j,k}^{(n)} \geq 0, d_{i,j,k}^{(n)} = 1\right]$ and (ii) $\left[w_{i,j,k}^{(n)} < 0, d_{i,j,k}^{(n)} = 0\right]$. While (i) implies that a trustable sensor (which perceives the channel as busy in the current sensing interval $\mathrm{QP}^{(t=n)}$) has a direct role in declaration of a busy channel, (ii) implies that an unreliable sensor (which discerns the channel state as idle) makes a reverse influence. In other words, with MC-LDS, both unreliable and reliable sensors constructively contribute towards an accurate

final decision. Similar deductions are made for negative $X_{i,j,k}^{(n)}$ resulting from two other cases: (iii) $\left[w_{i,j,k}^{(n)} < 0, d_{i,j,k}^{(n)} = 1\right]$ and (iv) $\left[w_{i,j,k}^{(n)} \geq 0, d_{i,j,k}^{(n)} = 0\right]$. The second fascinating attribute of MC-LDS is the relativity of the sensors' reliabilities emerged from the fact that $X_{i,j,k}^{(n)}$ takes values from a continuum. This creates a differentiating property that was inherited by designating reward-penalty scores in Table 5.3.

The block diagram in Fig. 5.7 summarizes the proposed sensing fusion mechanism for IEEE 802.22 WRANs.

5.3 Simulation and Validation Results

5.3.1 Simulator Platform

To verify the performance of the MC-LDS, a discrete-event simulator was built in the C++ and Maple programming languages. In the simulator, all the relevant management and controlling functions of WRAN were implemented. This chiefly includes the QP and sensing management procedures (including inter-frame and intra-frame sensing, synchronization and reporting), channel management functions (including channel list updating, channel switching scheduling, etc.), and an incumbent protection module. Coexistence among WRAN cells was implemented through the CBP protocol.

The topology in each cell is realized in $2D$ space, meaning that CPEs are assumed to be at the same height, which is close to reality for rural areas where WRAN antennas are normally mounted few meters above the ground level at almost the same elevation. All the communications settings, including transmit power, WRAN cell's diameter, SNR threshold, incumbent activity pattern, channel width, interference temperature, maximum allowed interference time, etc. were chosen based on recommendations and mandates of [8] (in particular *Annex A*) and, if not alluded, typical values were chosen. The channel degrading factors, noise and fading, were implemented for both the sensing and the reporting

channels. Also, channel correlation [94] was taken into account.[13] Mobility is not considered in the simulator since WRANs are aimed for static household rather than mobile users. This is apart from the fact that the antenna size on UHF/VHF frequencies poses serious practical considerations for mobile applications. Further, the incumbent activity pattern is assumed to be independent across all available channels, where the activity ratio (i.e., IAR=ON/OFF) on each channel is a mix of bursty and non-bursty patterns with the possibility to adjust the IAR.

A number of channels are randomly distributed among ISs where two or more different ISs may broadcast in the same channel to their respective users. In the latter case, the footprint of ISs broadcasting in the same channel can overlap to have the whole geographical area under coverage of the incumbent service. This may lead to an occasion where a CPE that accidentally located within the transmission range of two or more active ISs receives a primary signal with a large superimposed power from all sources, an incidence that can influence the false-alarm and mis-detection rates significantly. This also has been realized in the simulator.

5.3.2 Performance Measures

In order to evaluate the performance of the MC-LDS, several performance measures are introduced in this section. In what follows, we describe these measures.

False Alarm Rate

Though it was theoretically described according to (5.3), the false-alarm rate can be experimentally estimated, for an individual sensor as well as for the entire cell, using the

[13] For a proper realization of the correlated shadow fading, in the simulator, we restrict the CPEs that happen to be in $\mathrm{Wavelength}/4$ vicinity to experience similar channel intensity.

following formulae,

$$\begin{matrix} P_{\mathrm{FA},i,j}^{(n),k} = \dfrac{1}{\nu_k} \displaystyle\sum_{t=n-\nu_k}^{n-1} d_{i,j,k}^{(t)} \mathcal{Z}_k^{(t)} \\ P_{\mathrm{FA,T},j}^{(n),k} = \dfrac{1}{\nu_k} \displaystyle\sum_{t=n-\nu_k}^{n-1} D_{j,k}^{(t)} \mathcal{Z}_k^{(t)} \end{matrix} \quad , \quad \begin{matrix} 0 \leq i \leq m_{j,k} \\ j \in \mathbb{M} \\ k \in \mathbb{B} \end{matrix} \quad , \tag{5.15}$$

where $\mathcal{Z}_j^{(t)}$ represents the factual status of channel CH_k at $\mathrm{QP}^{(t)}$, whereof the BS has no deterministic knowledge but only estimates $D_{j,k}^{(t)}$ and the database readings $R_{j,k}^{(t)}$. Finally, ν_k is the number of samples taken from CH_k involved in the above averaging.

Mis-Detection Rate

Again, given its probabilistic expression in (5.3), the mis-detection rate is experimentally estimated for an individual CPE as well as for the entire cell using the following formulae,

$$\begin{matrix} P_{\mathrm{MD},i,j}^{(n),k} = \dfrac{1}{\nu_k} \displaystyle\sum_{t=n-\nu_k}^{n-1} \overline{d_{i,j,k}^{(t)}} \mathcal{Z}_k^{(t)} \\ P_{\mathrm{MD,T},j}^{(n),k} = \dfrac{1}{\nu_k} \displaystyle\sum_{t=n-\nu_k}^{n-1} \overline{D_{j,k}^{(t)}} \mathcal{Z}_k^{(t)} \end{matrix} \quad , \quad \begin{matrix} 0 \leq i \leq m_{j,k} \\ j \in \mathbb{M} \\ k \in \mathbb{B} \end{matrix} \quad . \tag{5.16}$$

Successful Discovery Rate

The successful probability of discovery quantifies the receiver ability in correct estimation of the channel state, and is experimentally calculated as

$$\begin{matrix} P_{\mathrm{SD},i,j}^{(n),k} = \dfrac{1}{\nu_k} \displaystyle\sum_{t=n-\nu_k}^{n-1} d_{i,j,k}^{(t)} \odot \mathcal{Z}_k^{(t)} \\ P_{\mathrm{SD,T},j}^{(n),k} = \dfrac{1}{\nu_k} \displaystyle\sum_{t=n-\nu_k}^{n-1} D_{j,k}^{(t)} \odot \mathcal{Z}_k^{(t)} \end{matrix} \quad , \quad \begin{matrix} 0 \leq i \leq m_{j,k} \\ j \in \mathbb{M} \\ k \in \mathbb{B} \end{matrix} \quad . \tag{5.17}$$

By probing these statistics for every sensing interval, the WRAN BS remains vigilant

as to whether certain thresholds are exceeded or not,[14] whereupon appropriate actions such as channel switching, cell shutdown, etc., may be taken.

Correlation Point Indicator

The false-alarm and mis-detection rates are intrinsically conflicting measures where an increase of one inevitably results in a decrease of the other. Thus, when comparing different sensing fusion mechanisms with these measures, no conclusive decision can be made about the superiority or the inferiority of one mechanism w.r.t. another. This fact, also known as bias-variance trade-off in machine learning, was illustrated in Fig. 5.5a, where the confined areas underneath the Gaussian PDF for null ($\mathcal{H}_0$) and alternative ($\mathcal{H}_1$) hypotheses signify the paradoxical fact on how the optimal choice of the threshold SNR, of which we have no knowledge, matters in having the right balance between utilization of the spectrum and radiated interference.

Consequently, the need for a performance measure that jointly and fairly encompasses P_{FA} and P_{MD} is felt. The correlation matrix $C = [\mathrm{NWCF}_{j,k}^{(n)}]_{\mathfrak{M}\times\mathfrak{B}}$ serves this purpose where $\mathrm{NWCF}_{j,k}^{(n)}$ quantifies the amount of statistical similarity between two data streams, namely the factual $(\{\mathcal{Z}_k^{(t)}\}_{t=n-\nu_k}^{n})$ status of CH_k and its estimation by WRAN_j $(\{D_{j,k}^{(t)}\}_{t=n-\nu_k}^{n})$, over the most recent ν_k QPs. In order to work with a scalar indicator, a metric called network-wide correlation factor (NWCF) is obtained by weighted-averaging the correlation matrix over its rows and columns.

Goodness of Fit

The next statistical test used to check the accuracy of MC-LDS is the *Pearson* chi-square test for the goodness-of-fit. As one of the most popular and prevalent statistical tests [95], the *Pearson* test is used to verify whether a given histogram (or estimated categorical data)

[14] Since said threshold was not explicitly stated in [8], $[P_{\mathrm{MD,T}}^{\max}, P_{\mathrm{FA,T}}^{\max}] = [0.1,\ 0.1]$ is chosen for the case study in this section.

matches the theoretical PDF (or factual data). Represented by $\left[\chi_{j,k}^{(n)}\right]^2$ and called Pearson's cumulative test statistic, a real number is obtained per cell, operating channel, and sensing interval, which can either be used directly to quantify the similarity level according to

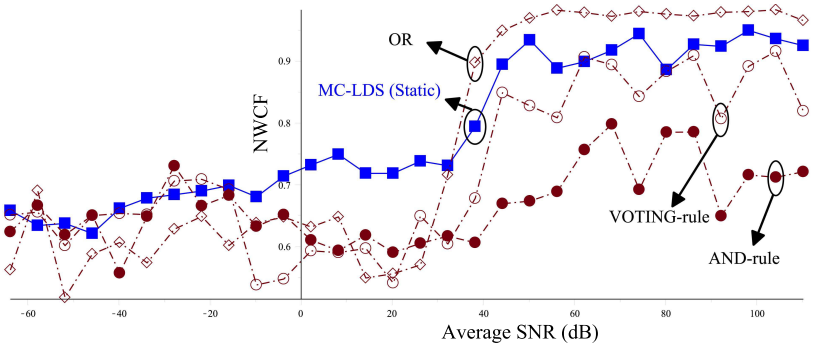
$$\left[\chi_{j,k}^{(n)}\right]^2 = \sum_{t=n-\nu_k}^{n-1} \frac{\left(D_{j,k}^{(t)} - \mathcal{Z}_k^{(t)}\right)^2}{\mathcal{Z}_k^{(t)}}, \qquad \begin{matrix} j \in \mathbb{M} \\ k \in \mathbb{B} \end{matrix} \tag{5.18}$$

or in conjunction with a *p-value* [95] to reject/accept whether $D_{j,k}^{(t)}$ is a good enough estimator for $\mathcal{Z}_k^{(t)}$. As introduced before, ν_k is an arbitrary channel-dependent figure that represents the number of samples involved in the above summation.

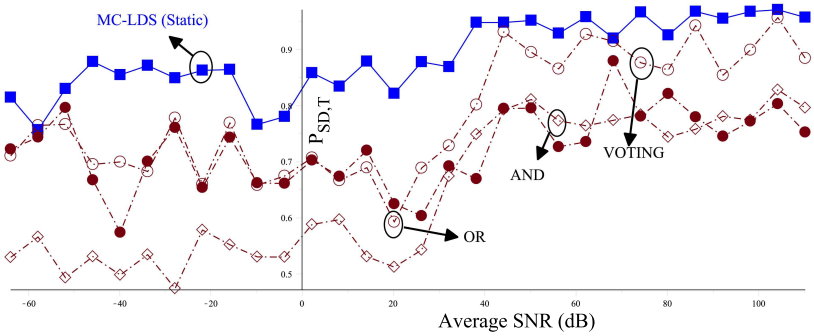
Now, given the two data streams, i.e., the factual channel state $[\mathcal{Z}_k]$ and the estimated channel state $[D_{j,k}]$ and considering that the network under study is composed of $\mathfrak{M}$ coexisting cells and $\mathfrak{B}$ channels, a single performance vector $[\mathrm{NWCF}, P_{\mathrm{SD,T}}, P_{\mathrm{MD,T}}, P_{\mathrm{FA,T}}, \chi^2]$ is obtained per cell, per channel, which delivers a $\mathfrak{M} \times \mathfrak{B}$ matrix to evaluate the performance of MC-LDS. An example of this matrix for the mis-detection rate is extracted and reproduced in Table 5.4.

5.3.3 Comparative Results

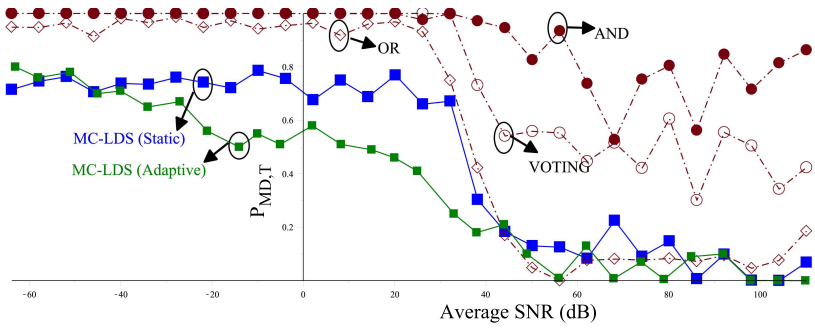
Numerous experiments were conducted to verify the performance of the MC-LDS mechanism. Comparative results were plotted versus average SNR (dB). For the sake of comprehensiveness, different circumstances were examined, including diverse SNR regimes, fading situations, incumbent activity pattern profiles, etc. Furthermore, the performance of MC-LDS was probed in the presence of hidden and exposed-terminal problems. The observations unanimously prove that MC-LDS is superior to AND, OR and VOTING sensing fusion mechanisms. Of particular importance is the performance uniformity of MC-LDS in different SNR regimes.



(a) Network wide correlation factor

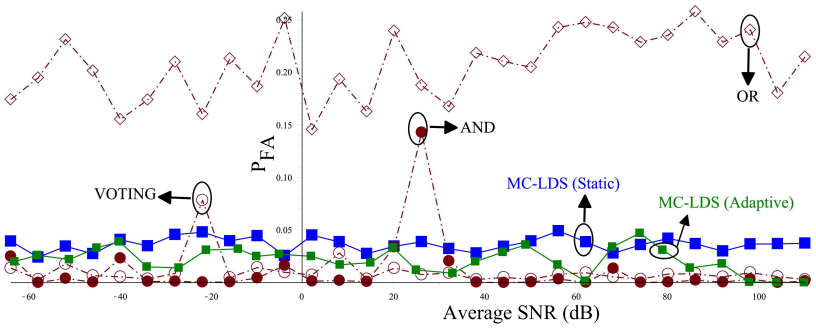


(b) Network wide successful discovery rate P_{SD}

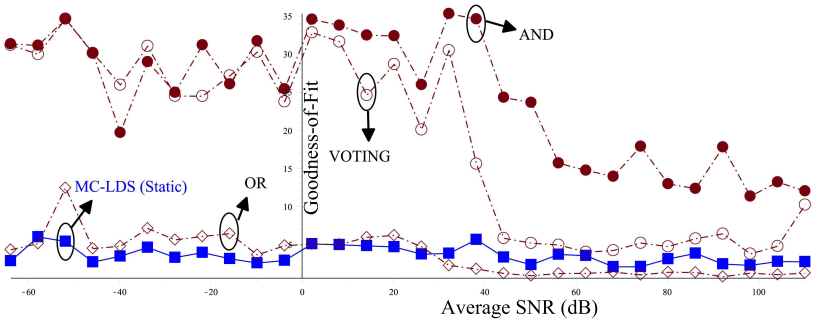


(c) Network wide mis-detection rate P_{MD}

Figure 5.8: Performance comparison of AND, OR, VOTING, and MC-LDS SMFs by different performance metrics.



(a) Network wide false-alarm rate P_{FA}



(b) Goodness of Fit

Figure 5.9: Performance comparison of AND, OR, VOTING, and MC-LDS SMFs.

Table 5.4: Mis-detection rate matrix $[P^{(n),k}_{\mathrm{MD,T},j}]_{k\in\mathbb{B},j\in\mathbb{M}}$. The appearance of NA in $[j,k]^{\mathrm{th}}$ entry indicates that during the simulator runtime, $k \notin \mathrm{OCL}_j$, hence, no conclusion can be made about CH_k in WRAN_j.

	CH$_1$	**CH$_2$**	**CH$_3$**	**CH$_4$**	**CH$_5$**	**CH$_6$**	**CH$_7$**	**CH$_8$**	**CH$_9$**	**CH$_{10}$**
WRAN$_1$	**NA**	0.0470	**NA**	0.173	0.225	0.071	0.052	0.038	0.16	**NA**
WRAN$_2$	0.141	0.321	**NA**	**NA**	0.02	0.165	0.043	0.0.72	0.123	0.022
WRAN$_3$	0.0083	0.258	0.0033	0.059	0.022	0.029	**NA**	0.0089	0.081	0.081
WRAN$_4$	0.25	0.0037	0.045	0.24	0.22	0.076	0.016	0.37	0.083	0.068
WRAN$_5$	0.016	0.049	0.049	**NA**	0.076	0.31	0.090	0.053	0.093	0.062
WRAN$_6$	0.066	0.29	0.11	0.40	0.083	0.058	0.27	0.29	0.085	0.25
WRAN$_7$	0.065	**NA**	0.12	0.20	0.096	0.077	0.34	**NA**	0.054	0.060
WRAN$_8$	0.086	0.34	0.076	0.071	0.089	0.17	0.060	0.079	0.055	**NA**
WRAN$_9$	0.089	0.073	0.39	0.060	0.072	0.090	0.11	0.17	0.26	0.40
WRAN$_{10}$	**NA**	0.28	0.051	0.085	0.20	0.080	0.31	0.12	0.13	0.26
WRAN$_{11}$	0.098	0.084	0.073	0.057	0.22	0.37	0.072	0.071	0.081	0.12
WRAN$_{12}$	**NA**	0.32	0.086	0.19	0.11	0.065	**NA**	0.10	0.052	0.094

Figure 5.8 proves the above claim for a wide SNR range (-70 to $110\,\mathrm{dB}$). Starting from low SNR regimes, wherein, principally, the real merits or weaknesses of any decision combining method lies, the MC-LDS sensing fusion mechanism is unequivocally superior to all its rivals for all the metrics. In high SNR regimes, MC-LDS still keeps the superiority in comparison to VOTING and AND rules, even though, correlation-wise, the OR-rule has the edge over MC-LDS. Notwithstanding, this slight superiority is severely compromised by the unacceptably high false-alarm rate that the OR-rule intrinsically exhibits at all the times. This is further to the fact that when IAR$<<$1 and the incumbent network's alteration frequency (IAF) is high,[15] the OR-rule exhibits extremely poor performance in all the metrics while the MC-LDS still keeps the edge over others. In summary, despite the fact that achieving low false-alarm and mis-detection rates at the same time is theoretically impossible, as discussed earlier, the MC-LDS mechanism still finds the best trade-off that

[15]The incumbent alteration rate is different from the incumbent activity rate (IAR) as the former represents the frequency at which ON and OFF periods alternate, whereas the latter indicates how relatively large ON periods are w.r.t. OFF periods.

maximizes the NWCF and minimizes the *Pearson*'s goodness-of-fit metric.

One should note that the superior behavior of the OR-rule in terms of mis-detection rate does not make it an efficient decision combining approach for the sake of the very high false-alarm rate it manifests, as shown in Figs. 5.8c, 5.9a. The same reasoning can be stated about the AND-rule, which obtains a low false-alarm rate but an unacceptably high mis-detection rate. More troublesome is the unstable behavior of these two sensing fusion mechanisms when IAR or IAF vary, exhibiting very different performance profiles.

5.3.4 Adaptive Tuning

The existence of two degrees of freedom, the temporal discount factor α and the historic count Σ in the previous section, gives MC-LDS a great deal of flexibility by assigning them to act as control variables. Our investigations confirm that in situations where the achieved performance vector is not satisfactory, for instance when there is no balance between $P_{\mathrm{MD},T}$ and $P_{\mathrm{FA},T}$, the dynamic adjustment of α and Σ can make drastic changes. One approach is to associate these parameters with the estimated false-alarm and mis-detection rates in a linear fashion, such that

$$\begin{aligned} \Sigma_{j,k} &= \left\lfloor \mathfrak{K}^2 - \mathfrak{K}^1 \cdot \overline{P^{(n),k}_{\mathrm{MD,T},j}} \right\rfloor && 0 \le i \le m_{j,k} \\ & , && j \in \mathbb{M} \quad , \\ \alpha_{j,k} &= \mathfrak{K}^3 + \mathfrak{K}^4 \cdot \overline{P^{(n),k}_{\mathrm{FA,T},j}} && k \in \mathbb{B} \end{aligned} \tag{5.19}$$

where $\lfloor\ \rfloor$ is the *floor* operation, $\mathfrak{K}^2 - \mathfrak{K}^1 > 1$, $\mathfrak{K}^3 + \mathfrak{K}^4 < 1$ and $\mathfrak{K}^1, \cdots, \mathfrak{K}^4$ are all positive real numbers that can be chosen statically. Green curves in Fig. 5.8c and 5.9a illustrates the obtainable gain, in terms of $P_{\mathrm{FA,T}}$ and $P_{\mathrm{MD,T}}$ by simply feeding the output of the system back to its input and substantiate the importance of feedback in stabilizing cognitive systems and improving their performance.

5.4 Summary

In this chapter, an efficient decision combining method, named multi-channel learning-based distributed sensing (MC-LDS) mechanism, was proposed for wireless regional area networks (WRANs). While conforming with the IEEE 802.22 standard directives, MC-LDS operates based on a simple learning concept that uses differentiation as well as reward-penalty procedures to intelligently combine current and past sensing measurements. The combination was done in such a way that both accurate and faulty sensors contribute constructively toward an accurate final decision. To substantiate the merits of MC-LDS, a WRAN simulator was built. Through precise probing of a comprehensive set of quantities such as false-alarm, misdetection and successful discovery rates, as well as correlation factor and goodness-of-fit metric, it was demonstrated that MC-LDS is considerably superior to AND, OR and VOTING rules. Among the advantages of MC-LDS, its accuracy, implementation ease, fairness, adjustability and stability are to be emphasized. Having shown the intrinsic inefficiencies and drawback of the SFMs suggested by the IEEE 802.22 standard, MC-LDS is recommended as a competent candidate for adoption in WRANs. Furthermore, MC-LDS can be integrated to boost the sensing performance in other promising technologies and standards such as White-Fi (IEEE 802.11af WLANs) wireless personal area networks and Zig-Bee (IEEE 802.15 family), cognitive WiMax (IEEE 802.16h) and the recent IEEE 1900.6b standard emerged to support spectrum databases using spectrum sensing information.

Chapter 6

Spectrum Sharing[1]

Futuristic networks are expected to have extensive coverage and extremely high speed, and be traffic-agnostic. This property has enormous implications. Generating hundreds of Exabytes of traffic with diverse characteristics every year, machines will be commonplace subscribers of 5G networks. They can be anything from a garage door controller, vacuum cleaner, car and city lamp posts, power meters, traffic lights, etc. The nature and scale of these machine-originated traffic are essentially different from traditional voice/video/web-browsing traffic created by hand-held mobile devices. In fact, while our existing networks are optimized for the latter types, networks carrying the former traffic type have different prerequisites. To live up to such growing demands and stringent requirements, allocation of the spectrum resources, as the beating heart of a network architecture, is to take a new shape. Obviously, static allocation is no more acceptable due its inherent inefficiency and stagnation. Instead, spectrum is to be allocated more dynamically, faster, and for shorter intervals. Having a glance at the evolution of cellular networks proves the dominance of this line of thought. From 2G cellular networks (GSM), where the architecture was connection-

[1]N. Tadayon and S. Aïssa, "Resource allocation in cognitive radio networks: Pricing and utility derivation", Submitted to *IEEE Journal on Selected Areas in Communications*, Apr. 2016.

[1]N. Tadayon and S. Aïssa, "Resource allocation in cognitive radio networks: Spectrum Assignment", Submitted to *IEEE Journal on Selected Areas in Communications*, Apr. 2016.

oriented with non-adaptive session-based scheduling, to the all-IP packet-switched architecture of 4G networks (LTE), where the resource scheduling is load and channel dependent and is done as frequent as every 1 msec, ever since spectrum allocation has become more agile and adaptive. 5G networks will have to allocate/deallocate/reallocate spectrum in the μsec scale for their near-zero delay tolerance. The shorter this allocation interval becomes, the more difficult it gets to formulate the resource allocation as a globally centralized optimization problem and find a solution to it. Adhering to the same line of thought and acknowledging that smartness will become a common attribute of our digital societies in coming decades [96, 97], adaptation in allocation is to take a different direction, where the network subscribers (machines, mobiles, laptops, etc.) who best know how much resource they need on the basis of their internal states and predictions they make, play a fundament role. This approach of liberating the allocating entities, such as BSs, from dealing with unessential details will certainly simplify the resource allocation problem to a large extent.

For the constant growth of intelligence in consumer electronics, mechanisms that proved efficient in allocating commodities to humans are worth consideration for allocating resources to wireless devices. Auctions are among these mechanisms [98]. They have been studied by game theorists for decades and the auction theory is one of the well-developed branches of microeconomics science [99]. Auctions are built on the foundation that the valuation of a commodity and its allocation efficiency are interrelated factors. In effect, valuation is a *demand-dependent* quantity increasing when demand rises. This dynamicity was independently discovered in the wireless communications field as the solution to static allocation of spectrum. Known as DSA, this new architecture allows wireless nodes to quickly adapt their behavior based on the cognition they learn from their surrounding environment (channel state, own traffic needs, energy state, sensing outcome, etc.) in order to optimize their utilities, e.g., spectrum efficiency, by using spectrum holes in the time, frequency and space dimensions. When comparing CR definition with that of “competitive market", one realizes the analogy between the two, where observations, actions, and best

responding in the latter corresponds to sensing, reconfiguration, and deciding in the former. This philosophy is better illustrated in Fig. 6.1.

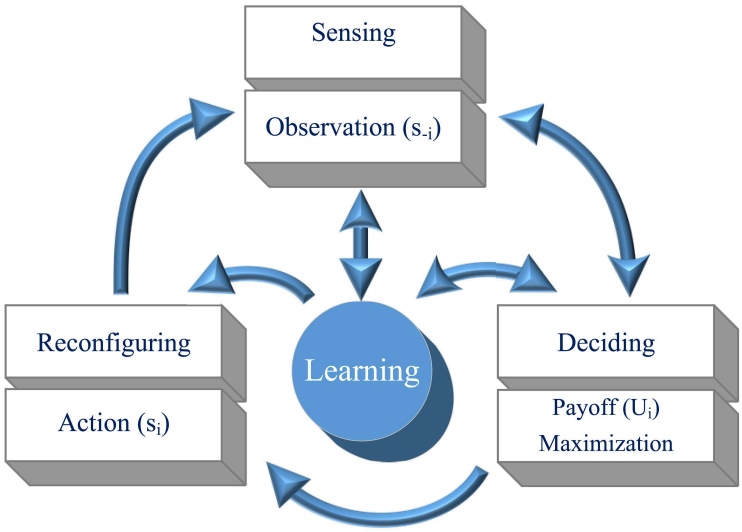


Figure 6.1: Analogy between the elements of a game and DSA architecture.

Along the investigations, compelling reasons were found in regard to the aptness of auctions in modeling and design of MC-CRNs. First, the abilities of machines in analyzing, deducing and making decisions, and their similarities with corresponding human thinking models, is unceasingly increasing. Second, auctions produce more efficient outcomes attained by imposing charges (prices). Motivated by the latter facts, the last part of this thesis is dedicated to designing a *resource allocation* mechanism for MC-CRNs using the auction theory. This design problem is fragmented into two different components, namely *spectrum pricing* and *spectrum assignment*. Each component is dealt with, separately.[2] The spectrum pricing component is characterized in Section 6.2 and the spectrum assignment component is dealt with in Section 6.3. From a different perspective, the proposed resource allocation mechanism satisfies six axioms. Whereas truthfulness, fairness, and individual rationality, are the axioms that will be instilled into the pricing component,

[2]Please note that, in this thesis, “resource allocation" is the general term embracing both “spectrum pricing" and “spectrum assignment" problems.

in Section 6.2, the spectrum allocation component and its associated axioms (efficiency, revenue maximization and computational manageability) are addressed in the Section 6.3. The work in both sections is built upon the seminal recent findings of the Nobel-laureate winning economist, Eric Maskin, in the area of optimal multi-unit auctions [100].

6.1 Background

6.1.1 Literature Review

To address the resource allocation problem among rational and self-interested SUs, the non-cooperative game theory that includes static, dynamic, spatial, evolutionary, and auction game models is the proper framework. The question on which game model is suitable for a problem depends heavily on the characteristics of the corresponding network and its components. In the *static* model, a game is only played once by the players, which is fitting for situations where the channel state changes, sporadically. Studies [101–104] model their allocation problems as static games in FDMA-based/cellular networks choosing total interference/throughput as the utility function under different power and rate constraints. Static game is also used in solving the channel allocation problem in OFDMA-based multi-interface networks [105, 106] and the power allocation problem in CDMA-based networks [107]. Unlike the static case, *dynamic* games are played repetitively. In the context of CRNs, [108–110] address the resource allocation problem where throughput is chosen as the utility function. The third category of non-cooperative games that find applications in biological systems is the *evolutionary* game model. [111] is among the handful of studies that investigates the application of evolutionary game models to wireless communications.

Emerged from the theory of microeconomics, the application of *auctions* to wireless networks were only limited to large-scale long-term allocation of spectrum to mobile/network operators. Realizing the fact that auctions deliver much more efficient outcomes and con-

verge quickly, they can be exploited to allocate spectrum to end-users on a small-scale short-term basis. Inspired by these prospects, [112] and [113] model the allocation problem in TDMA-based networks as a 2^{nd} price *Vickery* auction. Due to their similarities with auction games, *oligopolistic* competition (a contention where a few main players determine price and allocation) is exploited in [114, 115] in the context of CRNs.

Importantly, among more than a hundred different auction games, only a handful are known to be maximally efficient or optimal [98, 116]. In the latter group, nearly all of them are single-unit auction games. Indeed, due to the intricacies associated with characterization of optimal multi-unit auctions (a.k.a. combinatorial auction), this branch has evidenced very limited progress. In fact, what exists in the literature on optimal multi-unit auctions is only due to the contemporary endeavors of few economists/mathematicians, amongst whom Eric Maskin et al.[3] can be mentioned for magnificent findings in [100, 117].

6.1.2 Fundamentals of Game Theory

Solution Concepts

Before introducing the most commonly known solution concept of games, another general concept known as best response is to be established. In a formal language, *best response* (BR) of SU_i[4] with interdependent *payoffs* to a complementary strategy s_{-i} is the set of strategies that bring the maximum utility to SU_i, i.e.,

$$BR_i\left(s_{-i}\right)=\left\{s_i\in\mathcal{S}_i|u_i\left(s_i,s_{-i}\right)\geq u_i\left(s_i',s_{-i}\right),\ \forall s_i'\in\mathcal{S}_i,\ s_i'\neq s_i\right\},\qquad i\in\mathcal{M},\quad(6.1)$$

where $\mathcal{M}=\{1,\cdots,M\}$ is a finite set of SUs, $\mathcal{S}_i$ is the strategy profile of SU_i, $u_i:\mathcal{S}\rightarrow\mathbb{R}^{\geq 0}$ is SU_i's utility function, $\mathcal{S}=\mathcal{S}_1\times\cdots\times\mathcal{S}_N$ is the strategy space of all SUs, and

[3] For their significant contributions in the field of mechanism design, Eric Maskin, Roger Myerson and Leonid Hurwicz were jointly awarded the 2007 Nobel *Laureate* prize in economy.

[4] Terms "*agent*", "*user*" and "*player*" all denote an SU and may be interchangeably used, hereinafter.

$s_{-i} \in (\mathcal{S}_1 \times \cdots \times \mathcal{S}_{i-1} \times \mathcal{S}_{i+1} \times \ldots \mathcal{S}_N)$ is the complementary strategy space of SU_i.

Nash Equilibrium: The Nash equilibrium (NE) is a state $\mathbf{s}^* = (s_1^*, \cdots, s_N^*) \in \mathcal{S}$ of the game where no agent can improve its utility by, unilaterally, deviating to a different strategy (action), knowing that other agents stick to their current strategies. Formally speaking,

$$u_i\left(s_i^*, s_{-i}^*\right) \geq u_i\left(s_i, s_{-i}^*\right), \ \ \forall s_i \in \mathcal{S}_i, \ i \in \mathcal{M} \tag{6.2}$$

There are many issues to be investigated about the NE before making a conclusive decision for its adoption. In particular, *existence*, *uniqueness*, *stability*, and *efficiency* of NE, are of paramount importance. An alternative definition based on the BR concept introduces the NE as a strategy profile s^*, where each agent's strategy is the BR to its complementary strategy, i.e.,

$$s_i^* \in BR_i\left(s_{-i}^*\right), \ \forall i \in \mathcal{M}. \tag{6.3}$$

Types of Games

Before proceeding with the introduction of the two main classes of games, it is to be noted that any discussion, hereinafter, is related to the bigger class of *continuous* games, where the strategy profiles $\mathcal{S}_i, \ i \in \mathcal{M}$ are continuous kernels with uncountably infinite members. This is generally the case in telecommunications problems, which is contrary to the classic game theory where the strategy spaces are of discrete nature.

Non-Cooperative Games: These games emerged to model interaction among conflicting agents. Within the context of non-cooperative games, agents make their decisions (actions) independently of the others, though the resulting payoff of every single agent is affected by the actions of all. This class is intrinsically the proper framework for modeling distributed

wireless networks, as opposed to a cooperative game theory (also known as coalitional game) which is more fitting for centralized networks. The normal (strategic) representation of a non-cooperative game consists of a triplet $G = \langle \mathcal{M}, (\mathcal{S}_i)_{i\in\mathcal{M}}\,,\ (u_i)_{i\in\mathcal{M}} \rangle$.

From another perspective, games are classified into *static* and *dynamic* depending on whether *time* exists or not. While each agent in a static game takes one and only one action, many consecutive actions may be taken in rounds (sub-games) in a dynamic game by learning through time.

Bayesian Games: The class of non-cooperative games is divided into *complete* and *incomplete* games. While in a complete information game, all game components such as payoffs, strategy spaces, strategies, etc., are a-priori known to each and every player, in an incomplete information game, players lack knowledge about one or more of the aforementioned elements. These uncertainties are mathematically represented by RVs $\nu_i \in V_i$, known as *types* and quantified through forming beliefs, where V_i is an agent's type space. The agent's actions are represented by an internal rule $s_i : V_i \rightarrow A_i$ that transforms its private valuation ν_i into a declarable quantifiable action $a_i \in A_i$. Depending on actions a_i taken by the agent and its rivals, different utilities are attained according to $u_i : \mathcal{S} \times V \rightarrow \mathbb{R}^{\geq 0}$, where $\mathcal{S} = \prod_{j\in\mathcal{M}} \mathcal{S}_j$ and $V = \prod_{j\in\mathcal{M}} V_j$ are the global strategy and valuation spaces, respectively. After observing its own true valuation ν_i, each agent forms beliefs about other agents' valuations, denoted by ν_{-i}, which are quantified by a PDF $\Pr(\nu_{-i}|\nu_i)$ over the complementary valuation space $\nu_{-i} \in \prod_{j\in\mathcal{M}\setminus\{i\}} V_j$. This belief PDF is, then, used by the agent to determine the action that maximizes its Bayesian utility u_i, defined by

$$\underset{\nu_{-i}\in V_{-i}}{\mathcal{E}} \left\langle u_i \left(s_i \left(\nu_i \right), s_{-i} \left(\nu_{-i} \right),\ \nu_i \right) \right\rangle, \tag{6.4}$$

where $\mathcal{E}\langle\cdot\rangle$ is the expectation operator. The above definition of Bayesian utility is a statement in an ex-interim sense, signifying that the agent is certain about its own valuation but

is uncertain about its opponents' valuations. The NE corresponding to this new definition is called Bayesian Nash Equilibrium (BNE).

(Bayesian) Nash Equilibrium: Inherent Drawbacks

Efficiency: The selfish and overly rational behavior of agents (players) in most game models result in an equilibrium that is far from social optimality gained when solving the same problem in a centralized manner. Among other approaches, pricing and penalization rules were proven to be effective in enhancing the allocative efficiency by pushing the game (auction) equilibrium closer to the Pareto optimal frontier [118]. It is for the latter advantage that pricing rules have been applied to the game-theoretic resource allocation problems in CRNs, such as linear pricing for overlay power allocation [119] and nonlinear pricing for underlay power allocation [120].

To measure the magnitude of efficiency of the NE, the price of anarchy ζ, defined as the ratio between the value of social welfare function (SWF) (i.e., sum of utility functions of all agents) at its maximum (achievable through centralized decision mechanism) and NE, can be used. If there are multiple NEs (represented by the set $\{\mathrm{NE}\}$), then

$$\zeta = \frac{\max\limits_{s \in \mathcal{S}} \sum\limits_{i \in \mathcal{M}} u_i(s)}{\min\limits_{s \in \{\mathrm{NE}\}} \sum\limits_{i \in \mathcal{M}} u_i(s)} . \tag{6.5}$$

Convergence: Even though game theory delivers an analytical tool that empowers us to understand and model strategic interaction among wireless nodes and find the equilibrium point (if one exists), there is still left an important issue concerning the convergence to such equilibrium given the practical aspects of the network under consideration. In the perfect information scenario, where all players fully observe the actions and utilities of others, any game eventually converges to its equilibrium if one exists. However, it is too optimistic to assume that such information is available to all wireless nodes. A good example is a

power control game where transmit power is the action space of nodes. In such a situation, nodes' precise actions are not known to one another. Instead, each node can only measure the signal-to-noise-plus-interference ratio (SINR) as a lump quantity [121]. This inevitably puts the game in the realm of imperfect information (or Bayesian) games where convergence becomes an issue. Therefore, learning becomes an elemental task of each node, whereof the precision impacts the rate of convergence.

As an applied branch of economic theory, auctions emerged to effectively tackle the above-mentioned issues. Due to their distributed nature, stability, computational manageability and operational efficiency, auctioning is adopted in this thesis to allocate resources to CR nodes. Next, we delve into the pricing rule for the proposed auction-based resource allocation mechanism.

6.2 Spectrum Pricing and Utility Derivation

Auctions are games of incomplete information (Bayesian), whereby rational and self-motivated bidders (called SUs hereafter), strategically compete to maximize their own profits. They are generic allocation mechanisms applicable to different situations. Auction's outcomes are quickly obtainable. The scenario-agnostic nature of auctions is shown in Fig. 6.2 where primary, secondary, and autonomous wireless devices contend to access transmission resources.

In the proposed auction-based resource allocation mechanism, an SU's strategy is to submit the demanded bandwidth $a_i = \widetilde{w}_i$ that maximizes its utility to the secondary operator (SO). On the other side, the SO's role is to design the allocation mechanism components, such as different auction stages, pricing rule, etc., in a way that the SO's accrued revenue is maximized. As depicted in Fig. 6.3, the goal, in this section, is to define the pricing rule of the spectrum allocation auction game such that the axioms fairness, incentive compatibility (IC), and individual rationality (IR) are incorporated. The **fairness** axiom is satisfied by

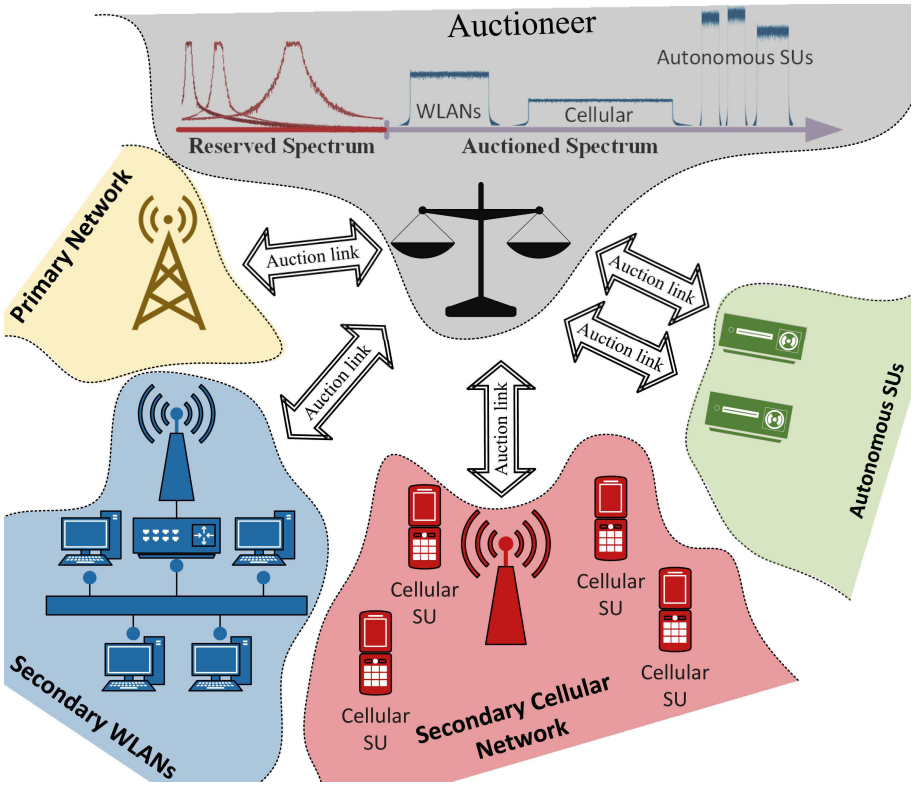


Figure 6.2: Illustration of the traffic-agnostic and platform-independent nature of auctions in assigning resources in wireless networks.

adhering to the logic that an SU that receives more spectrum is to be charged higher. In the presence of self-centered and intelligent SUs, **truthfulness** means that SUs have no incentive to confuse the allocating entity (called auctioneer) by lying or perturbing others' actions. This important property is realized through delicate selection of a payment rule that also enforces the **individual rationality**. The latter property guarantees that SUs' average utilities never become negative. Due its natural role in quantifying the preferences of SUs, spectral efficiency is chosen as the utility function that SUs constantly strive to maximize and is characterized in closed-form. Under the truthfulness-imposing payment rule introduced in this section, the spectrum allocation determined in the next section is shown to maximize total revenue and enhance efficiency in polynomial time. These properties

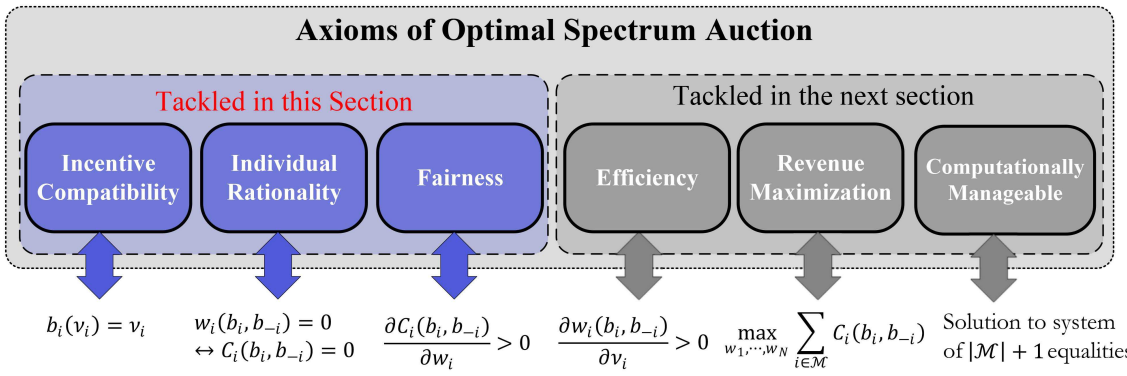


Figure 6.3: The axioms of the proposed spectrum allocation mechanism.

make the overall proposed resource allocation mechanism an ideal solution for deployment in 4G and 5G networks.

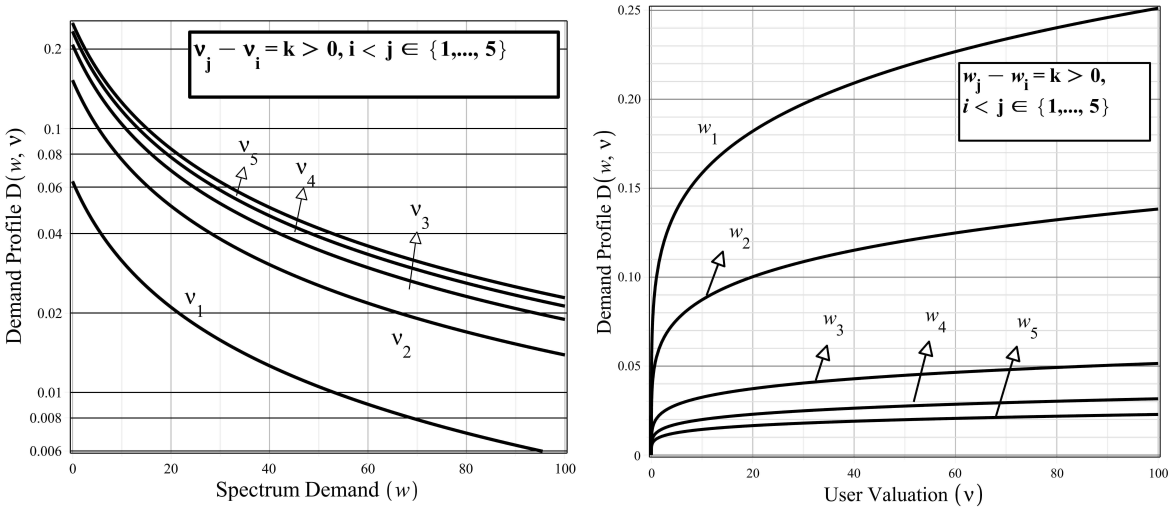
6.2.1 Demand Profile and Utility Function

The precise characterization of utility (payoff) function is a crucial and delicate task. Such function, which is intertwined with users' internalized preferences and network objectives, cannot be chosen arbitrarily. To have a well-defined and all-inclusive utility function, reverse engineering (bottom-up approach) from the lowest levels is to be done.[5]

With this vision, demand profile, $\mathcal{D}_{\nu_i}(w_i) = \mathcal{D}(w_i, \nu_i)$, a *common* feature of the allocation game, is a relationship between the demanded spectrum quantity w_i[6] and the unit price that the SU is is expected to pay, provided that the user privately values it as ν_i. We assume that the spectrum is infinitesimally granular and demand curves are smooth. In the case of limited gross resource w_0, where $w_i \in [0, w_0]$ and $\nu_i \in [\underline{\nu}_i, \overline{\nu}_i]$ are taken as fixed parameters, demand curves are sloped downwards, as shown in Fig. 6.4a. This is an im-

[5]Unfortunately, there have been some misapprehensions in this area having led to the emergence of game-theoretic studies and models that are based on obscure mathematical operations, e.g., defining a utility function by subtracting two quantities of different natures and dimensions [122]. [123] offers a comprehensive survey on different choices of utility functions in CRNs.

[6]We draw reader's attention to the distinction between the usage of symbol w in Chapter 5 as a dimensionless "confidence metric" and here as the "demanded/allocated bandwidth"



(a) Demand curves for fixed valuation.

(b) Demand curves for fixed quantities.

Figure 6.4: Demand profile complying with the law-of-demand in a setting with $M = 5$ SUs.

portant trait of markets and a.k.a. *law-of-demand*.[7] Despite the fact that the demand profile is common knowledge, each SU is spotted on a different curve of $\mathcal{D}(w_i, \nu_i)$, depending on the magnitude of its valuation ν_i of which no other agent is aware. The area under a demand curve confined to $[0, w_i]$, for a given valuation, is, indeed, the SU's revenue for leasing $\mathtt{W} = w_i$ units of spectrum bandwidth (BW).

The SO gains the non-exclusive access to a frequency band with non-deterministic width $w_0(X, Y, t)$. The latter is a stochastic process that varies in time, t, and space, X and Y. The randomness of w_0 is mathematically represented with the CDF $\mathcal{F}_{w_0}(\mathtt{W}; t, X, Y) = \Pr(w_0(X, Y, t) < \mathtt{W})$. Hereafter, w_0 is shown without its arguments, for brevity. To maximize its profit, the SO is willing to share w_0 among a number of SUs (M) whose propensities to buy are diverse and summarized by the demand profile $\mathcal{D}_{\nu_i}(w_i)$. With this in mind,

[7] The three robust explanations as to why demand curves are downward sloping are: diminishing marginal utility effect, income effect and substitution effect.

the utility function of the i^{th} SU is defined as

$$u_i\left(w_i, C_i, \nu_i\right) = \int_0^{w_i} \mathcal{D}\left(z, \nu_i\right) \mathrm{d}z - C_i, \tag{6.6}$$

where the first term in the RHS is the revenue that SU accrues from receiving $\mathtt{W} = w_i$ units of BW, and C_i is the corresponding payment that SU has to make in exchange for receiving w_i. SUs may have different demand profiles. The model adopted in this thesis is the private value model, where the valuation of each SU for spectrum resource is private information to that SU (not known to others) and SUs' valuations are diverse. However, the beliefs that SUs form about their opponents' valuations is a common knowledge and is represented with a CDF $\mathcal{F}_{\nu_1,\cdots,\nu_M}\left(\nu_1, \cdots, \nu_M\right)$.[8] In deriving the payment rule, the assumption made on the choice of the demand profile is as follows:

Assumption 1. *The demand profile $\mathcal{D}\left(w_i, \nu_i\right)$ is monotonically increasing w.r.t. the SU valuation, ν_i, that is, $\partial\mathcal{D}\left(w_i, \nu_i\right)/\partial\nu_i > 0$, and is monotonically decreasing w.r.t. the demanded resource, w_i, that is, $\mathcal{D}\left(w_i, \nu_i\right)/\partial w_i < 0$. Besides, $\mathcal{D}\left(w_i, \nu_i\right)$ is to be finite and twice continuously differentiable.*

6.2.2 Problem Formulation

As the first rule of allocation optimality, the network operator cannot coerce SUs to submit bids, or to allocate bandwidth portions to them that exceed or fall short of their demands. With these constraints in mind, the resource allocation task consists of (i) the assignment of $\left(w_1, \cdots, w_M\right)$ to SUs, where w_i is the allocated BW to i^{th} SU, $i \in \mathcal{M} = \{1, \cdots, M\}$ and (ii) the imposition of prices that SUs have to pay for what they were assigned. This allocation is realized through an auction, started with SUs signaling their demanded BWs $\widetilde{w}_i$s in the form of bid submissions $b_i\left(\widetilde{w}_i\right) \in B_i$ to the SO, followed by the SO assigning the

[8] From another vantage point, the private value model can be construed as if SUs' valuations are derived from the same distribution function.

allocation and pricing pairs $\left[w_i\left(b_i, b_{-i}\right), C_i\left(b_i, b_{-i}\right)\right],\ \forall i \in \mathcal{M}$ and signaling them back to the SUs. Here, b_{-i} is the complementary strategy of SU$_i$'s $M-1$ other opponents. This allotment should always satisfy the following hard constraints:

$$\begin{aligned} &\sum_{i=1}^{M} w_i\left(\mathbf{B}\right) \leq w_0, \\ &w_i\left(\mathbf{B}\right) \geq 0,\ C_i\left(\mathbf{B}\right) \geq 0,\ \forall i \in \mathcal{M}, \end{aligned} \tag{6.7}$$

where $\mathbf{B} = \left(b_1, \cdots, b_M\right) \in \prod_{i=1}^{M} B_i$ is the strategy (bidding) profile of all the SUs and $B_i = \left[\underline{b}_i, \overline{b}_i\right]$ is the strategy space of the i^{th} user.[9]

Among other choices, the strategy of the i^{th} SU is to announce b_i which results in an allocation $\left[w_i\left(b_i, b_{-i}\right), C_i\left(b_i, b_{-i}\right)\right]$ that, in turn, maximizes the utility in (6.6). However, such an allocation is also dependent on other SUs' strategies, i.e., b_{-i}, which are not known to this SU$_i$. Therefore, the allocation game has to be formulated in a Bayesian sense (cf. subsection 6.1.2), whereby the expected utility of the SU, in a private-value setting and ex-interim sense, is defined as

$$u_i\left(x, \nu_i\right) = \underset{\nu_{-i} \in V_{-i}}{\mathcal{E}} \left\langle \mathcal{R}_i\left(w_i\left(b_i\left(x\right), b_{-i}\left(\nu_{-i}\right)\right), \nu_i\right) - C_i\left(b_i\left(x\right), b_{-i}\left(\nu_{-i}\right)\right)\right\rangle, \tag{6.8}$$

where the revenue function $\mathcal{R}_i$ is given by

$$\mathcal{R}_i\left(w_i\left(b_i\left(x\right), b_{-i}\left(\nu_{-i}\right)\right), \nu_i\right) = \int_0^{w_i\left(b_i(x), b_{-i}(\nu_{-i})\right)} \mathcal{D}\left(z, \nu_i\right) \mathrm{d}z. \tag{6.9}$$

As a more elaborated version of (6.6), the representation in (6.8) embodies a number of points: (i) all strategies vary with SUs' valuations, (ii) the i^{th} SU disguises its true valuation

[9] Although for the purpose of conformance with the prevalent game theoretic notation, players' strategy space was represented by $\mathbf{S}$ in Section 6.1.2, hereafter, we adopt the notation $\mathbf{B}$. Moreover, b_i may be interchangeably used instead of $\widetilde{w}_i$ to indicate SU strategy. Particularly, the latter notation is used when it is required to emphasize that SUs' bids are measured in Hz.

ν_i by bidding $b_i(x)$, and (iii) SUs are not necessarily truthful, i.e., $b_i(\nu_i) \neq \nu_i$. This can further be simplified by dissolving b functions into $\mathcal{R}_i$ as

$$u_i(x, \nu_i) = \underset{\nu_{-i} \in V_{-i}}{\mathcal{E}} \langle \mathcal{R}_i(w_i(x, \nu_{-i}), \nu_i) \rangle - C_i(x). \tag{6.10}$$

To have a truthfulness inducing mechanism, whereby SUs have no motivation to confuse the SO, the allocation and pricing rules shall be defined such that revealing the true valuations as bids is always the Nash equilibrium of SUs. In other words,

$$u_i(x = \nu_i, \nu_i) = \max_x u_i(x, \nu_i). \tag{6.11}$$

The similar function arguments on the RHS of (6.11) need careful interpretation. In fact, while the first argument $x = \nu_i$ is the SU's bid that is equal to its valuation by virtue of truthful bidding, the second argument ν_i is the SU's valuation itself which appears directly inside the integrand $\mathcal{D}(z, \nu_i)$ in (6.9). For example, in case of linear revenue profile where $\mathcal{R}_i(w_i(x, \nu_{-i}), \nu_i) = \nu_i \cdot w_i(x, \nu_{-i})$, and once truthfulness is enforced by the proper choice of $[w_i, C_i]$, $u_i(x = \nu_i, \nu_i) = \max_x (\nu_i \cdot w_i(x, \nu_{-i}) - C_i(x)) = \nu_i^{2^{\text{st}}} \cdot w_i\left(\nu_i^{1^{\text{nd}}}, \nu_{-i}\right) - C_i(\nu_i^{1^{\text{nd}}})$, where the notational misuse $\nu_i^{1^{\text{nd}}} = \nu_i^{2^{\text{st}}} = \nu_i$ is only for clarification purpose. Equations (6.10) and (6.11) suggest that truthfulness (a.k.a. IC criterion) may be induced to SUs by the subtle choice of an allocation-pricing pair such that $(w_i, C_i)_{i \in \mathcal{M}}$ is a revelation mechanism. The striking findings in [100] equip us with the powerful tool that the allocation w_i and nonlinear charges C_i in any truth-telling enforcing mechanism are *necessarily* to be related by

$$u_i(x = \nu_i, \nu_i) = u_i(x = 0, \nu_i) + \underset{\nu_{-i} \in V_{-i}}{\mathcal{E}} \langle \int_0^{\nu_i} \mathcal{R}_i^{\text{II}}(w_i(y, \nu_{-i}), y) \mathrm{d}y \rangle, \tag{6.12}$$

which is the total utility that the SU accumulates. The integrand $\mathcal{R}_i^{\text{II}}(\cdot, \cdot)$ is the first derivative of the revenue function $\mathcal{R}_i$ w.r.t. its second argument. Equation (6.12) is the *necessary*

condition for having a truthful allocation mechanism. Now, by plugging (6.9) into (6.12), we have

$$u_i\left(x=\nu_i,\nu_i\right)=u_i\left(x=0,\nu_i\right)+\underset{\nu_{-i}\in V_{-i}}{\mathcal{E}}\langle\int_0^{\nu_i}\int_0^{w_i(y,\nu_{-i})}\frac{\partial\mathcal{D}\left(z,y\right)}{\partial y}\mathrm{d}z\mathrm{d}y\rangle. \tag{6.13}$$

Among the desirable allocation properties (axioms) introduced in at the beginning of this section, IR[10] is fulfilled if $u_i(0,\nu_i)=0$.

The other desirable property is efficiency. Between numerous definitions for allocation efficiency, we would like to provide more bandwidth to an SU who values it more than the one who values it less. Mathematically,

$$\frac{\partial w_i\left(\nu_i,\nu_{-i}\right)}{\partial\nu_i}\geq 0. \tag{6.14}$$

The reason for this choice is that this definition of efficiency features some notion of fairness with itself, as well.

The necessity condition in (6.13) is of little practical use for designing a truthful mechanism as it is already predicated on the truthfulness of the mechanism. Nonetheless, the *sufficient* conditions for an allocation mechanism that comply with (6.13) to be IC is that (i) Assumption 1, (ii) equation (6.14), and (iii) $u_i\left(0,0\right)\geq 0$, are simultaneously satisfied. Therefore,

$$\begin{cases} u_i(x,\nu_i) \text{ given by } (6.12) \\ u_i\left(0,\nu_i\right)=0 \\ \dfrac{\partial\mathcal{D}\left(w_i,\nu_i\right)}{\partial\nu_i}>0 \\ \dfrac{\partial w_i\left(\nu_i,\nu_{-i}\right)}{\partial\nu_i}\geq 0 \end{cases} \iff \begin{cases} \text{Incentive Compatibility} \\ \text{Individual Rationality} \\ \text{Fairness} \\ \text{Efficiency} \end{cases} \tag{6.15}$$

[10]This means that an SU whose valuation of BW is zero and announces its true valuation shall never be penalized by the mechanism through the imposition of a price such that it ends up with negative utility.

The reason that truthfulness is an important property is entangled with enhancing the spectral usage as the underlying objective of DSA. This objective, which is aligned with the incentives of the SO, who seeks to maximize its profit by selling more of the resources it possesses, cannot be realized through a mechanism whereby a number of self-motivated and rational (but dishonest) SUs do their best to gain advantages. In this situation, the SO has very little knowledge and limited control that can be leveraged to improve the efficiency of the resource allocation. Therefore, it is important for the SO to incite SUs to truthful behavior and strive to gain understanding about their private valuations. For that to be possible, the nature of a SU's valuation ν_i needs to be better understood. This matter is tackled next.

6.2.3 Secondary User's Valuation

The valuation ν_i varies depending on the task to be fulfilled. For instance, while ν_i of a CR terminal (SU) may be characterized with QoS metrics such as time delay inverse, throughput and frame success rate, a CR relay's aptness is centered around routing-table convergence rate, path shortness, hop-count inverse, etc. Nevertheless, one thing is known for sure: the valuation ν_i of an SU is not a stand-alone independent variable but a function of a number of underlying qualities, most notably the allocated BW w_i. Therefore, the solution to an SU's local maximization problem (not to be confused with the SO's revenue maximization) is to bid $b_i = \widetilde{w}_i$ that maximizes $\nu_i\left(w_i\left(\widetilde{w}_i\left(\cdots\left(\cdots\right)\right)\right)\right)$. This composite infinite interdependency of variables is descriptive of the sequential and chronological game-theoretic reasoning process which happens between players.

With this exposition in mind, a very natural choice in DSA context is the *instantaneous* spectral efficiency $k_i(\mathrm{bit/s/Hz})$ or the *average modified* spectral efficiency $\mathcal{K}_i(\mathrm{bit/s/Hz})$.[11]

[11] Note that $\mathcal{K}_i < k_i$ as $\mathcal{K}_i$ accounts for the fraction of information that is successfully transmitted, which is why the term $(1 - \mathrm{SER}_{m,i})$ appears inside the summation in (6.17). Nonetheless, from now on, we simply call it spectral efficiency.

Optimizing the latter utility function is an expectation maximization (EM) problem. The idea is clarified with two examples here.

Case I: $\mathcal{D}_i\left(w_i, \nu_i\right) = g_i\left(\nu_i\right)$

If the demand profile is only a function of the valuation, the revenue function becomes

$$\mathcal{R}_i = \underset{\nu_{-i}}{\mathcal{E}} \left\langle w_i\left(\nu_i, \nu_{-i}\right) \cdot g_i\left(\nu_i\right)\right\rangle = g_i\left(\nu_i\right) \cdot \underset{\nu_{-i}}{\mathcal{E}} \left\langle w_i\left(\nu_i, \nu_{-i}\right)\right\rangle, \tag{6.16}$$

which is the model assumed in many papers, such as [114], where the revenue function is defined as $\mathcal{R}_i = \mathcal{K}_i r_i w_i$ with $r_i(\$/\text{bit/s})$ denoting the monetary valuation per bit rate. Obviously, by taking $g_i\left(\nu_i\right) \equiv \mathcal{K}_i r_i$, and naming $\mathcal{K}_i$ as the average spectral efficiency, (6.9) boils down to (6.16).

In a cognitive environment, where SUs change their coding/modulation transmission mode in response to varying traffic and channel conditions, $\mathcal{K}_i$ can be mathematically stated as

$$\mathcal{K}_i = \sum_{m=1}^{\mathfrak{M}} k_{m,i} \Pr\left(m\right)\left(1 - \text{SER}_{m,i}\right), \tag{6.17}$$

where $\mathfrak{M}$ is the number of modulation/coding levels[12] and where $k_{m,i}$, $\Pr\left(m\right)$ and $\text{SER}_{m,i}$ represent, respectively, the instantaneous spectral efficiency, its corresponding occurrence probability and the symbol error rate (SER) of SU_i, when the m^{th} transmission mode is used. For coherent modulation, the approximation $\text{SER}_{m,i} \cong \alpha_m \mathbb{Q}\left(\sqrt{\beta_m \text{SINR}_i}\right)$ [124] can be used to obtain a closed-form expression for $\mathcal{K}_i$, whereby α_m and β_m depend on the modulation type[13] and $\mathbb{Q}(\cdot)$ is the Gaussian Q-function. Also, SINR_i symbolizes the SINR,

[12] Please note the distinct usage of $\mathfrak{M}$ as the number of "WRAN cells" in Chapter 5 and as the number of "coding/modulation" combinations in this chapter.

[13] Please note the distinct usages of α as the "mixing level" in modeling Chapter 2, as the "discount factor" for the calculation of confidence metrics in Chapter 5 and as the signal "modulation parameter" in this chapter.

which, in the most general form, is expressed by

$$\mathrm{SINR}_i = \frac{h_{i,i}^2 \mathcal{S}_i^{(\mathrm{Tr.})}}{\sum\limits_{j \in \mathcal{M}_{-i}} h_{j,i}^2 \mathcal{S}_j^{(\mathrm{Tr.})} + h_{0,i}^2 \mathcal{S}_{0,i}^{(\mathrm{Tr.})} + \mathcal{S}_{N,i}}, \tag{6.18}$$

where $\mathcal{M}_{-i} = \mathcal{M} \setminus \{i\}$ and symbols $\mathcal{S}_i^{(\mathrm{Tr.})}$, $\mathcal{S}_j^{(\mathrm{Tr.})}$ and $\mathcal{S}_{0,i}^{(\mathrm{Tr.})}$ denote the transmit power from the desired transmitter, from the j^{th} interferer, and from the PU, respectively. Also, $h_{k,i},\ k \in \mathcal{M} \cup \{0\} = \mathcal{M}^{+0}$ accounts for the long- and short-term channel fading and free-space propagation loss (FSPL) in the direct channel between the i^{th} transmitter/receiver pair ($h_{i,i}$), the cognitive interference channel between the j^{th} interferer and the i^{th} receiver ($h_{j,i},\ j \in \mathcal{M}_{-i}$), and the primary interference channel ($h_{0,i}$). These coefficients are RVs with PDFs that depend on the propagation properties of the transmission environment. Moreover, the noise power is $\mathcal{S}_{N,i} = \mathtt{N}_0 w_i$ where $\mathtt{N}_0$ is the noise spectral density and w_i is the SU_i allocated BW.

Case II: $\mathcal{R}_i\left(w_i, \nu_i\right) = \underset{\nu_{-i}}{\mathcal{E}} \left\langle w_i\left(\nu_i, \nu_{-i}\right) \log(1 + \frac{\mathcal{S}_i}{\mathtt{N}_0 w_i\left(\nu_i, \nu_{-i}\right)}) \right\rangle$

In this case, the revenue of the SU is proportional to its Shannon capacity. It can be shown that the demand profile corresponding to this revenue function is $\mathcal{D}\left(w_i, \nu_i\right) \approx \log(1 + \mathcal{S}_i / \mathtt{N}_0 w_i\left(\nu_i, \nu_{-i}\right)) - 1$, which satisfies the condition $\partial \mathcal{D}\left(w_i, \nu_i\right) / \partial w_i < 0$ in Assumption 1.

$\mathcal{K}_i$ in Rayleigh Channel

In a rich scattering environment, the histograms of channel gains $h_{k,i},\ k \in \mathcal{M}^{+0}$ in (6.18) fit a Rayleigh distribution with PDF

$$f_{h_{k,i}}(h) = \frac{h}{\overline{h}_{k,i}^2} e^{-\frac{h^2}{2\overline{h}_{k,i}^2}}, \qquad h \in [0, \infty],\ \forall k \in \mathcal{M}^{+0},\ \forall i \in \mathcal{M}, \tag{6.19}$$

where $\overline{h}_{k,i}$ is the average fading level. In this situation, obtaining a closed-form expression for $\mathcal{K}_i$ in (6.17) requires precise characterization of $\mathrm{SER}_{m,i}$ and $\Pr(m)$. Conditioned on the interference channel coefficients $h_{k,i}$ in the denominator of (6.18), SINR_i follows a *chi-square* distribution with two degrees of freedom. Thus, the SER at the receiver in a Rayleigh fading environment is given by

$$\begin{aligned}\underset{h_{i,i}}{\mathcal{E}}\left\langle \mathrm{SER}\left(h_{i,i}|h_{k,i},\forall k\in\mathcal{M}_{-i}^{+0}\right)\right\rangle =& \int_0^{\infty}\alpha_m\mathbb{Q}\left(\sqrt{\beta_m\mathrm{SINR}_i}\right)\frac{1}{\overline{\mathrm{SINR}}_i}e^{-\frac{\mathrm{SINR}_i}{\overline{\mathrm{SINR}}_i}}\mathrm{d}\mathrm{SINR}_i\\ =&\frac{\alpha_m}{2}\left(1-\sqrt{\frac{0.5\beta_m\overline{\mathrm{SINR}}_i}{1+0.5\beta_m\overline{\mathrm{SINR}}_i}}\right)\\ \approx&\frac{\alpha_m}{2\beta_m\overline{\mathrm{SINR}}_i}\end{aligned}, \tag{6.20}$$

which signifies that SER is inversely proportional to the received $\overline{\mathrm{SINR}}_i = 2\overline{h}_{i,i}^2\mathcal{S}_i^{(\mathrm{Tr.})}/(\mathcal{S}_{N,i}+\sum_{k\in\mathcal{M}_{-i}^{+0}}h_{k,i}^2\mathcal{S}_k^{(\mathrm{Tr.})})$. The notation $\mathcal{M}\cup\{0\}\setminus\{i\}=\mathcal{M}_{-i}^{+0}$ is chosen for brevity. By eliminating the conditions in (6.20) w.r.t. $h_{k,i},\forall k\in\mathcal{M}_{-i}^{+0}$ and performing the change of variable $H_{k,i}=h_{k,i}^2$, the SER in (6.17) can be written as

$$\mathrm{SER}_{m,i}=\int_0^{\infty}\cdots\int_0^{\infty}\frac{\alpha_m}{4\beta_m\overline{h}_{i,i}^2}\left(\frac{\sum_{j\in\mathcal{M}_{-i}^{+0}}H_{j,i}\mathcal{S}_j^{(\mathrm{Tr.})}+\mathcal{S}_{N,i}}{\mathcal{S}_i^{(\mathrm{Tr.})}}\right)f_{H_{0,i}}(\theta_{0,i})\cdots f_{H_{M,i}}(\theta_{M,i})\,\mathrm{d}\theta_{0,i}\cdots\mathrm{d}\theta_{M,i}, \tag{6.21}$$

where it can be demonstrated, using (6.19), that the $H_{k,i}$s are exponentially distributed RVs with mean $2\overline{h}_{k,i}^2$. Thereby, (6.21) is simplified as

$$\mathrm{SER}_{m,i} = \frac{\alpha_m}{4\beta_m \overline{h}_{i,i}^2} \left(\sum_{j \in \mathcal{M}_{-i}^{+0}} \frac{2\mathcal{S}_j^{(\mathrm{Tr.})}}{\mathcal{S}_i^{(\mathrm{Tr.})}} \overline{h}_{j,i}^2 + \frac{\mathcal{S}_{N,i}}{\mathcal{S}_i^{(\mathrm{Tr.})}} \right). \tag{6.22}$$

Given that $\mathcal{S}_{N,i} = \mathtt{N}_0 w_i$, SER is an increasing function of the SU's BW.

We can also find a mathematical expression for $\Pr(m)$ in (6.17), which is the probability that the m^{th} modulation/coding transmission mode with instantaneous spectral efficiency $k_{m,i}$ is chosen. It should be noted that the above mode selection is merely done based on the SINR, according to

$$\Pr(m) = \begin{cases} \Pr\left(\mathrm{SINR}_m^{\min} \leq \mathrm{SINR}_i < \mathrm{SINR}_{m+1}^{\min}\right), & m \in \{0, \cdots, \mathfrak{M}-1\} \\ 1 - \Pr\left(\mathrm{SINR}_i < \mathrm{SINR}_{\mathfrak{m}+1}^{\min}\right), & m = \mathfrak{M}-1 \end{cases}, \tag{6.23}$$

wherein $\Pr\left(\mathrm{SINR}_m^{\min} \leq \mathrm{SINR}_i < \mathrm{SINR}_{m+1}^{\min}\right) = \Pr\left(\mathrm{SINR}_i < \mathrm{SINR}_{m+1}^{\min}\right) - \Pr\left(\mathrm{SINR}_i < \mathrm{SINR}_m^{\min}\right)$ and $\mathrm{SINR}_m^{\min}$s' are ordered SINR thresholds. This can be further simplified if the channel is Rayleigh fading, as

$$\begin{aligned} \Pr\left(\mathrm{SINR}_i < \mathrm{SINR}_{m+1}^{\min}\right) &= \Pr\left(h_{i,i}^2 < \frac{\sum\limits_{j \in \mathcal{M}_{-i}} h_{j,i}^2 \mathcal{S}_j^{(\mathrm{Tr.})} + h_{0,i}^2 \mathcal{S}_{0,i}^{(\mathrm{Tr.})} + \mathcal{S}_{N,i}}{\mathcal{S}_i^{(\mathrm{Tr.})}} \mathrm{SINR}_{m+1}^{\min} \right) \\ &= 1 - e^{-\frac{\mathrm{SINR}_{m+1}^{\min} \mathcal{S}_{N,i}}{2\overline{h}_{i,i}^2 \mathcal{S}_i^{(\mathrm{Tr.})}}} \prod_{j \in \mathcal{M}_{-i}^{+0}} \frac{1}{\frac{\mathcal{S}_j^{(\mathrm{Tr.})}}{\mathcal{S}_i^{(\mathrm{Tr.})}} \left(\frac{\overline{h}_{j,i}}{\overline{h}_{i,i}} \right)^2 \mathrm{SINR}_{m+1}^{\min} + 1}. \end{aligned} \tag{6.24}$$

A similar result is obtained for $\Pr\left(\mathrm{SINR}_i < \mathrm{SINR}_m^{\min}\right)$. By plugging both expressions into

(6.23), $\Pr(m)$ is obtained. If the PU and SUs operate in the same environment ($\overline{h}_{j,i} = \overline{h}$) transmitting with the same power ($\mathcal{S}_j^{(\mathrm{Tr.})} = \mathcal{S}^{(\mathrm{Tr.})}$) and noise is negligible ($\mathcal{S}_{N,i} = 0$), we have

$$\Pr(m) = \begin{cases} \left(\dfrac{1}{1+\mathrm{SINR}_m^{\min}}\right)^{M-1} - \left(\dfrac{1}{1+\mathrm{SINR}_{m+1}^{\min}}\right)^{M-1}, & m \in \{0, \cdots, \mathfrak{M}-1\} \\ 1 - \left(\dfrac{1}{1+\mathrm{SINR}_m^{\min}}\right)^{M-1}, & m = \mathfrak{M}-1 \end{cases}. \tag{6.25}$$

Knowing that $\mathrm{SINR}_m^{\min} < \mathrm{SINR}_{m+1}^{\min}$, one can verify that $0 < \Pr(m) < 1$. By plugging (6.22) and (6.25) into (6.17) in order to derive the expression for $\mathcal{R}_i = g_i(\nu_i)\, w_i$, and assuming $g_i(\nu_i) \equiv \mathcal{K}_i$ (Case I), we get

$$\begin{gathered} \mathcal{R}_i = \textstyle\sum_{m=1}^{\mathfrak{M}} L_{m,i}(w_i) \\ L_{m,i}(w_i) = k_{m,i}\left(\mathfrak{K}_{m,i}^3 e^{-\mathfrak{K}_{m,i}^4 w_i} - \mathfrak{K}_{m+1,i}^3 e^{-\mathfrak{K}_{m+1,i}^4 w_i}\right)\left(\mathfrak{K}_{m,i}^2 - \mathfrak{K}_{m,i}^1 w_i\right) w_i \end{gathered}, \tag{6.26}$$

where

$$\begin{aligned} \mathfrak{K}_{m,i}^3 &= \prod_{j \in \mathcal{M}_{-i}^{+0}} \frac{1}{\dfrac{\mathcal{S}_j^{(\mathrm{Tr.})}}{\mathcal{S}_i^{(\mathrm{Tr.})}}\left(\dfrac{\overline{h}_{j,i}}{\overline{h}_{i,i}}\right)^2 \mathrm{SINR}_m^{\min} + 1}, \quad \mathfrak{K}_{m,i}^4 = \frac{\mathrm{SINR}_m^{\min} \mathtt{N}_0}{2\overline{h}_{i,i}^2 \mathcal{S}_i^{(\mathrm{Tr.})}}, \\ \mathfrak{K}_{m,i}^2 &= 1 - \frac{\alpha_m}{2\beta_m} \sum_{j \in \mathcal{M}_{-i}^{+0}} \frac{\mathcal{S}_j^{(\mathrm{Tr.})}}{\mathcal{S}_i^{(\mathrm{Tr.})}} \frac{\overline{h}_{j,i}^2}{\overline{h}_{i,i}^2}, \qquad \mathfrak{K}_{m,i}^1 = \frac{\alpha_m}{4\beta_m \overline{h}_{i,i}^2} \frac{\mathtt{N}_0}{\mathcal{S}_i^{(\mathrm{Tr.})}}. \end{aligned} \tag{6.27}$$

In reality, SUs do not have much difficulty to find w_is that maximize their revenues. In the special case above, $\mathcal{R}_i$ is a concave function (w.r.t. w_i) because summand terms $L_{m,i}(w_i)$ in (6.26) are concave (sum of a number of concave functions is concave). Even though this is not too easy to demonstrate, there is a crude way to verify it by using the

approximation $e^{-\mathfrak{K}^4_{m,i}w_i} \cong 1-\mathfrak{K}^4_{m,i}w_i$ (which is tight enough for $\mathfrak{K}^4_{m,i} \ll 1$) and, then, differentiating $L_{m,i}(w_i)$ twice w.r.t. w_i. Hence,

$$\frac{\partial^2 L_{m,i}(w_i)}{\partial w_i^2} = 2\left(\underbrace{\mathfrak{K}^3_{m,i}\mathfrak{K}^4_{m,i} - \mathfrak{K}^3_{m+1,i}\mathfrak{K}^4_{m+1,i}}_{<0}\right)\left(3\mathfrak{K}^1_{m,i}w_i - \mathfrak{K}^2_{m,i}\right) - 2\underbrace{\mathfrak{K}^1_{m,i}}_{>0}\left(\underbrace{\mathfrak{K}^3_{m,i} - \mathfrak{K}^3_{m+1,i}}_{<0}\right) . \tag{6.28}$$

The RHS of (6.28) is always negative when $3\mathfrak{K}^1_{m,i}w_i - \mathfrak{K}^2_{m,i} > 0$ (cf. Eq. (6.29)), which is a valid predicate for a practical choice of parameters.

$$3\mathfrak{K}^1_{m,i}w_i - \mathfrak{K}^2_{m,i} = \frac{\alpha_m}{2\beta_m}\left(\sum_{\substack{j\in\mathcal{M}\cup\{0\}\\ j\neq i}} \frac{\mathcal{S}_j^{(\mathrm{Tr.})}}{\mathcal{S}_i^{(\mathrm{Tr.})}}\frac{\overline{h}^2_{j,i}}{\overline{h}^2_{i,i}} + \frac{3\mathtt{N}_0}{2\mathcal{S}_i^{(\mathrm{Tr.})}\overline{h}^2_{i,i}}w_i\right) - 1 > 0. \tag{6.29}$$

This signifies that a single revenue maximizing point always exists for each SU, which can practically be found. It should be emphasized that the concavity of the valuation function as verified above is not a requirement for the validity of the allocation mechanism proposed in this chapter.

6.2.4 Secondary Operator's Perspective

From the vantage point of the SO, the truncated version of the sequential reasoning effect is represented by

$$\begin{matrix} w_i^* = w_i^*\left(\widetilde{w}_i\left(\nu_i\left(w_i,\cdots\right)\right), \widetilde{w}_{-i}\left(\nu_{-i}\left(w_{-i},\cdots\right)\right)\right) \\ C_i^* = C_i^*\left(\widetilde{w}_i\left(\nu_i\left(w_i,\cdots\right)\right), \widetilde{w}_{-i}\left(\nu_{-i}\left(w_{-i},\cdots\right)\right)\right) \end{matrix}, \quad \forall i \in \mathcal{M} , \tag{6.30}$$

where $\widetilde{w}_i$ and w_i^* are SU$_i$'s bidden and allocated BW, respectively. The innermost arguments in (6.30) signify the earliest stage of auction, where valuation ν_i of the SU$_i$ depends on many factors, prominently, on its BW needs w_i. As discussed in the previous subsection, this valuation function $\nu_i(w_i)$ has its maximum at $w_i = \breve{w}_i$ of which only SU$_i$ is aware. We recall from (6.17) that $\nu_i(w_i)$ is also represented by the spectral efficiency $\mathcal{K}_i$. Since the allocation mechanism enforces truthfulness, then $\widetilde{w}_i = \breve{w}_i$. Once the SO receives all bids $\widetilde{w}_i$, it determines allocations $[w_i^*,\ C_i^*]$ for all the SUs.

6.2.5 Nonlinear Pricing Rule

Based on the discussion in subsection 6.2.2, to induce *incentive compatibility*, the payment functions $C_i\left(\nu_i\right), \forall i \in \mathcal{M}$, are to be determined such that SUs' utilities always obey (6.13). Obviously, these payments are nonlinear w.r.t. the SUs' demanded bandwidth. By plugging (6.10) into (6.12), we get

$$C_i\left(\nu_i\right) = \underset{\nu_{-i}\in V_{-i}}{\mathcal{E}} \langle \mathcal{R}_i\left(w_i\left(\nu_i,\nu_{-i}\right),\nu_i\right) - \int_{\epsilon_i^+}^{\nu_i} \mathcal{R}_i^{\mathrm{II}}\left(w_i\left(y,\nu_{-i}\right),y\right)\mathrm{d}y\rangle - u_i\left(\epsilon_i^-,\nu_i\right), \quad \forall i \in \mathcal{M}, \tag{6.31}$$

where $\mathcal{R}_i^{\mathrm{II}}\left(\cdot,\cdot\right)$ is the first derivative of the revenue function $\mathcal{R}_i$ w.r.t. its second argument. Symbol ϵ_i in (6.31) satisfies the *individual rationality* (IR) condition that acts as a clearing reserved price under which the SO is not willing to allocate spectrum to SU$_i$ [116]. In other words, if $\nu_i \leq \epsilon_i^+$, then $w_i^*(\nu_i,\nu_{-i}) = 0$ and $C_i(\nu_i) = 0$. Note that ϵ_i^+ and ϵ_i^- are the one-sided limits of ϵ_i from above and below, respectively.

The expression in (6.31) has an intuitive interpretation. The first term on the RHS is the amount that a fully discriminating monopolist SO would charge an SU when perfect information is available. The second term is the revenue loss that the SO, with imperfect information, should undergo in exchange of diminishing uncertainties, by enforcing a payment

that incentivizes the SU reveal its true valuation. Further, the formula in (6.31) is the generalization of known truthful pricing mechanisms, such as Vickrey-Clarke-Groves (VCG) auction [98]. This claim needs some elaboration. In the VCG mechanism, the non-divisible BW w_0 is to be allocated to the SU with the highest valuation who has to pay the bid of the runner-up SU (i.e., the 2^{nd} highest-price SU). To verify this using (6.31), it should be noted that the demand profile in VCG auction is a delta function, i.e., $\mathcal{D}(w_i, \nu_i) = \nu_i \delta(w_i - w_0)$. Since SUs other than the one with the highest valuation (say SU_j) are not to receive any allocation, i.e., $w_i = 0 \ \ \forall i \in \mathcal{M}_{-j}$, then (6.31) becomes

$$\begin{aligned} C_i(\nu_i) = & \underset{\nu_{-i} \in V_{-i}}{\mathcal{E}} \langle \int_0^0 \nu_i \delta(z - w_0)\, \mathrm{d}z \rangle - \\ & \underset{\nu_{-i} \in V_{-i}}{\mathcal{E}} \langle \int_{\epsilon_i^+}^{\nu_i} \int_0^{w_i(y,\nu_{-i})} \delta(z - w_0) \mathrm{d}z \mathrm{d}y \rangle - u_i\left(\epsilon_i^-, \nu_i\right), \ \ \forall i \in \mathcal{M}_{-j} \end{aligned} . \tag{6.32}$$

In (6.32), the first term of the RHS is zero. Concerning the second term, the efficiency criterion in (6.14) enforces that a bidder who already lost by bidding ν_i should have lost again had he bidden with a lower valuation, i.e. $\epsilon_i^+ \leq y < \nu_i \Rightarrow w_i(y, \nu_{-i}) = 0$. Hence, the second term would also be zero. As will be explained shortly, satisfying the IR property requires the third term to be zero as well. Therefore, the losers of the auction gain and pay nothing. On the contrary, SU_j with the highest valuation receives the entire BW for the price

$$\begin{aligned} C_j(\nu_j) = & \underset{\nu_{-j} \in V_{-j}}{\mathcal{E}} \langle \int_0^{w_0} \nu_j \delta(z - w_0)\, \mathrm{d}z - \int_{\nu_k}^{\nu_j} \int_0^{w_0} \delta(z - w_0) \mathrm{d}z \mathrm{d}y - \\ & \int_{\epsilon_i^+}^{\nu_k} \int_0^0 \delta(z - w_0) \mathrm{d}z \mathrm{d}y \rangle, \end{aligned} \tag{6.33}$$

where $k = \underset{l \neq j,\, l \in \mathcal{M}}{\text{argmax}}\ \nu_l$. Of course, (6.33) indicates that $C_j(\nu_j) = \nu_k$. That is, in any single-unit truthful mechanism, the winner of the competition should always pay the 2^{nd} highest price. A similar implication can be made for a multi-unit VCG substantiating that (6.31) is

a generalized truthfulness inducing payment rule.

6.2.6 Pricing Rule: Digest of Findings

The bidding strategy of SUs in an optimal allocation mechanism is to bid $\widetilde{w}_i$s they need. On a different note, since it is postulated that the SUs have distinct valuations and that there is no cooperation or collusion among them, they cannot, logically, have any information about each other's bids, about the total number of participating SUs, about the net BW budget w_0 of the SO, etc. Subsequently, the SUs would only be able to announce their $\widetilde{w}_i$s on the basis of their local knowledge. Worse comes to worse, in *interweave* cognitive scenarios, SO should remain transparent to the primary network, thus, the amount of resource available to the SO is not even known with certainty. For these two reasons, at every allocation auctioning round, which may be held hundred times in a second, it is not a-priori known whether $\sum_{i=1}^{M} \widetilde{w}_i$ exceeds w_0 or falls short of it. Once the bids are submitted, the SO calculates the allocation pairs $[w_i^*, C_i^*]$ for all the SUs. If $w_i^* = \widetilde{w}_i$, then SU$_i$ cannot reject the offer of the SO. However, if $w_i^* \neq \widetilde{w}_i$, then SU$_i$ may reject the offer as it is not what he demanded. This allocation procedure, which is schematically depicted in Fig. 6.5, is described as follows:

(S.l) SO generates the demand profile $\mathcal{D}(w_i, \nu_i)$ and updates it, progressively, based on the history of allocations.

(S.2) SUs submit their demanded BW $\widetilde{w}_i$ simultaneously, during intervals that are exclusively assigned for the auctioning purpose. This requires that the MAC sublayer divides time to alternating access (payload transmission) and allocation (auctioning) intervals. This can be handled in a variety of ways but are not within the scope of this thesis.

(S.3) SO determines the optimal allocations w_i^*s based on the bid submissions $\widetilde{w}_i$s. The

optimal assignment problem is addressed in Section 6.3.

(S.4) Based on the amount of the allocated resources, payments C_i^* are calculated for each SU according to (6.31), which requires estimating the SUs' valuations $\widetilde{\nu}_i$ as well as a ϕ function. The approach for finding these estimates is discussed in Section 6.3.

$$C_i^*(\nu_i) = \underset{\nu_{-i}\in V_{-i}}{\mathcal{E}} \langle \int_0^{w_i^*} \mathcal{D}(z, \widetilde{\nu}_i)\,\mathrm{d}z - \int_{\epsilon_i^+}^{\widetilde{\nu}_i} \int_0^{\psi(z)} \frac{\partial \mathcal{D}(y,z)}{\partial z}\mathrm{d}y\mathrm{d}z \rangle, \quad \forall i \in \mathcal{M}, \tag{6.34}$$

(S.5) Allocation pair $[w_i^*, C_i^*]$ is immediately announced to SU_i, whereupon he can use w_i^* for the duration specified by the MAC mechanism in exchange for C_i^* payable to the SO.

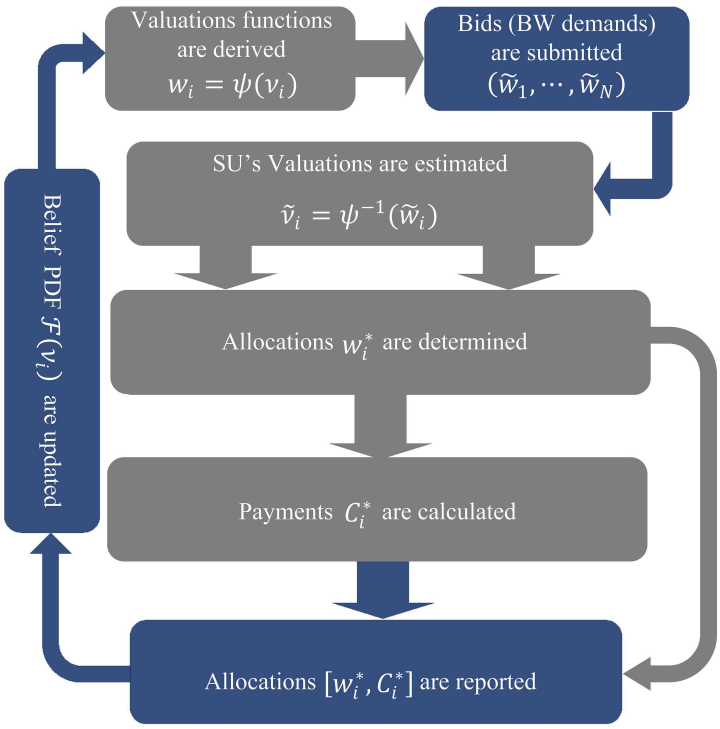


Figure 6.5: Resource allocation block diagram. While the blue-colored blocks in this figure are dealt with in this section, the grey-colored ones are treated in Section 6.3.

The above steps will be fully delineated in subsection 6.3.5 once we know how the solution to the spectrum assignment problem, as well.

6.3 Spectrum Assignment

In the previous section, a nonlinear spectrum pricing rule was defined to incorporate axioms incentive compatibility (IC), individual rationality (IR) and fairness into the allocation mechanism. In this section, we integrate the last three axioms, i.e., efficiency, revenue maximization (RM) and computational manageability (CM), into the allocation mechanism using a particular spectrum assignment algorithm. The overall allocation mechanism is **efficient**, as it allocates spectrum with a higher likelihood to those SUs who value it more, and it also maximizes the operator's total revenue, a desirable property known as **revenue-maximization**. The solution to the spectrum allocation problem having these properties is obtained by root-finding operations and solving almost linear system of equations. This means that the proposed spectrum allocation mechanism is **computationally manageable**, thus, highly scalable. For better clarification, Fig. 6.6 shows these six axioms wherein those approached in this section are colored in blue. These six attributes make the proposed resource allocation mechanism an ideal solution for deployment in 5G networks.

6.3.1 Problem Restatement

We restate the formulation of the auction-based allocation problem initiated in Section 6.2 as follows: The strategies of SUs are to submit payoff-maximizing bids $\widetilde{w}_i, \cdots, \widetilde{w}_M$ to the SO. We mentioned that a bid is, per se, a mapping from the SU's valuation space V_i to its action space B_i. The SU has full knowledge of its own type and forms beliefs about other SUs' types. The purpose of the SO is to design the auction mechanism such that its accrued revenue is maximized. This being the case, an allocation of BW in MC-CRNs is an assignment $(w_1^*, \cdots, w_M^*)$ of spectrum to SUs and enforcement of charges $(C_1^*, \cdots, C_M^*)$, in exchange. SUs' allocations shall entirely depend on their willingness to participate and the behavior they exhibit through submitting their best responses. A number of hard constraints were introduced in (6.7).

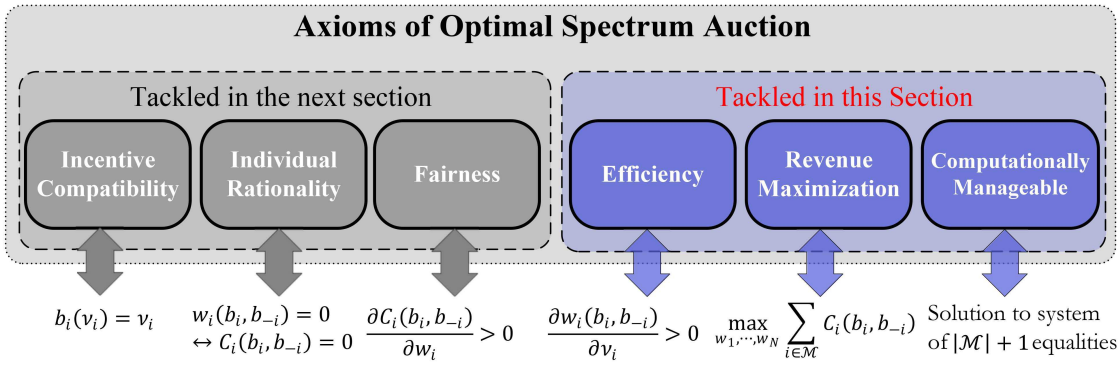


Figure 6.6: The axioms of the proposed spectrum allocation mechanism.

The SU_i plays b_i such that the succeeding allocation $\left[w_i^*\left(b_i, b_{-i}\right), C_i^*\left(b_i, b_{-i}\right)\right]$ maximizes its utility function defined in (6.6). Despite that, this choice may not, necessarily, maximize SU_i's utility as every allocation depends on other SUs' strategies, i.e., b_{-i}, too, which is not known to SU_i in a *deterministic* manner. Therefore, the allocation problem has to be formulated as a Bayesian game as per (6.8). Earlier in this chapter, we claimed that if the prices are derived from (6.31), SUs will act truthfully by revealing their true valuations.

Besides the first assumption made in subsection 6.2.1, the following three assumptions on the choice of the demand profile $\mathcal{D}\left(w_i, \nu_i\right)$ are leveraged in this section.

Assumption 2. *The SUs' belief CDF is increasing and continuously differentiable. That is, given that* $\nu_i \in [0, \overline{\nu}_i]$, *then* $\partial \mathcal{F}_{\nu_i}\left(\nu_i\right)/\partial \nu_i > 0$, $\mathcal{F}_{\nu_i}\left(0\right) = 0$ *and* $\mathcal{F}_{\nu_i}\left(\bar{\nu}_i\right) = 1$.

Assumption 3. *The SUs have non-decreasing demand elasticity, i.e.,* $\partial[(w_i/\mathcal{D}(w_i, \nu_i)) \cdot \partial \mathcal{D}(w_i, \nu_i)/\partial w_i]/\partial \nu_i \geq 0$, *hence, are risk-averse. It can also be shown that*

$$\frac{\partial}{\partial \nu_i}\left(\frac{\frac{\partial^2 u_i}{\partial w_i^2}}{\frac{\partial u_i}{\partial w_i}}\right) \geq 0 \rightarrow \frac{1}{w_i}\frac{\partial}{\partial \nu_i}\left(-\frac{w_i}{\mathcal{D}\left(w_i, \nu_i\right)}\frac{\partial \mathcal{D}\left(w_i, \nu_i\right)}{\partial w_i}\right) \leq 0, \tag{6.35}$$

where $\text{RA}\left(w_i, \nu_i\right) = -(\partial^2 u_i/\partial w_i^2)/(\partial u_i/\partial w_i)$ *on the LHS is known as the risk aversion*

factor.

Obviously, for concave utility functions, where $\partial^2 u_i/\partial w_i^2 \leq 0$, if $\partial u_i/\partial w_i > 0$ (as shown in Fig. 6.4), then $\mathrm{RA}\,(w_i, \nu_i) \geq 0$ [100].

Assumption 4. *The demand profile is a concave function w.r.t. the valuation,* $\partial^2 \mathcal{D}\,(w_i, \nu_i)/\partial \nu_i^2 \leq 0$.

6.3.2 Revenue Maximization

The fairness, efficiency, IC, and IR axioms were already integrated into the proposed allocation mechanism. To incorporate the next property, that is revenue maximization (RM), BW allocations $w_i\,(\nu_i, \nu_{-i})\,, i \in \mathcal{M}$, should be chosen such that the *total* payment $\mathcal{U}^{\mathbf{P}}$ collected from all SUs is maximized. Please note that the RM axiom is one of the strong definitions of optimality in auction theory. These allocations are also needed for the exact determination of payments $C_i^*(\nu_i)$, as it is evident from (6.31). Mathematically speaking, the optimal allocation $\widehat{\mathbf{w}}^* = [w_1^*, \cdots, w_M^*]$ is the solution to the following optimization problem.

$$\begin{aligned} \text{Maximize} \sum_{i\in\mathcal{M}} C_i\,(\nu_i) \\ \text{s.t.} \quad \sum_{i\in\mathcal{M}} w_i \leq w_0,\ \frac{\partial w_j}{\partial \nu_j} \geq 0,\ w_j \geq 0,\ \ \forall j \in \mathcal{M} \end{aligned}\ , \tag{6.36}$$

where w_0 is a stochastic process representing the total instantaneous spectrum width (bandwidth *budget*) in the possession of the SO. Since SO has no deterministic information about SUs' valuations, the optimization in (6.36) is to be conducted in *ex-ante* sense. Thereby, after plugging (6.31) into (6.36), the unconstrained problem becomes

$$\mathcal{U}^{\mathbf{P}} = \max_{w_1,\cdots,w_M} \sum_{i=1}^{M} \left(\underset{\nu_i,\nu_{-i}}{\mathcal{E}} \langle \mathcal{R}_i\,(w_i, \nu_i) - \int_{\epsilon_i^+}^{\nu_i} \mathcal{R}_i^{\mathrm{II}}\,(w_i\,(y, \nu_{-i})\,, y)\mathrm{d}y \rangle - u_i\left(\epsilon_i^-, \nu_i\right) \right), \tag{6.37}$$

which, after integration-by-part, transforms to

$$\mathcal{U}^{\mathbf{P}}=\max_{w_1,\cdots,w_M}\ \sum_{i=1}^{M}\left(\underset{\nu_i,\nu_{-i}}{\mathcal{E}}\langle\mathcal{R}_i\left(w_i,\nu_i\right)-\ \mathcal{R}_i^{\mathrm{II}}\left(w_i\left(y,\nu_{-i}\right),y\right)\big|_{y=\nu_i}\cdot\frac{1}{\Omega(\nu_i)}\rangle-u_i\left(\epsilon_i^{-},\nu_i\right)\right), \tag{6.38}$$

where $\Omega(\nu_i)$, a.k.a. the hazard or failure rate of belief, is defined by

$$\Omega\left(\nu_i\right)=\frac{f_{\nu_i}\left(\nu_i\right)}{1-\mathcal{F}_{\nu_i}\left(\nu_i\right)}, \tag{6.39}$$

with $f_{\nu_i}\left(\nu_i\right)$ and $\mathcal{F}_{\nu_i}\left(\nu_i\right)$ being, respectively, the PDF and CDF of the SO's belief about the SU$_i$'s valuation. The very first step towards solving (6.38) is to satisfy the IR axiom by choosing $u_i\left(\epsilon_i^{-},\nu_i\right)=0,\forall i\in\mathcal{M}$, knowing that these terms are independent of the optimization variables $w_1,\cdots,w_M$. Back to the constrained optimization problem in (6.36), replacing its objective function by (6.38), we obtain

$$\begin{aligned}\mathcal{U}^{\mathbf{P}}=&\max_{w_1,\cdots,w_M}\sum_{i=1}^{M}\left(\underset{\nu_i,\nu_{-i}}{\mathcal{E}}\langle J_i\left(w_i,\nu_i\right)\rangle\right),\\ \text{s.t.}\ \ &\textstyle\sum_{i=1}^{M}w_i\leq w_0,\\ &J_i\left(w_i,\nu_i\right)=\mathcal{R}_i\left(w_i,\nu_i\right)-\frac{\mathcal{R}_i^{\mathrm{II}}\left(w_i\left(y,\nu_{-i}\right),y\right)\big|_{y=\nu_i}}{\Omega\left(\nu_i\right)},\\ &w_i\geq 0,\\ &\frac{\partial w_i}{\partial\nu_i}\geq 0,\qquad\forall i\in\mathcal{M}.\end{aligned} \tag{6.40}$$

The SO has to quantify two things appearing in (6.40), namely, valuations ν_i, and hazard rates $\Omega\left(\nu_i\right)$, $\forall i\in\mathcal{M}$. Intuitively speaking, learning ν_i is important as a quantity that the rational and self-motivated SU is painstakingly hiding since it is directly (often monotonically) intertwined with its utility. The nature and range of the valuation, that is $\nu_i\in V_i=[\underline{\nu}_i,\overline{\nu}_i]$, can be easily identified by inspecting the demand profile $\mathcal{D}\left(w_i,\nu_i\right)$. Such

basic knowledge helps the SO formulate a hazard rate $\Omega(\nu_i)$ for each SU, which, according to (6.39), requires forming belief and generating its estimated PDF $f_{\nu_i}(\nu_i)$. The closer ν_i's estimated PDF is to its factual PDF (whereof only SU$_i$ is aware), the more efficient the allocation mechanism becomes. The first step towards generating $f_{\nu_i}(\nu_i)$ is to understand how ν_i is inherently related to SU$_i$'s underlying preferences. This question was answered in subsection 6.2.3 where ν_i was chosen proportional to the instantaneous spectral efficiency of the SU$_i$. This being the case, $f_{\nu_i}(\nu_i) \propto \Pr(m)$. Now, the objective is to find the answer to (6.41) which is a restatement of (6.40).

$$\begin{aligned}\widehat{\mathbf{w}}^* = \underset{[w_1,\cdots,w_M]}{\text{argmin}} \; & -\sum_{i=1}^{M} \mathcal{R}_i(w_i,\nu_i) - \frac{\mathcal{R}_i^{\text{II}}(w_i(y,\nu_{-i}),y)|_{y=\nu_i}}{\Omega(\nu_i)}, \\ \text{s.t.} \quad & \Omega(\nu_i) = \frac{f_{\nu_i}(\nu_i)}{1-\mathcal{F}_{\nu_i}(\nu_i)}, \\ & \sum_{i=1}^{M} w_i - w_0 \leq 0, \\ & -w_i \leq 0, \\ & \frac{\partial w_i}{\partial \nu_i} \geq 0, \qquad \forall i \in \mathcal{M}.\end{aligned} \tag{6.41}$$

In fact, as it becomes clear soon, (6.41) is a convex problem.

6.3.3 Convexity of the Allocation Problem

Treating the efficiency axiom $\partial w_i(\nu_i,\nu_{-i})/\partial \nu_i > 0$ in (6.14) as a constraint in (6.41) creates a *non-convex* problem, which is not desirable. To induce convexity, we restrict ourselves to a class of distributions for ν_i, known as regular distributions, that inherently satisfies the efficiency axiom [117]. Hence, the last constraint in (6.41) can be eliminated.

A RV ν_i is said to have a regular distribution if

$$\frac{\partial \zeta(\nu_i)}{\partial \nu_i} > 0, \quad \zeta(\nu_i) = \nu_i - \frac{1}{\Omega(\nu_i)}. \tag{6.42}$$

The above condition is normally satisfied for distributions which tail is not so heavy. Table 6.1 shows that many of the well-known distributions are, indeed, regular. A more relaxed version of the regularity condition is $\partial\Omega\left(\nu_i\right)/\partial\nu_i > 0$.

Table 6.1: Verification of the regularity condition for several well-known distributions.

Distribution	$f_\nu(\nu)$	$\zeta\left(\nu\right)$	Regular
Exponential	$\lambda e^{-\lambda\nu}$, $\nu \in [0,\infty]$	$\nu - 1/\lambda$	always
Gamma	$\frac{\lambda^{k-1}}{\Gamma\left(k\right)}\nu^{k-1}e^{-\lambda\nu}$, $\nu \in [0,\infty]$, $k,\,\lambda > 0$	$\nu - \frac{\gamma\left(k,\lambda\nu\right)}{\lambda^k\nu^{k-1}e^{-\lambda\nu}}$	always
Pareto	$\frac{\alpha}{\nu}\left(\frac{\nu^0}{\nu}\right)^\alpha$, $\nu \in \left[\nu^0,\infty\right]$, $\alpha,\,\nu^0 > 0$	$\left(1-\frac{1}{\alpha}\right)\nu$	for $\alpha > 1$
Uniform	$\frac{1}{b-a}$, $\nu \in [a,b]$	$2\nu - b$	always
Normal	$\frac{1}{\sqrt{2\sigma^2}}e^{-\frac{\left(\nu-\mu\right)^2}{2\sigma^2}}$, $\nu \in [-\infty,\infty]$	$\nu - \frac{\pi}{2}\sigma e^{\frac{\left(\nu-\mu\right)^2}{2\sigma^2}}\left(1-\mathrm{erf}\left(\frac{\nu-\mu}{\sqrt{2}\sigma}\right)\right)$	always

Provided that Assumption 1 in previous section, Assumptions 2–4 in this section, and the regularity condition in (6.42) are all satisfied, then, based on [100], it can be verified that

1. Terms $J_i\left(w_i,\nu_i\right)$ in the summation in (6.41) are concave, i.e.,

$$\frac{\partial^2 J_i\left(w_i,\nu_i\right)}{\partial w_i^2} < 0 \quad \forall i \in \mathcal{M}; \tag{6.43}$$

2. $\partial J_i\left(w_i,\nu_i\right)/\partial w_i$ is increasing in the valuation, i.e.,

$$\frac{\partial^2 J_i\left(w_i,\nu_i\right)}{\partial w_i \partial \nu_i} > 0 \quad \forall i \in \mathcal{M}. \tag{6.44}$$

Owing to these results, (6.41) becomes a *convex* optimization problem as the objective function is concave (linear combination of some concave functions is convex) and inequality constraints are linear. Consequently, the Karush-Kuhn-Tucker (KKT) conditions are sufficient criteria [125] to obtain the solution of (6.41).

6.3.4 Solution of the Allocation Problem

Let $\widehat{\mathbf{x}}^*$ be the optimal solution vector of a generalized non-linear minimization problem with the objective function $g_0 : \widehat{\mathbf{x}} \rightarrow \mathbb{R}$, inequality constraints $g_i^{\text{neq.}}(\widehat{\mathbf{x}}) \in \mathbb{R}^{\leq 0}$, $\forall i \in \{1, \cdots, p\}$, and equality constraints $g_j^{\text{neq.}}(\widehat{\mathbf{x}}) = 0$, $\forall j \in \{1, \cdots, q\}$. According to the KKT theorem, $\widehat{\mathbf{x}}^*$ satisfies the following criteria, known as KKT conditions:

(1) $g_i^{\text{neq.}}(\widehat{\mathbf{x}}^*) \leq 0 \quad \forall i \in \{1, \cdots, p\}$,

(2) $g_j^{\text{eq.}}(\widehat{\mathbf{x}}^*) = 0 \ \forall i \in \{1, \cdots, q\}$,

(3) $\alpha_i^* \geq 0 \ \forall i \in \{1, \cdots, p\}$,

(4) $\alpha_i^* g_i^{\text{neq.}}(\widehat{\mathbf{x}}^*) = 0 \ \forall i \in \{1, \cdots, p\}$ (complimentary slackness condition),

(5) $\nabla_{\widehat{\mathbf{x}}} L\left(\widehat{\mathbf{x}}^*, \widehat{\alpha}^*, \widehat{\beta}^*\right) = \nabla g_0\left(\widehat{\mathbf{x}}^*\right) + \sum_{i=1}^{p} \alpha_i^* \nabla g_i^{\text{neq.}}\left(\widehat{\mathbf{x}}^*\right) + \sum_{j=1}^{q} \beta_j^* \nabla g_j^{\text{eq.}}\left(\widehat{\mathbf{x}}^*\right) = 0$ (zero-Lagrangian gradient condition),

where $L\left(\widehat{\mathbf{x}}, \widehat{\alpha}, \widehat{\beta}\right)$ is the Lagrangian of the *primal* problem, and α_i^*, β_j^* are the solutions of the *dual* problem.

Applying these results to the problem in (6.41), the KKT conditions corresponding to our problem are

$$\sum_{i=1}^{M} w_i^* - w_0 \leq 0, \tag{6.45}$$

$$\alpha_0^* \geq 0, \tag{6.46}$$

$$\alpha_0^* \left(\sum_{i=1}^{M} w_i^* - w_0 \right) = 0, \tag{6.47}$$

$$-w_i^* \leq 0, \ \ \forall i \in \mathcal{M}, \tag{6.48}$$

$$\alpha_i^* \geq 0, \ \ \forall i \in \mathcal{M}, \tag{6.49}$$

$$\alpha_i^* w_i^* = 0, \ \ \forall i \in \mathcal{M}, \tag{6.50}$$

$$-\frac{\partial J_i\left(w_i^*,\nu_i\right)}{\partial w_i}-\alpha_i^*+\alpha_0^*=0,\ \ \forall i\in\mathcal{M}. \tag{6.51}$$

Equations (6.45)-(6.51) consist of $2M+1$ equalities, $2M+2$ inequalities, and $2M+1$ variables of which we are only interested in M of them, i.e., $\widehat{\mathbf{w}}^*=[w_1^*,\cdots,w_M^*]$. From (6.9) and (6.40), the zero-Lagrangian gradient condition in (6.51) can be expressed in terms of the SU's demand profile as follows

$$-\mathcal{D}\left(w_i^*,\nu_i\right)+\frac{\mathcal{D}^{\text{II}}\left(w_i^*\left(y,\nu_{-i}\right),y\right)\big|_{y=\nu_i}}{\Omega\left(\nu_i\right)}-\alpha_i^*+\alpha_0^*=0\ \ \forall i\in\mathcal{M}, \tag{6.52}$$

where $\mathcal{D}^{\text{II}}\left(\cdot,\cdot\right)$ is the first derivative of the demand profile $\mathcal{D}\left(\cdot,\cdot\right)$ w.r.t. the second variable (i.e., ν_i). By eliminating $\alpha_i^*,\ i\in\mathcal{M}$, the KKT conditions in (6.45)-(6.51) reduce in rank to (6.53) which consists of $M+1$ variables, $M+1$ equalities, and $2M+1$ inequalities.

$$\begin{gathered}
-w_i^*\leq 0,\ \ \forall i\in\mathcal{M},\\
\sum_{i=1}^{M}w_i^*-w_0\leq 0,\\
\alpha_0^*\left(\sum_{i=1}^{M}w_i^*-w_0\right)=0,\\
\alpha_0^*\geq\max\left(\mathcal{D}\left(w_i^*,\nu_i\right)-\frac{\mathcal{D}^{\text{II}}\left(w_i^*\left(y,\nu_{-i}\right),y\right)\big|_{y=\nu_i}}{\Omega\left(\nu_i\right)},0\right),\\
\left(-\mathcal{D}\left(w_i^*,\nu_i\right)+\frac{\mathcal{D}^{\text{II}}\left(w_i^*\left(y,\nu_{-i}\right),y\right)\big|_{y=\nu_i}}{\Omega\left(\nu_i\right)}+\alpha_0^*\right)w_i^*=0,\ \ \forall i\in\mathcal{M}.
\end{gathered} \tag{6.53}$$

6.3.5 Allocation Mechanism: Digest of Findings

In this section, we recap the entire allocation mechanism proposed in Sections 6.2 and 6.3. Recall that the step-wise delineation of the mechanism that was initiated in subsection 6.2.6 was left incomplete since, at that point, we did not know the solution to the optimal assignment problem.

Once again, in the proposed allocation mechanism, an SU's bidding strategy is to an-

nounce the amount of BW he needs, i.e., $\widetilde{w}_i$. The SUs are not allowed to submit out-of-range bids, e.g., negative values. Postulating that SUs cannot cooperate or collude, they do not have any information about others' valuations, the number of participating SUs (M), the total BW budget (w_0), etc. In fact, in the *interweave* cognitive paradigm, such as IEEE 802.22 WRANs, where the SO shall remain transparent to the primary network, w_0 is not even known to the SO itself. Consequently, the SUs would only be able to announce their $\widetilde{w}_i$s on the basis of their local preferences. For these reasons, during spectrum auctioning rounds taking place once every hundred microseconds, it is not a-priori known whether $\sum_{i=1}^{M} \widetilde{w}_i$ exceeds w_0 or falls short of it. Once SUs submit their bids, the SO derives pairs $[w_i^*, C_i^*]$. If $w_i^* = \widetilde{w}_i$, then SU$_i$ must accept the offer. On the other hand, when $w_i^* \neq \widetilde{w}_i$, SU$_i$ may refuse the allocation since what he is allocated is different from he requested. For the fact that two distinct cases may arise (please follow through), the allocation mechanism is summarized as the following six steps:

(S.l) The SO generates and, progressively, updates the demand profile $\mathcal{D}\left(w_i, \nu_i\right)$ based on the allocation history.

(S.2) The SUs submit their BW needs $\widetilde{w}_i$s, simultaneously, during intervals that are exclusively assigned for the auctioning purpose. This requires that the medium access control (MAC) sublayer divides time to alternating access (payload transmission) and allocation (auctioning) intervals. This can be handled in a variety of ways but the details are out of the scope of this thesis.

From this point forward, the procedure depends on how $\widetilde{w}_i$s pull together and is dealt with, separately.

Bandwidth Budget Surplus

If $w_0 - \sum_{i=1}^{M} \widetilde{w}_i > 0$, it is always in the best interest of the SO to set $w_i^* = \widetilde{w}_i$, $\forall i \in \mathcal{M}$, resulting in $\sum_{i=1}^{M} w_i^* \neq w_0$. Hence, the third constraint in (6.53) suggests $\alpha_0^* = 0$. This

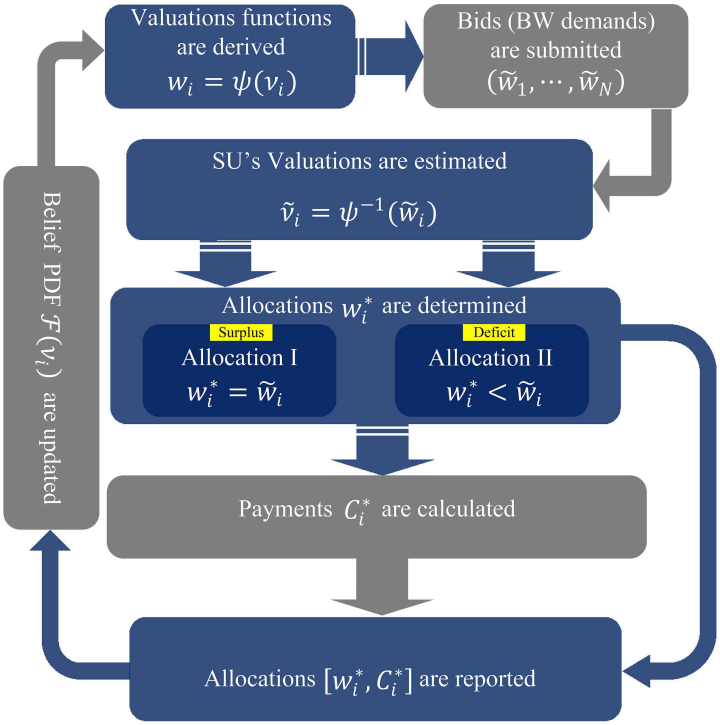


Figure 6.7: Auction-based resource allocation block diagram.

simplifies the problem to the following system of equations with M variables, M equalities and $M+1$ inequalities.

$$\begin{gathered} \sum_{i=1}^{M} w_i^* - w_0 \leq 0, \\ -w_i^* \leq 0, \quad \forall i \in \mathcal{M}, \\ \left(-\mathcal{D}\left(w_i^*, \nu_i\right) + \frac{\mathcal{D}^{\mathrm{II}}\left(w_i^*\left(y, \nu_{-i}\right), y\right)\big|_{y=\nu_i}}{\Omega\left(\nu_i\right)} \right) w_i^* = 0, \quad \forall i \in \mathcal{M}. \end{gathered} \tag{6.54}$$

As such, the allocation procedure resumes as follows:

(S.3) With $w_i^* = \widetilde{w}_i$, the inequalities in (6.54) are satisfied. Hence, the non-trivial solution

to the problem would be

$$-\mathcal{D}\left(\widetilde{w}_i\,,\widetilde{\nu}_i\right)+\frac{\mathcal{D}^{\mathrm{II}}\left(\widetilde{w}_i\,,y\right)\big|_{y=\widetilde{\nu}_i}}{\Omega\left(\nu_i\right)}=0,\ \ \forall i\in\mathcal{M}. \tag{6.55}$$

Contracting (6.55) as $\widetilde{w}_i = \psi\left(\widetilde{\nu}_i\right)$, the fixed-point $\widetilde{\nu}_i$ is attained for SU_i who has bidden $\widetilde{w}_i$. We call $\widetilde{\nu}_i$ the SO's estimation of the SU_i's true valuation ν_i. Thus,

$$\widetilde{\nu}_i=\psi^{-1}\left(\widetilde{w}_i\right)\ \ \forall i\in\mathcal{M}. \tag{6.56}$$

Such fixed-point $\widetilde{\nu}_i$ always exists as is proven next.

Proof. Let's denote $\phi_1\left(\nu_i\right) = \partial\mathcal{D}(\widetilde{w}_i\ ,\nu_i)\ /\partial\nu_i$ and $\phi_2\left(\nu_i\right) = \Omega\left(\nu_i\right)\mathcal{D}\left(\widetilde{w}_i\,,\nu_i\right)$. With this change of notation, (6.55) becomes $\phi_1\left(\nu_i\right) = \phi_2\left(\nu_i\right)$. A fixed-point $\widetilde{\nu}_i$ exists in this equation, if and only if $\partial\phi_2\left(\nu_i\right)/\partial\nu_i > 0$ given that, according to Assumption 4 (cf. Section 6.3), $\partial\phi_1\left(\nu_i\right)/\partial\nu_i < 0$. The increasing monotonicity of $\phi_2\left(\nu_i\right)$ is a true postulation and can be proven by taking its derivative, applying the knowledge about $\partial\Omega\left(\nu_i\right)/\partial\nu_i > 0$ (from the relaxed regularity condition), Assumption 1, and the fact that both $\Omega\left(\nu_i\right)$ and $\mathcal{D}\left(\widetilde{w}_i\,,\nu_i\right)$ are confined to take positive values. Knowing that $0 < \xi \ll 1 \rightarrow \phi_2\left(\nu_i = \underline{\nu}_i - \xi\right) = 0$ and $\phi_1\left(\nu_i = \underline{\nu}_i - \xi\right) > 0$, then $\exists\ \widetilde{\nu}_i \geq \underline{\nu}_i$ such that $\phi_1\left(\widetilde{\nu}_i\right) = \phi_2\left(\widetilde{\nu}_i\right)$. Put into words, there always exists a fixed-point situated in either region II or III of Fig. 6.8. In order to prove that $\widetilde{\nu}_i$ can only happen inside region II, it would suffice to show that $\phi_1\left(\nu_i = \overline{\nu}_i\right) \leq \phi_2\left(\nu_i = \overline{\nu}_i\right)$, which can also be expressed by

$$\left(1-\mathcal{F}_{\nu_i}\left(\overline{\nu}_i\right)\right)\frac{\partial\mathcal{D}\left(\widetilde{w}_i\,,\nu_i\right)}{\partial\nu_i}\Bigg|_{\nu_i=\overline{\nu}_i}\leq f_{\nu_i}\left(\overline{\nu}_i\right)\mathcal{D}\left(\widetilde{w}_i\,,\overline{\nu}_i\right). \tag{6.57}$$

This inequality is always valid as the LHS is zero, because $\mathcal{F}_{\nu_i}\left(\overline{\nu}_i\right) = 1$, and the RHS is non-negative. Therefore, the fixed-point $\widetilde{\nu}_i$ is unique and lies within the second region of

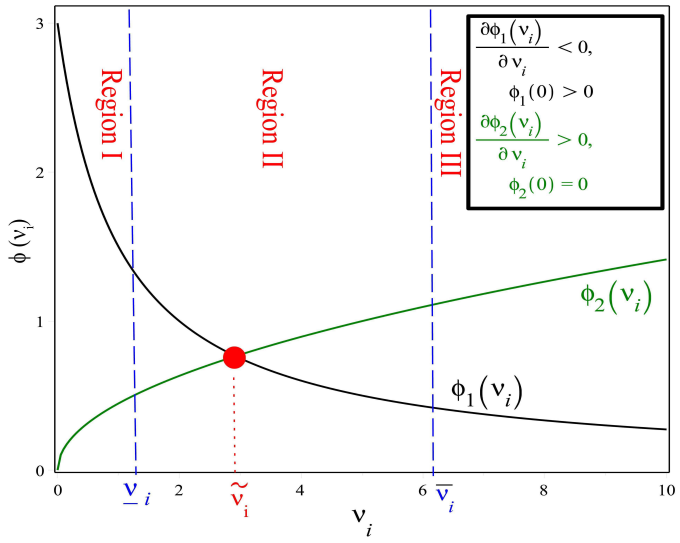


Figure 6.8: Illustration of the existence of fixed-point in (6.56).

Fig. 6.8. □

Bandwidth Budget Deficit

When $w_0 - \sum_{i=1}^{M} \widetilde{w}_i < 0$, the SO has to opt for $w_i^* < \widetilde{w}_i$ such that $\sum_{i=1}^{M} w_i^* = w_0$. To fulfill this, further to steps (S.1)-(S.3) above, the following step has to be taken.

(S.4) The valuation estimate $\widetilde{\nu}_i$ obtainable from (6.56) in (S.3) is to be plugged into the zero-Lagrangian gradient constraint in (6.53) as follows:

$$\left(-\mathcal{D}\left(w_i^*, \psi^{-1}\left(\widetilde{w}_i\right)\right) + \frac{\mathcal{D}^{\mathrm{II}}\left(w_i^*, y\right)\big|_{y=\psi^{-1}(\widetilde{w}_i)}}{\Omega\left(\psi^{-1}\left(\widetilde{w}_i\right)\right)} + \alpha_0^*\right) w_i^* = 0, \quad \forall i \in \mathcal{M}. \tag{6.58}$$

Along with the resource constraints $\sum_{i=1}^{M} w_i^* = w_0$ (while accounting for the inequalities $w_i^* \geq 0$ and $\alpha_0 \geq 0$), we are left with a system of $M+1$ equations and $M+1$ unknowns, which guarantees the existence of a solution.

Finally, the next two steps are to be executed in both cases (surplus or deficit):

(S.5) The payment of SU_i is obtained by substituting $\widetilde{\nu}_i = \psi^{-1}(w_i^*)$ and $\widetilde{w}_i(z, \nu_{-i}) = \psi(z)$ into (6.31) as follows:

$$C_i^*(\nu_i) = \underset{\nu_{-i} \in V_{-i}}{\mathcal{E}} \langle \int_0^{w_i^*} \mathcal{D}\left(z, \psi^{-1}(w_i^*)\right) \mathrm{d}z - \int_{\epsilon_i^+}^{\psi^{-1}\left(w_i^*\right)} \int_0^{\psi(z)} \frac{\partial \mathcal{D}(y,z)}{\partial z} \mathrm{d}y \mathrm{d}z \rangle, \quad \forall i \in \mathcal{M}, \tag{6.59}$$

where, in case of surplus $w_i^* = \widetilde{w}_i$ and, under deficit, w_i^* is the solution of (S.4).

(S.6) Allocation pair $[w_i^*, C_i^*]$ is immediately announced to SU_i, whereupon he can access w_i^* for the duration specified in exchange for C_i^* payable to the SO.

As the parameter chosen by the SO, the lower integral limit $\epsilon_i = \nu_{0,i}$ in (6.59) denotes the minimum valuation that the SU_i should have to be able to participate in the auction. To enforce higher fairness, the SO may choose $\epsilon_i = \epsilon_j$, $\forall i, j \in \mathcal{M}$. Fig. 6.7 summarizes the above allocation mechanism.

6.4 Application Scenarios and Case Study

The next generation of wireless networks (5G) is expected to be an ultra fast, ubiquitous, and reliable communications infrastructure. The machine-centric architecture in these networks enables any two machines to talk to each other no matter where they are and what task they are assigned to accomplish. Due to the emergence of numerous attractive use cases, traffic volume will be huge and the nature of traffic will be highly diversified. Therefore, to make 5G a reality, not only sufficient capacity is needed by utilizing spectrum much more efficiently but also traffic differentiation is a crucial undertaking. While the native optimal auction-based allocation mechanism introduced so far can partly address the capacity crunch problem, the latter objective is addressed by the following extension of the mechanism.

6.4.1 Allocation in Prioritized Access Networks

Let us concentrate on a cellular network with $\mathcal{P}$ critical traffic classes and M SUs associated with a base station (BS). Each class corresponds to a number, called *criticality* degree. In the case of budget surplus, SUs from all classes receive what they demand, hence, differentiation is pointless. The problem arises when the BW budget is in the deficit state. In this case, the criticality degrees $(\mathcal{P}_1, \cdots, \mathcal{P}_M)$ of all the SUs, which can be real-positive numbers $\mathcal{P}_i \in [1, \infty], i \in \mathcal{M}$ or the same as their service level agreements (SLAs), are known to the BS. In fact, as long as their rational relativities w.r.t. each other are held, it does not matter what values are chosen. Now, to differentiate the SUs, the BS overrides the valuation estimates $\widetilde{\nu}_i$s using the following rule

$$\left[\psi^{-1}(\widetilde{w}_i)\right]_{1\times M} \leftarrow \left[\psi^{-1}(\widetilde{w}_i)\right]_{1\times M} \times \mathfrak{P}_{M\times M}, \tag{6.60}$$

where $\mathfrak{P} = \text{diag}(\mathcal{P}_1, \cdots, \mathcal{P}_M)$ is a $M \times M$ diagonal matrix whose main diagonal entries are the criticality degrees of the SUs, and $[\psi^{-1}(\widetilde{w}_i)]_{1\times M}$ is the row primitive valuation vector whose i^{th} element is SU$_i$'s valuation estimation obtainable from (6.56).

Using (6.60), each SU's priority and valuation are merged into a single number, in a rational and stable way. These overridden valuations are, then, plugged into the zero-Lagrangian gradient condition (6.58), to find the differentiated allocation vector $(w_1^*, \cdots w_M^*)$, and into (6.59) to calculate the corresponding payments. In an exemplary network with $\mathcal{P} = 2$ and $M = 2$, where the normative SU ($\mathcal{P}_1 = 1$) and critical SU ($\mathcal{P}_2 = 1.5$) demand the same amount of BW ($\widetilde{w}_1 = \widetilde{w}_2$), according to (6.60), the latter is assigned 1.5 times larger overridden valuation than the former even though the primitive valuations are the same. If $\partial(\widetilde{\nu}_i = \psi^{-1}(\widetilde{w}_i))/\partial\widetilde{w}_i > 0$, then, under the deficit situation, (6.58) allocates the critical SU a BW share that is closer to what he demanded. In other words, $\widetilde{w}_1 - w_1^* > \widetilde{w}_2 - w_2^* > 0$. To prove that $\psi^{-1}(\widetilde{w}_i)$ is a monotonically increasing function of bids, we start from (6.55) restating it as

$$\lim_{\epsilon\to 0} -\Omega\left(\widetilde{\nu}_i\right)\mathcal{D}\left(\widetilde{w}_i\,, \widetilde{\nu}_i\right) + \frac{\partial\mathcal{D}\left(\widetilde{w}_i\,, \widetilde{\nu}_i\right)}{\partial\widetilde{\nu}_i} - \epsilon\Omega\left(\widetilde{\nu}_i\right) = 0. \tag{6.61}$$

The last term on the LHS becomes zero as $\epsilon \to 0$. Next, we take the derivatives of this equation w.r.t. $\widetilde{w}_i$,

$$\begin{aligned} &-\frac{\partial\Omega\left(\widetilde{\nu}_i\right)}{\partial\widetilde{\nu}_i}\frac{\partial\widetilde{\nu}_i}{\partial\widetilde{w}_i}\mathcal{D}\left(\widetilde{w}_i, \widetilde{\nu}_i\right) - \Omega\left(\widetilde{\nu}_i\right)\left(\frac{\partial\mathcal{D}\left(\widetilde{w}_i, \widetilde{\nu}_i\right)}{\partial\widetilde{w}_i} + \frac{\partial\mathcal{D}\left(\widetilde{w}_i, \widetilde{\nu}_i\right)}{\partial\widetilde{\nu}_i}\frac{\partial\widetilde{\nu}_i}{\partial\widetilde{w}_i}\right) \\ &+ \frac{\partial^2\mathcal{D}\left(\widetilde{w}_i, \widetilde{\nu}_i\right)}{\partial\widetilde{\nu}_i^2}\frac{\partial\widetilde{\nu}_i}{\partial\widetilde{w}_i} + \frac{\partial^2\mathcal{D}\left(\widetilde{w}_i, \widetilde{\nu}_i\right)}{\partial\widetilde{\nu}_i\partial\widetilde{w}_i} = 0, \end{aligned} \tag{6.62}$$

which, after factorization, delivers

$$\underbrace{\frac{\partial \widetilde{\nu}_i}{\partial \widetilde{w}_i}}_{\geq 0} = \frac{\overbrace{\Omega\left(\widetilde{\nu}_i\right)\frac{\partial \mathcal{D}\left(\widetilde{w}_i,\widetilde{\nu}_i\right)}{\partial \widetilde{w}_i} - \frac{\partial^2 \mathcal{D}\left(\widetilde{w}_i,\widetilde{\nu}_i\right)}{\partial \widetilde{\nu}_i \partial \widetilde{w}_i}}^{\leq 0 \text{ (Assump. 1 \& RA condition)}}}{\underbrace{\frac{\partial^2 \mathcal{D}\left(\widetilde{w}_i,\widetilde{\nu}_i\right)}{\partial \widetilde{\nu}_i^2}}_{\leq 0 \text{ (Assump. 4)}} - \underbrace{\frac{\partial\left(\Omega\left(\widetilde{\nu}_i\right)\mathcal{D}\left(\widetilde{w}_i,\widetilde{\nu}_i\right)\right)}{\partial \widetilde{\nu}_i}}_{\geq 0 \text{ (Assump. 1 \& Regularity)}}} \tag{6.63}$$

Thus, the estimated valuation function $\psi^{-1}(\widetilde{w}_i)$ is monotonically increasing in the demanded bandwidth (bids), hence, (6.58) allocates the critical SU a BW share that is closer to what he demanded.

6.4.2 Application in LTE and WRANs

Application of the proposed optimal resource allocation scheme goes beyond FDMA-based networks, since no restriction was imposed on the nature of the allocation units. In fact, this proposal is applicable to any access mechanism with mutually exclusive and collectively exhaustive (MECE principle [126]) resource units. It can be utilized for BW allocation, time-slot allocation, admission control, and power control. Particularly, the allocation unit can be deemed as physical resource blocks (PRB) in networks with multi-carrier modulation such as LTE and IEEE 802.22 WRANs. Assuming MECE principle is satisfied, a PRB in the latter networks is a segment from the time-frequency grid with a time span of seven symbols (0.5 msec) and a frequency span of 12 subcarriers (180 KHz) [127].

In reality, PRBs get allocated to user equipment (UEs), or CPEs in the case of WRANs, according to round-robin, max C/I, or proportional fairness scheduling policies. A simple network with two UEs is shown on the left-hand-side (LHS) of Fig. 6.9, where the UE whose channel quality in a PRB is stronger gets allocated that resource block. Setting aside round-robin as an inefficient approach, the last two policies are channel-dependent,

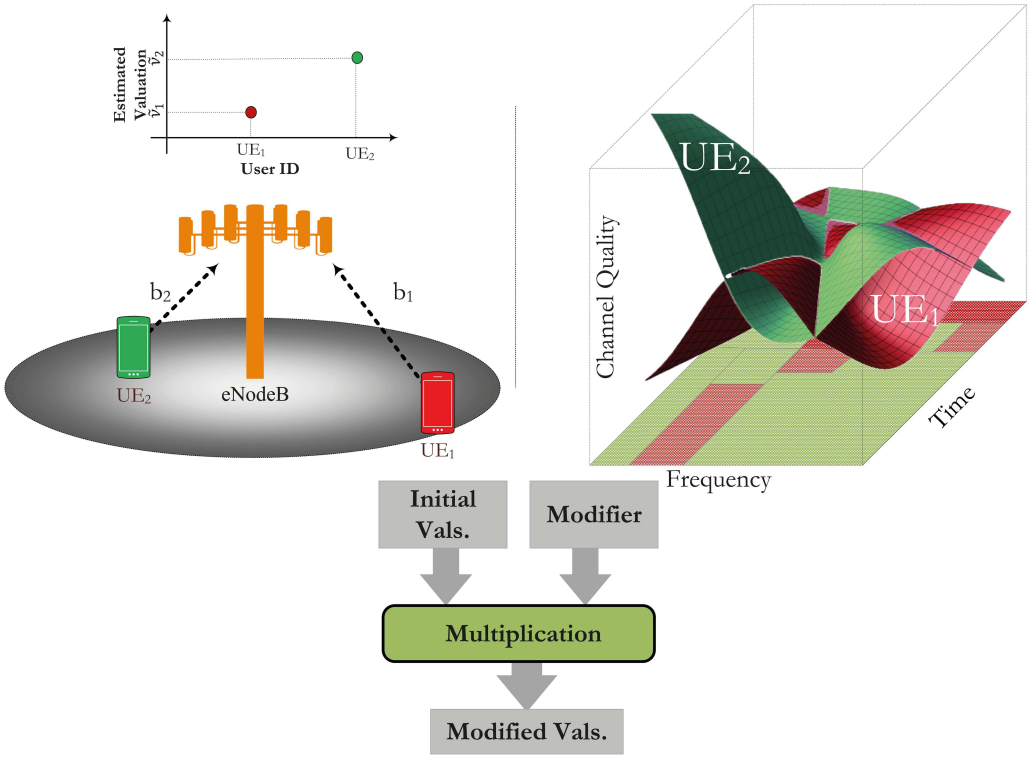


Figure 6.9: Application of the proposed auction-based allocation scheme to LTE cellular networks. On the left, two UEs submit their bandwidth demand whereupon the network estimates their valuations. On the right, the variation of those UEs' CQI on the resource grid is shown.

meaning their outcomes depend on the channel state information (CSI) that is received in uplink direction by LTE-BS, a.k.a. eNodeB, over physical uplink shared channel (PUSCH). Briefly speaking, CSI, which comprises the channel quality indicator (CQI), precoding matrix indicator (PMI) and rank indicator, is incorporated into the most essential LTE functionalities and is generated by measuring reference signals (RS) that eNodeB periodically sends at regular frequencies and intervals. However, CSI is not the only useful information needed to efficiently schedule resources. Among other factors, traffic/load type, intensity and variation play important roles in how optimal PRBs are allocated. To better clarify this point, allocating all PRBs to a UE with large CQI and an empty buffer is equivalent to depriving other UEs from accessing those PRBs, i.e., resources remain unutilized for the rest

of the allocation interval. Clearly, this signifies that there are many factors that pique UEs' interests and impact their valuations ν_i. This logic undermines the optimality of scheduling methods that are solely dependent on channel quality, and suggests that better can be done.

The alternative solution for the allocator is to figure out what *internal* factors participate in UEs' utility function, calibrate the corresponding demand profile accordingly (step (S.1) in Section 6.2.6), derive estimates of UEs' valuations $\tilde{\nu}_i$, readjust these estimates by engaging the *external* factors using a matrix called Modifier, and then progress through steps (S.4) and (S.5). In a mathematical language, matrix $\mathfrak{P}_{M\times M}$ in (6.60) becomes

$$\mathfrak{P}_{M\times M} = \underbrace{\begin{pmatrix} \zeta_{1,1} & \cdots & \zeta_{1,M} \\ \vdots & \ddots & \vdots \\ \zeta_{M,1} & \cdots & \zeta_{M,M} \end{pmatrix}}_{\text{ModifierMatrix}}. \tag{6.64}$$

If CSI is the only impacting factor, the modifier matrix elements $\zeta_{i,i} = \text{CQI}_i$. Whereas the off-diagonal entries $\zeta_{i,j} = 0 \; \forall \; j \neq i$ for the case where UEs' private valuations ν_i are mutually independent, the BS may choose otherwise if it realizes that some sort of collusion is taking place among UEs to gain access to more resources at reduced charges. In fact, by choosing $\zeta_{i,j} \neq 0 \; \forall \; j \neq i$, the BS gains a leverage to dissuade colluding UEs from continuing such behavior.

The implementation of this allocation mechanism is a straightforward task for the fact that the mechanism does not require the exchange of so much information between eNodeB and UEs. For the sheer similarity between LTE and WRANs, in the following, we only discuss the integration of the proposed allocation mechanism into the WRAN frame structure. This is accomplished by adding an extra interval to the onset of each frame, namely, the *allocation subframe*. As illustrated in Fig. 6.10, each CPE submits a bid (BW demand) during the BID sub-interval embedded at the beginning of each subframe. Based on these bids, and using the steps introduced in subsection 6.3.5, a tentative allocation (TA) is calculated

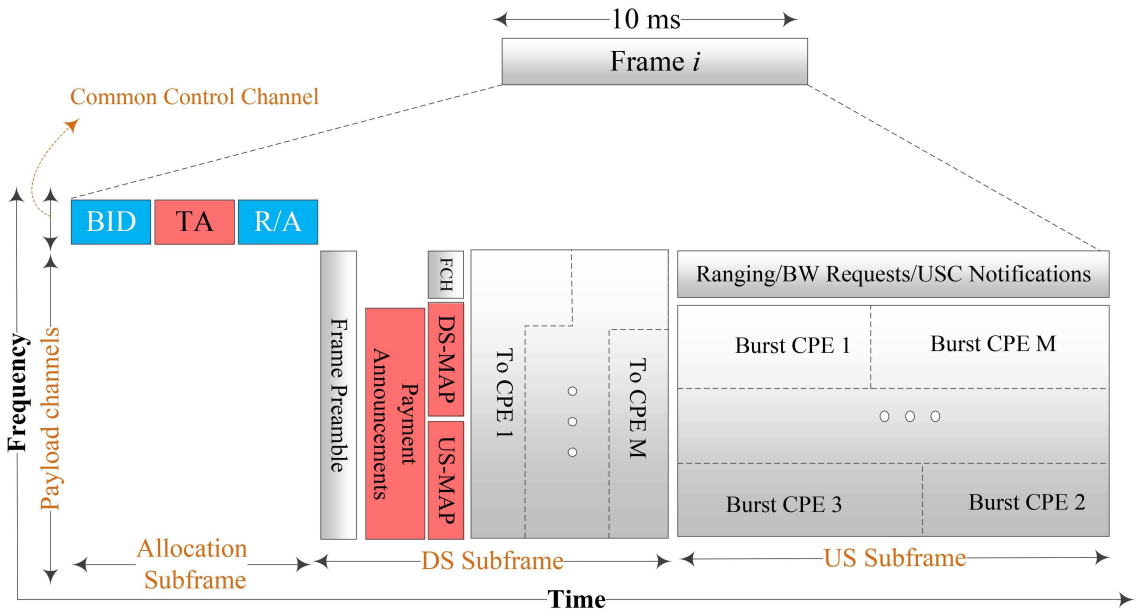


Figure 6.10: Modified frame structure of IEEE 802.22 integrated with the proposed allocation mechanism.

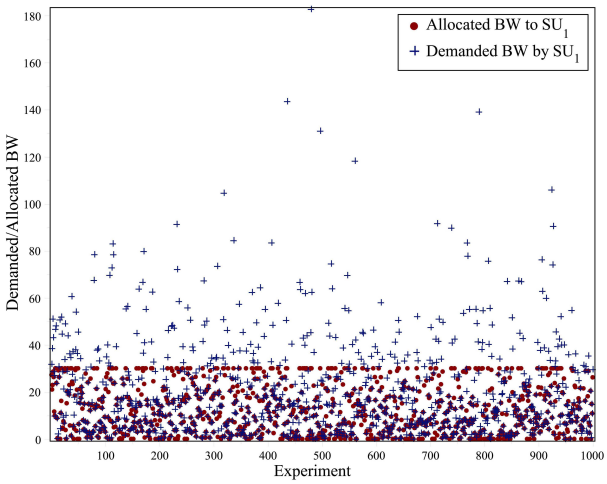
and declared to CPEs within the TA sub-interval. According to the allocation rule described before, those CPEs that were allocated exactly what they requested have no right to decline the offer. On the other hand, those that were assigned less than what they demanded (say, due to deficit of net resource) may refuse or accept the offer, which should be declared back to the BS immediately, within the refusal/acceptance (R/A) sub-interval. For transmissions in the BID and R/A sub-intervals, CPEs can use the existing CDMA-based or contention-based access mechanisms [8]. The entire negotiation takes place over a CCC that is one of the mandatory features in [8]. The rest of the operations remain intact and identical to the IEEE 802.22 specification, including deceleration of the finalized allocation map to the CPEs during upstream map (US-MAP) and downstream map (DS-MAP) in DS subframe. In comparison to the amount of control information that gets transferred between CPEs and the BS in each frame, the overhead associated with appending the allocation subframe is negligible.

6.4.3 Case Study

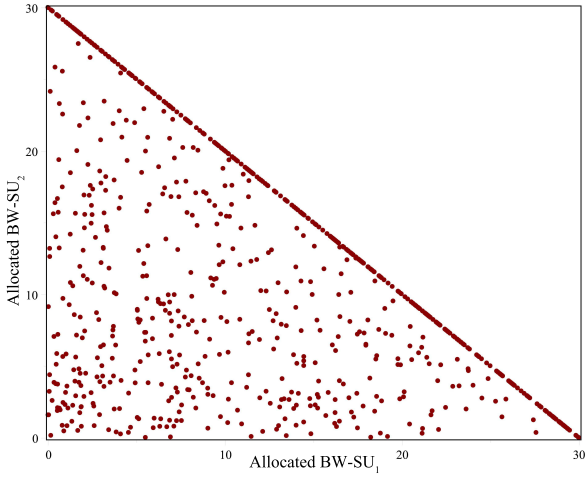
As a special case, we study the performance of the proposed allocation mechanism in a simple network with $M = 2$ SUs whose demand profile is a linear function, i.e., $\mathcal{D}(w_i, \nu_i) = \nu_i - \sigma w_i$. The simulation is run for several hundred milliseconds, corresponding to almost 1000 experiments, and the allocation outcomes $(C_1^*, w_1^*; C_2^*, w_2^*)$ are recorded. The total spectrum available to the SO for allocation to SU_1 and SU_2, i.e., w_0, is an exponential stochastic process whose average simply changes during the day as $\mu(t) = \{15^{(\text{t=Morning})}, 30^{(\text{t=Noon})}, 45^{(\text{t=Night})}\}$ MHz due to the varying activity of the primary network. We assume that SU_i's valuation ν_i is a private information drawn independently from an exponential PDF with parameter $1/\overline{\nu}$. Moreover, we consider that $\widetilde{w}_i = \varrho_i \nu_i$, where $\varrho_i \in \mathbb{R}^{>0}$ is the privately-known demand-valuation coefficient. Over the simulation run, the surplus and deficit situations both take place, frequently. Fig. 6.11 presents the simulation results. In Fig. 6.11a, the matching of SU_1's demand with supply for 1000 experiments is shown. The corresponding empirical PDF (histogram) is depicted in Fig. 6.11c. The allocation scatterplot is illustrated in Fig. 6.11b from the same simulation. Finally, Fig. 6.11d compares the histogram of SU_1's true valuation ν_1 with the estimated valuation $\widetilde{\nu}_1$. Next, we investigate the average revenue C^{net} collected by the SO, which is equal to the sum of payments C_i^* made by each SU (cf. Eq. (6.36)). It was postulated in subsection 6.3.2 that C^{net} is the optimality metric of the allocation mechanism. In the above experiment, C^{net} can be derived in closed-form as

$$C^{\text{net}} = \underbrace{\Pr\left(\widetilde{w}_1 + \widetilde{w}_2 < w_0\right)}_{\Pr(\text{Sur.})} C^{\text{Sur.}} + \underbrace{\Pr\left(\widetilde{w}_1 + \widetilde{w}_2 \geq w_0\right)}_{\Pr(\text{Def.})} C^{\text{Def.}}, \tag{6.65}$$

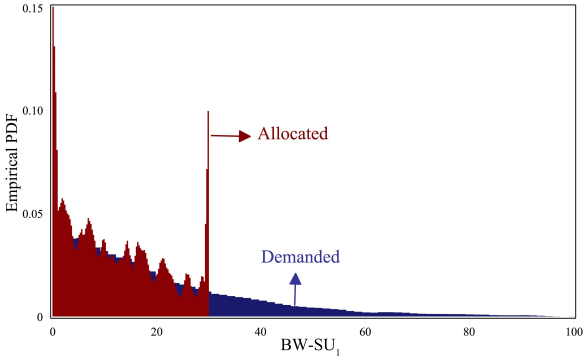
where the first (second) term in the RHS is the total collected revenue in the case of BW surplus (Deficit). Further, $C^{\text{Def.}} = C_1^{\text{Def.}} + C_2^{\text{Def.}}$ and $C^{\text{Sur.}} = C_1^{\text{Sur.}} + C_2^{\text{Sur.}}$ are the payment sums the two SUs make when deficit and surplus occur, respectively. For the linear demand



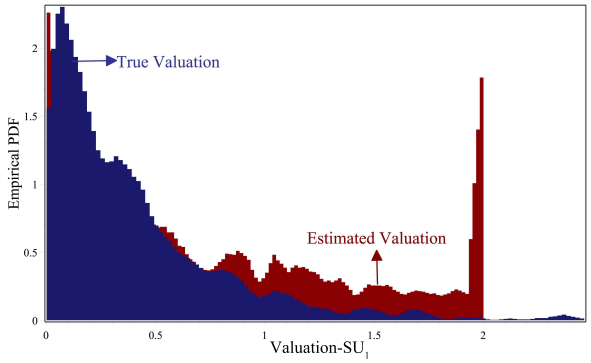
(a) Demand/supply matching.



(b) Scatter-plot of allocated BW for both SUs.



(c) Histogram: allocated BW w_1^* vs. demand $\widetilde{w}_1$.

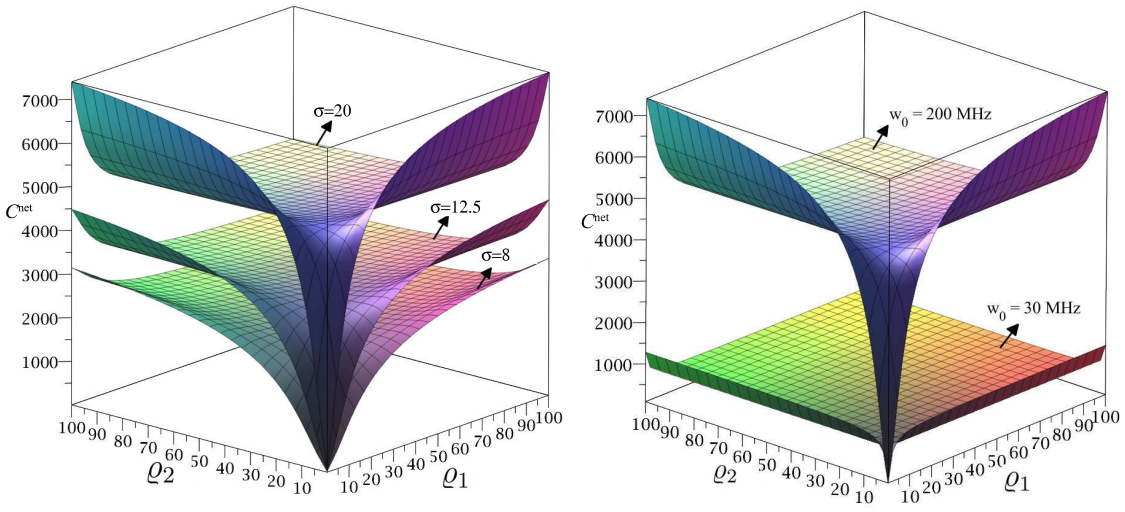


(d) Histogram: true valuation $\breve{\nu}_1$ vs. estimated valuation $\widetilde{\nu}_1$.

Figure 6.11: Simulation experiments for $M = 2$ SUs.

profile, linear valuation-demand relationship, and exponential valuation distribution, the total accrued revenue would be equal to

$$C^{\text{net}} = \Pr(\text{Sur.})\overline{\nu}^2(\varrho_1 + \varrho_2 + \sigma^{-1}) + \Pr(\text{Def.})\overline{\nu}^2(\frac{w_0}{\overline{\nu}} + \sigma^{-1}), \quad \varrho_1 \neq \varrho_2, \tag{6.66}$$



(a) Variable mean valuation $\overline{\nu}$.

(b) Variable total available resource budget w_0.

Figure 6.12: Total revenue collected by the SO from two SUs vs. the demand-valuation coefficients, ϱ_1 and ϱ_2, when the demand profile and demand-valuation are both linear functions and the SUs' true valuations are exponentially distributed.

where

$$\begin{aligned}\Pr(\text{Sur.}) &= \underset{w_0}{\mathcal{E}}\left\langle 1-\overline{\nu}^{-1}(\varrho_1-\varrho_2)^{-1}(\varrho_1\overline{\nu}\mathrm{e}^{-\frac{w_0}{\overline{\nu}\varrho_1}}-\varrho_2\overline{\nu}\mathrm{e}^{-\frac{w_0}{\overline{\nu}\varrho_2}})\right\rangle \\ &= \frac{\overline{\nu}^{-2}}{\mu(t)^2\varrho_1\varrho_2+\overline{\nu}^{-1}\mu(t)(\varrho_1+\varrho_2)+\overline{\nu}^{-2}},\end{aligned} \tag{6.67}$$

and $\Pr(\text{Def.}) = 1 - \Pr(\text{Sur.})$.

Results are presented in Fig. 6.12. According to this figure, the SO accrues low (high) revenue if both SUs have small (large) demand-valuation coefficients for the BW. Surprisingly, the largest revenue is gained when there exist asymmetry between the SUs' demands ($\varrho_1 \gg \varrho_2$ or $\varrho_1 \ll \varrho_2$). Obviously, the larger the auctioned resource w_0 is, the more revenue is accrued (Fig. 6.12b). When the common belief PDF $f_{\nu_i}(\nu_i) = \overline{\nu}^{-1}\exp(-\nu_i/\overline{\nu})$ has higher mean $\overline{\nu}$, it is equivalent to having higher ϱ_1 and ϱ_2, leading to higher revenue (Fig. 6.12a). It should be noted that other demand profiles, beliefs, etc., may generate 3D

revenue surfaces that look different from those in Fig. 6.12.

6.5 Summary

This chapter proposed a solution framework to the resource allocation problem in MC-CRNs. The contribution is based on the thriving theory of auctions from economic science. The aptness of auctioning for allocation of resources in micro-scale is justified by the similarities that futuristic wireless terminals share with human brain, in terms of reasoning and decision-making abilities. Moreover, auctions are well-known for their allocation efficiency and higher revenue generation properties. Built upon the recently-emerged theory of optimal multi-unit auctions, the allocation mechanism, this contribution resolves several deep-rooted issues in wireless networks: On the user side, the fairness-optimality trade-off was settled in an elegant way; on the network operator side, the total collected revenue was maximized by the imposition of payments. To diminish uncertainties of the network operator and improve the stability of resource assignments, the aforementioned payment rule was delicately designed to induce truthfulness among self-centered and rational users. These properties, along with the computational manageability of the formulated problem and the existence and stability of its solution, renders the proposed allocation mechanism an ideal scheme for current and future networks.

Chapter 7

Thesis Conclusion

It is beyond any doubt that 5G wave will be hitting us soon. This wave is forceful and dissimilar to what we are familiar with. Machines will become the dominating network subscribers pushing aside the human-operated devices. They will surround us everywhere we go; they can be anything from our home appliances, the autonomous traffic controlling apparatuses, self-ruling energy distribution gauges, environmental sensors to machineries, cars, so on and so forth. It is hardly possible to draw a limit for the possible scenarios; once this door is open, new use cases will emerge day by day. Triggered by this revolution, the aggregated traffic is expected to be immense in size and diverse in nature posing unprecedented problems ahead of us whereof spectrum insufficiency is at the very core. Concretely, while transmitting a single bit of information requires bandwidth, there is, basically, not enough of the latter to offer. Reports from UMTS and European Commission (EC) predict that, by 2020, the total spectrum need for mobile systems will be between $1.6\ \mathrm{GHz}$ and $2.6\ \mathrm{GHz}$. This is almost $1\ \mathrm{GHz}$ to $2\ \mathrm{GHz}$ deficit. In finding a solution to this problem, several experimental studies were conducted unveiling that, to a large extent, the spectrum is not utilized efficiently. In recent years, it became clear that static allocation of spectrum is the reason for its inefficient utilization. Therefore, to improve spectral in-

sufficiency, improving spectral efficiency is inevitable. Consequently, dynamic spectrum allocation (DSA) was proposed as the solution to the spectral inefficiency problem. The work in this PhD dissertation has been motivated by the same ideas. The overall goal has been to substantiate cognitive radio (CR) as the solution to the problem of spectrum inadequacy. Simply uttered, to make CR better known to the future generation of researchers. To fulfill this objective the extensive research approach (horizontal) was taken, instead of deeply focusing on a very narrow subject (vertical). As a result, different research facets of CR were explored including modeling, analysis, synthesis, and design.

The first part of this thesis was devoted to building robust mathematical models that can predict the behavior of CR networks (CRNs) under diverse circumstances with high accuracy. The proposed models are based on the solid mathematical framework of queuing theory. These models act as abstract descriptions for the CRNs. They are pre-deployment benchmarks that help us understand how CRNs operate and whether or not they meet the expectations w.r.t. service requirements. Moreover, sufficient attention in the first part was dedicated to deriving closed-form expressions for different metrics such as user's throughput, queue occupancy, and delay. The analytical and modeling approach taken in this thesis is more realistic compared to any previous approaches for single-channel and multi-channel CRNs, which are mainly grounded upon simplifying assumptions. Besides modeling CRNs in generic settings, the wireless regional area network (WRAN), as the most auspicious CR technology known to us, is modelled and its performance was evaluated. In an attempt to answer a question of major practical importance, the synthesis problem was tackled too. The latter problem shifts the focus to the application side of the CRNs, namely, whether an application requirements can be satisfied in a given network with a-priori known network/traffic conditions, and if not, what adjustments/sacrifices should be made so that these requirements get satisfied.

Despite its importance, the investigations in the area of designing mechanisms for MC-CRNs have been less abundant. In filling this void, the second part of this PhD thesis

is dedicated to designing smart, efficient, and scalable mechanisms for MC-CRNs. The first chapter proposes a spectrum sensing and decision combining mechanism for WRANs. Termed MC-LDS, this protocol operates based on a simple learning concept that uses differentiation as well as reward-penalty procedures to intelligently combine current and past sensing measurements. The combination was done in such a way that both accurate and faulty sensors contribute constructively toward an accurate final decision. Through precise probing of a comprehensive set of quantities such as false-alarm, misdetection and successful discovery rates, as well as correlation factor and goodness-of-fit metric, it was demonstrated that MC-LDS is considerably superior to AND, OR and VOTING rules featured in standard. Among the advantages of MC-LDS, its accuracy, implementation ease, fairness, adjustability and stability are to be emphasized. Most importantly, MC-LDS can be integrated to boost the sensing performance in other promising technologies and standards such as White-Fi (IEEE 802.11af WLANs) wireless personal area networks and Zig-Bee (IEEE 802.15 family), cognitive WiMAX (IEEE 802.16h) and the recent IEEE 1900.6b standard emerged to support spectrum databases using spectrum sensing information.

The last chapter in the mechanism design part addresses the important issue of spectrum sharing in CRNs. The work is erected on the thriving theory of multi-unit auctions from microeconomics. The aptness of auctioning to allocate resources in wireless networks on a micro-scale basis is justified by the similarities that futuristic wireless terminals share with human reasoning and decision-making abilities. Moreover, auctions are well-known for their allocation efficiency and higher revenue generation properties. In a systematic way to deal with this problem, an efficacious allocation is characterized by six axioms, namely, incentive compatibility (IC), individual rationality (IR), fairness, efficiency, revenue maximization (RM) and computational manageability (CM). While the first three axioms were satisfied through a non-linear pricing rule, the last three axioms were integrated into the mechanism through a particular spectrum assignment. Moreover, due to the importance of the accurate characterization of utility function as a quantity that SUs strive to maximize,

spectral efficiency is selected and a closed-form expression is obtained for it. Besides satisfying the optimality and revenue maximization criteria, the allocation mechanism is highly scalable as the solution is obtained by root-finding operations and solving almost linear system of equations and inequalities. These properties make the proposed resource allocation mechanism an ideal candidate for deployment in 5G networks as planned from the beginning. This chapter is summarized by discussing the application and implementation aspects of the proposed sharing mechanism in LTE and WRANs and conducting numerical analysis.

For all we know, the orientation towards DSA cannot happen in one shot. Such approach is neither feasible not admissible. The truth of the fact is that protecting the interest of licensed (incumbent) spectrum users is the biggest impediment towards the faster adoption of DSA paradigm. They are the legitimate owners of the spectrum and violating their rights has serious legal consequences. This makes network operators extremely reluctant to move towards a dynamic spectrum model before ample regulations is in place to shield them and sufficient standardizations to lead them. Regrettably, the synergy between these latter has gotten caught up in a vicious cycle. In fact, while both understand the importance of moving towards the open spectrum model, the regulators need technical insight from standardization bodies to set out the mandates and no standard will release before a regulation is in effect. There are hopes that this paradox will be broken soon. A European regulatory framework, called WAPECS, is one of the pioneering efforts along this path. The next barrier towards adopting DSA is tied to the fundamental role that spectrum plays in radiocommunications. The existing communications systems are built to operate in predefined spectrum bands. This means that moving on from the static to dynamic spectrum model is a revolutionary shift needful of redesigning devices, protocols, architectures and network structures from scratch. Such fundamental changes would be impossible without shutting down everything and building up everything. During this process billions of dollars will be lost deterring solution providers, network operators, and tech companies from

taking such risk. Instead, there are good reasons for advancing towards DSA in an evolutionary manner. First, it motivates the involved parties to take the risk without having to worry about service interruption. As a matter of fact, telecommunications is a concentrated market wherein operators abruptly changing to new solution, like DSA, may be doomed to lose their markets to those who choose to stand. Second, given that DSA is an entirely new practice, its probability of success will be higher if the step-by-step, construction by learning, approach is adopted. Third, given the fundamental wideband limitations of the existing electronics and RF technologies, a fully technology-and service-neutral paradigm, where the entire spectrum is open to all communicating devices, seems more like a dream for now. Actually, to get to that point may take couple of decades. Therefore, a more realistic approach would be to use dynamic spectrum model in different isolated segments of spectrum and have the boundaries gradually removed as the hardware technology evolves. The widespread popularity and deep penetration of the license-exempt wireless local area networks (WLANs) is an advocate for the success of the evolutionary transition to DSA. With all the due considerations, DSA has only one realistic path to adoption and that is the gradual path. This path may seem long and torturous but the fruits are worth it. During this time, patience and perseverance is needed.

A faster transition to DSA hinges on finding the right balance between technological efficiency, economic efficiency, and flexibility in using spectrum. This should be the main focus of researchers who see DSA as the solution to the problem. This multi-faceted problem can only be addressed within an interdisciplinary research. What has been done in this dissertation is a step in this path by utilizing queuing theory, auction theory, and learning from other branches of science to model and design CRNs.

7.1 Publications

The findings of this thesis have already resulted in several published and submitted transactions and conference papers as listed in the following. .

I. N. Tadayon and S. Aïssa, “Modeling and analysis of cognitive radio-based IEEE 802.22 wireless regional area networks", *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4363–4375, Sep. 2013. [Available here]

II. N. Tadayon and S. Aïssa, “Multi-channel cognitive radio networks: modeling, analysis, and synthesis", *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 11, pp. 2065–2074, Feb. 2014. [Available here]

III. N. Tadayon and S. Aïssa, “Modeling and analysis framework for multi-interface multi-channel cognitive radio networks", *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 935–947, Oct. 2014. [Available here]

IV. N. Tadayon and S. Aïssa, “A learning-based distributed spectrum sensing mechanism for IEEE 802.22 wireless regional area networks", IEEE GLOBECOM, San Diego, CA, Dec. 2015, pp. 1–6. [Available here]

V. N. Tadayon and S. Aïssa, “A multi-channel spectrum sensing fusion mechanism for cognitive radio networks: Design and application to IEEE 802.22 WRANs", Accepted at *IEEE Transactions on Cognitive Communications and Networking*, vol. 1, no. 4, pp. 359–371, Dec. 2015. [Available here]

VI. N. Tadayon and S. Aïssa, “Resource allocation in cognitive radio networks: Pricing and utility derivation", Submitted to *IEEE Journal on Selected Areas in Communications*, Apr. 2016.

VII. N. Tadayon and S. Aïssa, “Resource allocation in cognitive radio networks: Spectrum Assignment", Submitted to *IEEE Journal on Selected Areas in Communications*, Apr. 2016.

Part III

Résumé Français

Appendix A

Introduction et Objectifs

A.1 Introduction à la Radio Cognitive

La prolifération des technologies sans fil dans tous les aspects de la vie moderne a connu une augmentation sans précédent au cours de la dernière décennie. Les appareils sans fil constituent maintenant un aspect indispensable de la vie moderne, avec plus de trois milliards d'entre eux en usage aujourd'hui, un nombre qui devrait dépasser 25 milliards d'ici 2025 [1]. L'impulsion du déclenchement pour cette accélération peut être directement attribuée à la demande sans cesse croissante pour la communication humaine rapide, fiable et illimitée sous la forme de trafic numérique. Fusionné avec les progrès parallèles dans le développement des applications (logiciel) et la fabrication dispositif (Matériel), aujourd'hui, la communication sans fil exige les opérateurs de réseau qui sont très capables. Surement, la base de ces réseaux est la disponibilité des ressources spectrales. Même si l'étroitesse de la bande passante (BW) a gardé le rythme avec les besoins ci-dessus mentionnés jusqu'ici, l'allocation spectrale courante basse sur le commandement et le contrôle philosophique, qui attribue le spectre aux applications qu'aux machines, basée sur la demande, est sur le point de ne plus satisfaire les besoins de la future en communication

massive de la société moderne. Les études statistiques récentes sur l'utilisation du spectre dévoilent qu'une partie importante du spectre est temporellement et géographiquement non utilisée tandis que certaines portions sont très surchargées [2–6]. Ceci a motivé la communauté des chercheurs de prendre l'initiative et d'introduire un nouveau paradigme pour l'utilisation spectrale, connue sous le nom d'allocation dynamique du spectre (DSA) ou radio cognitive (CR), renforcée par l'architecture radio logicielle (SDR). CR s'est maintenant imposé comme la solution ultime pour remédier à la sous-utilisation courante et l'allocation inefficiente du spectre. En étant rapide et conscient, un nœud de CR est capable d'adapter efficacement ses paramètres tels que la puissance, la fréquence, le débit de données, etc., à l'évolution des circonstances afin d'exploiter au maximum les possibilités de spectre disponible dans les domaines du temps, de l'espace, et la fréquence.

A.2 Motivation et Vision

Le plus grand avantage des réseaux de radio cognitive (CNRs) sur les réseaux existants est l'efficacité spectrale améliorée qu'elle promet. CR est réputée avoir des effets élémentaires sur toutes les parties impliquées et à différents niveaux. Tout d'abord, elle offre à la société plus de chances pour un développement durable car elle libère les potentiels de communication et ouvre de nouvelles portes. Deuxièmement, il donne aux individus une meilleure qualityé de service (QoS) et une pléthore de nouveaux services, en payant moins pour ce qu'ils reçoivent car maintenant, ils ont affaire à un marché plus concurrentiel. Troisièmement, elle fournit aux opérateurs de réseaux plus d'utilité, car les clients ont des exigences élevées et en constante évolution. Enfin, elle aide les organismes de réglementation et les gouvernements directement et indirectement, car ils peuvent en tirer plus d'impôts en louant plus de licences, et en ayant de nouveaux marchés et d'emplois créés, respectivement. Néanmoins, en dépit des avantages et applications mentionnés, il y a des défis qui y sont impliqués. Malheureusement, après des décennies d'étude et de mise au point, les

obstacles et les incertitudes dans ce domaine dépassent encore la détermination et les certitudes. Dans un sens large, ce qui a été accompli par la communauté des chercheurs à ce jour, a été étendu, mais désorienté avec déficiences telles que indiquées ci-dessous:

1. Choix des hypothèses: les modèles de réseau qui sont créés sont pour la plupart basés sur les hypothèses qui sont soit trop simpliste ou irréaliste.

2. Validation: Les modèles présentés à ce jour ont parfois été validés par les prototypes implémentés ou, au moins, les plates-formes de simulation.

3. Canale-unique vs. Multicanal: les modèles de réseau qui ont été créés sont majoritairement orientés sur CRNs à canal unique, alors que les principaux potentiels de CR peuvent être exploités dans CRNs multicanal, où le spectre est à considérer comme une ressource divisible.

4. Synthèse vs. Modélisation: Une question importante qui est restée sans réponse jusqu'ici est de loin le problème de modélisation inverse. Ce problème, qui est également connue comme problème de synthèse, répond à la question sur si les exigences d'un service/application peuvent être satisfaites, sachant que la quantité d'interférences sur les utilisateurs primaires (PUS) est limitée à un certain degré.

5. Distributions vs Statistiques: Même si un RV est complètement caractérisée par sa fonction de densité de probabilité (PDF), la plupart des études existantes utilisent les moments et les statistiques, qui fournissent des informations partielles sur les caractéristiques d'un phénomène.

6. Régimes à forte circulation et de stabilité: Les réseaux ne fonctionnent toujours pas dans les régimes à faible trafic et leur performance lorsque l'intensité du trafic est élevée et la diminution de la réactivité du réseau sont d'une importance primordiale. C'est dans cette dernière situation que la situation d'instabilité augmente, comme prédit par la théorie des queues, qui ont été à peine traitée dans les études précédentes.

7. Modélisation entre-couche: L'approche de la modélisation qui est confinée à une seule couche sans tenir compte de l'influence que les différentes couches présentent l'une sur l'autre, ne peut pas être précise.

Le même ensemble de lacunes est observé dans le domaine de la conception du mécanisme. Cette désorientation a abouti à des modèles économiques ainsi que des normes techniques moins développés. Pour toutes ces raisons, les partenaires industriels ont été jusqu'ici réticents à prendre des mesures pour la commercialisation de CR pour le fait qu'ils ne sont pas sûrs que la DSA sera une solution durable et rentable qui peut en réalité éviter les besoins de communication éminemment déferlantes des clients. Conformément à ce raisonnement, et ayant à l'esprit les défauts mentionnés ci-dessus, je l'ai trouvé une tâche impérative de prouver la supériorité des CNRs multi-canal en établissant des modèles bien définis, bien structurés et plus précis dans la première étape de mon programme de doctorat.

Au-delà de la modélisation du système et l'analyse des performances, il existe d'autres questions importantes en CR qui doivent être adressées. Mettre sous le titre de mécanisme la conception et le développement des protocoles, dont la gestion du spectre, la détection du spectre, la mobilité du spectre et le partage du spectre, sont optimistes, les enquêtes ont été moins productives, en particulier pour les CNRs multi-canal. Une telle pénurie n'est pas une question de discussion et est proclamée dans la norme 802.22 [8] sous forme de questions ouvertes qui sont destinés à traiter et améliorer collectivement par la communauté. Inspiré par ces faits, j'ai décidé de consacrer la deuxième phase de mon programme de doctorat à la conception des mécanismes de partage du spectre et de répartition efficace du spectre et des protocoles pour les CRNs multi-canal.

A.3 Objectifs de Recherches

Après avoir discuté les motivations pour dédier ma thèse aux CRNs multi-canal, les objectifs de recherche de mon doctorat sont centrés autour de l'analyse de la performance, la modélisation et le mécanisme de conception tel que détaillé successivement.

A.3.1 Modélisation et Analyse de Performances

Basé sur le point de vue que j'exposais dans la section précédente, un outil de modélisation mathématique pour les CNRs est important de différentes façons: Tout d'abord, il assure l'adéquation du CR comme une approche durable pour l'amélioration de l'efficacité spectrale; deuxièmement, il peut être utilisé comme un indice de référence des moindres coûts pour des fins de pré-évaluation. Ayant cette vision à l'esprit, j'ai concentré la première partie de mon voyage de recherche à proposer des approches de modélisation solides qui pourraient être utilisées pour évaluer la performance des CNRs multi-canal. Conformément à cet objectif général, mes objectifs de recherche de modélisation sont tirés de trois points de vue différents. Au cours de la première perspective, je concentre mes efforts sur

- l'évaluation de la modélisation et de la performance des CNRs Single-interface multi-canal (SIMC): (1) CRNs génériques; (2) les réseaux régionaux sans fil (WLAN).
- l'évaluation de la modélisation et de la performance des CNRs multi-interface multi-canal (CMIM).

Dans la seconde perspective, mon effort de modélisation incarne quatre aspects différents, à savoir

- Modélisation
- Analyse

- Synthèse
- Optimalité

Alors que dans les deux premiers aspects, notre accent serait mis sur l'établissement d'un cadre mathématique qui peut être utilisé pour l'évaluation de la performance des CNRs multi-canal, la partie de synthèse répond à la question si les exigences d'un service /application peut on être satisfaite, sachant la contrainte du réseau et des opérateurs. Celui-ci devient une force motrice pour les discussions d'optimalité. Notre approche proposée est différente des études antérieures, aussi bien dans la méthodologie que dans l'étendue. À partir de la dernière perspective, j'essaye de remplir mes objectifs en trois directions distinctes, à savoir

- Modélisation numérique
- Dérivation forme-fermée
- Validation de la simulation

En raison, il fait que nos modèles proposés soient fondés sur des chaînes dynamiques de Markov; parfois, la complexité de la résolution numérique devient extrêmement importante ce qui pourrait diminuer leurs mérites pratiques. Par conséquent, des expressions à formes fermées pour les paramètres pertinents qui peut en estimer les performances du réseau très rapide et sans frais étaient également développées. Puisque l'assurance de la validité des modèles est d'une importance primordiale, je me mets dans ma deuxième phase de recherche à l'écriture de simulateurs pour voir si les modèles analytiques sont correctes, et si oui, de les valider et d'observer les différences d'erreur.

A.3.2 Conception de mécanisme

Le premier chapitre dans la deuxième partie vise la détection du spectre, les problèmes de décision et de mobilité. L'accent sera mis sur la norme IEEE 802.22 WRAN comme étant

la plus importante norme apparue sur les CRNs. Depuis sa première version en 2011, cette norme a été révisée plusieurs fois. Néanmoins, il y a encore beaucoup de travail qui reste à faire étant donné que certaines fonctionnalités de WRAN sont explicitement déclarées comme questions ouvertes. La détection et le mécanisme de fusion (SFM) font partie de ces questions. Étant donné l'importance du processus de détection du spectre en DSA, les règles de combinaison proposées dans la norme WRAN ne sont pas de fiable, précise et stable de manière satisfaisante. Motivé par ces faits, un système distribué efficace (SFM) qui est conforme aux directives standard, et qui peut être facilement intégré dans le réseau est proposé dans cette thèse pour éviter les inconvénients ci-dessus mentionnés.

Le deuxième chapitre dans la deuxième partie de cette thèse aborde le problème de partage du spectre dans les CNRs. Le problème de partage des ressources dans les CNRs est formulé sous la forme d'un jeu d'enchères. Construit sur la théorie récemment émergée de multi-unité optimale de ventes aux enchères, le mécanisme d'allocation proposé dans cette thèse résout plusieurs problèmes profondément enracinés dans les réseaux sans fil. Du côté de l'utilisateur, le compromis équité-optimalité est réglé d'une manière élégante; du côté de l'opérateur de réseau, le total du chiffre d'affaires recueilli est maximisé par l'imposition des paiements. Ces propriétés, avec la facilité de gestion de calcul du problème formulé, ainsi que l'existence et la stabilité de sa solution rendent l'allocation du mécanisme proposée un système idéal pour son exploitation dans les réseaux 5G à base de CR.

A.4 Methodologie

A.4.1 Modélisation de Performance

Parmi les différents ensembles des outils d'analyse, la théorie des files d'attente a montré son efficacité dans la capture des dynamiques des systèmes. En dépit de son importance,

l'application de cet outil de modélisation dans l'analyse des CRNs est nouvelle. Même si les enquêtes sur un seul canal CRNs ont été large, très peu a été découvert sur la classe impérative des CRNs multicanaux en raison du fait que la théorie des files d'attente multi-serveurs est à peine traitée par les mathématiciens, sauf pour des cas très particuliers. Par conséquent, notre méthodologie dans cette thèse est l'exploitation de deux disciplines importantes dans la théorie de file d'attente multiserveurs, qui est

- File d'attente autonomes
  - FIFO: First-in First-out
  - PQ: Priority Queuing
- Réseau de file d'attente

Alors que nous exploitons l'approche autonome de file d'attente FIFO pour la modélisation des CMIM-CRNs, l'approche PQ est utilisée pour l'analyse de SIMC-CRN générique. En fin de compte, nous tirons parti de la théorie du réseau de files d'attente pour modéliser les performances de WRANs. Nos méthodes de modélisation sont basées sur les chaines de Markov stables et dynamiques qui sont facilement compréhensibles et plus logiquement fort dû à la méticuleuse et du fait des choix des transitions, des états et des taux.

A.4.2 Conception du Mécanisme

Le mécanisme de combinaison de décision proposé dans cette thèse est un algorithme de détection multi-canal distribué (MC-LDS) qui est basé sur l'apprentissage et est renforcé à partir du dessous par une pénalité rationnellement récompensée. Ces deux concepts, lorsque combinés, devraient augmenter la précision de détection et de prise de fiabilité. D'un autre point de vue, les logiques de différenciation dans différents niveaux sont intégrées dans le mécanisme, ce qui signifie une différenciation temporelle et spatiale, afin d'adapter le mécanisme de détection à des conditions variables.

Motivé par les similitudes entre les réseaux CRNs et le marché concurrentiel et permis par les progrès continus et les avancements rapides dans le domaine de l'apprentissage machine, la prise de décision, l'optimisation et le traitement du signal, la théorie de la vente aux enchères a été choisie pour l'attribution du spectre entre les utilisateurs sans fil. En structurant le problème de l'allocation des ressources comme un jeu d'enchères d'informations incomplètes, le problème est formulé comme un jeu bayésien où le raisonnement séquentiel d'égoïste, mais les joueurs rationnels (nœuds) convergent vers un équilibre de Nash qui est stable, en convergence rapide, et efficace.

Appendix B

Modelisation, Analyse et Synthèse

Même si beaucoup a été découvert sur l'analyse des réseaux CR mono-canal (CRNs), la modélisation et l'analyse des CRNs multi-canaux ont été à peine découvertes, dont [9, 10, 128] peut en citer. Motivé par ce fait, nous avons décidé de combler cette lacune en établissant un cadre de modélisation pour les CRNs multi-canaux à l'aide d'un modèle de file d'attente stable et bien défini, puis, en développant une approche globale d'analyse et de synthèse pour ce modèle afin d'obtenir des mesures de performance clés, région réalisable, et des solutions optimales.

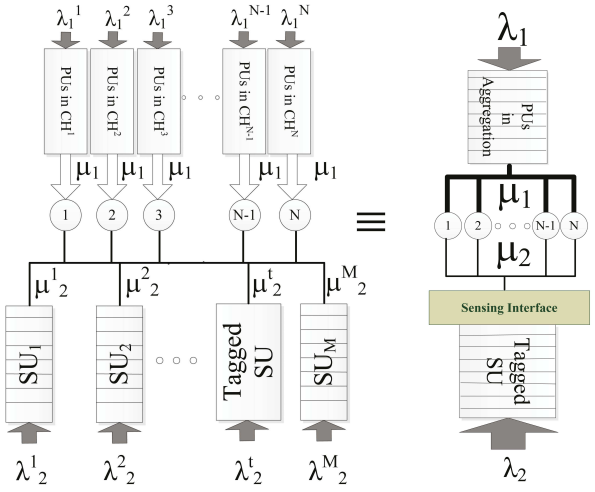
B.1 CRNs à Interface-Unique: Analyse Priority-Queue

Notre premier travail est différent de toute étude précédente dans la méthodologie que nous avons prise ainsi que par la manière avec laquelle nous avons procédée. Alors que notre modélisation dans cette section fournit la distribution de la longueur de la file d'attente des SUs sur la base d'un modèle dynamique, [9] (ou [10]) qui travaille sur les moments de retard (ou la distribution de la queue) selon l'une des méthodes d'approximation. Pour cette raison, nous avons recours à la théorie des files d'attente prioritaires. Bien que cet outil de

modélisation ait trouvé une large application dans d'autres domaines au cours des années, son bénéfice pour l'analyse des CRNs à canal unique n'a été découvert que récemment. Par exemple, dans [12–14], les auteurs utilisent la théorie des files d'attente prioritaires pour trouver plusieurs statistiques de réseaux. Dans la section suivante, nous développons notre modèle de réseau pour des CRNs à une seule interface multi-canal.

B.1.1 Modèle de Développement

L'analogie entre les CNRs et les files d'attente prioritaires tourne autour du fait que, dans les deux cas, les ressources sont partagées entre les classes de trafic avec les droits préférentiels différents. Par conséquent, en représentant chaque canal avec un serveur, le problème de la modélisation CRNs multi-canal peut entrer dans le cadre des files d'attente prioritaires multi-classes multiserveurs, comme illustré sur la Fig. B.1a



(a) Modèl original.

(b) Modèl découplé.

Figure B.1: Modélisation du comportement interactif des SUs et PUs dans le cadre des files d'attentes prioritaires résumé-préemption multi-classe multi-serveur.

Répresentation CTMC

Sur le côté primaire, le réseau de N files d'attente à serveur unique indépendants, avec les taux d'arrivé λ_1^i, peuvent être estimés avec une seule file d'attente N-server, comme le montre la Fig. B.1b, avec un taux d'arrivée équivalent $\lambda_1 = \sum_{i=1}^{N} \lambda_1^i$. Compte tenu de l'hypothèse de découplage, l'un des SUs (appelées SU étiqueté ci-après) peut être détaché du reste du système tel que représenté sur la Fig. B.1b si son taux de service est correctement modifié pour refléter de façon significative la caractéristique du réseau multi-utilisateur (c.a.d $\mu_2^t \Rightarrow \mu_2$, avec t désignant le SU marqué sur la Fig. B.1).

Avec le découplage des files d'attente sur la Fig. B.1, nous pouvons mathématiquement caractériser le CRN dans la Fig. B.1a en utilisant l'outil de modélisation de la chaîne de Markov. Ici, la quantité d'intérêt est le nombre moyen de paquets dans les files d'attente de PU et SU du modèle découplé sur la Fig. B.1b (i et j, respectivement) représenté par la paire (i, j) comme cité dans la chaine de temps continu de Markov 2-D (CTMC) représentée sur la Fig. B.2. Pour économiser de l'espace tout en préservant la clarté, nous avons représenté l'état générique de ce CTMC dans le coin inférieur droit de la Fig. B.2. Maintenant, à partir de l'état (i, j), un déplacement horizontale (verticale) vers la droite (vers le bas) représente l'ajout d'un paquet à la classe SU (PU) au taux $\lambda_2(\lambda_1)$, et le déplacement horizontale (verticale) vers la gauche (vers le haut) représente le départ d'un paquet à partir de l'état (i, j) avec un taux $\mu_2(i, j) = \mu_2 \min(j, \max(N - i, 0))$ $(\mu_1(i, j) = \mu_1 \min(i, N))$.

De toute évidence, cet état infini de CTMC doit avoir une condition de stabilité pour assurer un fonctionnement stable. Cette condition de stabilité, comme indiqué dans [24–28], concerne les grandes quantités de réseau et est inchangé pour toutes sortes d'ordonnancement de priorité, comme suit:

$$0 < \rho = \sum_{i=1}^{r} \frac{\lambda_i}{\mu_i} < N, \tag{B.1}$$

où N est le nombre de serveurs (ou les files d'attente PU), r est le nombre de classes de

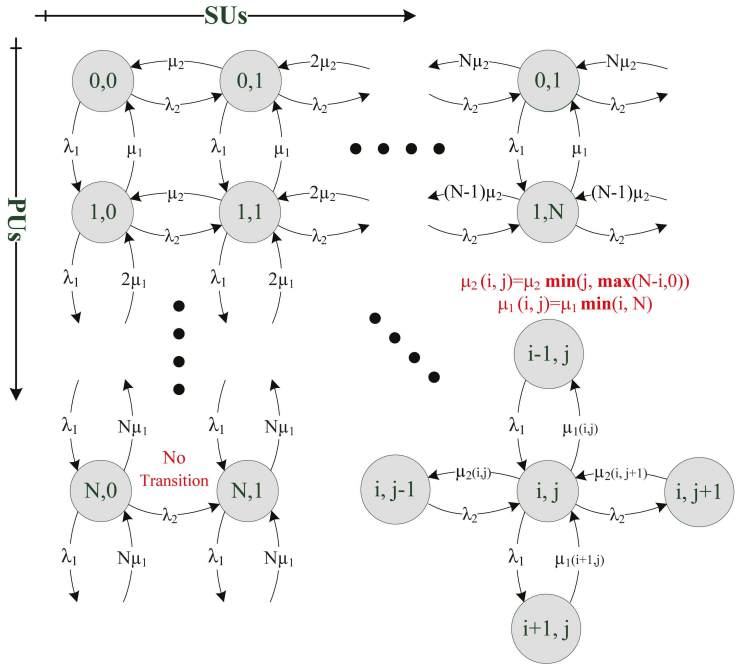


Figure B.2: Représentation du CRN multi-canal avec CTMC 2-D.

priorité et $\rho_i = \lambda_i/\mu_i$ représente le facteur d'utilisation de la i^{me} classe.

Compte tenu de cette explication, nous résolvons cette CTMC en utilisant l'approche de la transformé Z. Plusieurs observations intuitives peuvent être faites à ce point. Tout d'abord, même si le petit changement $\rho_1 \rightarrow \rho_1 + \Delta\rho$ ne change pas sensiblement la PMF marginale primaire, il provoque un changement considérable pour la secondaire, en enroulant et le déplacent son centre rapidement. En second lieu, les examens statistiques réalisés démontrent que la PMF secondaire a une caractéristique de queue lourde, devient plus lourd que le facteur d'activité PUs augmente (ou de manière équivalente parce que ρ_1 devient plus grand).

Fig. B.3 illustre le retard moyen total D_i (file d'attente + Service) dans les deux classes pour le réglage de HTR décrits auparavant. Comme observé, l'augmentation de ρ_1 de 0.6 à 5.46 résulte dans le retard total primaire (D_1) pour augmenter seulement 13% (à droite des ordonnées) tout en tirant le retard total du secondaire (D_2) à environ 600% (gauche des

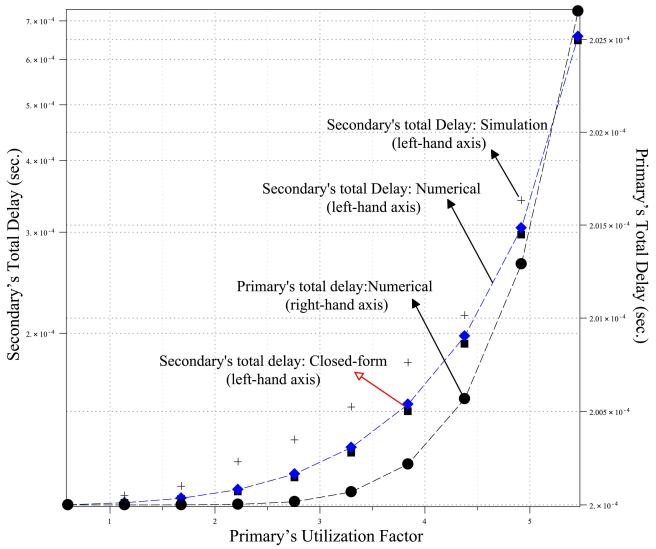


Figure B.3: Numérique, forme fermée, et les résultats de simulation obtenus pour les retards totaux moyens D_1, D_2 (file d'attente + Service) en 2 classes sous le régime de HTR.

ordonnées) ce qui signifie l'importance des mécanismes tels que le contrôle d'admission et le contrôle de congestion. À cette fin, les expressions mathématiques qui caractérisent à la fois les deux classes sont nécessaires, ce qui peut en être utilisés en tant que fonction coût (utilitaire) dans le problème d'allocation de ressource ou d'une fonction de seuil dans le problème de contrôle d'admission. C'est ce que nous poursuivons dans la section suivante.

B.1.2 Synthèse du Réseau

La réponse à la question à savoir si la performance demandée est réalisable ou non, compte tenu des caractéristiques du trafic et les attributs de la file d'attente, et si oui, sous quelle politique, est essentielle à trouver. Le long de la procédure d'analyse de nos résultats, nous sommes arrivés à une expression à forme fermée pour la loi de conservation dans le travail multi-serveurs en conservant les files d'attente prioritaires qui généralisent le travail

dans [38] pour les distributions de temps de service non-identiques, comme suit:

$$\rho_1 D_1 + \rho_2 D_2 = \frac{1}{1 + \sum_{k=0}^{N-1} \frac{N!\,(N-\rho_1-\rho_2)}{k!N\,(\rho_1+\rho_2)^{N-k}}} \cdot \frac{\frac{\rho_1}{\mu_1} + \frac{\rho_2}{\mu_2}}{N-\rho_1-\rho_2} + \frac{\rho_1}{\mu_1} + \frac{\rho_2}{\mu_2}. \tag{B.2}$$

La comparaison du retard total numériquement tracé en fonction du retard total dérivé de (B.2) confirme la validité de ce résultat, avec une précision de plus de 99.5% (cf. Fig. B.3). En outre, un simulateur d'événements discrets a été écrit pour le modèle de réseau dans la Fig. B.1, pour vérifier la précision des traitements analytiques et des simplifications et la validité des hypothèses formulées. Ces résultats prouvent que le modèle de réseau sur la Fig.B.1, son associé CTMC à la Fig. B.2 et l'expression à forme fermée dans l'expression (B.2), tous sont en conformité stricte dans la valeur absolue et le taux du changement. S'il vous plaît noter que le retard total primaire (tracé en noir avec des marques en pointillés) est tout simplement la formule classique pour $\mathrm{M/M/c}$.

Appendix C

Conception du Mecanisme

Dans le chapitre précédent, cette proposition a été consacré à la modélisation et à l'analyse des performances des MC-CRNs. Ce chapitre se concentre sur un autre problème dans les CRN, c'est à dire la conception du mécanisme.

C.1 Mecanisme de Detection

La version de la norme IEEE 802.22, dont nous avons parlé dans le chapitre précédent et que nous avons modélisée à travers des outils théoriques de jeu, a été la nouvelle la plus propice pour ceux qui croient que l'allocation dynamique du spectre est la seule solution viable au problème de l'allocation inefficace des ressources. Cette norme qui résulte d'une décennie de laborieux efforts dans le milieu de la recherche, est encore considérée comme une création pour un long chemin, un fait qui a été bien perçue par les groupes de travail concernés. Une indication de cette affirmation est l'existence des sujets qui sont explicitement laissé comme des questions ouvertes dans [8]. Étant donné que le choix d'un mécanisme de détection distribué omniprésent est parmi ces questions ouvertes, des efforts limités ont été dédiés dont [81–83] peuvent être mentionnés. Dans cette section,

nous proposons un mécanisme de fusion efficace, nommé apprentissage multi-canal basé sur la détection distribuée (MC-LDS), qui évite les inconvénients ci-dessus, est conforme aux directives standard, et peut facilement être intégré comme partie du système.

C.1.1 IEEE 802.22 Spectrum Sensing Fonction

Dans le mécanisme de détection de collaboration proposée dans la norme IEEE 802.22 [8], tous les CPE sont obligés d'effectuer les tâches de détection individuels pendant les périodes calmes (QPS) et les rapportent à la station de base par intermittence. Ce mécanisme est appelé la détection in-band (IBS). Une fois que les résultats de détection des CPE sont signalés à la BS, il est de la responsabilité de la BS de résumer ces informations pour voir si l'un des canaux est temporairement occupé par le réseau primaire ou non. Si inférer à être occupé, ce canal i ainsi que ses deux canaux adjacents $i \pm 1$ doivent être libérés, retirés de la liste des canaux d'exploitation (OCL), et ajoutés à la liste des canaux protégés (PCL). D'autre part, il existe une autre procédure de détection obligatoire, nommé détection out-of-band (OBS) ([7], la clause 10.3.1) qui est nécessaire pour les canaux de détection dans le canal protégé, le candidat (CCL) et le canal de sauvegarde (BCL) ([7], clause 10.2.3). La Fig. C.1 délimite le diagramme de transition de la procédure de mise à jour du canal.

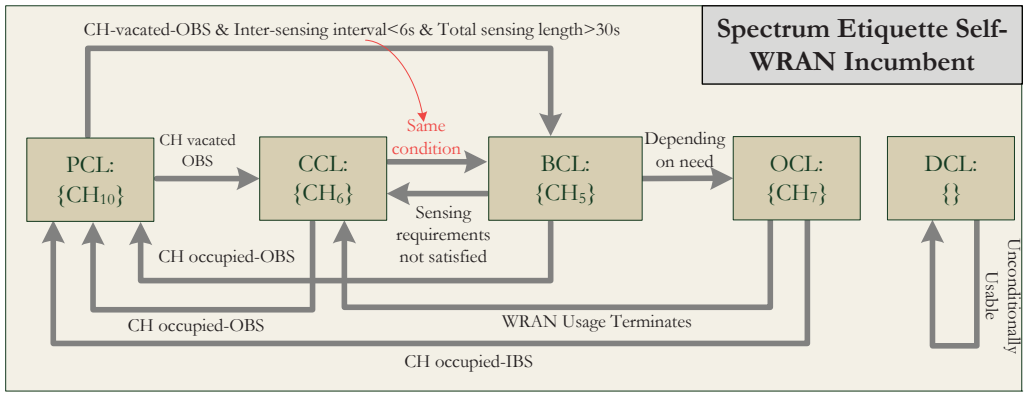


Figure C.1: Diagramme de transition de la liste canal.

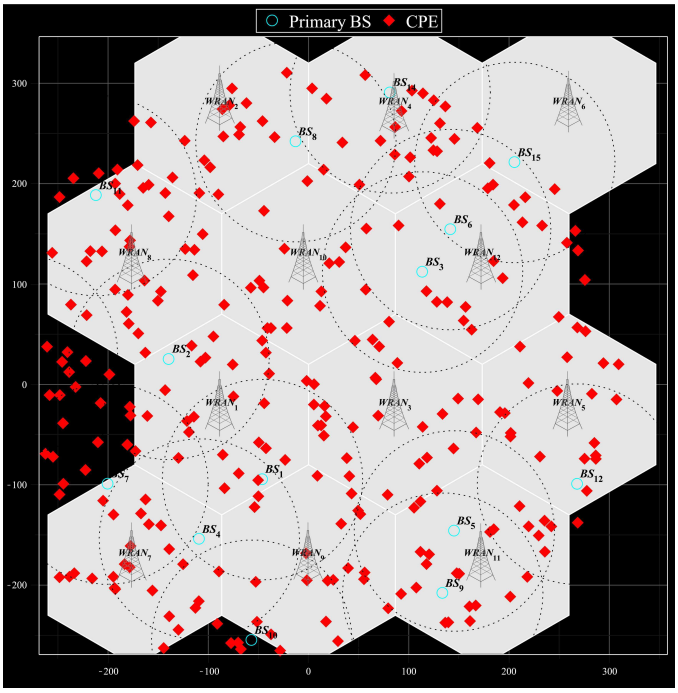


Figure C.2: Un aperçu des 12 cellules WRAN co-existantes.

La norme IEEE 802.22 permet deux modes de fonctionnement: le mode normal (seulement un seul WRAN utilise un seul canal) et le mode de coexistence (un ou plusieurs canaux peuvent être partagés entre les cellules WRAN qui se chevauchent). Dans ce dernier cas, les exigences sans interférences dans le standard IEEE 802.22 WRANs impose des contraintes sur le choix des canaux d'exploitation dans les cellules WRAN voisins. Nous renvoyons le lecteur à [8] pour de plus amples détails sur les mécanismes de gestion des canaux pertinents tels que la commutation de canaux, la mise à jour de la liste de canal, la gestion QP, etc.

C.1.2 Détection Distribuée Basée sur L'Apprentissage Multi-canal

Considérons la réalisation dans la Fig. C.2, qui est une copie d'écran d'une zone à large service avec 12 cellules WRANs co-existantes. La méthode de décision combinée in-

troduite dans la présente proposition est une approche de détection distribuée basée sur l'apprentissage multi-canal qui est renforcée par une récompense de rationalité-pénalité. Dans une autre perspective, la compétence de ce mécanisme découle de deux niveaux de différenciations qu'ils entretiennent: géographiques et temporelles. Dans la différenciation géographique, les opinions des CPE avec des résultats de détection plus fiables seront prises en compte avec des poids plus importants et vice versa. Pour quantifier ce niveau de différenciation, l'approche de la peine de récompense suivante est adoptée en fonction de différentes situations qui peuvent se présenter.

$$\begin{cases} d_i = R, \begin{cases} d_i = D, \text{ Recompense } (+1) \\ d_i \neq D, \text{ Recompense } (+2) \end{cases} \\ \\ d_i \neq R, \begin{cases} d_i = D, \text{ Pénalité } (-2) \\ d_i \neq D, \text{ Pénalité } (-1) \end{cases} \end{cases} \tag{C.1}$$

où d_i représente la mesure la plus récente à la i^{me} CPE, R est l'état du canal reçu de la base de données historique (DB), et D est la décision finale prise par la BS sur l'état du canal. Cette rationalité se résume à la formule logique suivante pour le poids du i^{me} CPE dans l'intervalle k^{me} de détection pour un canal donné.

$$\begin{aligned} L_i(k) = &- d_i^{(k)} \oplus R^{(k)} + d_i^{(k)} \odot R^{(k)} - \\ &\left(\overline{d_i^{(k)}} - d_i^{(k)}\right)\left(\overline{D^{(k)}} R^{(k)} - D^{(k)} \overline{R^{(k)}}\right), \quad i = 1, \cdots, m \end{aligned} \tag{C.2}$$

où m est le nombre de CPE qui détectent un canal donné.

L'autre type de différenciation est de type temporel, de sorte que les mesures antérieures sont pondérés pour avoir moins d'impact que les récentes. En combinant ces différenciations en une seule mesure, chaque CPE (dit i^{me}) dans chaque intervalle de détection (ici

n^{me}) doit générer une métrique de confiance $w_i(n)$ comme suit:

$$w_i\left(n\right) = \sum_{k=n-N}^{n-1} \alpha^{n-k} L_i\left(k\right). \tag{C.3}$$

Dans l'étape suivante, une variable non binaire est calculée en utilisant le détecteur de seuil suivant:

$$X_{i\,(BS)}\left(n\right) = \begin{cases} w_{i\,(BS)}\left(n\right), & d_{i\,(BS)}^{(n)} = 1 \\ \\ -w_{i\,(BS)}\left(n\right), & d_{i\,(BS)}^{(n)} = 0 \end{cases} \tag{C.4}$$

où $d_{i\,(BS)}^{(n)}$ est la dernière décision de détection binaire du i^{th} CPE (BS) au n^{me} QP, alors que $w_{i\,(BS)}\left(n\right)$ est donnée en (C.3). S'il vous plaît noter que, l'indice de la $_{BS}$ est pour tenir compte du fait que en plus d'être le centre de décision de fusion, la station de base devrait également agir comme un capteur indépendant pendant la procédure de détection de canal, tel que mandaté par [8], et, par la suite, ceci est la notation qui sera utilisée. Une fois que les vecteurs $\left[d_i^{(n)}, w_i\left(n\right), X_i\left(n\right)\right]$ sont signalés à la BS, la somme pondérée suivante est calculée à la BS:

$$D_j(n) = \begin{cases} 1, & X_{BS}\left(n\right) + \sum_{i=1}^{m_j} \beta_i X_i\left(n\right) > 0 \\ \\ 0, & X_{BS}\left(n\right) + \sum_{i=1}^{m_j} \beta_i X_i\left(n\right) \leq 0 \end{cases} \tag{C.5}$$

où m_j est le nombre de CPE qui a détecté le j^{me} canal et β_i sont les rapports de métrique d'état de canal. L'implication des coefficients β_i dans (C.5) est importante du point de vue que la précision de la détection est une question différente de la précision de réception. A partir de (C.4), il n'est sous-entendu qu'un $X_i\left(n\right)$ positif dans (C.4), qui renforce la

possibilité de déclarer qu'un canal occupé, est atteint lorsque l'un de ces cas se produit: $[w_{i\,(BS)}\,(n) > 0, d^{(n)}_{i\,(BS)} = 1]$ ou $[w_{i\,(BS)}\,(n) < 0, d^{(n)}_{i\,(BS)} = 0]$. Ce premier cas implique qu'un CPE fiable, qui croit que le canal est occupé dans le QP courant (n^{me}), a un impact direct dans la déclaration d'un canal occupé. De l'autre côté, le second cas implique qu'un CPE fiable, qui croit actuellement que le canal est inactif, a un impact inverse et déclare un canal occupé. En d'autres mots, en utilisant ce mécanisme, les deux fiables et non fiables sont mis à profit pour contribuer de manière constructive dans la décision finale, ce qui améliore évidemment la précision de la décision de sortie. Des déductions similaires peuvent être faites lorsque $X_i\,(n)$ est négatif. Un autre attribut important de ce mécanisme est la relativité de la fiabilité causé par le mécanisme de pointage, qui se traduit par une plage de valeurs pour $X_i\,(n)$.

C.2 Mecanisme d'allocation de Ressource

Les réseaux futuristes sont censés avoir une couverture étendue et un débit extrêmement élevée, et doivent être agnostique au trafic. En générant des centaines d'Exabyte de trafic avec diverses caractéristiques chaque année, les machines seront un abonnés communs de la cinquième génération de réseaux cellulaires (5G). Pour vivre des telles demandes croissantes et des exigences strictes, l'allocation des ressources du spectre est en train de prendre une nouvelle forme. De toute évidence, l'allocation statique n'est plus acceptable. Au contraire, le spectre est à allouer de façon plus dynamique, plus rapide, et pendant des intervalles plus courts.

Pour la croissance constante de l'intelligence dans l'électronique grand public, des mécanismes qui ont prouvés l'efficacité dans l'allocation des produits de base pour l'homme est digne de considération pour allouer des ressources aux appareils sans fil. Les enchères sont parmi ces mécanismes [98]. Parmi les investigations, des raisons impérieuses ont été trouvées à l'égard de la pertinence de ventes aux enchères dans la modélisation et la con-

ception du MC-CRN. Par conséquent, la dernière partie de cette thèse est consacrée à la conception d'un mécanisme *d'allocation des ressources* pour les MC-CRNs en utilisant la théorie des enchères. Ce problème de conception est fragmenté en deux différentes composantes, à savoir *la tarification du spectre* et *l'assignation des fréquences*.

Les algorithmes de partage de spectre conduit DSA sont censés améliorer l'efficacité du spectre en exploitant toutes les informations disponibles (cognition) qu'ils apprennent de l'environnement (par exemple l'information de l'état du canal, la température d'interférence, etc.) par le biais de la résolution d'un problème d'optimisation complexe qui implique l'ajustement des paramètres du système de toute part (par exemple la modulation, le codage, la puissance de transmission, etc.). Résoudre ces problèmes d'une manière centralisée exige une énorme puissance de calcul qui rend l'allocation centralisée impraticable, non évolutive et coûteuse. Donc, l'approche décentralisée semble être une architecture plus appropriée, dans laquelle le gros problème d'optimisation est fragmenté pour les petits problèmes qui doivent être résolus séparément par les SUs basée sur la connaissance partielle que chaque nœud apprend de l'environnement. Avec cette idée, des méthodologies différentes ont été adoptées par les chercheurs jusqu'à présent pour modéliser l'allocation dans les CRNs, comme la théorie de la coloration de graphe [129, 130], la théorie des jeux [74, 131], la théorie stochastique [49], les algorithmes génétiques [132], et les algorithmes intelligents de swarm [133]. Parmi tous ceux-ci, il semble que l'interaction entre les nœuds cognitifs est de meilleures captures dans le cadre de la théorie des jeux. Plus précisément dans le cadre distribué, le rôle que la cognition a dans toutes les étapes du processus de décision de prise de nœuds cognitifs, apporte l'idée de marché parfait concurrentiel en idée, dans laquelle chaque concurrent a des incitations à maximiser sa propre utilité par un ensemble d'actions qu'il peut faire indépendamment par interaction stratégique (Thèse), l'apprentissage et l'adaptation. Cela est tout simplement pourquoi les chercheurs ont commencé d'appliquer la théorie des jeux pour comprendre l'interaction entre les nœuds sans fil intelligents avec des intérêts contradictoires totale ou partielle.

L'analogie étroite entre la compétition stratégique dans les réseaux cognitifs et les jeux non-coopératifs en général, m'a assuré que la théorie des jeux est une alternative cadre pour l'analyse et la modélisation des performances des CRNs (et la théorie du jeu inverse, pour la conception et la mise en œuvre de la CRN avec des propriétés souhaitables et de qualités), par rapport aux solutions développées jusqu'à présent dans cette thèse par le biais de la théorie des files. Mes investigations préliminaires me montrent que l'utilisation de la théorie des jeux permettrait non seulement de valider mes résultats actuels d'un point de vue différent, mais aussi, de concevoir des mécanismes d'allocation efficaces en utilisant des branches dans le vaste domaine de la théorie des jeux inverses, à savoir des ventes aux enchères.

La dernière partie de la thèse sera consacrée à l'étude des CRNs en utilisant la théorie des ventes aux enchères. La raison de ce choix est claire; la ventes aux enchères sont connus comme l'une des plus efficaces mécanismes macroéconomiques pour les scénarios de répartition des ressources limitées non coopératives et distribuées, lorsque la valeur des produits de base (BW cette situation) n'est pas fixe et est déterminée par la demande, plutôt que par les vendeurs. Avec cette idée, une allocation de BW dans un cadre cognitif est une affectation d'un vecteur de ressources $(w_1, w_2, ..., w_N)$ à SUs, où w_i est la BW allouée au i^{me} SU, $\forall i \in \mathcal{N} = \{1, ..., N\}$, et une règle de tarification qui charge ces SUs pour l'attribution qu'ils reçoivent. Cette attribution a lieu sous la forme d'une vente aux enchères, par laquelle après que les demandes de SUS ont été signalées dans la forme de présentation des soumissions $b_i \in B_i$, l'attribution et la tarification des paires correspondantes sont calculées et déclarées séparément à chaque retour SU $[w_i(b_i, b_{-i}), C_i(b_i, b_{-i})]$, $\forall i \in \mathcal{N}$, où b_{-i} est la notation théorique du jeu commun pour représenter la stratégie de l'adversaire de i^{me} SU.

Dans ce contexte, mon objectif est de concevoir un mécanisme d'allocation des ressources basé sur les enchères qui est,

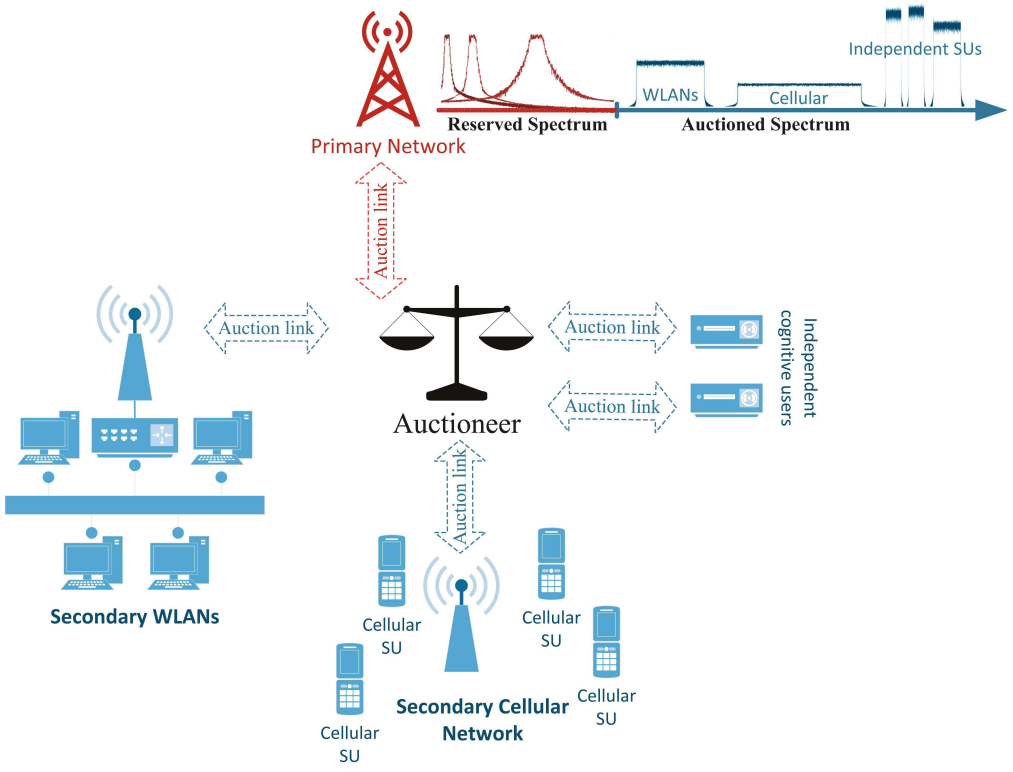


Figure C.3: Allocation de la bande passante aux enchères pour MC-CRNs.

- Efficace: choisit le résultat qui maximise la fonction de protection sociale (Utilité totale).
- Juste: sélectionne le résultat qui satisfait un certain critère d'équité.
- Maximisation du chiffre d'affaires: maximise les revenus des vendeurs.
- Pareto efficace: le mécanisme est efficace en ce sens que, en choisissant un résultat différent, au moins un nœud cognitif se blesse.
- Motivation compatible: motive les nœuds cognitifs de déclarer leurs véritables évaluations que leurs offres.
- Individuellement rationnelle: est non dommageable, ce qui signifie que l'utilité moyenne

des nœuds cognitifs n'est jamais inférieur à zéro. Une conséquence directe de cette propriété est ce nœud qui participe à la compétition, le gain utilitaire qui n'est jamais inférieur à ce qu'il gagne en ne participant pas, indépendamment du fait qu'il perd ou gagne.

- Calcul efficace: converge en très peu de temps et il est possible donc pour des scénarios en temps réel.

La Fig. C.3 est une représentation du cadre d'allocation des ressources basé sur les enchères pour les CRNs multi-canal.

C.3 Conclusion

Il ne fait aucun doute que la vague de la 5G va nous frapper bientôt. Les machines deviendront les abonnés du réseau dominant poussant de côté les dispositifs opérés par les humains. Poussé par cette révolution, le trafic agrégé devrait être immense en taille et de type différent ce qui nous posera des problèmes sans précédent parmi lesquels l'insuffisance du spectre est au cœur même. Dans la recherche d'une solution à ce problème, plusieurs études expérimentales ont été réalisées et ont dévoilé que, dans une large mesure, le spectre n'est pas utilisé de manière efficace. Au cours des dernières années, il est devenu clair que l'allocation statique du spectre est la raison de son utilisation inefficace. L'allocation dynamique du spectre (DSA) a été proposée comme solution au problème de l'inefficacité spectrale. Les travaux dans cette thèse de doctorat ont été motivés par les mêmes idées. L'objectif global a été de justifier la radio cognitive (CR) comme la solution au problème de l'insuffisance du spectre. Pour remplir cet objectif, l'approche de la recherche étendue (Horizontal) a été prise, au lieu de se concentrer profondément sur un sujet très étroit (vertical).Différentes facettes de recherche du CR ont été explorées, dont la modélisation, l'analyse, la synthèse et la conception.

References

[1] K. Fitchard, "Wireless 2025: A look at wireless in the year 2025," Telephony Online, http://telephonyonline.com/, Tech. Rep., Apr. 2009.

[2] FCC, "Notice of proposed rulemaking and order," Federal Communications Commission, Tech. Rep. ET Docket No. 03-108, Dec. 2003.

[3] P. Kolodzy, "Spectrum policy task force: Findings and recommendations," Federal Communications Commission, Tech. Rep., Nov. 2002.

[4] M. McHenry, P. Tenhula, D. McCloskey, D. Roberson, and C. Hood, "Chicago spectrum occupancy measurements, analysis and a long-term studies proposal," in *ACM TAPAS*, Boston, USA, Aug. 2006, pp. 1–5.

[5] M. Islam, "Spectrum survey in Singapore: Occupancy measurements and analyses," in *CrownCom*, Singapore, May 2008, pp. 1–7.

[6] S. D'Iitri and M. McHenry, "Dynamic spectrum access moves to the forefront," DefenseElectronics, Tech. Rep., Apr. 2008.

[7] V. Valenta, R. Marsalek, G. Baudoin, M. Villegas, and M. Suarez, "Survey on spectrum utilization in Europe: Measurements, analyses and observations," in *CrownCom*, Cannes, France, June 2010, pp. 1–5.

[8] IEEE 802.22, "Part 22: Cognitive wireless RAN medium access control (MAC) and physical layer (PHY) specifications: Policies and procedures for operation in the TV bands," *IEEE Standard*, pp. 1–680, July 2011.

[9] S. Wang, J. Zhang, and L. Tong, "Delay analysis for cognitive radio networks with random access: A fluid queue view," in *IEEE INFOCOM*, San Diego, CA, Apr. 2010, pp. 1–9.

[10] A. Laourine, S. Chen, and L. Tong, "Queueing analysis in multichannel cognitive spectrum access: A large deviation approach," in *IEEE INFOCOM*, San Diego, CA, Mar. 2010, pp. 1–9.

[11] L. Wang, C. Wang, and K. Feng, "A queuing-theoretical framework for QoS-enhanced spectrum management in cognitive radio networks," *IEEE Commun. Mag.*, vol. 18, no. 6, pp. 18–26, Dec. 2011.

[12] I. Suliman and J. Lehtomaki, "Queueing analysis of opportunistic access in cognitive radios," in *IEEE CogART Workshop*, Aalborg, Denmark, May 2009, pp. 153–157.

[13] C. Zhang, X. Wang, and J. Li, "Cooperative cognitive radio with priority queuing analysis," in *IEEE ICC*, Dresden, Germany, June 2009, pp. 1550–3607.

[14] C. Do, N. Tran, and C. Hong, "Throughput maximization for the secondary user over multi-channel cognitive radio networks," in *IEEE ICOIN*, Bali, Feb. 2012, pp. 65–69.

[15] H. Tran, T. Duong, and H. Zepernick, "Average waiting time of packets with different priorities in cognitive radio networks," in *IEEE ISWPC*, Modena, Italy, May 2010, pp. 122–127.

[16] L. Zhang, T. Song, M. Wu, J. Guo, D. Sun, and B. Gu, "Modeling for spectrum hand-off based on secondary users with different priorities in cognitive radio networks," in *IEEE WCSP*, Huangshan,China, Oct. 2012, pp. 1–6.

[17] A. Cobham, "Priority assignment in waiting line problems," *Operations Research*, vol. 2, no. 1, pp. 70–76, Feb. 1954.

[18] J. Holley, "Waiting line subject to priorities," *Operations Research*, vol. 2, no. 3, pp. 341–343, Aug. 1954.

[19] H. White and L. Christie, "Queuing with preemptive priorities or with breakdown," *Operations Research*, vol. 6, no. 1, pp. 79–95, Feb. 1958.

[20] B. Avi-Itzhak, "Preemptive repeat priority queues as a special case of the multi-purpose server problem-I," *Operations Research*, vol. 11, no. 4, pp. 597–609, July 1963.

[21] D. Gaver, "A waiting line with interrupted service including priorities," *Journal of the Royal Statistical Society*, vol. 24, no. 1, pp. 73–90, 1962.

[22] N. Jaiswal, "Preemptive resume priority queue," *Operations Research*, vol. 9, no. 5, pp. 732–742, Sep. 1961.

[23] F. Stephan, "Two queues under preemptive priority with poisson arrival and service rates," *Operations Research*, vol. 6, no. 3, pp. 399–418, May 1958.

[24] B. Ngo and H. Lee, "Analysis of a pre-emptive priority M/M/c model with two types of customers and restriction," *IEEE Electronics Letters*, vol. 26, no. 15, pp. 1190–1192, July 1990.

[25] E. Kao and S. Wilson, "Analysis of non-preemptive priority queues with multiple servers and two priority classes," *European Journal of Operational Research*, vol. 1, no. 18, pp. 181–193, Oct. 1999.

[26] I. Mitrani and P. King, "Multiprocessor systems with preemptive priorities," *Performance Evaluation Journal*, vol. 1, no. 2, pp. 118–125, May 1981.

[27] A. Bondi and J. Buzen, "The response times of priority classes under preemptive resume in M/G/m queues," *Operations Research*, vol. 31, no. 3, pp. 456–465, May 1983.

[28] T. Williams, "Non-preemptive multi-server priority queues," *Operations Research*, vol. 31, no. 12, pp. 1105–1107, Dec. 1980.

[29] P. Hokstad, "Approximations for the M/G/m queue," *Operations Research*, vol. 26, no. 3, pp. 510–523, May 1978.

[30] M. Rashid, J. Hossain, E. Hossain, and V. Bhargava, "Opportunistic spectrum access in cognitive radio networks: A queueing analytic model and admission controller design," in *IEEE GLOBECOM*, Vancouver, BC, Nov. 2007, pp. 4647–4652.

[31] M. Rashid, J. Hossain, E. Hossain, and V. Bhargava, "Opportunistic spectrum scheduling for multiuser cognitive radio: A queueing analysis," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5259–5269, Oct. 2009.

[32] T. Apostolopoulos and E. Protonotarios, "Queueing analysis of buffered slotted multiple access protocols," *Computer Communications Journal*, vol. 8, no. 1, pp. 9–21, Feb. 1985.

[33] T. Wan and A. Sheikh, "Performance and stability analysis of buffered slotted ALOHA protocols using tagged user approach," *IEEE Trans. Veh. Technol.*, vol. 49, no. 2, pp. 582–593, Mar. 2000.

[34] M. Carvalho and J. Garcia-Luna-Aceves, "Delay analysis of IEEE 802.11 in single-hop networks," in *IEEE ICNP*, Atlanta, USA, Nov. 2003, pp. 146–155.

[35] G. Bolch, S. Greiner, H. de Meer, and K. S. Trivedi, *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*, 2nd ed. John Wiley and Sons, 2006.

[36] E. Coffman and I. Mitrani, "A characterization of waiting time performance realizable by single-server queues," *Operations Research*, vol. 28, no. 3, pp. 810–821, May 1980.

[37] L. Kleinrock, "A conservation law for a wide class of queueing disciplines," *Naval Research Logistics Quarterly*, vol. 12, no. 2, pp. 181–192, June 1965.

[38] A. Federgruen and H. Groenevelt, "Characterization and optimization of achievable performance in general queuing systems," *Operations Research*, vol. 36, no. 5, pp. 733–741, Sep. 1988 1988.

[39] L. Kleinrock, *Queueing Systems. Volume 1: Theory*. John Wiley and Sons, 1980.

[40] J. Shantikumar and D. Yao, "Multiclass queueing systems: Polymatroid structure and optimal scheduling control," *Operations Research*, vol. 40, no. 2, pp. 293–299, June 1992.

[41] C. Cordeiro, K. Challapali, D. Birru, and S. Shankar, "IEEE 802.22: The first worldwide wireless standard based on cognitive radios," in *IEEE DySPAN*, Baltimore, MD, USA, Nov. 2005, pp. 328–337.

[42] C. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. Shellhammer, and W. Caldwell, "IEEE 802.22: The first cognitive radio wireless regional area network standard," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 130–138, Jan. 2009.

[43] F. Digham, M. Alouini, and M. Simon, "On the energy detection of unknown signals over fading channels," in *IEEE ICC*, Anchorage, USA, May 2003, pp. 3575–3579.

[44] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio networks," in *IEEE DySPAN*, Baltimore, USA, Nov. 2005, pp. 137–143.

[45] A. Sahai, N. Hoven, and R. Tandra, "Some fundamental limits in cognitive radio," in *Conf. Commun., Control and Comput.*, Monticello, Oct. 2004, pp. 1–11.

[46] D. Cabric, S. Mishra, and R. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *IEEE Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, USA, Nov. 2004, pp. 772–776.

[47] A. Fehske, J. Gaeddert, and J. Reed, "A new approach to signal classification using spectral correlation and neural networks," in *IEEE DySPAN*, Baltimore, MD, USA, Nov. 2005, pp. 144–150.

[48] H. Tang, "Some physical layer issues of wide-band cognitive radio system," in *IEEE DySPAN*, Baltimore, MD, USA, Nov. 2005, pp. 151–159.

[49] J. Zhao, H. Zheng, and G. Yang, "Distributed coordination in dynamic spectrum allocation networks," in *IEEE DySPAN*, Baltimore, USA, Nov. 2005, pp. 259–268.

[50] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environment," in *IEEE DySPAN*, Baltimore, MD, USA, Nov. 2005, pp. 131–136.

[51] J. Jackson, "Networks of waiting lines," *Operations Research*, vol. 5, no. 4, pp. 518–521, Aug. 1957.

[52] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Prentice Hall, 1992.

[53] R. Miller, “Priority queues,” *The Annals of Mathematical Statistics*, vol. 31, no. 1, pp. 86–103, Mar. 1960.

[54] B. Avi-itzhak and P. Naor, “On a problem of preemptive priority queuing,” *Operations Research*, vol. 9, no. 3, pp. 664–672, Sep. 1961.

[55] I. Mitrani and B. Avi-Itzhak, “A many-server queue with service interruptions,” *Operations Research*, vol. 16, no. 3, pp. 628–638, May 1968.

[56] G. Cooper, “The computational complexity of probabilistic inference using Bayesian belief networks, journal = "artificial intelligence journal", year = "1990", volume = "42", number = "2-3", pages = "393–405", month = "mar.", note = "", abstract = "", keywords = "", source = "",.”

[57] A. Economou, “A characterization of product-form stationary distributions for queuing systems in random environment,” *International Journal of Simulation*, vol. 4, pp. 24–31, 2003.

[58] B. Avi-itzhak and P. Naor, “Some queuing problems with service station subject to breakdown,” *Operations Research*, vol. 11, no. 3, pp. 303–320, May 1963.

[59] A. Maaref and S. Aïssa, “Exact error probability analysis of rectangular QAM for single and multichannel reception in Nakagami-m fading channels,” *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 214–221, Jan. 2009.

[60] A. Yau, P. Komisarczuk, and P. Teal, “On multi-channel MAC protocols in cognitive radio networks,” in *IEEE ATNAC*, Australia, Dec. 2008, pp. 300–305.

[61] J. Jia, Q. Zhang, and X. Shen, “HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 106–117, Jan. 2008.

[62] C. Cordeiro and K. Challapali, "C-MAC: A cognitive MAC protocol for multichannel wireless networks," in *IEEE DySPAN*, Dublin, Ireland, Apr. 2007, pp. 147–157.

[63] K. Ghaboosi, M. Latva-aho, and Y. Xiao, "A distributed multi-channel cognitive MAC protocol for IEEE 802.11s wireless mesh networks," in *CrownCom*, Singapore, May 2008, pp. 1–8.

[64] L. Ma, X. Han, and C. Shen, "Dynamic open spectrum sharing MAC protocol for wireless ad-hoc networks," in *IEEE DySPAN*, Baltimore, USA, Nov. 2005, pp. 203–213.

[65] H. Salameh, M. Krunz, and O. Younis, "MAC protocol for opportunistic cognitive radio networks with soft guarantees," *IEEE Trans. Mobile Comput.*, vol. 8, no. 10, pp. 1339–1352, Oct. 2009.

[66] Y. Kondareddy and P. Agrawal, "Synchronized MAC protocol for multi-hop cognitive radio networks," in *IEEE ICC*, Beijing, China, May 2008, pp. 3198–3202.

[67] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisioning over cognitive radio wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 118–129, Jan. 2008.

[68] B. Hamdaoui and K. Shin, "OS-MAC: An efficient MAC protocol for spectrum-agile wireless networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 8, pp. 915–930, Aug. 2008.

[69] M. Khabazian, S. Aïssa, and N. Tadayon, "Performance modeling of a two-tier primary-secondary network operated with IEEE 802.11 DCF mechanism," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3047–3057, Sep. 2012.

[70] I. Akyildiz, W. Lee, M. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks Journal*, vol. 50, no. 13, pp. 2127–2159, Sep. 2006.

[71] T. Krishna and A. Das, "A survey on MAC protocols in OSA networks," *Computer Networks Journal*, vol. 53, no. 9, pp. 1377–1394, June 2009.

[72] A. Domenico, E. Strinati, and M. Benedetto, "A survey on MAC strategies for cognitive radio networks," *Commun. Surveys Tuts.*, vol. 14, no. 1, pp. 21–44, Feb. 2010.

[73] K. Chowdhury and I. Akyildiz, "Cognitive wireless mesh networks with dynamic spectrum access," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 168–181, Jan. 2008.

[74] C. Zou and C. Chigan, "A game theoretic DSA-Driven MAC framework for cognitive radio networks," in *IEEE ICC*, Beijing, China, May 2008, pp. 4165–4169.

[75] E. Askari and S. Aïssa, "Full-duplex cognitive radio with packet fragmentation," in *IEEE WCNC*, Istanbul, Turkey, Apr. 2014, pp. 1502–1507.

[76] E. Askari and S. Aïssa, "Single-band full-duplex MAC protocol for distributed access networks," *IET Communications Journal*, vol. 8, no. 10, pp. 1663–1673, July 2014.

[77] T. Hassen, "Synchronization in cognitive overlay systems," MSc thesis, Aalto University.

[78] I. Akyildiz, W. Lee, and K. Chowdhury, "CRAHNs: Cognitive radio ad-hoc networks," *Ad Hoc Networks Journal*, vol. 7, no. 5, pp. 810–836, July 2009.

[79] D. Cox and V. Isham, *Point Processes*. Chapman and Hall, 1980.

[80] V. Asghari and S. Aïssa, "Resource sharing in cognitive radio systems: Outage capacity and power allocation under soft sensing," in *IEEE GLOBECOM*, New Orleans, USA, Nov. 2008, pp. 1–5.

[81] S. Lim, H. Jung, and S. Song, "Cooperative spectrum sensing for IEEE 802.22 WRAN system," in *IEEE ICCCN*, San Francisco, USA, Aug. 2009, pp. 1–5.

[82] H. Kim and K. Shin, "In-band spectrum sensing in IEEE 802.22 WRANs for incumbent protection," *IEEE Trans. Mobile Comput.*, vol. 9, no. 12, pp. 1766–1779, Dec. 2010.

[83] A. Min, K. Shin, and X. Hu, "Secure cooperative sensing in IEEE 802.22 WRANs using shadow fading correlation," *IEEE Trans. Mobile Comput.*, vol. 10, no. 10, pp. 1434–1447, 2011.

[84] A. Sahai, R. Tandra, S. M. Mishra, and N. Hoven, "Fundamental design tradeoffs in cognitive radio systems," in *ACM TAPAS*, Boston, USA, Aug. 2006.

[85] S. Mishra, A. Sahai, and R. Brodersen, "Cooperative sensing among cognitive radios," in *IEEE ICC*, Istanbul, Turkey, June 2006, pp. 1658–1663.

[86] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio-part I: Two user networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2204–2213, June 2007.

[87] J. Unnikrishnan and V. Veeravalli, "Cooperative sensing for primary detection in cognitive radio," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 18–27, Feb. 2008.

[88] J. Ma and Y. Li, "Soft combination and detection for cooperative spectrum sensing in cognitive radio networks," in *IEEE GLOBECOM*, Washington, USA, Nov. 2007, pp. 4502–4507.

[89] E. Vistotsky, S. Kuffner, and R. Peterson, "On collaborative detection of TV transmissions in support of dynamic spectrum sharing," in *IEEE DySPAN*, Baltimore, USA, Nov. 2005, pp. 338–345.

[90] X. Huang, N. Han, G. Zheng, S. Sohn, and J. Kim, "Weighted-collaborative spectrum sensing in cognitive radio," in *CHINACOM*, Shanghai, China, Aug. 2007, pp. 110–114.

[91] T. Qin, H. Yu, C. Leung, Z. Shen, and C. Miao, "Towards a trust aware cognitive radio architecture," in *ACM SIGMOBILE*, ser. 2, vol. 13, NY, USA, Apr. 2009, pp. 86–95.

[92] E. F. I. Witten and M. Hall, *Data Mining Practical Machine Learning Tools and Techniques*, 1st ed. Morgan Kaufmann, 2011.

[93] FCC, "Before the Federal Communications Commission," https://www.ic.gc.ca, Washington, D.C., Nov. 2008, eT Docket No. 04-186 and 02-380.

[94] M. Gudmundson, "A correlation model for shadow fading in mobile radio," *IEEE Electronic Letters*, vol. 27, no. 23, pp. 2145–2146, Nov. 1991.

[95] H. Chernoff and E. Lehmann, "The use of maximum likelihood estimates in χ^2 tests for goodness of fit," *The Annals of Mathematical Statistics*, vol. 25, no. 3, pp. 579–586, 1954.

[96] Ericsson, "On the Pulse of Networked Society," Ericsson Mobility Report, June 2015.

[97] Huawei, "5G: A Technology Vision," Huawei Technologies, Shenzhen, China, Technical Report, 2013.

[98] V. Krishna, *Auction Theory*, 2nd ed. Academic Press, 2009.

[99] R. Myerson, *Game Theory: Analysis of Conflict*, 1st ed. Harvard University Press, 1991.

[100] E. Maskin, J. Riley, F. Hahn, "Optimal multi-unit auctions," *Economics of Missing Markets, Information, and Games*, pp. 312–335, Mar. 1989, reprinted in P. Klemperer, The Economic Theory of Auctions, London: Edward Elgar, 2000.

[101] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y. Yao, "Opportunistic spectrum access in cognitive radio networks: Global optimization using local interaction games," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 2, pp. 180–194, Mar. 2012.

[102] Z. Han, C. Pandana, and K. Liu, "Distributive opportunistic spectrum access for cognitive radio using correlated equilibrium and no-regret learning," in *IEEE WCNC*, Hong Kong, Mar. 2007, pp. 11–15.

[103] F. Wu, S. Zhong, and C. Qiao, "Globally optimal channel assignment for non-cooperative wireless networks," in *IEEE INFOCOM*, Phoenix, USA, Apr. 2008, pp. 1543–1551.

[104] Z. Han, Z. Ji, and K. Liu, "Non-cooperative resource competition game by virtual referee in multi-cell OFDM networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 6, pp. 1079–1090, Aug. 2007.

[105] M. Felegyhazi, M. Cagalj, S. Bidokhti, and J. Hubaux, "Non-cooperative multi-radio channel allocation in wireless networks," in *IEEE INFOCOM*, Anchorage, USA, May 2007, pp. 1442–1450.

[106] H. Kwon and B. G. Lee, "Distributed resource allocation through non-cooperative game approach in multi-cell OFDMA systems," in *IEEE ICC*, Turkey, Istanbul, June 2006, pp. 434–4350.

[107] F. Meshkati, M. Chiang, H. Poor, and S. Schwartz, "A game-theoretic approach to energy-efficient power control in multicarrier CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 6, pp. 1115–1129, June 2006.

[108] Y. Xu, J. Wang, and Q. Wu, "Opportunistic spectrum access in unknown dynamic environment: A game-theoretic stochastic learning solution," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1380–1391, Apr. 2012.

[109] H. Li, "Multi-agent Q-learning of channel selection in multi-user cognitive radio systems: A two by two case," in *IEEE SMC*, San Antonio, USA, Oct. 2009, pp. 1893–1898.

[110] ——, "Multi-agent Q-learning for competitive spectrum access in cognitive radio systems," in *IEEE SECON*, Boston, USA, June 2010, pp. 1–6.

[111] X. Chen and J. Huang, "Evolutionarily stable spectrum access," *IEEE Trans. Mobile Comput.*, vol. 12, no. 7, pp. 1281–1293, Apr. 2012.

[112] A. Hoang and Y. Liang, "Dynamic spectrum allocation with second-price auctions: When time is money," in *CrownCom*, Singapore, May 2008, pp. 1–6.

[113] J. Sun, E. Modiano, and L. Zheng, "Wireless channel allocation using an auction algorithm," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 1085–1096, May 2006.

[114] D. Niyato and E. Hossain, "Competitive spectrum sharing in cognitive radio networks: A dynamic game approach," *IEEE Trans. on Wireless Commun.*, vol. 7, no. 7, pp. 2651–2660, July 2008.

[115] J. Jia and Q. Zhang, "Competitions and dynamics of duopoly wireless service providers in dynamic spectrum market," in *ACM MobiHoc*, Hong Kong, May 2008, pp. 313–322.

[116] R. Myerson, "Optimal auction design," *Operations Research*, vol. 6, no. 1, pp. 58–73, Feb. 1981.

[117] E. Maskin and J. Riley, "Monopoly with incomplete information," *Rand Journal of Economics*, vol. 15, no. 2, pp. 171–196, Summer 1984.

[118] F. Wang, M. Krunz, S. Cui, "Price-based spectrum management in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 74–87, Feb. 2008.

[119] J. Mwangoka, K. Letaief, Z. Cao, "Joint power control and spectrum allocation for cognitive radio networks via pricing," *Physical Communication Journal*, vol. 2, no. 1-2, pp. 103–115, Mar. 2009.

[120] W. Wang, Y. Cui, T. Peng, and W. Wang, "Non-cooperative power control game with exponential pricing for cognitive radio network," in *IEEE VTC*, Dublin, Ireland, Apr. 2007, pp. 3125–3129.

[121] S. Kucera, S. Aïssa, and S. Yoshida, "Adaptive channel allocation for enabling target SINR achievability in power-controlled wireless networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 833–843, Feb. 2010.

[122] C. S. Jean and B. Jabbari, "Bayesian game-theoretic modeling of transmit power determination in a self-organizing CDMA wireless network," in *IEEE VTC*, vol. 5, Los Angeles, USA, Sep. 2004, pp. 3496–3500.

[123] Y. Zhao, S. Mao, J. Neel, and J. Reed, "Performance evaluation of cognitive radios: Metrics, utility functions, and methodology," *Proc. IEEE*, vol. 97, no. 4, pp. 642–659, Apr. 2009.

[124] A. Goldsmith, *Wireless Communications*, 1st ed. 2005, Cambridge University Press.

[125] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. Cambridge University Press, 2004.

[126] E. Rasiel, *The McKinsey Way*, 1st ed. McGraw-Hill, 1999.

[127] J. S. E. Dahlman, S. Parkvall, *4G: LTE/LTE-Advanced for Mobile Broadband*, 2nd ed. Academic Press, 2013.

[128] N. Tadayon and S. Aïssa, "Modeling and analysis framework for multi-interface multi-channel cognitive radio networks," *IEEE Trans. Wireless Comm.*, vol. 14, no. 2, pp. 935–947, Oct. 2014.

[129] H. Zheng and C. Peng, "Collaboration and fairness in opportunistic spectrum access," in *IEEE ICC*, vol. 5, Seoul, South Korea, May 2005, pp. 3132–3136.

[130] D. Willkomm, M. Bohge, D. Hollos, J. Gross, and A. Wolisz, "Double hopping: A new approach for dynamic frequency hopping in cognitive radio networks," in *IEEE PIMRC*, vol. 5, Cannes, France, Sep. 2008, pp. 1–6.

[131] F. Wang, O. Younis, and M. Krunz, "GMAC: A game-theoretic MAC protocol for mobile ad hoc networks," in *IEEE WiOpt*, Boston, USA, Apr. 2006, pp. 1–9.

[132] T. Rondeau, B. Le, C. Rieser, and C. Bo, "Cognitive radios with genetic algorithms: Intelligent control of software defined radios," in *SDR Forum Tecn. Conf.*, Phoenix, USA, Nov. 2004, pp. 3–8.

[133] B. Atakan and O. Akan, "Biologically-inspired spectrum sharing in cognitive radio networks," in *IEEE WCNC*, Kowloon, Hong Kong, Mar. 2007, pp. 43–48.