Quality of service in heterogeneous mobile constrained networks

Guillaume Artero Gallardo

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Réseaux, Télécommunications, Systèmes et Architecture

Présentée et soutenue par :

M. GUILLAUME ARTERO GALLARDO

le lundi 2 mars 2015

Titre :

QUALITE DE SERVICE DANS DES ENVIRONNEMENTS RESEAUX MOBILES, CONTRAINTS ET HETEROGENES

Ecole doctorale :

Mathématiques, Informatique, Télécommunications de Toulouse (MITT)

Unité de recherche :

Institut de Recherche en Informatique de Toulouse (I.R.I.T.)

Directeur(s) de Thèse :

M. ANDRE LUC BEYLOT

M. GENTIAN JAKLLARI

Rapporteurs :

M. HERVE RIVANO, INRIA

Mme ISABELLE GUERIN LASSOUS, UNIVERSITE LYON 1

Membre(s) du jury :

M. THOMAS NOEL, UNIVERSITE STRASBOURG 1, Président

M. ANDRE LUC BEYLOT, INP TOULOUSE, Membre

M. GENTIAN JAKLLARI, INP TOULOUSE, Membre

Mme LUCILE CANOURGUES, ROCKWELL COLLINS STE, Membre

M. PAUL MUHLETHALER, INRIA, Membre

à Henri-Pierre

Acknowledgements

Les voici enfin, les fameuses pages de remerciements, celles dont on songe tout au long de la thèse et qui reflètent les sentiments du jeune docteur.

Je commence par remercier mon directeur de thèse, André-Luc Beylot, pour sa confiance et son encouragement dans la réalisation de mes projets et plus généralement la qualité de son encadrement. Je le remercie de m'avoir proposé mon sujet de thèse et avoir constitué dans ce cadre une très belle équipe d'encadrants. Fortement demandé, il se rend toujours disponible pour venir en aide à ses thésards et participe activement à la bonne ambiance qui règne au laboratoire. À ce titre, je le remercie d'avoir organisé les différentes activités de cohésion d'équipe qui ont contribué au bon déroulement et au succès de cette thèse.

Je remercie ensuite mon encadrant Gentian Jakllari pour m'avoir guidé et m'avoir toujours poussé à me depasser. Je le salue pour sa très grande implication dans la réalisation de nos articles. En plus de m'avoir toujours littéralement "boosté" dans mon travail, il m'a apporté beaucoup d'enseignements et a notamment contribué à l'amélioration de mon niveau d'anglais ! En tant que premier thésard encadré par Gentian, j'espère bien evidemment avoir été à la hauteur.

Je remercie enfin mon encadrante Lucile Canourgues, qui m'a élevé au sein de la société Rockwell Collins en m'enseignant le fonctionnement de l'entreprise ainsi que le métier d'ingénieur systèmes. Je la remercie d'avoir été à l'initiative de cette thèse et m'avoir toujours montré sa confiance et son intérêt dans la réalisation de mes recherches. Un grand merci pour avoir été aussi impliquée dans mon intégration au sein du groupe avant de me laisser voler de mes propres ailes.

Je remercie Isabelle Guérin-Lassous et Hervé Rivano pour m'avoir fait l'honneur de rapporter ma thèse. Je remercie également Thomas Noel pour avoir accepté de présider mon jury ainsi que Paul Mühlethaler pour en avoir fait partie. Je les remercie pour leurs remarques faites au cours de la soutenance qui m'ont permis d'améliorer mon travail.

Je remercie mes collègues de Rockwell Collins qui ont œuvré à la bonne réalisation de cette thèse et avec qui j'ai eu plaisir à partager des moments de discussion, des repas et des sorties. Mention spéciale pour la team Sécu et tous les stagiaires. Je remercie tout particulièrement mon chef de service, Olivier Le Toumelin, qui s'est toujours rendu disponible pour satisfaire mes requêtes et m'a donné l'opportunité d'encadrer

deux stages d'élèves ingénieurs, une expérience qui a beaucoup compté dans le succès de ma thèse.

Je remercie ensuite tous mes collègues de l'IRIT sans qui cette belle histoire ne se serait pas écrite. Je m'adresse tout d'abord aux permanents de l'équipe IRT : Manu qui a été mon co-bureau, Julien, Riadh, Jérôme, Carlos, Katia, Béa, Jean-Luc et Christian. Je remercie également Nathalie, Benoît, Corinne, Martial, Jean-Yves et Xavier pour m'avoir encouragé depuis que je suis étudiant TR. Merci aussi aux trois secrétaires de l'IRIT : Sylvie, SAM et Audrey pour leur efficacité et les services qu'elles m'ont rendu.

Je remercie à présent mes amis thésards et ex-thésards qui se seront confrontés au même défi. D'abord les anciens : Razvan, Fabrice, Fabian, Gabriel, Tony, Michaël, Adnane, Xiaoting, Mauricio, Vincent. Merci ensuite aux moins anciens qui ont su me montrer le chemin : Micka, Clément, Patou et ses défis exceptionnels et Renaud pour ses réunions de thèse. Je les remercie même s'ils auraient bien mérité quelques petits coups de clochette supplémentaires ! Merci enfin à ceux qui sont toujours dans la bataille : Nesrine, Aziz, JB et ses sessions "juste une !", Farouk, les Mohammed, Éric, Émilie, Élie, Maia et Ane. Merci aussi à la team signal avec Bouchra, Jeannot, Tarik et Raoul. Mention spéciale pour JGK qui m'a aidé dans la réalisation de mes travaux et m'a apporté de nouvelles connaissances sur le vélo et la Corse.

C'est avec grand plaisir que je remercie mes amis Toulousains de l'ENVT ainsi que les membres de la "Vic'Team" parmi lesquels figurent Nico, Walid, Lala, JK, TDU, Delphine, Damien, JB, Iliès. Je pense aussi à tous les enseeihtiens avec qui j'ai passé de belles années d'école, les belles rencontres faites à Dublin, mes amis fermateux ainsi que mes vieux copains du TKD !

J'adresse maintenant une pensée affectueuse pour mes "copains de Perpi" : Renaud, Grégoire, Marwan, Hiep, Anne-Laure, Florine, Fanny, Mathilde, Émilie. Je les remercie pour leur amitié qui dure depuis tant d'années. J'ai une pensée profonde pour Henri-Pierre, ami de toujours, dont les encouragements et conseils m'auront permis d'en arriver là où je suis aujourd'hui.

Je m'adresse enfin à ma famille et belle famille en les remerciant pour tout leur amour et leurs encouragements. Merci à mes parents, Martine et Martin, mon frère Benjamin, mes grands-parents, oncles, tantes, cousins qui ont toujours cru en moi et ont manifesté la plus grande des fiertés. Merci à mes beaux-parents Isabelle et Philippe ainsi que mon beau-frère Vincent pour l'accueil qu'ils m'ont réservé dans leur famille.

Souhaitant garder le meilleur pour la fin, je remercie enfin ma compagne Mélodie pour son soutien tout le long de ces trois années de thèse et l'amour qu'elle me porte depuis que nous nous sommes rencontrés. Petite de taille mais grande d'esprit elle a su m'accompagner dans les moments les plus difficiles.

Résumé

Les télécommunications sans fil ont connu ces dernières années un immense succès à tel point que le spectre des fréquences est désormais surchargé. *A contrario*, certaines bandes de fréquence telles que celles allouées à la télévision analogique ne sont quasiment plus utilisées. À cet effet, des techniques de réutilisation dynamique du spectre ont vu le jour sous la dénomination de radio cognitive. Il s'agira de tester la disponibilité du support afin de ne pas gêner les communications prévues dans ces bandes (celles des utilisateurs dits primaires) et de les partager de manière opportuniste et efficace.

Cette thèse se place dans le contexte de réseaux sans fil tactiques hétérogènes comportant des segments de radios cognitives. Ils peuvent être organisés en une multitude de sous-réseaux caractérisés par des technologies d'accès, des topologies ou des niveaux d'utilisation différents. La difficulté provient alors de la garantie de qualité de service de bout en bout : respect du débit négocié, du délai et de la gigue.

Nous nous sommes tout d'abord intéressés au contrôle d'admission dans ce type de réseaux. La littérature fait apparaître que les algorithmes généralement proposés reposent sur des méthodes d'accès de type aléatoire, qui ne correspondent pas aux techniques mises en œuvre dans ce type de réseaux, ou bien des solutions ne pouvant pas être implantées de manière distribuée. Nous proposons alors une méthode de calcul de bande passante résiduelle de bout en bout s'appuyant sur un algorithme de complexité polynomiale et pouvant être implanté de manière distribuée.

Nous nous sommes ensuite concentrés sur le routage en proposant une nouvelle métrique tenant compte des particularités de ce type de réseaux. Nous revisitons le calcul du nombre moyen de transmissions par paquet (ETX) en y intégrant les caractéristiques des utilisateurs primaires. En effet, ETX, de par sa facilité d'implantation, a été largement utilisée pour le routage dans les réseaux sans fil modernes. Cette métrique ne peut cependant pas être directement adaptée au contexte des radios cognitives.

Enfin, les besoins en termes de qualité de service peuvent s'exprimer au travers de multiples critères. Nous nous focalisons alors sur la thématique du routage à contraintes multiples. Le problème sous-jacent étant NP-complet, nous étudions les algorithmes d'approximation proposés dans la littérature. Il s'agit alors de choisir entre des solutions à faible complexité mais dont les performances ne sont pas garanties, ou bien des solutions possédant ces garanties mais dont la complexité est importante.

Souhaitant évaluer ce type d'algorithmes, nous avons implanté en environnement réel les propositions les moins coûteuses. En effet elles présentent de bonnes performances en pratique. La contrainte exhibée est alors la nécessité de mettre en œuvre du routage par la source.

Abstract

The unprecedented success of wireless telecommunication systems has resulted in the wireless spectrum becoming a scarce resource. At the same time, extensive measurements conducted in the early 2000's have shown that a significant part of the licensed spectrum, for instance that dedicated to TV broadcast services, is under-utilized. Cognitive Radio systems have been proposed as the enabling technology allowing unlicensed equipments, referred to as secondary users, to opportunistically access the licensed spectrum when not in use by the licensed users, referred to as primary users. Obviously, changing decades-old policies on spectrum access in the civilian and military domain will not happen overnight and many issues, technological and legal, will have to be ironed out first. However, the fundamental principle that we expect to underlie all solutions is that of access while doing no harm – secondary users should not interfere with primary users.

The focus of this thesis is on heterogeneous tactical networks deploying cognitive radios in parts or in their entirety. Such networks can be organized in multiple sub-networks, each characterized by a specific topology, medium access scheme and spectrum access policy. As a result, providing end-to-end Quality of Service guarantees in terms of bandwidth, delay and jitter, emerges as a key challenge.

We first address the admission control in multi-hop cognitive radio networks. We show that for this type of networks there is no algorithm capable of estimating the available end-to-end bandwidth in a distributed fashion. Therefore, we fill the gap by introducing a polynomial time algorithm that lands itself to a distributed implementation.

Then, we focus on routing and propose a new metric that takes into account the specifics of such networks. Using empirical data from a USRP testbed we show that ETX, the de facto standard metric for wireless networks, fails in the cognitive radio context. Therefore, we revisit ETX by considering the effects of primary user activity and the implications of the principle of access while doing no harm.

Finally, as quality of service requirements can be expressed using multiple metrics, we turn our attention to multi-constrained quality of service routing. With the underlying problem being NP-complete, we review the proposed heuristics and approximation algorithms. Our research reveals a trade-off between utilizing theoretically-proven but computably expensive algorithms and solutions that are fast but have poor worst-case bounds. To test the feasibility of solving multi-constrained routing in practice,

we implement on a real testbed low complexity algorithms that extend the Dijkstra's shortest path algorithm. We show that these algorithms can be incorporated in link state routing protocols, such as OSPF and OLSR. While these solutions suffer from poor worst-case performance bounds, in practice, they lead to satisfactory results when compared to exact but non tractable solutions.

Contents

Résumé v

Abstract vii

Contents ix

List of Figures xiii

List of Tables xv

List of Acronyms xvii

1 Introduction 1

1.1 Motivation . . . 2

1.2 Contributions and Organization . . . 4

2 Networking with Cognitive Radios 7

2.1 A brief overview of cognitive radio systems . . . 7

2.1.1 Dynamic spectrum sharing in TV white-spaces . . . 7

2.1.2 Cognitive Radio as the solution to the spectrum scarcity problem . . . 8

2.2 Deployment of cognitive radio systems . . . 10

2.2.1 Standardization processes . . . 10

2.2.2 Practical systems . . . 11

2.2.3 Future applications and propositions . . . 12

2.3 Challenges in cognitive radio ad hoc networks . . . 12

2.3.1 Spectrum Sensing . . . 13

2.3.2 Spectrum Analysis . . . 14

2.3.3 Spectrum Adaptation . . . 15

2.3.4 Spectrum Sharing . . . 15

2.3.5 Routing . . . 16

2.3.6 Transport protocol . . . 16

2.4 Strategies for estimating the end-to-end bandwidth . . . 17

2.4.1 Probing . . . 17
2.4.2 Using MAC layer information . . . 17
2.4.2.1 Random-access networks . . . 18
2.4.2.2 Time-slotted systems and the capacity region problem . . . 19
2.4.3 Concluding remarks . . . 22
2.5 Routing metrics for cognitive radio networks . . . 24
2.5.1 Single-path routing metrics . . . 24
2.5.2 Expected Transmission Count (ETX) . . . 24
2.5.3 High-throughput routing metrics in legacy networks . . . 25
2.5.3.1 Expected Transmission Time (ETT) . . . 25
2.5.3.2 Weighted Cumulative Expected Transmission Time (WCETT) . 25
2.5.3.3 Metric of Interference and Channel-switching (MIC) . . . 26
2.5.3.4 Available bandwidth . . . 26
2.5.4 ETX adaptation to the cognitive context . . . 27
2.5.5 Summary . . . 27

3 Admission control in cognitive radio ad hoc networks 29
3.1 Preliminaries . . . 29
3.1.1 Network model . . . 30
3.1.2 Channel access . . . 30
3.1.3 Model of interference . . . 31
3.1.3.1 Secondary-to-Secondary Interference . . . 32
3.1.3.2 Quantifying the Primary User Interference . . . 32
3.2 BRAND: An approach for estimating the bandwidth . . . 35
3.2.1 Problem Definition . . . 35
3.2.2 Fundamental principles of BRAND . . . 38
3.2.3 Computing the Average End-to-End Throughput with Random Scheduling 39
3.2.3.1 The Path's Slots Availability Table (PSAT) . . . 39
3.2.3.2 Fundamental principles of the method . . . 40
3.2.3.3 Interference-Clique sliding approach . . . 43
3.3 Distributed implementation of BRAND . . . 45
3.4 Performance evaluation of Brand . . . 48
3.4.1 Simulation Parameters . . . 48
3.4.2 BRAND parameters . . . 49
3.4.3 Basis for comparison . . . 49
3.4.4 End-to-End Bandwidth with BRAND . . . 49
3.4.5 End-to-End Throughput with Randomized Scheduling . . . 50
3.4.6 Admission Control Performance . . . 51
3.4.7 Cognitive Effect: Primary Users and Multi-Rate Links . . . 52
3.4.8 Distributed BRAND: Mobility, Channel Fading and Multiple Flows . . . 53

3.5 Conclusions . . . 55

4 Revisiting Transmission Count for cognitive radio networks 57
4.1 Questioning the basic concepts . . . 58
4.2 Primary user impact on transmission count . . . 58
4.2.1 Experimental setup . . . 59
4.2.2 Impact of primary users on transmission count . . . 59
4.2.3 Capturing the impact of primary users . . . 60
4.2.4 Would a straightforward solution work? . . . 61
4.2.5 Summary . . . 62
4.3 COExiST: Revisiting the Expected Transmission Count . . . 63
4.3.1 Model and preliminaries . . . 64
4.3.2 Computing COExiST using an absorbing discrete-time Markov chain . . . 65
4.3.3 Computing COExiST using a decomposition approach . . . 71
4.3.4 Practical application of COExiST . . . 75
4.4 Testbed implementation . . . 76
4.4.1 CSMA/CA implementation . . . 77
4.4.2 COExiST implementation . . . 77
4.5 Performance evaluation . . . 77
4.5.1 Experimental setup . . . 79
4.5.2 Accuracy of COExiST . . . 80
4.5.3 Sensitivity of COExiST to input errors . . . 81
4.5.4 Transmission Count Accuracy & Throughput . . . 81
4.6 Discussion . . . 83
4.7 Conclusions . . . 83

5 Greedy routing: a promising solution for Quality of Service routing 85
5.1 Preliminaries . . . 85
5.1.1 Routing algebra . . . 85
5.1.2 Multi-constrained QoS routing . . . 86
5.2 Motivation . . . 87
5.2.1 Strategies for finding feasible paths . . . 88
5.2.2 Strategies for approximating optimal MCOP paths . . . 88
5.2.2.1 K-Approx . . . 89
5.2.2.2 Greedy . . . 89
5.2.3 Application in practice . . . 90
5.3 Evaluation of Greedy for practical QoS routing . . . 91
5.3.1 OLSR and K-constrained routing . . . 91
5.3.2 Selecting Greedy for computing the MCOP paths . . . 92
5.3.3 Greedy requires source routing . . . 94

5.3.4 Summary . . . 94
5.4 Testbed implementation . . . 95
5.4.1 QoS metric measurements . . . 96
5.4.2 Multi-metric consideration . . . 97
5.4.3 Source routing implementation . . . 97
5.4.4 QoS flow differentiation & proof of concept . . . 98
5.4.5 Summary . . . 98
5.5 Perspectives for evaluating the performance of greedy strategies . . . 99
5.5.1 Conditions that impact the performance of Greedy . . . 99
5.5.2 Obtaining a simple approximation algorithm based on Greedy . . . 99
5.5.3 Insight for reducing the bound . . . 101
5.6 Conclusions . . . 102

6 Conclusions and Perspectives 103
6.1 Conclusions . . . 103
6.2 Perspectives . . . 104

A Complements for computing BRAND 107
A.1 General expression of the update equation . . . 107
A.2 Approximation in the update equation . . . 109
A.3 Integer program formulation . . . 110

Publications 113

International Conferences 113

International Conferences (under review) 113

International Journals (under revision) 113

Bibliography 115

List of Figures

1.1 QoS routing over heterogeneous networks 5
1.2 QoS routing in the overlay network 6

2.1 Cognitive Radio cycle 9
2.2 IEEE 802.22 cognitive radio network 11
2.3 Cognitive Radio ad hoc networks architecture 13
2.4 Probing modal distribution 18
2.5 WCETT fails in selecting the best path. 26

3.1 Cognitive Radio ad hoc networks model 31
3.2 TDMA frame structure 31
3.3 Illustration of the SU-to-SU interference 32
3.4 Possibility of Primary-to-Secondary interference due to sensing imperfection . . . 33
3.5 The Path's Slots Availability Table 39
3.6 3-link available slot set decomposition 41
3.7 Distributed implementation of BRAND 47
3.8 Distributed BRAND and the Clock Condition 47
3.9 BRAND vs Kodialam 50
3.10 BRAND accuracy 51
3.11 BRAND admission performance 52
3.12 Effect of Primary Users and Multi-rate links 53
3.13 BRAND is resilient to sensing errors. 54
3.14 Mobile scenario simulation 55
3.15 BRAND performance with mobility, channel fading and multiple concurrent flows 56

4.1 Experimental setup 59
4.2 Sampling of the primary user activity 60
4.3 Performance of ETX relative to PU/SU interference 61
4.4 Impact of primary users on T_r and T_t. 62
4.5 Impact of $\bar{T_{on}}$ on the packet delivery ratio 63

4.6 Impact of T_{on} on the transmission count 63
4.7 Modelling assumptions 64
4.8 Absorbing Discrete-Time Markov Chain for computing COExiST 66
4.9 Detailed transmission/retransmission scheme 71
4.10 Software architecture 76
4.11 Accuracy of COExiST. 78
4.12 COExiST captures the variation of $\bar{T}_{on}$ 79
4.13 COExiST vs ETX and SAMER 79
4.14 COExiST is close to the actual count time-average 80
4.15 Sensitivity to input errors 82
4.16 Better accuracy on transmission count leads to better throughput 83
4.17 COExiST correctly estimates that 1-2-4 is the best path. 84

5.1 In MCOP routing, subpaths of optimal paths are not necessarily optimal. 87
5.2 Greedy and K-Approx do not necessarily find the optimal MCOP path 93
5.3 Impact of weight distributions on the performance of Greedy 94
5.4 Impact of the number of constraints on the performance of Greedy 95
5.5 Greedy must be coupled with source routing. 95
5.6 Software architecture for QoS routing 96

A.1 Clique reduction approximation 110

List of Tables

2.1 Review of strategies for estimating the available bandwidth 23

5.1 Worst-case computational complexity of the most relevant multi-constrained approximation algorithms 89

List of Acronyms

5G Fifth Generation

ABE Available Bandwidth Estimation

AODV Ad hoc On demand Distance Vector

ARCEP Autorité de Régulation des Communications Électroniques et des Postes

ARPANET Advanced Research Projects Agency NETwork

BRAND available Bandwidth with RANDomized scheduling

BS Base Station

CCC Common Control Channel

COExiST COgnitive radio EXpected transmISsion Count

CPE Customer Premises Equipment

CR Cognitive Radio

CREW Cognitive Radio Experimentation World

CRAHN Cognitive Radio Ad Hoc Network

CSMA-CA Carrier Sense Multiple Access - Collision Avoidance

CTS Clear To Send

DCF Distributed Coordination Function

DCLC Delay Constrained Least Cost

DSR Dynamic Source Routing

ETT Expected Transmission Time

ETX Expected Transmission Count

FCC Federal Communications Commission

FPTAS Fully Polynomial Time Approximation Scheme

GDB Geolocation DataBase

GPS Global Positioning System

HF High Frequency

IEEE Institute of Electrical and Electronics Engineers

IETF Internet Engineering Task Force

IP Internet Protocol

IRU Interference-aware Resource Usage

ISM Industrial, Scientific and Medical

LIBRA Load and Interference Balanced Routing Algorithm

MAC Medium Access Control

MANET Mobile Ad Hoc NETwork

MCOP Multi-Constrained Optimal Path problem

MCP Multi-Constrained Path problem

MIC Metric of Interference and Channel switching

MPR MultiPoint Relay

NSFNET National Science Foundation NETwork

OFCOM OFfice of COMmunications

OFDMA Orthogonal Frequency-Division Multiple Access

OLSR Optimized Link State Routing

OSPF Open Shortest Path First

PHY PHYsical

PRR Packet Reception Ratio

PSAT Path Slot Availability Table

PU Primary User

QoS Quality of Service

RABE Retransmission-based Available Bandwidth Estimation

RF Radio Frequency

RSSI Received Signal Strength Indication

RTS Request To Send

RTT Round Trip Time

SAMCRA Self-Adaptive Multiple Constraints Routing Algorithm

SAMER Spectrum-Aware MEsh Routing

SDR Software-Defined Radio

STOD-RP Spectrum-Tree based On-Demand Routing Protocol

SU Secondary User

TDMA Time Division Multiple Access

TC Topology Control

TCP Transmission Control Protocol

TOS Type of Service

TV TeleVision

UDP User Datagram Protocol

UHF Ultra High Frequency

US United States of America

USRP Universal Software Radio Peripheral

VHF Very High Frequency

WCETT Weighted Cumulative Expected Transmission Time

WLAN Wireless Local Area Network

WRAN Wireless Regional Area Network

1 Introduction

The radio transmission systems were invented in the late 1800's following the Hertzian waves discovery. The area has seen tremendous progress over the last century, providing the society with novel telecommunication means, from wireless telegraphy to mobile telephony. Today, wireless telecommunication systems constitute a major economic sectors, in particular with the development of mobile applications. With the growing deployment of fourth generation systems and WiFi networks, people can now connect to the Internet on the go and run multimedia applications and business services on mobile devices such as laptops, tablets and smartphones.

Most architectures for commercial wireless communication systems are infrastructure-based, that is, the mobile devices connect to the rest of the network via an access point, which is often static and provides coverage to a limited geographical area. However, starting as early as the 1970's, a lot of effort has been put into the research and development of self-configuring, wireless mobile networks that do not rely on a fixed infrastructure. Instead, end-to-end communications are established in multi-hop fashion, with every node acting as both host and router. This kind of networks are commonly referred to as mobile ad hoc networks and find applications in areas such as environmental monitoring in regions where no communication infrastructure exist, disaster recovery situations where local infrastructures have collapsed, community mesh networks [1], etc.

In addition to the civilian applications, mobile ad hoc networks are particularly important to military operations. They provide the troops with mobile and self-deployable communication means that can be used at various theaters, including marine, ground and airborne deployments. Nevertheless, despite the simplicity of deployment of mobile ad hoc networks, human intervention is still required for configuring the radios according to a given frequency plan – adding cost and complexity to the process of deployment at a given military theater. Ideally, the transmission/reception frequencies should be assigned automatically and without human intervention based on spectrum utilization policies and in compliance with well defined hierarchies between different users in the modern battlefield. This vision requires more "intelligent" communication radios, or as they are commonly referred to, cognitive radios, capable of adapting their communication parameters based on their surroundings.

The emergence of cognitive radios has generated a great deal of interest in the consumer market as well, albeit for slightly different reasons. The wireless spectrum for the consumer market is treated as a scarce commodity that has been tightly regulated by centralized governmental

authorities, such as the FCC in the United States. Operating on wireless spectrum requires either applying for a license by the respective governmental authority or using select frequencies, like those in the ISM band, that are freely available to everyone. The latter approach is followed by many popular wireless technologies, such as IEEE 802.11, Bluetooth, sensor networks, etc, which has led to the ISM frequencies often being overcrowded.. At the same time, extensive measurements in the TV broadcast services bands have shown that a significant part of the licensed spectrum remains underutilized [2]. As a result, the FCC issued a historic ruling permitting unlicensed devices to use unutilized licensed spectrum [3, 4]. This ruling, coupled with the development of the cognitive radio concept [5], has ignited a lot of interest in the research and development of cognitive radio networks capable of exploiting the best spectrum available [2]. The central governments around the world are also starting to get involved with the US President's council of advisors on science and technology in its 2012 report [6] recommending that, when it comes to the wireless spectrum, "the norm should be sharing, not exclusivity", and estimating that a new architecture and a corresponding shift in practices could multiply the effective capacity of the spectrum by a factor of 1,000.

1.1 Motivation

Mobile ad hoc networks have been subject of numerous research studies and standardization efforts with varying degrees of success. However, the cognitive radio represents a paradigm shift in how mobile nodes access and utilize the wireless spectrum and thus necessitates revisiting and/or redesigning many of the previously proposed solutions. This is reflected in many recent research works in the subject. Nevertheless, we are still far from having all the answers. Particular questions that have not yet been addressed are that of end-to-end QoS provisioning, and specially admission control, and QoS routing in multi-hop software-defined/cognitive radio networks. These questions play a significant role in deployed networks on the battlefield, such as the Tactical Internet [7, 8] that interconnects wired/wireless networks of different technologies.

The Tactical Internet, illustrated in Figure 1.1, integrates both military and commercial systems. It can be deployed for multiple networking scenarii such as for tactical operations or disaster recovery situations. The heterogeneous network components are interconnected through intelligent gateways equipped with different front-ends and capable of collecting QoS metrics on the different media they are connected to. In particular, each gateway is assumed to collect metrics relative to a path connecting it to a neighboring gateway. In such an heterogeneous network, the end-to-end flows have to be routed according to agreed upon QoS requirements. These requirements are specified in terms of end-to-end metrics such as the achievable throughput, the packet delivery delay or simply the cost of the communication. For instance, a critical data flow relative to a tactical situational/positional awareness application should be transmitted through a highly-secured VHF network, while less critical applications would rather select high-throughput media. An example of such routing decisions is available in Figure 1.1.

However, the definition of an end-to-end routing protocol is missing [7, 9]. With the variety of flows presenting different QoS requirements, and every transit network having specific delivery characteristics, the key challenge motivating our work consists of providing an end-to-end QoS routing solution. Formally speaking, it consists of finding the sequence of gateways a flow should go through to satisfy its associated end-to-end QoS requirements.

The path connecting two successive gateways is not necessarily optimized and can be modeled as a virtual link/tunnel characterized by a set of quality of service routing metrics (bandwidth, delay, jitter, cost). Hereafter, the objective is equivalent to performing QoS routing in the overlay network [10] illustrated in Figure 1.2. The desired routing solution has to be scalable, consistent, land itself to a distributed implementation and generate little overhead. As the paths have to be computed according to multiple metrics, a good candidate for realizing this promise would be the recently proposed OLSRv2 protocol [11] augmented with the multi-metric capability [12]. However, the mechanisms proposed for standardization perform routing according to a single metric only. That is, the multi-metric aspect only provides one routing per metric, but does not simultaneously take them into account. A multi-constrained QoS routing approach could be envisioned at the overlay level. Computing the best multi-constrained paths, however, is a NP-hard problem, which has been widely studied in the literature [13, 14, 15, 16, 17]. Unfortunately, these studies only deal with the algorithmic efficiency of the proposed heuristics/approximation algorithms. They do not provide a thorough analysis of the challenges arising when implementing these solutions in practice. Only seletive works have been concerned with the implementation aspects [18]. However, they reveal the importance of accurately estimating the underlying QoS metrics and show that higher performance is achieved when routing is performed as a function of the available bandwidth.

The design of a QoS routing protocol in the Tactical Internet is thus subject to the capability of measuring the available end-to-end bandwidth in every interconnected network. A significant challenge is that the available bandwidth calculation problem has not been fully solved. Indeed, while the traditional probing techniques [19] work well in the wired domain, they perform poorly for the case of wireless networks. The problem has been well addressed for the case of 802.11 networks [20, 21, 22, 23, 24, 25] where the estimation methods are based on MAC layer information. However, while the medium access method might be of more deterministic nature in tactical and software-defined/cognitive radio networks, the question here remains open.

Provided with an accurate estimation of the available end-to-end bandwidth in a given MANET, the operator can try to additionally optimize the intra-networks routing. However, the straightforward approach, consisting of adapting the classical path calculation algorithms and replacing the path length metric with the bandwidth estimation is not applicable. Due to the high complexity of the possible algorithmic solutions, the recommended approach consists of using an alternative additive routing metric that characterizes high-throughput paths. The baseline metric for this purpose is ETX: the Expected Transmission Count [26, 27]. This metric aims at determining the expected number of transmission attempts necessary to successfully

transmit a packet over a given link. Unfortunately, measurements on a practical cognitive radio network have demonstrated that ETX fails to accurately capture the actual transmission count in the cognitive radio context.

Finally, routing mechanisms based on the available bandwidth metric or multiple additive constraints suffer from the same drawbacks. In particular, they share the same problems when the routing is done greedily [28], that is, as a simple extension of the Dijkstra's shortest path algorithm: 1) the greedy approach is not optimal, 2) hop-by-hop routing is not applicable. The reason why greedy routing is not optimal and cannot be run on a hop-by-hop fashion is the lack of the sub-optimality property: "sub-paths of an optimal path are optimal". To overcome this challenge that puts into question the applicability of most modern routing protocols, researchers have focused on the design of approximation algorithms. While the question of near-optimal algorithms is well-addressed, provisioning the greedy routing approach with performance guarantees remains largely open. In spite of relatively good practical performance of this easy-to-implement solution, the available theoretical bounds are not sufficiently tight for adopting such a solution for future deployments.

1.2 Contributions and Organization

The remainder of the dissertation is organized as follows.

Chapter 2 In this chapter we review the state-of-the-art on networking with cognitive radios. We first give an overview of cognitive radio systems by explaining the reasons why they respond to the the spectrum scarcity problem. Then, we present their main technical capabilities, the associated standardization efforts as well as their deployments in current systems. Most proposed solutions are communications systems using an infrastructure-based architecture. In this thesis we consider distributed ad hoc architectures. Therefore, we present the challenges arising in cognitive radio ad hoc networks and relate them to the admission control and quality of service routing problems. As a consequence we review the strategies for estimating the available end-to-end bandwidth as well as the metrics used for cognitive routing.

Chapter 3 In this chapter we present BRAND, a polynomial-time algorithm for computing the end-to-end bandwidth in cognitive radio ad hoc networks. The proposed scheme can be implemented in a distributed fashion and returns an online computation of the available bandwidth, an essential requirement for doing admission control and estimating link metrics for QoS routing in the overlay network. The algorithm is based on a careful analysis of the intra-path interference as well as that caused by primary user activity.

Chapter 4 In this chapter we revisit the expected transmission count calculation problem for cognitive radio networks. We present COExiST, a metric that accurately computes the

transmission count by taking into account the characteristics of the primary users. Using a USRP testbed, we show that ETX, the *de facto* metric used in legacy networks, fails in estimating the actual transmission count in cognitive radio networks. We mathematically derive the expression of COExiST through two different methods. As the metric is additive, we show that it can be coupled with OLSR to identify high-throughput paths which state-of-the art cognitive routing approaches fail to capture.

Chapter 5 This chapter is dedicated to the study of the applicability of greedy routing algorithms in the context of multi-constrained quality of service routing. With the underlying multi-constrained path finding problem being NP-complete, we review the approximation algorithms proposed in literature. We show that the greedy solutions, for which routing efficiency is not well established, achieve satisfactory results in practice, either in terms of constraint validation or execution time. Thus, we implemented one of the greedy solutions on a USRP testbed by integrating it into the OLSR routing protocol. We show that, for keeping routing efficiency, packets belonging to the same QoS flow have to be source routed. We then provide perspectives for finding better performance bounds for such routing strategies.

Appendix A: In this appendix we provide additional material that complements Chapter 3. In particular, we explain in detail how we obtained the approximation scheme permitting to reduce the BRAND complexity. Furthermore, we provide the Integer Program formulation we utilized for evaluating the performance of the BRAND slot allocation process.

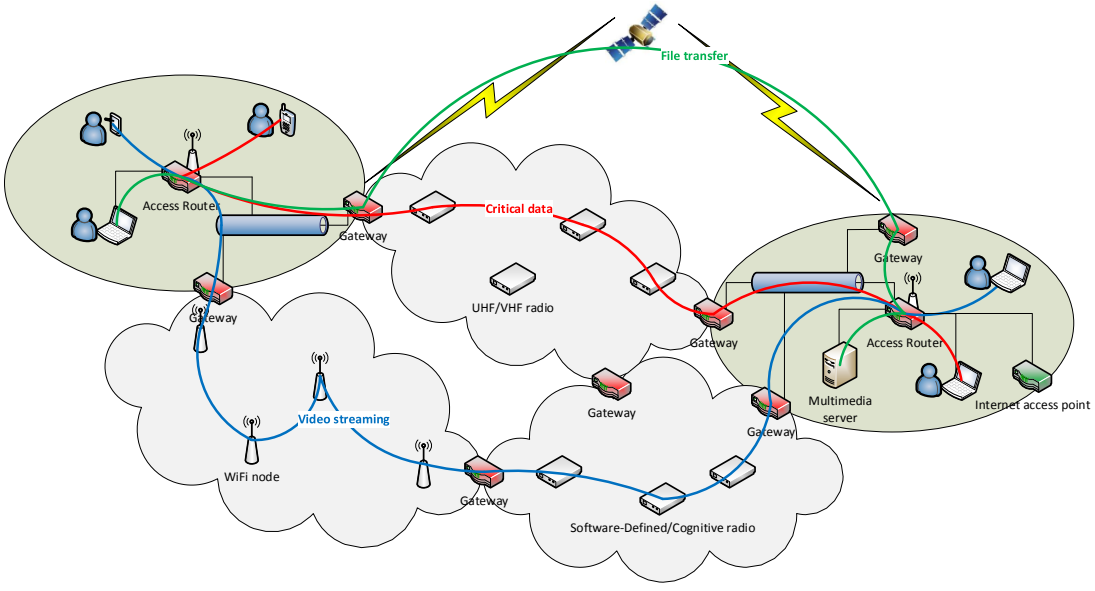


Figure 1.1 – QoS routing over heterogeneous networks

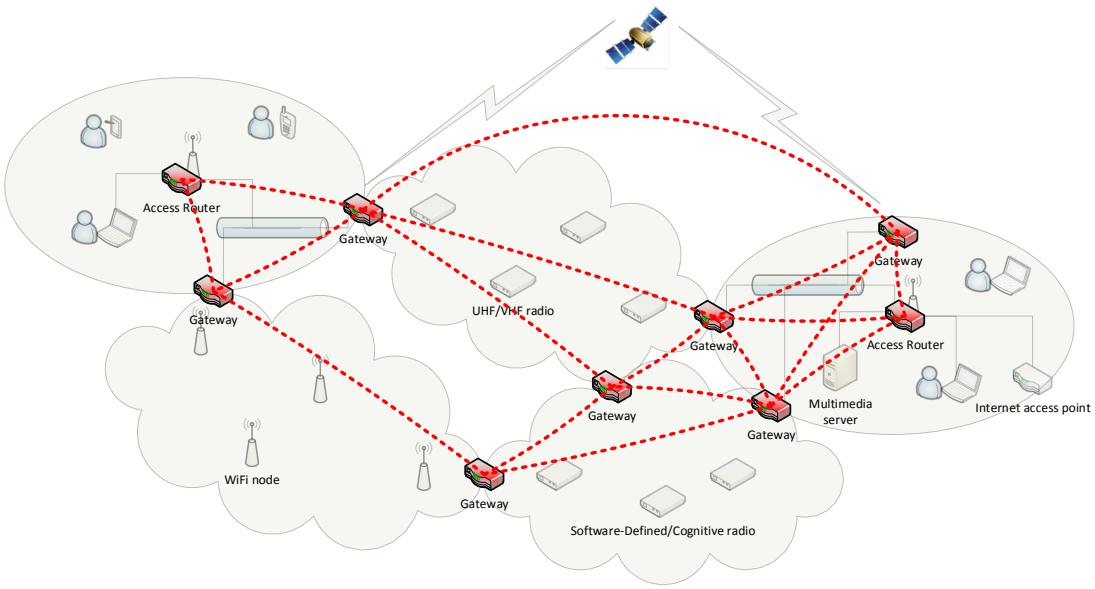


Figure 1.2 – QoS routing in the overlay network

2 Networking with Cognitive Radios

This chapter aims at clarifying the context of this thesis by giving an overview of cognitive radio network systems. In detail, we first introduce such networks and their relation to TV whitespaces in Section 2.1. Then, we review the main technical challenges encountered in the design of cognitive radio ad hoc networks in Section 2.3. Because the development of cognitive radio networks is at its nascent stage, techniques for estimating the available bandwidth are inherited from traditional multi-channel multi-hop network technologies. Therefore, we dedicate Section 2.4 to a review of research advances in estimating the available bandwidth in such legacy wireless multi-hop networks. A review of routing metrics for cognitive radio networks is then available in Section 2.5.

2.1 A brief overview of cognitive radio systems

2.1.1 Dynamic spectrum sharing in TV white-spaces

The wireless spectrum has been tightly regulated by centralized authorities, such as the FCC in the United States, OFCOM in the United Kingdom and ARCEP in France. Therefore, operating on wireless spectrum requires either applying for a license by the respective governmental authority or, requires using select frequencies, like those in the ISM band, that are freely available to everyone [29]. The latter approach is followed by many popular wireless technologies, such as IEEE 802.11, Bluetooth, sensor networks which has led to the ISM frequencies often being overcrowded.

At the same time, extensive measurements conducted in the early 2000's have shown that a significant part of the licensed spectrum is under-utilized [2], and revealed an inefficient usage of the spectrum resource. Among those, the under-utilized regions of VHF/UHF bands dedicated to TV broadcast services (470-800MHz in the US) have been referred to as TV white-spaces. The level of utilization of such TV white spaces is not constant and depends on the location as well as the observation time.

Initiated in 2003, the ruling permitting an opportunistic access on the TV white-spaces has been finally issued in 2008 [2] and the relative frequencies opened in 2010. This ruling is historic as it permits unlicensed equipments, referred to as secondary users, to opportunistically access

the TV broadcast bands let vacant. More importantly, it constitutes a solution to the spectrum scarcity problem and will enable the provision of wireless connectivity to currently undeserved populations. Indeed, in addition to delivering extra 200MHz spectral bandwidth resources in average, the targeted UHF bands present superior line-of-sight propagation characteristics than in the GHz band. Therefore, they permit a wider coverage at fixed transmitting power. Coupled with some equalization techniques, the transmitted signals can resist to the degrading effects of hills, forest. They can thus provide a larger and cheaper wireless connectivity in rural regions that are not provided with broadband access, where traditional technologies such as WiMax perform poorly and/or are expensive to deploy. Hereafter, taking advantage of these TV white spaces opens a new door for novel business applications.

However, for ensuring that the TV broadcast service will not suffer from such emergent technologies, the relative opportunistic access must guarantee that no harmful interference will be caused to the licensed equipments, also referred to as primary users, that can either transmit or receive the service. In conclusion, the unlicensed TV white space usage must be realized through a dynamic spectrum sharing technology, such as cognitive radio, that can reconfigure its transmission scheme on-demand.

2.1.2 Cognitive Radio as the solution to the spectrum scarcity problem

The cognitive radio technology has been identified as the key enabling solution for exploiting such TV white-spaces in particular, and addressing the spectrum scarcity problem in general. Introduced by Joseph Mitola and defined by the FCC as a "radio that can change its transmitter parameter based on the interaction with the environment in which it operates", such a technology is characterized by the two following features [2, 30, 31]:

- **Cognitive capability:** The radio can perceive the characteristics of its surrounding radio environment. In particular, it can determine the level of utilization of a given spectrum band at specific time and location. This feature can be achieved by whether performing spectrum sensing, or querying an appropriate database through the Internet.
- **Reconfigurability:** Once an available channel is detected, the radio can adapt its transmitter/receiver parameters to exploit this channel. This capability is typically provided by software-defined radios that can reprogram the transmitter/receiver frequencies at the software level.

The global functioning of a cognitive radio can thus be summarized in the cognitive radio cycle illustrated in Figure 2.1. Such a cycle is composed of four main inter-operating capabilities. The way these four features inter-operate is detailed in Section 2.3.

- **Spectrum Sensing:** The radio must detect the presence of primary user communications in the exploited licensed bands. To this end, when such an information is not provided by a suitable database, it must perform real-time wideband spectrum sensing.
- **Spectrum Analysis:** This module is responsible for determining the characteristics of the sensed channels and the specificities of the incumbent primary users. Its must provide

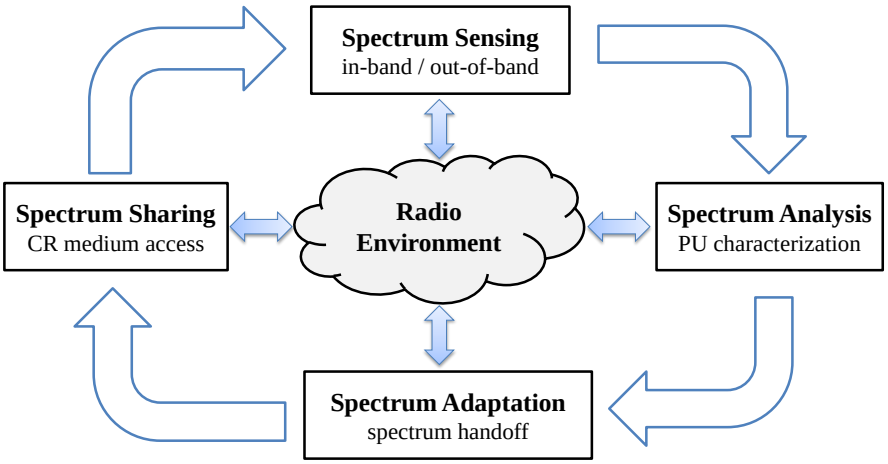


Figure 2.1 – Cognitive Radio cycle

the radio with sufficient information for evaluating a set of primary user channels, also referred to as primary channels in the remaining of this thesis.

- **Spectrum Adaptation:** Also called *Spectrum Mobility*, it provides the radio with the ability to switch its transmission/reception schemes from one primary channel to another. This typically happens when a primary user shows up in the currently used frequency channel. It is a key point of dynamic spectrum access.
- **Spectrum Sharing:** Once a primary channel is selected by the radio, it needs to coordinate with other cognitive radios so that they do not interfere with each others. For this purpose, it requires the design of specific medium access control schemes.

In general, Spectrum sensing is part of a more global spectrum monitoring operation and is complemented with the information provided by an approved database. For complying with the FCC requirements, cognitive radio systems have to minimize the amount of interference caused to the primary users. At the same time, they must exploit the best available spectrum to achieve a certain QoS level. A requirement for a TV white-space broadband technology to make it on the market is for instance to provide high-speed Internet connectivity. This requires thus new physical, link layer and routing developments for achieving the intended dynamic spectrum sharing and QoS expectations. For this purpose, standardization frameworks have been defined in the early 2010's. They are overviewed in the following section.

2.2 Deployment of cognitive radio systems

2.2.1 Standardization processes

The architecture of cognitive radio systems is not unique and has already been the subject of two IEEE standards. Both stantards present centralized architectures for WRAN and WLAN networks.

IEEE 802.22[32]: The IEEE 802.22 standard has been released in July 2011. It envisages a centralized infrastructure-based architecture wherein user stations, referred to as customer-premises equipments (CPE), are attached to a fixed base station. It aims at enabling the deployment of cognitive Wireless Regional Area Network (WRAN), as illustrated in Figure 2.2. These networks will operate on the VHF/UHF TV white-spaces to provide an alternative broadband access in sparsely populated regions. The standardization work particularly aims at providing high-speed Internet access and enabling the deployment of services with different QoS levels. The intended geographical coverage must reach 30 km. However, with adapted PHY and MAC mechanisms, it is expected to extend from 30 km to 100 km.

In IEEE 802.22 WRAN, the base station is responsible for the spectrum sharing and adaptation. The spectrum decision is based on the combination of two spectrum monitoring schemes. The first one consists in the base station querying a geolocation database. It contacts the database with its GPS coordinates to obtain a list of available primary channels as well as the maximum transmission power allowed in these channels ([32], p. 137). This database is administrated by the local regulatory entities and updated every day by the FCC. It can at least indicate a first overview of the available channels in the region. It is then coupled with local spectrum sensing performed at each station. The presence of the incumbent primary users is therefore reported periodically to the base station, that can decide to switch to another unoccupied channel. Indeed, with the protection of the incumbents being a priority, this *in-band* sensing procedure is part of the strict recommendations of the 802.22 standard [33].

Then, the standard recommends that the available channel discovery does not last more than few milliseconds. It also defines the main components of the physical and link layers. At the PHY, an adaptive OFDMA modulation scheme has been selected. Then, a point-to-multipoint medium access control approach has been envisioned at the MAC. It is connection-oriented and presents different mechanisms for downstream/upstream medium access. Downstream and upstream transmissions share the same channel and are scheduled at different time. For the downstream access, it uses a time-division multiplexing, while for upstream transmissions, the bandwidth resources are provided to CPE on a on-demand basis. CPE access requests are transmitted to the base station via a polling policy. The access is therefore based on a synchronized scheme with time division multiple access.

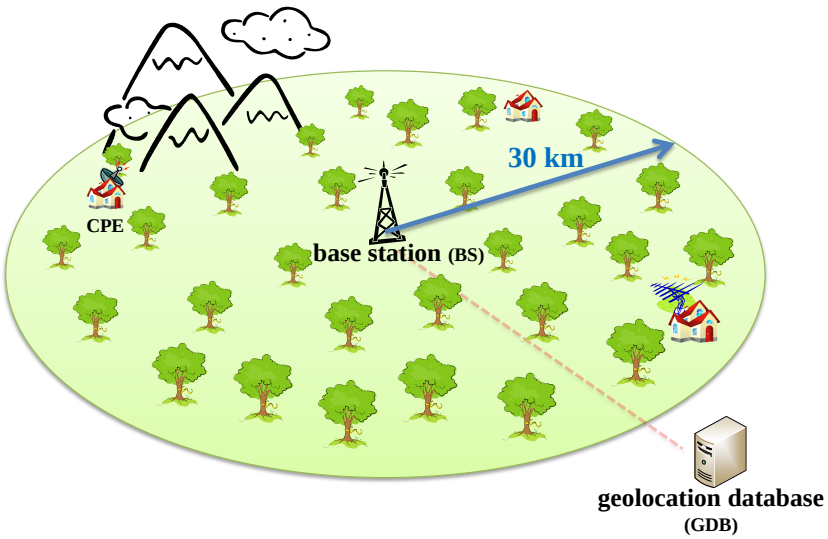


Figure 2.2 – IEEE 802.22 cognitive radio network

IEEE 802.11af [34]: The 802.11 standards family has also been extended with the recent issue of the cognitive 802.11af standard, in February 2014. For guaranteeing the protection of the primary users, the approach is similar to the 802.22 standard and uses a database strategy. This is the main feature that distinguishes 802.11af from the rest of the 802.11 family. Then, the available spectrum is shared among the secondary users using the traditional 802.11 random access scheme. Note that alternative adaptations of the WiFi, that have not been subjected to standardization processes, are available in the literature. One of them is the WhiteFi system [4] that also proposes opportunistic access on the UHF TV bands, but does not query an FCC database.

2.2.2 Practical systems

With the FCC opening the TV white-spaces frequencies in 2010 and the standardization efforts provided for sharing the relative spectrum, many cognitive radio products have been engineered over the recent few years. For complying with the FCC requirements and having a chance to succeed on the market place, they must integrate a database-driven channel discovery scheme. Then, they can innovate on other cognitive radio technical challenges. Among these products, we can take as an example the RuralConnect system from the US Carlson Wireless Technologies company. The RuralConnet system[35] is one of the first TV white-space WRAN system that have been approved by the FCC for unlicensed commercial use in the UHF TV channels. It has been certified by the FCC on January, 15^{th} 2014 after two years of intensive trials worldwide. RuralConnect utilizes the cognitive radio technology coupled with spectrum shar-

ing databases. While worldwide companies have already targeted the database market (Google, Nominet, LS Telcom, iconectiv, Key Bridge, Fairspectrum and Spectrum Bridge) RuralConnect, located in the United States of America, has been authorized by the FCC to use the Spectrum Bridge database. It promises performing applications with the provision of high-speed broadband access in undeserved rural regions. Other big companies such as Texas Instruments, that commissioned the *WhiteSpace Alliance* for providing broadband wireless and machine-to-machine solutions [36], have also been involved in the TV white-spaces market.

2.2.3 Future applications and propositions

The US example has initiated a growing interest in exploiting the TV white-spaces in Europe. The OFCOM launched in 2013 an important campaign for studying the applicability of dynamic spectrum sharing strategies in Great Britain, making the United-Kingdom being the European leader in this domain. In 2014, the French government has ordered an equivalent study through the so-called *Mission Spectre* [37] conducted by Professor Joëlle Toledano and commissioned by Fleur Pellerin, Minister Delegate with responsibility for Small and Medium-sized Enterprises, Innovation and the Digital Economy. This study concludes on the great benefit that would offer dynamic spectrum sharing in TV white-spaces for the future 5G networks and the Internet of Things. A set of research and organizational recommendations has been established to promote the development of cognitive radio network technologies over the next few years, following the actions led in United-Kingdom and the United States of America.

From a general point of view, the establishment of cognitive radio networks is still at its nascent stage and we are far from having seen all their applications. Cognitive radio systems cannot be limited to the 802 standards and TV white-spaces systems, presented so far, that are infrastructure-based. They can be organized as well in distributed ad hoc networks. For these typical systems, the first implemented platforms are mainly part of cognitive radio research/industrial testbeds [4]. A European project, called CREW [38], has been started in October 2010 for this special purpose. With such distributed architectures, the potential of cognitive radio networks could be dramatically increased [31]. The resulting networks, using relaying nodes, could enlarge the wireless connectivity coverage. At the same time, they could reduce the transmission power and mitigate the interference to the primary users. However, for these promising networks to be established, novel PHY, MAC and routing mechanisms need to be ironed out. The associated technical challenges are presented in the following section.

2.3 Challenges in cognitive radio ad hoc networks

Designed on top of traditional software-defined radio technologies, cognitive radio ad hoc networks should inherit from a common core architecture at the physical, link, and routing layers. However, such architectures must be augmented with capabilities that are specific to the cognitive context. In particular, a main requirement for any cognitive radio architecture is to

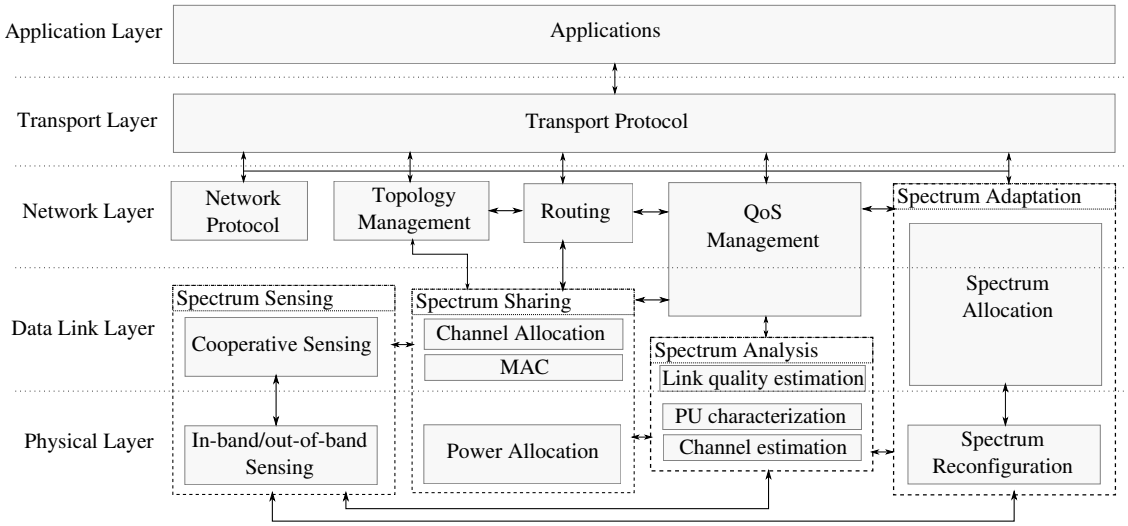


Figure 2.3 – Cognitive Radio ad hoc networks architecture

protect the primary users from interference [3, 39], and then exploit the best available spectrum. The architecture of cognitive radio ad hoc networks is provided in Figure 2.3. As depicted in Figure 2.1, such a cross-layer architecture is composed of the main components composing the cognitive cycle. It is then complemented with a QoS management block as well as a specific transport protocol.

2.3.1 Spectrum Sensing

One of the major tasks at the physical layer is to perform efficient spectrum sensing. Such a scheme is segmented in three categories, as detailed in the following:

- **In-band sensing:** In-band sensing consists of monitoring the presence of primary user communications on the channel currently used for opportunistic secondary user transmissions [33]. The primary user detection can be achieved through three different schemes: match filter detection, feature detection and energy detection [31, 40]. The match filter technique requires the primary and secondary networks to be synchronized with each others as well as prior knowledge of the primary signal pattern. Differently, the feature detection analyzes the spectral characteristics of the primary user signals. However, such an analysis is computationally expensive. For these reasons, energy detection, that simply consists in measuring the primary user signal power (RSSI), has been adopted as the reference sensing technique [41].
- **Out-of-band sensing:** Cognitive radio ad hoc networks are not provided with a centralized entity that query a database for discovering the available primary channels. Each node must carry out all the operations defined in the architecture depicted in Figure 2.3.

Among those, out-of-band sensing is dedicated to the discovery of alternative unoccupied primary channels [42]. Such a sensing procedure should be executed as fast as possible to minimize the available channel discovery latency. However, one main technical problem is to determine the set of channels to be sensed as well as the associated sensing durations and order [40].

- **Cooperative sensing:** For improving the detection accuracy, as well as the associated detection and false alarm probabilities, neighboring secondary nodes can cooperate and exchange local sensing information. This can serve not only for improving the performance of the secondary network but also reducing the level of interference to the primary users. Indeed, while the detection of primary transmitters performs well, it is likely that a cognitive radio does not perceive a neighboring primary receiver [31]. This phenomenon, amplified with channel fading and noise uncertainty [43], is referred to as the *receivers uncertainty* problem and can be addressed by state-of-the-art cooperative schemes [40].

As part of the FCC priority recommendations, spectrum sensing has received a lot of attention by the research community. The sensing performance actually depends on many parameters such as the spectrum access scheme. In this thesis, we do not propose a new spectrum sensing scheme and rather consider the sensing as a black box that is provided with relative detection and false alarm probabilities.

2.3.2 Spectrum Analysis

In ad hoc networks, a given channel is likely to present varying characteristics at different locations. It is thus essential to define and quantify some performance metrics that are link-specific, such as the packet transmission delay. Another feature consists of collecting statistics on the primary users activity. Indeed, when the primary and secondary networks are not synchronized, it is likely that a primary user shows up once cognitive radio communications have been established. A stochastic representation of the primary user activity can thus be leveraged for the design of efficient medium access schemes, as well as the parameterization of the sensing [44]. It can also provide the necessary information for switching from one primary channel to another.

Most of the research in this domain have modelled primary user activity using an alternative ON/OFF process with exponentially distributed ON/OFF period durations [40, 42, 44]. However, in practice, such period durations are not exponentially distributed and depend on the primary network characteristics, as well as the sensing time scale [45]. Nevertheless, as pointed out in the recent literature [45], the exponential distribution suitably fits some empirical distributions.

In this thesis, we do not intend to design a new primary user activity model. Instead, we leverage the exponential model for the design of new cognitive link/path quality of service metrics.

2.3.3 Spectrum Adaptation

Once the radio is aware of the available primary channels and their characteristics, it can adapt it transmission/reception scheme by allocating new channels on the available spectrum bands. Note that the terms *channel* and *band* are often used interchangeably. However, in the literature, a channel rather refers to a sub-portion of a given spectrum band. The channel assignment operation is part of the Spectrum Sharing block. Spectrum adaptation is often performed in conjunction with the routing and the MAC [46, 47]. It may follow some specific optimization strategies, whether the objective should be primary user protection, end-to-end delay minimization or throughput optimization. This is mainly due to the fact that a given channel can perform differently from one hop to another and affect the incumbent users in a varying way.

In this thesis we focus on capturing link/path quality of service metrics given an arbitrary spectrum/channel allocation scheme.

2.3.4 Spectrum Sharing

The key novel challenge when designing a channel access protocol with cognitive radios is maximizing the realized capacity of the secondary users without adversely affecting the primary user transmissions [42, 44, 48]. For this purpose, power allocation can be of great interest and serve to reduce the amount of interference in the secondary (as well as primary) network [46, 49]. Another decisive response to this challenge is the design of new MAC protocols for cognitive radio networks [50, 51]. For all the diversity in the proposed solutions, one thing underlying all protocols is the need for a sensing module whose responsibility is identifying when the cognitive radio may be interfering with a primary user. In its basic form, this module relies on physically sensing the channel periodically [4] to look for primary user activity. When possible, the physical sensing can be complemented by a database of well known primary users [52]. But, given the functionality of the sensing module, a MAC protocol for cognitive radio networks needs to provide periods of network silence to be dedicated to sensing for primary user activity. This means that, at given time intervals, all cognitive radios in the network will stop from generating any traffic, and instead, focus on sensing. Such requirement can be easily accommodated by a TDMA protocol with sensing being performed at the beginning of each frame/time-slot. Indeed, a majority of the MAC protocols proposed for cognitive radio networks [32, 53, 54, 55, 56], including the IEEE 802.22 MAC [32], are based on TDMA. Therefore, sensing efficiency is related to spectrum access [40, 44, 48].

Nevertheless, some solutions based on random access have also been proposed [57, 58, 59, 60]. Such MAC protocols assume no time synchronization among the secondary users. The sensing is performed on a dedicated common control channel (CCC). Such an approach can typically apply in cognitive radio technologies built on top of the IEEE 802.11af standard [34].

This thesis does not intend to propose a new MAC for cognitive radio networks. However, as further explained in Section 2.4, the MAC plays a significant role in estimating the available bandwidth. While there is no clear winner yet among the MAC protocols proposed so far, it is likely that a deterministic medium access protocol, typically used in tactical networks, will better serve an architecture where multiple technologies share the same spectrum and synchronization is required for the sensing.

2.3.5 Routing

For complying with the FCC primary protection constraint, a secondary node is required to adapt its path computations according to the primary users activity. To this end, it can either route around the primary users, thus potentially increasing the path length, or, switch its transmission channel on the affected links [61]. Obviously, both strategies will increase the end-to-end delay, which is aimed at being minimal. In [62], for instance, a geographic routing solution is proposed. It selects next hops and operating channels so as to avoid regions of primary users activity, while minimizing the end-to-end path latency. An optimal routing metric focusing on delay is proposed in [63]. The authors analytically demonstrate its optimality and accuracy for the cases of mobile and static networks. Then, several works have focused on optimizing other criteria than delay. For example, joint route selection and spectrum decision is addressed in [46, 63]. While many solutions make use of a common control channel (CCC), an adaptation of the AODV protocol free of a CCC is proposed in [64]. In [65] an opportunistic scheduling, based on congestion control, that maximizes the overall capacity of secondary users while satisfying a constraint on time average collision rate at the primary users is proposed. Finally, [47] has proposed a distributed algorithm for jointly optimizing routing, scheduling, spectrum allocation and transmit power. The optimization objective is utility-based. The defined utility function couples dynamic congestion control schemes with spectrum availability information.

While the works presented so far are shown to handle well the primary users, none of them addresses the problem of admission control by estimating the end-to-end bandwidth of a given path.

2.3.6 Transport protocol

Transport layer protocols such as TCP have been historically designed assuming that packet losses are mainly due to router congestion rather than transmission errors. However, such a statement does not hold for ad hoc networks where the wireless channel might be of poor and varying quality. Moreover, cognitive radio networks are likely to present a variety of link capacities within the same network. The link quality can vary with sudden establishment of licensed users communications. Then, both spectrum sensing and channel switching delays may affect the TCP segment transmission delay. Indeed, while shorter sensing time can increase the TCP throughput by reducing the RTT, miss-detecting an interfering primary user can dramatically

degrade the performance of the secondary user connections. Cognitive radio networks require thus the development of new cross-layer transport protocols [66].

This thesis does not provide a new transport layer scheme. Instead, it concentrates on challenging issues at the MAC and network layers.

2.4 Strategies for estimating the end-to-end bandwidth

As pointed out in Section 2.3.5, no cognitive-specific mechanism has been proposed yet in the literature for estimating the available end-to-end bandwidth of a given path. We therefore review the techniques developed so far for non-cognitive (legacy) single/multi-radio ad hoc networks and discuss their applicability to the cognitive radio context.

2.4.1 Probing

The available end-to-end bandwidth can be measured by sending probing packets along a path and analizing the repartition of the inter-probe reception time at the destination. The relative techniques, referred to as probing, have been widely used for estimating the bandwidth in multi-hop wired networks. In particular, they have served for building overlay networks in the Internet [10]. Among them, packet-pairs and packet-trains techniques are presented in [19]. These techniques aim at identifying the available capacity of the path bottleneck link. They have led to the design of some tools such as *cprobe* [67] and *pathrate* [19] that have been widely used for estimating the capacity of hundreds of paths on the Internet. However, while these techniques can be deployed in a very simple way, the resulting estimates are often inaccurate. As illustrated in Figure 2.4, borrowed from [19], queueing effects in the wired networks make the bandwidth measurements following a complex modal distribution. One technical challenge consists in identifying the right mode for doing efficient admission control.

Such a phenomenon is amplified in wireless environments with varying link characteristics and interference [68]. Indeed, the problem is inherently different for the case of wireless multi-hop networks as the capacity of a path does not equal the one of its bottleneck link [21, 69]. Overall, the available bandwidth in wireless network is a scarce resource that will be affected by the probes themselves. Therefore, a challenging issue over the last past decade has been related to the development of adapted non-intrusive techniques for doing admission control and estimating the bandwidth in mobile ad hoc networks.

2.4.2 Using MAC layer information

In this section we review the techniques, developed so far in mobile ad hoc networks, that use MAC layer information for estimating the available bandwidth. While these methods do not send probes across the network, the nodes can exchange a limited amount of controlling

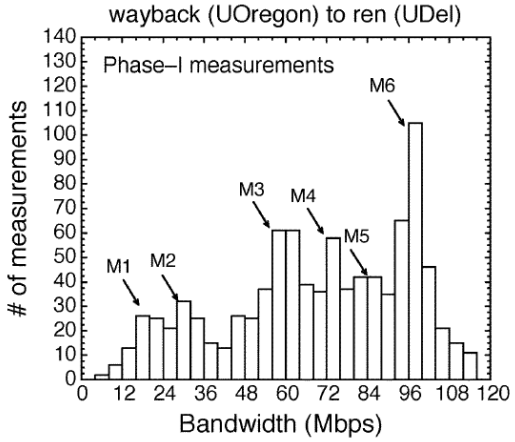


Figure 2.4 – Probing modal distribution

information for computing the bandwidth. In the following, we classify the designed solutions according to the related medium access control scheme.

2.4.2.1 Random-access networks

Many research efforts have been put on the performance evaluation of random access networks based on the 802.11 DCF protocol. For estimating the available end-to-end bandwidth, these works have considered the case of the RTS/CTS mechanism being disabled. The different propositions described in the literature consider the three following quantities: the fraction of time each node can occupy the medium, the fraction of time the medium is busy and the fraction of time spent doing back-off. They also take into account affecting phenomena such as frame collisions.

The ABE technique, presented in [23] estimates the frame collision probability through an online computation of the loss rate affecting the routing *HELLO* packets. As such a broadcast packets are of constant size, the authors use a Lagrange polynomial interpolation to approximate the collision probability affecting packets of arbitrary size. ABE was the first bandwidth calculation method to integrate the collision probability in the bandwidth estimation process. The authors have shown that once integrated in AODV it enhances the resulting per-flow performance. This framework approximates the available bandwidth by assuming that the idle periods observed at each node are uniformly distributed on a given time interval. The actual distribution of the idle period durations is studied in detail in [70].

Considering the IEEE 802.11 DCF protocol limitation on the number of packet retransmission attempts, RABE [22] provides a method for estimating the ratio of packet having reached the retransmission limit as well as the extra time needed when retransmitting a packet that collided. These considerations leads RABE being in average twice more accurate than ABE.

A recent hierarchical framework for evaluating the performance of 802.11 multi-hop networks based on a fixed-point solution has been proposed in [24]. It has been completed in [25] for integrating the case of hidden terminals. While the analysis returns accurate results on the achieved throughput and delay, the method cannot be applied to paths longer than three hops.

Another analytical framework that takes into account both inter-flow and intra-flow interference has been proposed in [20, 21]. It aims at predicting the throughput of each station in an arbitrary topology. By jointly considering the hidden terminal, information asymmetry and flow in the middle problems, the authors intend to examine all possible sources of packet losses and use a clique computation technique for reaching high prediction accuracy. The computed available bandwidth estimate, taken as a routing metric, has been experimentally demonstrated to perform well for identifying high-throughput paths [21]. The provided solution requires the use of some iterative procedures and motivated the work in [71] for designing a closed-form solution that incorporates the same affecting factors.

While the previous approaches are built on top of some specific distributions of the packet size, back-off durations as well as packet inter-arrival time, a novel framework that only depends on the relative average values is available in [72]. This analytical framework rather aims at determining the capacity region of the 802.11 ad hoc networks. It follows some recent results that permit to express such networks characterizations by using product-form expressions. However, it does not consider the case of hidden terminals and is built on top of a simplified network model.

Summary: The solutions presented so far have well addressed the problem of estimating the available end-to-end bandwidth in legacy 802.11 ad hoc networks. While these solutions can run on top of any routing protocol, they are not yet designed for the case of multi-channel networks and thus straightforwardly applicable to the cognitive radio context. However, the designed frameworks show high accuracy and constitute good promises for this purpose. Indeed, with most of these frameworks based on active measurements of the channel occupancy, it is likely that primary channel analysis results be leveraged for the design of bandwidth estimation technique that take into account the specificities of cognitive radio systems.

2.4.2.2 Time-slotted systems and the capacity region problem

The problem of admission control in time-slotted multi-hop networks has been largely studied in works that either addressed network throughput optimization or either investigated the capacity region problem. The former is addressed in [73], wherein the authors study the joint routing and channel assignment problem for the case of wireless mesh networks with multiple radios. They propose a constant approximation algorithm to the NP-complete problem of maximizing the overall network throughput subjected to fairness constraints. The relative computations are done offline and in a centralized fashion. This work is similar to more general approaches addressing the capacity region problem [68, 74], that consists of determining the maximum amount of flows that can be scheduled in a network. The flows are subjected to a set of constraints such as

intra/inter-link interference. In [68, 74], the authors propose a general framework for addressing the problem of joint routing and link scheduling given a set of flows to be admitted and a network interference model. Their approach consists in translating the set of network constraints into a linear integer program and then solving the relaxed version of this program. While the work in [74] considers only the use of unidirectional antennas, limiting the contribution to a certain class of networks, [68] gives a generalization of the method to any interference model. In the following, we give an overview of the underlying methods for computing the capacity region.

Sketch of Kodialam: The framework presented in [68], referred to as *Kodialam* in the remaining of this thesis, provides a recipe for evaluating the capacity region of a multi-radio multi-hop wireless network. The computation is done offline in a centralized fashion. It can apply to a variety of network interference models, traffic planning strategies, channel allocation and routing decision schemes. Indeed, with the joint routing, channel allocation and link scheduling problem being NP-complete, and formalized through a linear integer program, the method provides a succession of approximation schemes for each of these problems taken individually. Therefore, one can replace one of those with a mechanism of its own and use Kodialam as a global benchmarking tool. For example, the framework provides specificities for the cases of both static and dynamic channel assignment strategies. Moreover, while the method, as presented in [68], addresses the question of by how much a given set of flows across the network can be proportionally augmented such that an associated feasible schedule exists, the framework can be easily adapted to address the case of estimating the residual capacity on a given path. For this typical purpose, the link scheduling strategy is organized as follows:

1. It proposes solving a relaxed version of the initial linear integer problem. The associated constraints still depend on the set of link-channel, node-radio and interference constraints defined for the initial problem.
2. As these constraints are not sufficient for ensuring the solution to be provisioned with a feasible schedule, this solution gives an upper-bound of the maximal capacity. A feasible schedule is then obtained by running a greedy edge colouring algorithm, given the achieved upper bound: the achievable capacity equals the computed upper bound multiplied by the ratio between the TDMA frame size and the number of colors required to allocate the resulting flow.

Kodialam showed excellent results in approximating the capacity of multi-radio ad hoc networks. The main advantage of Kodialam is its extensibility to a variety of capacity estimation problems. The method can be easily adapted to the case of heterogeneous link rates and is thus a perfect candidate for evaluating the performance of channel access schemes, that be of random or deterministic nature. Unfortunately, Kodialam cannot be implemented distributively.

Similarly, in [75] the capacity region of multi-radio multi-channel wireless networks has been studied by introducing a multi-dimensional conflict graph characterizing the interferences between adjacent (radio, link, channel)-tuples. An admission control scheme is provided by deriv-

ing a set of local sufficient conditions for flow feasibility in such networks. It aims at providing an alternative and simpler solution than [68] for admission control. To this end, the authors use simple conditions based on the scaled clique constraint introduced in [76]. The proposed solution can identify feasible flows. However, among the feasible flows it can only provide a positive answer for admission to the ones that satisfy the sufficient conditions; there is no answer for the feasible flows that do not satisfy the sufficient conditions. Moreover, in this work as well as in [68, 74], the focus has been on the offline version of the admission control problem. That is, given a network with no prior allocations, the problem considered is that of computing the maximum rate that can be admitted between a source and a destination. In our work, we focus on the online version of the admission control problem. Given a live network, where capacity is allocated as traffic sessions arrive, the problem we tackle is that of computing the bandwidth available between a source and a destination *at the time* a new traffic session arrives.

An online and distributed admission control scheme for single-channel ad hoc networks has been proposed in [77]. In this work, it is shown that for a TDMA architecture, the problem of computing the residual end-to-end bandwidth on a given path is NP-complete. Intuitively speaking, the problem is hard because computing the residual end-to-end bandwidth is coupled with the problem of per-link slot assignments. With the problem being NP-complete, a greedy heuristic has been proposed and incorporated in the AODV routing protocol. However, this heuristic was designed for a single radio, non-cognitive radio architecture and cannot be readily applied to the cognitive context. In particular, it cannot handle the variability of link rates in the local optimization strategies.

Similarly, [78] provides a set of distributed, online and provably efficient algorithms for joint routing, channel assignment and scheduling in multi-hop multi-radio ad hoc networks. The proposed control algorithms guarantee of exploiting a fraction of the maximal capacity. This work is similar to the framework presented in [65]. However, both do not permit to compute the actual residual capacity on a given path; instead they admit traffic by controlling the network congestion level.

Summary: None of the reviewed schemes provide an online and distributed computation of the available bandwidth on a given path. The work closest to this problem is available in [77]. However, the underlying heuristic cannot incorporate the specifics of cognitive radio networks. On the other hand, the works presented in [68, 73] address a superset of the targeted issue, but cannot be implemented distributively. Unlike these works, our objective is not to propose a new joint routing and scheduling scheme that maximizes the network throughput. Instead, we focus on solving the problem of admission control once the paths are computed. The advantage of our approach is that it allows for a solution that can be computed online and adopted by already established routing protocols.

2.4.3 Concluding remarks

While probing can adapt to any network architecture, this estimation technique has been showed to perform poorly and degrade the wireless network performance. Therefore, we have considered approaches that are based on MAC layer information. The related works, summarized in Table 2.1, provide numerous frameworks designed for the cases of single-radio and multi-radio networks. Some of those can jointly optimize the routing and channel allocation scheme. However, all of them depend on the implemented MAC.

With a time-slotted structure envisioned at the MAC layer to address the sensing synchronization issue, this review reveals that at this time there is no distributed algorithm that permits to estimate, in an online and distributed fashion, the available end-to-end bandwidth in multi-hop cognitive radio networks. The conducted survey permits to identify the works that are the closest to this issue. First, the work provided by Zhu in [77], that presents a distributed online algorithm for achieving this purpose in single-channel constant-rate networks, but is based on a greedy heuristic that cannot be adapted to the cognitive radio context. Second, the method proposed by Kodialam in [68], that achieves well this objective through a centralized offline algorithm.

	Access type	Online/Offline computation	Implementation type	Bandwidth estimation	Multi-hop	Multi-rate	Multi-radio	Channel allocation	Routing optimization
Dovrolis [19]	Any	online	distributed	✓	✓	✓	✓	X	X
Sarr [23]	CSMA/CA	online	distributed	✓	✓	✓	X	X	X
Nguyen [22]	CSMA/CA	online	centralized	✓	X	✓	X	X	X
Abreu [25] from [24]	CSMA/CA	offline	centralized	✓	up to 3	✓	X	X	X
Garetto [20]	CSMA/CA	online	centralized	✓	X	✓	X	X	X
Salonidis [21] using [20]	CSMA/CA	online	centralized	✓	✓	✓	X	X	X
Nardelli [71]	CSMA/CA	online	centralized	✓	X	✓	X	X	X
Laufer [72]	CSMA/CA	offline	centralized	✓	✓	✓	X	X	X
Lin [78]	TDMA	online	distributed	X (control)	✓	✓	✓	✓	✓
Xue [65]	TDMA	online	distributed	X (control)	✓	✓	✓	✓	✓
Alicherry [73]	TDMA	offline	centralized	✓	✓	✓	✓	✓	✓
Kodialam [68] from [74]	TDMA	offline	centralized	✓	✓	✓	✓	✓	✓
Li [75]	TDMA	online	distributed	X (admission)	✓	X	✓	X	X
Zhu [77]	TDMA	online	distributed	✓	✓	X	X	X	✓

Table 2.1 – Review of strategies for estimating the available bandwidth

2.5 Routing metrics for cognitive radio networks

Routing in cognitive radio networks can be addressed through many aspects. It can consist of designing an efficient routing metric as well as a specific protocol. Both are often related so as to optimize several criteria such as primary user protection, delay minimization or throughput maximization. In the following, we present related works focused on the design of routing metrics for the case of cognitive radio networks. We show that most of these metrics are inherited from the metrics used in multi-hop multi-radio networks.

2.5.1 Single-path routing metrics

In cognitive radio networks, reflecting the unsettled nature of the field, there have been several proposed approaches to routing. Some have advocated for complete system solutions that address joint route-spectrum selection, protection to primary users [47], [62], [79], QoS provisioning [39] and route stability [80]. As a result many single-path cognitive routing metrics have been proposed in the literature. They can be actually classified into many categories, depending on the criteria to be optimized [81]:

- Hop-count [62, 64, 82]
- Delay [46, 62, 63, 83, 84, 85, 86]
- PU region avoidance / location-based decision [62, 80, 83, 84, 87, 88]
- Spectrum availability [64, 80, 82, 83, 84]
- Route stability [46, 62, 80, 83, 84]

When it comes to optimizing the end-to-end throughput, the metrics are inherited from ETX [26], the *de facto* routing metric for traditional ad hoc networks, as well as its extensions [27, 89, 90, 91].

2.5.2 Expected Transmission Count (ETX)

Estimating transmission count, the expected number of transmissions required for delivering a data packet over a link, as a means of improving routing performance in wireless networks was pioneered by De Couto et al. [26]. Their solution, ETX, has been modified (mostly augmented) many times, to include other features such as the physical bit-rate [27]. Its effectiveness and ease of implementation have made it a building block for most modern routing metrics. Indeed, ETX is calculated as the inverse of the packet reception ratio measured for broadcast probes transmitted periodically. Today, the broadcast probes remain the most effective and practical solution for a measurement-based link quality estimation. Moreover, ETX has been applied to contexts far beyond the original, including to sensor networks [92], backpressure routing [93], opportunistic routing [94, 95] or network coding [96].

2.5.3 High-throughput routing metrics in legacy networks

The design of efficient metrics for high-throughput routing in legacy ad hoc networks has been studied in several works over the last past decade. This can be achieved by using a metric that advantages high-bandwidth links, limits the amount of interference between secondary users and/or favors channel diversity along the same path. This is typically the purpose of ETT [27], WCETT [27] and MIC [90, 91].

2.5.3.1 Expected Transmission Time (ETT)

The ETT metric, for Expected Transmission Time, basically combines ETX with the physical link rate. For a given link, it can be expressed as follows:

$$ETT = ETX \times \frac{packet\_size}{link\_rate} \tag{2.1}$$

ETT estimates the time for a packet to be successfully transmitted over a link. ETT is additive, that is, the path weight is simply obtained by summing the ETT value computed along the path.

2.5.3.2 Weighted Cumulative Expected Transmission Time (WCETT)

The WCETT metric, for Weighted Cumulative Expected Transmission Time is aimed at reducing intra-path interference by favoring channel diversity. The metric is designed so as to minimize the number of link simultaneously transmitting over the same channel along the path. It comprises both additive and concave (max) terms. For a given path p:

$$WCETT = (1-\beta)\sum_{\forall l \in p} ETT_l + \beta \max_{1 \leq j \leq k} X_j \tag{2.2}$$

where X_j is the number of times channel $j \in \{1, ..., k\}$ is used along the path and β is a parameter that can be set in the interval $[0, 1]$.

While WCETT handles well the trade-off between transmission delay and interference between the secondary users, it presents two main drawbacks. On the one hand, it penalizes paths using the same channel several times, while the relative links might not interfere with each others if they are sufficiently spaced. On the other hand, routing according to WCETT might lead to non-optimal paths. This is demonstrated in [28] using the example illustrated in Figure 2.5 and parameter β set to 0.5. When the Dijkstra's shortest path algorithm is coupled with WCETT, it selects path $S_1 \rightarrow S_2 \rightarrow C \rightarrow D \rightarrow T$ as the best path from S_1 to T. However, the optimal path is $S_1 \rightarrow B \rightarrow T$. This is due to the sub-path $S_1 \rightarrow B$ being of higher weight than $S_1 \rightarrow A \rightarrow B$ and being not considered anymore for the remaining of the path calculation process.

In addition, an hop-by-hop routing strategy may cause forwarding loop as the path computed by Dijkstra's algorithm from S_2 to T is $S_2 \rightarrow S_1 \rightarrow B \rightarrow T$. Therefore, once a packet if forwarded from S_1 to S_2, S_2 immediately forwards it back to S_1.

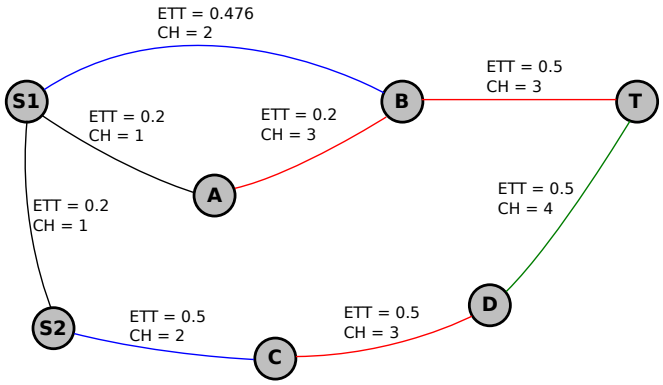


Figure 2.5 – WCETT fails in selecting the best path

WCETT can thus lead to forwarding loops and routing non-optimality. In [28], the authors provide a set of necessary and sufficient conditions on a routing metric for guaranteeing routing optimality when used in Dijkstra's, Bellman-Ford's or flooding-based routing algorithms. They demonstrate that WCETT does not satisfy these necessary conditions. In detail, WCETT is not isotonic.

2.5.3.3 Metric of Interference and Channel-switching (MIC)

The MIC metric, for Metric of Interference and Channel-switching was proposed in [91]. This metric aims at minimizing the level of intra-path interference by combining the IRU metric [91], for Interference-aware Resource Usage, which reflects the effect of inter-flow interference on packet transmission delay, with a channel switching cost. The channel switching cost is set locally and depends on the channel used for transmission over the previous link. The MIC metric cannot be used in Dijkstra-based routing algorithms as it suffers from the same drawback as WCETT: sub-paths of an optimal path are not necessarily optimal. The problem is due to the channel switching cost component. Fortunately, as each component of MIC is additive and can be iteratively computed using only the previous link information, the metric can be decomposed and used in an auxiliary network where virtual nodes have been added to consider any local channel assignment. This decomposition is operated by the LIBRA routing protocol [91] and permits to find MIC shortest paths without any issue. Note that such a transformation only applies to the MIC metric and cannot be generalized to any metric, and WCETT in particular.

2.5.3.4 Available bandwidth

Finally, one might think that routing according to an available end-to-end bandwidth metric might be optimal. Such a strategy has been adopted in [21] and led to satisfactory throughput results. However, the underlying metric does not satisfy the sub-path optimality property.

As demonstrated in [97], it is not isotonic [28] and cannot guarantee routing optimality when coupled with Dijkstra's or Bellman-Ford's algorithms. Such a phenomenon is mainly due to the specificities of intra-path interference.

2.5.4 ETX adaptation to the cognitive context

Most of the metrics presented so far for cognitive radio routing have been designed on top of ETX [26]. While cognitive radio networks present different link dynamics, with alternative primary user activity/non activity periods, ETX is computed exactly the same way as for traditional ad hoc networks, by measuring the packet reception ratio affecting broadcast probes sent periodically. Nevertheless, ETX was not designed to quantify the impact of primary users on transmission count. Only few strategies have intended to adapt the metric to the cognitive context. Two of those are SAMER [82] and STOD-RP [46]. They are built on basically multiplying ETX with a factor characterizing the primary user activity. SAMER essentially multiplies the packet reception ratio by the fraction of time the link is available, while STOD-RP combines link quality with spectrum availability and divides ETT [27] by the time duration of the link.

2.5.5 Summary

Available bandwidth metrics in multi-hop wireless networks can be leveraged to perform admission control. However, using them as routing metrics could lead to non-optimal path selection. It requires using alternative metrics that manage to capture high-throughput paths and possibly ensure routing optimality. Such metrics are basically built on top of the ETX metric, the *de facto* metric for estimating transmission count in traditional ad hoc networks.

However, in cognitive radio networks, the way ETX is computed remains the same as for legacy networks and no special study has demonstrated the actual effect of the primary users interference on transmission count.

3 Admission control in cognitive radio ad hoc networks

In this chapter we address admission control in multi-hop cognitive radio networks through the particular problem of computing the available bandwidth of a given path. In Section 3.1, we describe in detail the multi-hop cognitive radio network architecture under study. In particular, because the research and development of a widely accepted MAC protocol for these networks is still ongoing, we consider a bare-bones TDMA protocol at the link layer and show that for the system considered the problem of computing the available bandwidth of a given path is NP-complete. Instead of working on an approximation algorithm, we follow a different approach and use a simple scheduling heuristic. We adopt randomized scheduling due to its simplicity and efficiency. However, in spite of such a simple scheduling rule, computing the available bandwidth over a given multi-hop path remains an open problem. We solve this problem in Section 3.2 via BRAND, a polynomial-time algorithm for computing the *available Bandwidth with RANDomized scheduling*. We show in Section 3.3 that BRAND can be implemented in a distributed fashion and thus integrated by almost all routing approaches. An extensive numerical analysis led in Section 3.4 demonstrates the accuracy of BRAND and its enabling value in performing admission control in both static and mobile environments.

3.1 Preliminaries

As depicted in Section 2.4, computing the end-to-end bandwidth is not a new problem. However, there are two reasons for revisiting it. First, there is no practical solutions that permits to accurately estimate the bandwidth in a online and distributed fashion, for TDMA-based systems. Second, the cognitive radio architecture makes the problem non-trivially different. The main reason for this is the existence of the so called primary-secondary hierarchy [1]. In issuing its landmark ruling [3], permitting the use of unlicensed devices in the UHF spectrum, the FCC required that the unlicensed secondary users do not interfere with the incumbent primary

1. In the future, if the vision of a fully dynamic spectrum sharing becomes a reality, we will probably see more than two classes of users and respective channel access priorities. However, for the moment we focus on the basic case of a two-class hierarchy.

users. Thus, the end-to-end bandwidth in these networks will be conditioned not only by the interference between peers, as is the case in legacy wireless networks, but also by a different kind of interference: Primary users, as their prerogative, will access the channel without making any effort to avoid interfering with secondary users. In addition to the primary user interference, another distinguishing feature of these networks is that cognitive radios, in their pursuit of available spectrum, may access different frequency bands and use different channel widths [51]. This will lead to network links with widely varying capacities. Multi-rate links can also be seen in legacy networks, such as IEEE 802.11, but in cognitive networks they are a fundamental feature that has to be taken into account by any bandwidth calculation algorithm. For this purpose, in the following sections, we describe in detail how we model a cognitive radio ad hoc network as well as the inference caused by peers and primary users.

3.1.1 Network model

We model a multi-hop cognitive radio network as a graph $G = (V, E)$, where V is the set of nodes and E the links. We assume that the network is composed of only symmetric links, that is, there exists an edge between two vertices v_i and v_j if and only if nodes n_i and n_j are able to correctly communicate with each other. Every cognitive radio node is equipped with a constant number of half-duplex transceivers, each capable of sensing and transmitting on B predefined orthogonal wireless channels [98]. All the channels can offer different data rates. An additional transceiver could be used for control signaling. We assume that the channel assignment is performed by a spectrum allocation protocol [46] and focus on estimating the available end-to-end bandwidth once such assignment is completed. The only assumption we make about the frequency assignment algorithm is that only one frequency channel is assigned between a particular pair of neighboring nodes. Such a network model is illustrated in Figure 3.1 wherein a cognitive radio ad hoc networks operates on $B = 3$ orthogonal channels. The green channel is taken as an example of channel in the unlicensed ISM frequencies, while the red and blue ones refer to distinct TV white-spaces in the VHF/UHF bands. Note that the primary user activity pattern is not known *a priori* from the secondary users: the Secondary and Primary networks are assumed not synchronized.

3.1.2 Channel access

As a prerequisite for using licensed spectrum, cognitive radios are not to use the channel when it is in use by the respective licensed user. In literature, this is referred to as a secondary-primary [1] hierarchy, with the primary (licensed) user having strict priority in accessing the channel.

As stated in Section 2.3.4, there is no clear winner yet among the MAC protocols proposed for cognitive radio ad hoc networks. In addition, we believe a deterministic medium access protocol will better serve an architecture where multiple technologies share the same spectrum

1. We will use the terms primary user/secondary user, primary/secondary and PU/SU interchangeably.

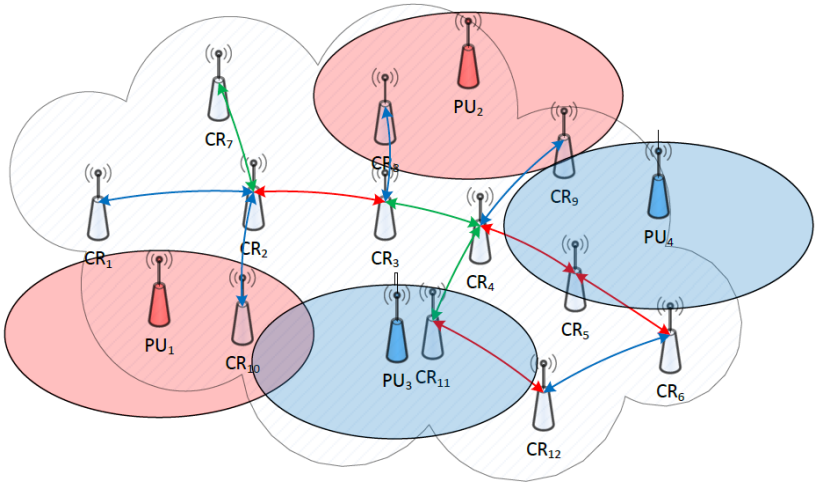


Figure 3.1 – Cognitive Radio ad hoc networks model

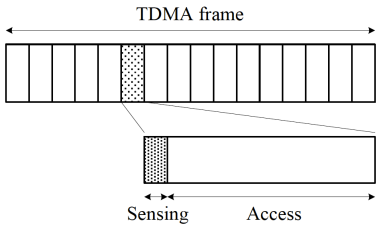


Figure 3.2 – TDMA frame structure

and synchronization is required for the sensing. Therefore, we adopt a system in which a TDMA MAC with frame size S is implemented on every assigned channel. Every time-slot is started by a in-band sensing period as illustrated in Figure 3.2. In this Figure, the TDMA frame is composed of $S = 16$ time-slots. During the in-band sensing period, the secondary users remain silent and use energy detection to determine the channel occupancy. When a node needs to transmit data to a neighboring node, it can access the medium by reserving time-slots on the frequency channel assigned to this particular link. For ease of presentation, we refer to the pair $(channel, timeslot)$ simply as, a slot.

3.1.3 Model of interference

There are two kinds of interference sources in a cognitive radio network. First, there is the interference from other cognitive radios in the same interference domain, usually referred to as secondary-to-secondary interference. Then, there is the interference from the Primary User. The first is not unlike the interference legacy wireless networks have to cope with: nodes running the same protocol contend for access to the same channel. The primary user interference, however,

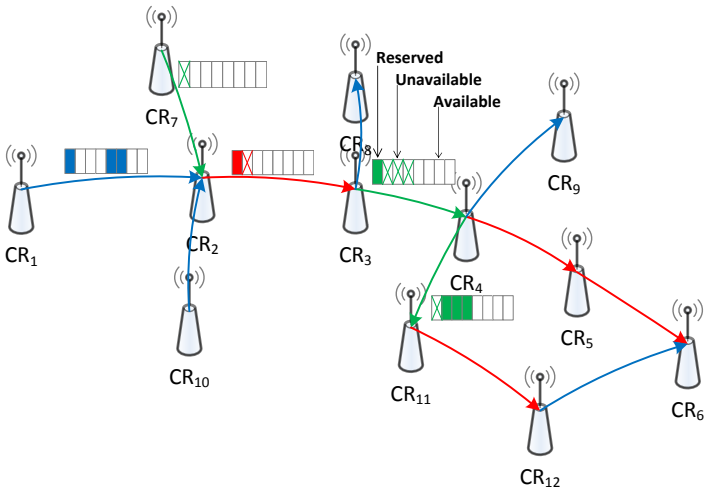


Figure 3.3 – Illustration of the SU-to-SU interference

is different: As its prerogative, the primary user can choose to access the channel at any given time with the expectation of no interference from any potential secondary users. In the following we discuss how these two sources of interference are modeled in this work.

3.1.3.1 Secondary-to-Secondary Interference

For the secondary-to-secondary interference we use the model usually employed in TDMA systems on half-duplex wireless transceivers. Specifically, should a particular node need to reserve a new time-slot to transmit data to a neighbor, it does so on the corresponding assigned channel. However, due to the potential interference from other cognitive radios operating on the same channel, for the time-slot to be selected, it needs to satisfy the following requirements:

1. ***This time-slot is not used on this channel by node itself for transmitting,***
2. ***It is not used on this channel by any one-hop neighbor for transmitting,***
3. ***It is not used on this channel by any two-hop neighbor for transmitting.***

We assume every node knows the slot allocations in its two-hop neighborhood and thus can check the satisfiability of the above constraints. This information can be easily obtained by sending beacons containing bitmaps with slots scheduled for transmission or reception for the node itself and its one-hop neighbors. Some of these constraints are illustrated in Figure 3.3.

3.1.3.2 Quantifying the Primary User Interference

Once the sensing module identifies a primary user, the cognitive radios are to stop all communications. Therefore, it is required that a cognitive radio spend part of the time sensing for

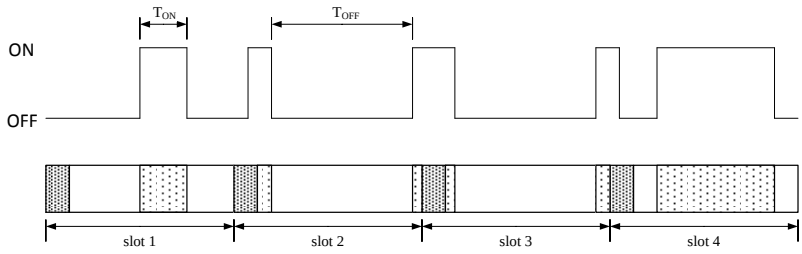


Figure 3.4 – Possibility of Primary-to-Secondary interference due to sensing imperfection

primary users and part of the time actually transmitting data. To accommodate this requirement, in our model, as shown in Figure 3.2, an amount of time in every slot is dedicated to sensing, while the rest for actual channel access. The optimal ratio between sensing and channel access will depend on several factors, including the Primary User activity, the traffic demands for the cognitive radio, etc. A thorough study of these factors for computing the optimal sensing time is beyond the scope of this thesis and related to the works presented in Section 2.3.1. However, the correctness of our scheme does not depend on the exact values of sensing and access times. Obviously, it will be impacted by the spectrum analysis performance. If during the sensing period a primary user is identified, no communication will take place in the access part of the slot. Otherwise, the cognitive radio is free to access the channel. However, sensing is not perfect and it can very well happen that, while no primary user is identified during the sensing period, a primary user does become active for the whole or part of the access time, as depicted in Figure 3.4. This is typically due to the fact that Primary and Secondary networks are not synchronized. When this happens, the exact consequences on whatever secondary user transmissions going on will vary depending on the location and the power strength of the primary user. We follow a somehow pessimistic assumption: A primary, when active, will interfere destructively with any secondary communication taking place in its range [1].

Computation of the fraction of slot duration available to secondary users: If we denote with η the part of the slot access time that will be available to the secondary user, based on the reasoning so far, we have:

$$\eta = \text{P[sensing the channel idle]} \times \text{(Fraction of Access Time Free of PU)}$$

Theorem 1 (Fraction of the slot access time available to the secondary users)**.** *Denoting with u_i the probability of a primary user becoming active on link l_i during a particular slot, the fraction*

1. We assume that if a secondary can sense a primary user then the particular secondary is in the interference range of the primary user.

of the slot access time available to the secondary on link l_i can be computed as follows:

$$\eta_i = (1 - u_i)^2 \tag{3.1}$$

Proof. The PU's channel state can be modeled by an alternative ON/OFF process [42, 44, 48]. The durations of the ON and OFF states related to link l_i, respectively T^i_{on} and T^i_{off} can be assumed to be exponentially distributed with probability density functions:

$$f^i_{on}(t) = \frac{1}{\overline{T^i_{on}}} e^{-t/\overline{T^i_{on}}}$$

$$f^i_{off}(t) = \frac{1}{\overline{T^i_{off}}} e^{-t/\overline{T^i_{off}}}$$

with parameters $\overline{T^i_{on}}$ and $\overline{T^i_{off}}$ estimated with maximum likelihood methods whatever the sensing duration is. This results in the famous formula $u_i = \overline{T^i_{on}}/[\overline{T^i_{on}} + \overline{T^i_{off}}]$. Jiang et al. [44, 48] have focused on computing the quantity of PU-SU interference accumulated during the access time for evaluating the impact of SU's communications on PU's Quality of Service and appropriately tuning the access time duration. Defining $I_0(T_A)$ as the expected length of all ON states within access time T_A given that T_A begins from the OFF state and $I_1(T_A)$ the same given that T_A begins from the ON state, they demonstrated that $I_0(T_A)$ and $I_1(T_A)$ satisfy two renewal equations and derived their closed-form expressions. Regarding the communication link l_i, as $\mathbb{P}$[occurrence of the OFF state] $= 1 - u_i$, we get:

$$I^i_0(T_A) = u_i T_A - u_i^2 \times \overline{T^i_{off}} \times \left[1 - e^{-T_A/(u_i \overline{T^i_{off}})}\right] \tag{3.2}$$

and

$$I^i_1(T_A) = u_i T_A + (1 - u_i)^2 \times \overline{T^i_{on}} \times \left[1 - e^{-T_A/(u_i \overline{T^i_{off}})}\right] \tag{3.3}$$

Although the sensing may declare the primary channel idle, it is still possible that PU's communications begin just before the sensing period ends. This situation can occur if the computed energy does not exceed the selected threshold. Therefore, we define the quantity of interference relative to a slot access time as:

$$Q_i = \frac{(1 - u_i) \times I^i_0(T_A) + u_i \times I^i_1(T_A)}{T_A} \tag{3.4}$$

where T_A corresponds to the slot access time. Hereafter, substituting Equations 3.2 and 3.3 in Equation 3.4 leads to $Q_i = u_i$ and $\eta_i = (1 - u_i)^2$. □

Then, by taking into account the sensing time, the *fraction of slot duration, f_l,* available for

secondary-to-secondary communication is:

$$f_l = \eta_l \times \frac{T_{access}}{T_{sensing} + T_{access}} \tag{3.5}$$

Equation 3.1 quantifies the effect of two things on the capacity for the secondary. First, the interference from the primary, who as the owner of the frequency is bound by no protocol to try to avoid interference with an ongoing secondary communication. And second, the mechanism put in place, i.e. sensing, for satisfying the requirement of doing no harm to the primary. Note that, for a primary activity of 10%, the secondary user will not realize more than 81% of the slot access time capacity. At first, one might think that the secondary should instead be able to reach 90%. The explanation for the 9% loss is the sensing. A primary could be active during the sensing period but not so during the access time and yet, the secondary will not use the access time, leading to unnecessary loss of capacity.

3.2 BRAND: An approach for estimating the bandwidth

3.2.1 Problem Definition

Let the demand d, expressed in bits per second, refer to the amount of end-to-end bandwidth required by an application. Before admitting to route this demand, for QoS purpose, we first would like to know whether this demand can be satisfied end-to-end. This question can be answered by simply computing the *currently* available end-to-end bandwidth of the path to the destination.

Definition 1. The available end-to-end bandwidth of a path *is the maximum amount of data, in bits per second, that can be currently transported over the path.*

Remark 1. *Unlike the maximum end-to-end bandwidth, the available bandwidth is time sensitive and depends on the current conditions and allocations in the network. If there is no other ongoing traffic in the network and there is no primary user activity, the available bandwidth is equivalent to the maximum path bandwidth.*

Remark 2. *The admission control problem could alternatively be framed as one of computing the path with the maximum available bandwidth. However, this would require implementing a new routing protocol for this purpose. What is more, given that we want to solve this problem online, as the traffic sessions arrive in real-life, it is not clear that, overall, it would lead to more sessions being admitted. Thus, we opted for an approach that can be added to any available routing protocol which, once computing the routes based on whatever metric it considers crucial, can simply asks us for the available bandwidth on a particular path.*

Computing the available end-to-end bandwidth obviously depends on the capacity of the links constituting the path, the ongoing traffic and the primary user activity. But what makes the

problem challenging, and as we show later in this section, NP-Complete, is the *self-interference* on the path. Slots allocated on a given link of a multi-hop path will not be available on other links on the path that are in the same interference domain (see Section 3.1.3.1). Thus, slot allocations made on a path link can condition subsequent slot allocations on other path links in the same interference domain. And since these subsequent allocations will similarly condition slot allocation following them, we end up with a domino effect across the whole path.

Constructing the formal problem definition: A path is modeled as a directed chain $n_1 \rightarrow n_2 \cdots \rightarrow n_{N_H+1}$ composed of N_H hops. For ease of presentation, we denote a link $n_i \rightarrow n_{i+1}$ as l_i. The bit-rate [1] of every link is denoted by ϕ_i and, for every link, the TDMA frame size is S slots. To take into account the effect of self-interference, that is, links on the same path interfering with each other, we use the exponential notation (j) to specify that the considered quantity is evaluated just before node n_j on the same path does its allocations. Using this convention, we define $A_i^{(j)}$ as the number of slots available at node n_i for communication on the link l_i just before n_j does its own allocations.

Let us analyze the network behavior when admitting a new flow with demand d. The first node on the path, n_1, converts the flow demand, d, to the required number of slots, r_1, to be allocated on the first path link, l_1. The number of required slots will depend on the demand, the link bit-rate, ϕ_1, the TDMA frame size, S, as well as the primary user interference (quantified in Section 3.1.3.2, Equation 3.5):

$$r_1 = \left\lceil \frac{d}{\phi_1 \times f_1} \times S \right\rceil$$

Let a_i denote the number of slots specifically allocated on every hop $i \in \{1, .., N_H\}$ for servicing this flow. For every hop this number will depend on both the demand and, how many slots are actually available for new allocations. Thus, for the first hop we have $a_1 = \min(r_1, A_1^{(1)})$. If $a_1 < r_1$ the demand on the second link will be lower than the original demand, d. To distinguish the two, we denote the demand on the second link, which depends on the allocation on the first link as, d_1. Rigorously speaking, $d_1 = \min\left(d, a_1 \times (\frac{\phi_1 \times f_1}{S})\right)$, where the quantity by which a_1 is multiplied is the capacity of a single slot on the first link. We can generalize these results for any hop, $i > 1$, as follows:

$$r_i = \left\lceil \frac{d_{i-1}}{\phi_i \times f_i} \times S \right\rceil \tag{3.6}$$

$$a_i = \min\left(r_i, A_i^{(i)}\right) \tag{3.7}$$

$$d_i = \min\left(d_{i-1}, a_i \times \left(\frac{\phi_i \times f_i}{S}\right)\right) \tag{3.8}$$

1. How the MAC layer computes the bit-rate on a particular link is orthogonal to our contribution. A potential approach could be combining the physical bit-rate with the ETX metric [26].

Thus, for a specific demand d, the **realized end-to-end throughput** is $\min(d_1, d_2, ..., d_{N_H}) = d_{N_H}$, since $d_i \geq d_{i+1}$. This analysis gives us a way for tackling the main problem, computing the available end-to-end bandwidth.

Problem 1. *Computing the available end-to-end bandwidth of a path is equivalent to solving the following optimization problem:*

$$\max_{d \in I_d} \quad d_{N_H}(d) \tag{3.9}$$

where $I_d = [0, \min(\phi_1, \phi_2, ..., \phi_{N_H})]$.

The optimization problem thus defined leads to two observations:

1. ***The realized end-to-end throughput,*** $d_{N_H}(d)$***, given a demand,*** d***, obviously depends on*** d***.***
2. $d_{N_H}(d)$ ***depends on how the slots are allocated on every hop.***

Why computing the available end-to-end bandwidth is challenging: To illustrate the implications of the above observations and how challenging the above optimization problem is, let us consider the following toy example. Consider the first three nodes, A, B, C of a multi-hop path and let us assume they all have assigned the same channel, which puts them in the same interference domain. The TDMA frame consists of ten slots. In node A slots 1,2,5,6 are available for transmitting and receiving, in node B, slots 2,5,6 are available for transmitting and receiving while slot 1 is available for receiving only [1], and in node C all slots are available for transmitting and receiving. If the demand d is such that four slots are required for satisfying it, the first node, A, will allocate slots 1,2,5,6. Since nodes B, C are in the same interference domain, node B will not be able to transmit to node C on any of its slots available for transmitting. This will result in a zero end-to-end throughput. Now let us consider that the demand d is such that two slots are required for satisfying it. In this case node A has $\binom{4}{2}$ ways to chose the two slots for allocation out of the four available. While for node A all choices are equivalent, that is not the case for nodes B and C and ultimately, the end-to-end throughput. If A allocates slots 2 and 5, node B will be left with only one slot, slot 6, to use for forwarding traffic to node C. If, instead, A allocates slots 1,2, node B will be left with two slots, slots 5,6, for forwarding traffic.

Clearly, the way slots are allocated on every node will have an impact on the realized end-to-end throughput. Because on a given channel a node cannot reserve a slot used for transmission by one of its 1-hop or 2-hop neighbors, $A_i^{(i)}$ is likely to decrease if allocations are done on links l_{i-2} and l_{i-1} [2]. Therefore, the number of slots allocated on every hop highly depends of the demand as well as the channels selected along the path. For instance when l_1 and l_2 operate on the same channel and all the slots are available for reservation on the first hop, a demand requiring the reservation of all these slots will leave zero slots available for communication on

1. It can happen that a node scheduled to receive on slot 1 is in the same interference domain with node B but not A.
2. We will not take into account the impact caused by allocations done on the subsequent links.

the second link and the resulting end-to-end bandwidth would be zero. Moreover, the number of possible schedules can grow exponentially with the number of nodes on the path. This leads to the following result.

Theorem 2. *Computing the available end-to-end bandwidth of a path in a TDMA-based multi-hop cognitive radio networks with multiple transceivers is NP-complete.*

Proof. The proof is straightforward so we provide a sketch. We show that our problem is NP-Complete by reducing the problem of computing the maximum path bandwidth in a single-channel TDMA-based multi-hop network, therein referred to as P_2, to our problem, therein referred to as P_1. To this end, we consider the instance of P_1 where a same channel with a constant data rate is assigned on every link along the path and the probability of primary activity on all links is zero. Solving P_1 actually consists of solving one instance of P_2. Since P_2 has been shown to be NP-complete[77], that concludes the proof. □

With the problem of computing the available end-to-end bandwidth being NP-Complete, the overwhelming approach in literature has been to design a scheduling heuristic. We follow a different approach. We select a specific slot scheduling algorithm and focus on computing the available end-to-end bandwidth resulting from applying this particular algorithm. As scheduling algorithm we select the randomized scheduling [99]: *when a node needs to assign a certain number of slots, it will select them at random among those available.* Randomized scheduling is widely used because of its simplicity and efficiency even though the worst case performance can be poor. In the following we present BRAND, our high-level algorithm for estimating the available end-to-end Bandwidth with RANDom scheduling.

3.2.2 Fundamental principles of BRAND

For every possible demand d, the necessary slots are allocated at random among those available on every link and the resulting end-to-end throughput is computed. By Equation 3.9, the available end-to-end bandwidth is simply the maximum end-to-end throughput realized over all possible demands d.

Algorithm 1: BRAND (Available **B**andwidth with **RAND**om Scheduling)

```
   Output : The available end-to-end bandwidth
1 : begin
2 :     for every possible demand d do
            //Use Random Scheduling
3 :         avThput ← Compute-averageThput (d);
4 :         if avThput > AvailBW then
5 :             AvailBW ← avThput;
6 :     Return AvailBW;
```

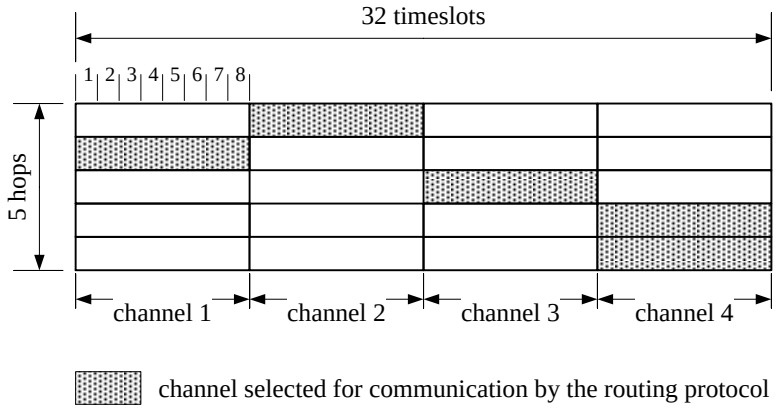


Figure 3.5 – The Path's Slots Availability Table

The value of the demand, d, is upper-bounded by the lowest radio bit-rate in the network and lower-bounded by 0. Since it is a slotted system, the values of d have to be multiples of the smallest slot capacity in the network. Thus, the possible values of d that need to be considered are bounded by a constant. The non-trivial step of BRAND, line 3 in Algorithm 1, is computing the average end-to-end throughput when the required slots for satisfying a particular demand, d, are assigned at random. Note that, because the slots are allocated at random, we can only compute the average and not the exact value of the resulting end-to-end throughput. In the following, we give a centralized approach that, given a demand d, computes the average end-to-end throughput realized. In Section 3.3 we propose a distributed approach.

3.2.3 Computing the Average End-to-End Throughput with Random Scheduling

In the following, we propose an analytical framework for computing the average throughput that would be achieved on every link of a path if a new flow with demand d were to be admitted and random scheduling is used on every link. This solution is centralized in that, the source node is assumed to have global knowledge of the network.

3.2.3.1 The Path's Slots Availability Table (PSAT)

For the sake of clarity, we define a new data structure indicating, for every link of a path, the slots available for reservation. We call this structure the path's slots availability table (PSAT). This table is composed of N_H lines and $S \times B$ columns. Each line i, composed of B sub-blocks of size S, indicates the available slots on link l_i. Then, each sub-block refers to the available time-slots on each sensed frequency channel. All the time-slots of each of the $B-1$ frequency channels not selected for communication on link l_i, are considered unavailable and the corresponding entries in the table are set to 0. A simple example of a PSAT table is depicted in

Figure 3.5 for the case of a 5-hop path. In this example, every link is assigned a channel among the four sensed by the spectrum decision module. Each TDMA frame is composed of eight time-slots. Whenever at a node the spectrum decision module carries out channel reassignment, the PSAT is updated accordingly. Therefore, the PSAT summarizes all the necessary information required for computing the average number of slots that would be allocated on every hop of the path if a new flow with demand d were to be admitted.

3.2.3.2 Fundamental principles of the method

To make the problem tractable, we relax it by working on average values. Even though mathematically speaking $\mathrm{E}[a_3] \neq \min\left(\mathrm{E}[r_3], \mathrm{E}[A_3^{(3)}]\right)$, we do such an approximation of the average number of slots allocated on the third hop to reduce the calculation complexity. As depicted in Section 3.4, our simulation results show that this approximation does not degrade the performance of the overall estimation process. Based on this assumption, a first solution consists of estimating for each hop i the quantity $A_i^{(i)}$ which is now considered an average for the remaining of this chapter.

l-link available slot set decomposition: From the PSAT, we can calculate for any communication link l_i the set $S_i^{(1)}$ containing the index of slots available for reservation on that link at the beginning of the estimation process. The indexes of slots are now taken in the set $\{1, 2, .., S.B\}$ related to the super-frame composing the whole line in the PSAT. It is also crucial to see that, when considering such an indexation, when a slot is allocated on l_i, it cannot be allocated anymore on both l_{i+1} and l_{i+2} as this would create interference. Thus, when this slot is only available for reservation on link l_i and neither on links l_{i+1} nor l_{i+2}, the sets $S_{i+1}^{(j)}$ and $S_{i+2}^{(j)}$ are not impacted. Inversely, if this slot is also available to one of these links, the corresponding sets are impacted and thus the number of slots that would be allocated on the next hops is likely to decrease. Thus, each slot belongs to a certain category depending on the links it appears to be available for reservation to.

Therefore, we propose to divide the set $\{1, ..., S.B\}$ in non overlapping subsets that cover $\{1, ..., S.B\}$ and permit to categorize every slot according to the links it is available for reservation to in the PSAT table. To be more precise, we define such a decomposition on a set of l consecutive links $\{i, i+1, ..., i+l-1\}$ along the path. For $k \in \{0, 1, ..., l\}$, the number of subsets characterizing the slots available to a set of k links but not the $l-k$ others is $\binom{l}{k}$. Thus, the total number of subsets in the decomposition is $\sum_{k=0}^{l}\binom{l}{k} = 2^l$. We refer to such a decomposition as a *l-link available slot set decomposition* and use the following notations $E^{(j)}_{a,\bar{b},c}$ to denote the set of slots available for reservation on both two links a and c but not b just before node n_j does its allocations. Its cardinality is written as $C^{(j)}_{a,\bar{b},c}$. To elucidate the meaning of these variables, let us consider a 3-hop path with the first channel selected for communication on every hop, $B = 2$, $S = 8$ and $S_1^{(1)} = \{2, 3, 4, 5\}$, $S_2^{(1)} = \{2, 6, 7\}$ and $S_3^{(1)} = \{1, 2, 4, 5, 6, 7\}$. This leads to the following eigth sets: $E^{(1)}_{1,2,3} = \{2\}$, $E^{(1)}_{1,2,\bar{3}} = \emptyset$, $E^{(1)}_{1,\bar{2},3} = \{4, 5\}$, $E^{(1)}_{\bar{1},2,3} = \{6, 7\}$, $E^{(1)}_{1,\bar{2},\bar{3}} = \{3\}$, $E^{(1)}_{\bar{1},2,\bar{3}} = \emptyset$, $E^{(1)}_{\bar{1},\bar{2},3} = \{1\}$ and $E^{(1)}_{\bar{1},\bar{2},\bar{3}} = \{8, 9, ..., 16\}$.

Achievable end-to-end throughput calculation: The random nature of the slot allocation process provides good properties to evaluate the average number of slots impacted in every subset as further depicted for the case of a 3-hop path in Algorithm 2 for which a flow with demand d needs to be relayed from the source to the destination. From now on, we work with average values. From the PSAT, we can compute the 3-link available slot set decomposition related to l_1, l_2 and l_3. At the same time, the initial number of available slots on every communication link $A_i^{(1)}$ can be calculated. To forward the new traffic flow to its next hop n_2, node n_1 reserves exactly $a_1 = \min(r_1, A_1^{(1)})$ additional slots on the first communication link. Among these slots, some might have also been available for reservation on links l_2 and l_3 but, due to interference, become unavailable after these allocations.

Let us consider a discrete random variable X_i taking its values in the set $\{0, 1, ..., a_i\}$ and representing the number of slots initially available for reservation on link l_i that have been reserved by node n_1 to relay the new incoming flow on l_1. X_i represents the number of slots in the set $S_1^{(1)} \cap S_i^{(1)}$ reserved by n_1 for communication on l_1. As the slots are allocated at random, we see that a proportion $a_1/A_1^{(1)}$ of the available slots on l_1 are likely to be reserved by n_1. Using the available slot set decomposition, we can measure the impact caused by these slot allocations on the slots remaining available on the consequent links. Mathematically speaking, X_i follows an hypergeometric distribution with parameters $(A_1^{(1)}, |S_1^{(1)} \cap S_i^{(1)}|, a_1)$. The expectation value of such a random variable is $\mathbb{E}[X_i] = |S_1^{(1)} \cap S_i^{(1)}| \times a_1/A_1^{(1)}$ and thus, in every set $S_1^{(1)} \cap S_i^{(1)}$, an average proportion $p_1 = a_1/A_1^{(1)}$ of slots is reserved by node n_1. Note that for the case of $A_1^{(1)} = 0$ we get $a_1 = 0$ and $p_1 = 0$.

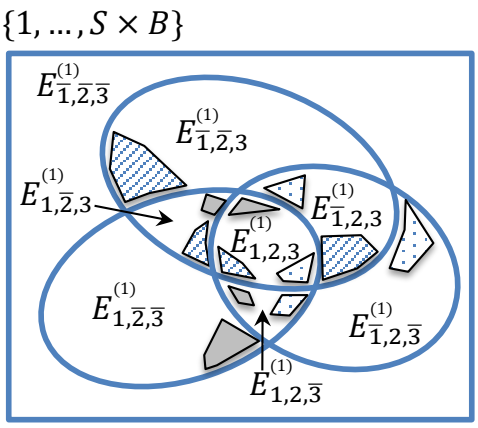


Figure 3.6 – 3-link available slot set decomposition

Exactly the same analysis can be carried out on every set resulting from the 3-link available slot set decomposition related to l_1, l_2 and l_3. This way, the average values $A_2^{(2)}$ and $A_3^{(2)}$ just after n_1 did its reservations can be computed as detailed in Algorithm 2 and illustrated in Figure 3.6. In this Figure, when slot allocations are carried out on l_1 the proportion of slots allocated, p_1, is

Algorithm 2: Computing the average end-to-end throughput on a 3-hop path.

input : d, S, ϕ_1, ϕ_2, ϕ_3, f_1, f_2, f_3, $S_1^{(1)}$, $S_2^{(1)}$, $S_3^{(1)}$
output : a_1, a_2, a_3, d_1, d_2, d_3

1 : **begin**
  //Initialization
2 : $\forall i \in \{1,2,3\}, \quad A_i^{(1)} \leftarrow |S_i^{(1)}|$;
  //Available slot set decomposition
3 : $C^{(1)}_{\overline{1},\overline{2},\overline{3}}, C^{(1)}_{1,\overline{2},\overline{3}}, C^{(1)}_{\overline{1},2,\overline{3}}, C^{(1)}_{\overline{1},\overline{2},3}, C^{(1)}_{1,2,\overline{3}}, C^{(1)}_{1,\overline{2},3}, C^{(1)}_{\overline{1},2,3}, C^{(1)}_{1,2,3}$;
  //Allocations on l_1
4 : $r_1 \leftarrow \lceil \frac{d}{\phi_1 \times f_1} \times S \rceil$;
5 : $a_1 \leftarrow \min(r_1, A_1^{(1)})$;
6 : $d_1 \leftarrow \min\left(d, a_1 \times \frac{\phi_1 \times f_1}{S}\right)$;
7 : $p_1 \leftarrow a_1 / A_1^{(1)}$;
8 : $A_2^{(2)} \leftarrow A_2^{(1)} - p_1.(C^{(1)}_{1,2,\overline{3}} + C^{(1)}_{1,2,3})$;
9 : $A_3^{(2)} \leftarrow A_3^{(1)} - p_1.(C^{(1)}_{1,\overline{2},3} + C^{(1)}_{1,2,3})$;
  //Allocations on l_2
10 : $r_2 \leftarrow \lceil \frac{d_1}{\phi_2 \times f_2} \times S \rceil$;
11 : $a_2 \leftarrow \min(r_2, A_2^{(2)})$;
12 : $d_2 \leftarrow \min\left(d_1, a_2 \times \frac{\phi_2 \times f_2}{S}\right)$;
13 : $p_2 \leftarrow a_2 / A_2^{(2)}$;
14 : $A_3^{(3)} \leftarrow A_3^{(2)} - p_2.[C^{(1)}_{\overline{1},2,3} + (1-p_1).C^{(1)}_{1,2,3}]$;
  //Allocations on l_3
15 : $r_3 \leftarrow \lceil \frac{d_2}{\phi_3 \times f_3} \times S \rceil$;
16 : $a_3 \leftarrow \min(r_3, A_3^{(3)})$;
17 : $d_3 \leftarrow \min\left(d_2, a_3 \times \frac{\phi_3 \times f_3}{S}\right)$;

removed from every subset that characterizes slots initially available for reservation on l_1, $E^{(1)}_{1,2,3}$ taken as an example. This proportion is represented by plain areas. The selected slots become unavailable on all three links and thus are transferred to the set of unavailable slots for those links. The same mechanism is then repeated for the allocations carried out on l_2. This time, the slots are selected at random among those remaining available for reservations. The proportion of slots selected, p_2, represented by dotted areas, is thus taken in every subset characterizing slots remaining available for allocation on l_2, such as $E^{(2)}_{1,2,3}$. Finally, the same mechanism is in place for allocations on link l_3, as depicted by the striped areas.

This approach still applies when increasing the path length. However, the impact of n_1 allocations is still required to be evaluated when computing the average number of slots that would be allocated on any further communication link l_i. Such an impact can be evaluated by first doing the i-link available slot set decomposition and then carefully measuring the dependence of each of the previous node allocations. This leads to an exponential number of sets to deal with which makes the solution not tractable. We refer to this phenomenon as the domino effect.

To address this issue, in the following we introduce the interference-clique sliding approach that breaks the domino effect and reduces the calculation complexity.

3.2.3.3 Interference-Clique sliding approach

The interference-clique [1] sliding approach breaks the domino effect by processing the end-to-end bandwidth estimation clique by clique while using only a polynomial number of variables. The basic idea consists of eliminating the dependence on the allocations that took place in the previous nodes. We define an interference-clique, or simply clique, as any set of three consecutive links on the path. For instance, a 4-hop path is composed of two cliques: $c_1 = \{l_1, l_2, l_3\}$ and $c_2 = \{l_2, l_3, l_4\}$ [2].

Initialization: Given a path of length N_H, we start by computing the available slot sets resulting from the 3-link available slot set decomposition of every clique. This leads to eight corresponding sets for each clique, that is, for the i^{th}: $E_{\overline{i},\overline{i+1},\overline{i+2}}$, $E_{i,\overline{i+1},\overline{i+2}}$, $E_{\overline{i},i+1,\overline{i+2}}$, $E_{\overline{i},\overline{i+1},i+2}$, $E_{i,i+1,\overline{i+2}}$, $E_{i,\overline{i+1},i+2}$, $E_{\overline{i},i+1,i+2}$ and $E_{i,i+1,i+2}$. The following describes how to extend the bandwidth estimation process when sequentially passing the calculation on to the next cliques.

Interference-Clique 1: The clique 1 is the easiest to process as it does not depend on any previous allocations. As described above, exactly $a_1 = \min(r_1, A_1^{(1)})$ slots are reserved for communication on link l_1. Then, the slots remaining available for communication on l_1 are not considered anymore and the calculation is passed on to clique 2.

Interference-Clique 2: To process any clique i, we calculate a_i by first estimating $A_i^{(i)}$, the average number of slots remaining available for reservation on link l_i just before n_i does its allocations. Given the 3-link available slot set decomposition of clique i, we get:

$$A_i^{(i)} = C^{(i)}_{i,\overline{i+1},\overline{i+2}} + C^{(i)}_{i,i+1,\overline{i+2}} + C^{(i)}_{i,\overline{i+1},i+2} + C^{(i)}_{i,i+1,i+2} \tag{3.10}$$

Indeed, all the resulting sets of the decomposition are disjointed and form a partition of the global slot set $\{1, ..., S.B\}$. Then, to correctly measure the impact caused on clique 2 sets by reservations done on l_1, we just extend the 3-link available slot set decomposition related to clique 2 to the 4-link decomposition including l_1. This way, we note that:

$$C^{(1)}_{2,\overline{3},\overline{4}} = \underbrace{C^{(1)}_{1,2,\overline{3},\overline{4}}}_{\substack{\text{impacted by} \\ \text{allocations on } l_1}} + \underbrace{C^{(1)}_{\overline{1},2,\overline{3},\overline{4}}}_{\text{not impacted}} \tag{3.11}$$

From this equation, we infer that an average proportion $p_1 = a_1/A_1^{(1)}$ of slots in $E^{(1)}_{1,2,\overline{3},\overline{4}}$ is likely to become unavailable for reservation on link l_2 after node n_1 performs its allocations for

1. For the rest of the document, we will use the terms clique and interference-clique interchangeably.
2. The index of a clique is equal to the index of the first link in the clique.

communication on link l_1. Using this principle and considering that there is no interference between l_1 and l_4, the clique 2 sets can be updated as follows:

$$\begin{aligned}
C^{(2)}_{2,\overline{3},\overline{4}} &= C^{(1)}_{2,\overline{3},\overline{4}} - p_1.C^{(1)}_{1,2,\overline{3},\overline{4}} \\
C^{(2)}_{2,\overline{3},4} &= C^{(1)}_{2,\overline{3},4} - p_1.C^{(1)}_{1,2,\overline{3},4} \\
C^{(2)}_{\overline{2},3,\overline{4}} &= C^{(1)}_{\overline{2},3,\overline{4}} - p_1.C^{(1)}_{1,\overline{2},3,\overline{4}} \\
C^{(2)}_{\overline{2},3,4} &= C^{(1)}_{\overline{2},3,4} - p_1.C^{(1)}_{1,\overline{2},3,4} \\
C^{(2)}_{2,3,\overline{4}} &= C^{(1)}_{2,3,\overline{4}} - p_1.C^{(1)}_{1,2,3,\overline{4}} \\
C^{(2)}_{2,3,4} &= C^{(1)}_{2,3,4} - p_1.C^{(1)}_{1,2,3,4} \\
C^{(2)}_{\overline{2},\overline{3},4} &= C^{(1)}_{\overline{2},\overline{3},4} + p_1.\left(C^{(1)}_{1,2,\overline{3},4} + C^{(1)}_{1,\overline{2},3,4} + C^{(1)}_{1,2,3,4}\right) \\
C^{(2)}_{\overline{2},\overline{3},\overline{4}} &= C^{(1)}_{\overline{2},\overline{3},\overline{4}} + p_1.\left(C^{(1)}_{1,2,\overline{3},\overline{4}} + C^{(1)}_{1,\overline{2},3,\overline{4}} + C^{(1)}_{1,2,3,\overline{4}}\right)
\end{aligned} \tag{3.12}$$

As depicted in the previous equations, some sets receive new slots. This phenomenon results from slot allocations on l_1 having a different impact on slots initially available for reservation on links l_2, l_3 and l_4. Indeed, due to the 2-hop nature of the interference, a proportion of slots that were initially available in common for l_1, l_2, l_3 and l_4 have become unavailable to l_2 and l_3 and thus become exclusively available to l_4. At this point, it is possible to correctly calculate the average values of $A^{(2)}_2$, r_2, a_2, d_2 and $p_2 = a_2/A^{(2)}_2$.

Interference-Clique 3: Exactly the same interference phenomenon occurs when processing the third clique. However, as this clique suffers from interference created by allocations on both previous links l_1 and l_2, the same approach needs to be followed by extending the available slot set decomposition including these two links. It is even more complex than that since (1) l_1 interferes only with l_3, not l_4 and (2) l_3 suffers from interferences created by allocations on both l_1 and l_2 as:

$$C^{(1)}_{3,4,5} = \underbrace{C^{(1)}_{1,2,3,4,5}}_{\substack{\text{impacted by allocations} \\ \text{on } l_1 \text{ and then } l_2}} + \underbrace{C^{(1)}_{1,\overline{2},3,4,5}}_{\substack{\text{impacted by allocations} \\ \text{on } l_1 \text{ but not } l_2}} + \underbrace{C^{(1)}_{\overline{1},2,3,4,5}}_{\substack{\text{impacted by allocations} \\ \text{on } l_2 \text{ but not } l_1}} + \underbrace{C^{(1)}_{\overline{1},\overline{2},3,4,5}}_{\substack{\text{not impacted}}} \tag{3.13}$$

that leads to:

$$C^{(3)}_{3,4,5} = C^{(1)}_{3,4,5} - \left[p_1 + p_2(1-p_1)\right].C^{(1)}_{1,2,3,4,5} - p_1.C^{(1)}_{1,\overline{2},3,4,5} - p_2.C^{(1)}_{\overline{1},2,3,4,5} \tag{3.14}$$

As described in Appendix A.1 when processing the i^{th} clique, the influence of allocations on the two previous links can be correctly considered by updating the sets resulting from its 3-link available slot set decomposition using the following equations:

$$\mathbf{C}_i^{(i)} = \mathbf{C}_i^{(1)} - \mathbf{p}_i \times \mathbf{I}_i + \mathbf{u}_i + \mathbf{v}_i \tag{3.15}$$

where

$$\mathbf{C}_i^{(j)} = \begin{pmatrix} C^{(j)}_{\overline{i},\overline{i+1},\overline{i+2}} & C^{(j)}_{i,\overline{i+1},\overline{i+2}} & \cdots & C^{(j)}_{i,i+1,i+2} \end{pmatrix}$$
$$\mathbf{p}_i = \begin{pmatrix} p_{i-2} & p_{i-1} & [p_{i-2} + p_{i-1}.(1-p_{i-2})] \end{pmatrix}$$

and

$$\mathbf{I}_i = \begin{pmatrix} C^{(i-2)}_{i-2,\overline{i-1},\overline{i},\overline{i+1},\overline{i+2}} & \cdots & C^{(i-2)}_{i-2,\overline{i-1},i,i+1,i+2} \\ C^{(i-2)}_{\overline{i-2},i-1,\overline{i},\overline{i+1},\overline{i+2}} & \cdots & C^{(i-2)}_{\overline{i-2},i-1,i,i+1,i+2} \\ C^{(i-2)}_{i-2,i-1,\overline{i},\overline{i+1},\overline{i+2}} & \cdots & C^{(i-2)}_{i-2,i-1,i,i+1,i+2} \end{pmatrix} \tag{3.16}$$

Once the clique sets are updated, the average values $A_i^{(i)}$, r_i, a_i, d_i and $p_i = a_i/A_i^{(i)}$ can be correctly evaluated and the calculation process can be passed on to the following clique.

Interference-Clique 4 and beyond: When processing the third clique, the entries of matrix $\mathbf{I}_i$ were strictly referring to sets that had not varied from the beginning of the estimation process. However, that is not the case when processing the fourth clique. Indeed, the corresponding sets are likely to have been impacted by allocations on previous links. Such a set, for instance $E^{(2)}_{2,\overline{3},4,5,6}$, has suffered from allocations on l_1 and thus needs also to be updated. A straight solution would consist in forming the sets resulting from the 6-link available slot set decomposition and identify the way every set is impacted. This method is correct but leads to the previously mentioned domino effect. Fortunately, the random nature of the slot allocation can simplify the analysis and bound the number of variables to deal with for each clique process. It consists in applying the approximation scheme detailed in Appendix A.2 for estimating the terms in matrix $\mathbf{I}_i$.

Then, no additional techniques are required to process the remaining cliques and the calculation can be completed by simply applying the approach given in algorithm 2 when processing the last clique of the path. The main advantage of this approach is that there is no domino effect and the resulting calculation complexity is $\mathcal{O}(N_H)$.

3.3 Distributed implementation of BRAND

The centralized version of BRAND can be easily used with source routing protocols, like DSR [100], where the path computation is centralized at the source. However, this is not the case for non-source routing protocols, like the popular OLSR [101], where the route computation is performed at every node in distributed fashion. To address this limitation we present a simple mechanism that enables the distributed execution of BRAND.

BRAND's interference-clique sliding approach (Section 3.2.3.3) is what makes it amenable

Algorithm 3: Distributed BRAND

```
//The pseudo-code is for the non-trivial case of paths longer than two hops. For
  paths up to two hops the source node has all the information necessary to compute
  BRAND.
Input  : Bandwidth Request Packet BP
1 : begin
2 :     if (I am the source node) then
3 :         Initialize (BP);
4 :         Transmit (BP, next_hop);
5 :         Return;
        //Look up the routing table
6 :     next_hop ← get_next_hop(BP.destID);
7 :     if (get_hopCount (BP) == 1) then
8 :         Transmit (BP, next_hop);
9 :     else
10 :        Compute_Bandwidth_Clique(BP, avail_BW);
11 :        if (next_hop == BP.destID) then
12 :            Transmit (avail_BW, BP.sourceID);
13 :        else
14 :            Update (BP);
15 :            Transmit (BP, next_hop);
```

to a distributed computation, for two reasons. The computation at every step requires only knowledge of the interference clique at hand and the three reduction factors, α, resulting from the computation on the previous clique (see Appendix A.2). Second, the *sliding* is strictly linear, needing a single pass from the source to the destination.

A straightforward solution would work as follows. The first node on the path would execute the computation for the first clique, comprising links l_2, l_3, l_4. Once the computation was done, the node would insert the results of the computation and the reduction factors in a control packet, call it the bandwidth packet (BP), and would transmit it to the next hop on the path asking it to perform the same for the second clique, comprising links l_2, l_3, l_4. The process would be repeated till the destination, which would complete the *sliding* and send the results back to the source node. Assuming the BP packets were all received, the result would be identical to centralized BRAND.

The problem with this solution is that, with non-source routing, a particular node does not know which direction a given packet is going beyond the next hop. That means the source node only knows the next hop for a given path, when doing the computation for the first clique of that path requires knowledge of the next three hops. To overcome this challenge, we propose a simple trick: Simply shift the interference-clique calculation by two hops down the path. As illustrated in Figure 3.7. The BRAND computation is executed clique by clique – first clique being l_1, l_2, l_3, second l_2, l_3, l_4, and so forth. In non-source routing, in addition to source/destination, a node only knows the previous and next hop for a given flow. Thus, n_1 despite knowing the two-hop

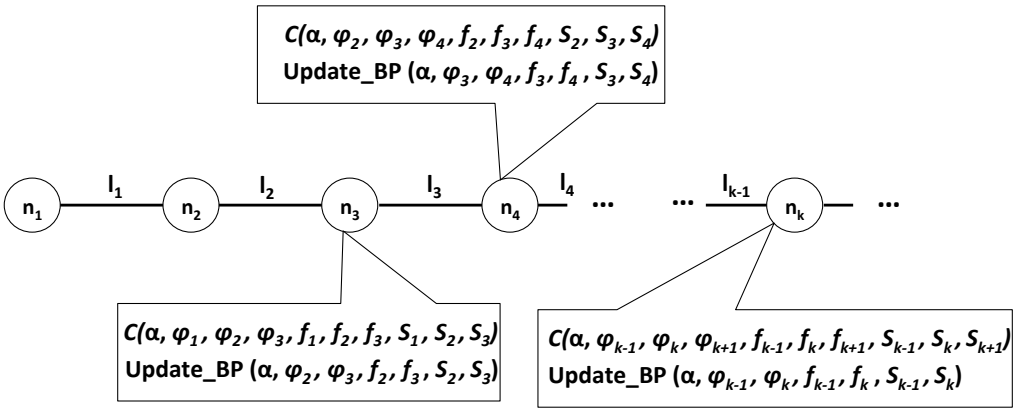


Figure 3.7 – Distributed implementation of BRAND

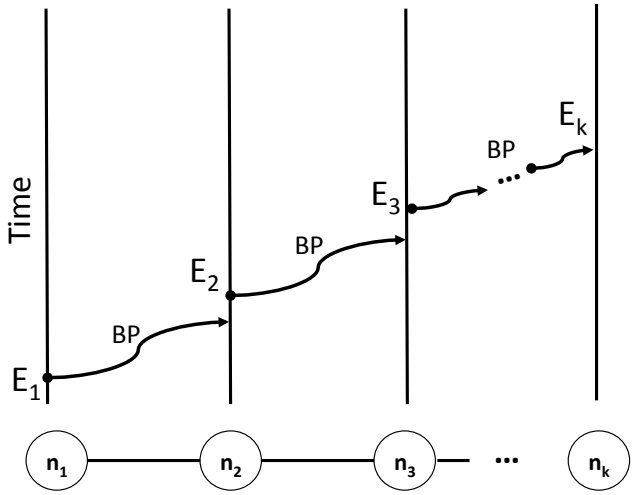


Figure 3.8 – Distributed BRAND and the Clock Condition

neighborhood does not know which one will be n_2, and n_2 does not know which one will be n_3. However, node n_3 knows the source, n_1, the previous hop, n_2, and next hop, n_3, which allow it to bootstrap the BRAND computation for the first clique [1]. Once done, it passes the necessary information using the BP packet to n_4 for performing the computation for the second clique. This is repeated for the rest of the path. Algorithm 3 gives the specification of the distributed algorithm in pseudo-code for the case of non-trivial (longer than two hops) paths.

Finally, in the following we show the correctness of distributed BRAND.

Lemma 1. *Assuming the bandwidth packet (BP) is eventually received by all nodes along the path and the communications follow the protocol model [102], distributed BRAND results in the*

1. A simple heart-beat protocol can give every node the quantities ϕ, f, p, S for all links in the two-hop neighborhood.

same computation as centralized BRAND.

Proof. For distributed BRAND to give the same result as centralized BRAND it suffices to show that clique computations take place in order. That is, computation in clique $i-1$ always happens-before computation in clique i. That is equivalent to showing that if $C(E_k)$ is the time [1] at which Algorithm 3 is executed at node n_k along the path, $C(E_{k-1}) < C(E_k)\ \forall k$ [103]. This is straightforward to see (illustrated in Figure 3.8) given that Algorithm 3 is executed only upon reception of a BP packet, and based on the protocol model assumption a BP packet for a given flow on a given path can be received by node k only once transmitted by node $k-1$. The events of computing BRAND for a given flow on all nodes along the path, E_i, satisfy the Clock Condition [103]. □

3.4 Performance evaluation of Brand

In this section, we evaluate BRAND numerically using MATLAB and compare it with the work of Zhu et al. [77] and Kodialam et al [68] introduced in Section 2.4.2.2. In summary, we make the following main observations: In Section 3.4.4, we demonstrate that although BRAND uses a very simple scheduling heuristic –randomized scheduling – its achieved end-to-end bandwidth is competitive when compared with the optimal policy and a centralized scheduling algorithm. In Section 4.5.4, we demonstrate the correctness of BRAND's algorithm for computing the average throughput of randomized scheduling. In Section 3.4.6, we demonstrate that BRAND delivers the bandwidth it promises in achieving an almost perfect admission control. In Section 3.4.7, we demonstrate that BRAND deals successfully with the challenges arising from the cognitive radio network architecture. Specifically, BRAND maintains an almost perfect admission control in face of Primary User interference and multi-rate links. In Section 3.4.8, we demonstrate that distributed BRAND deals successfully with the challenges arising from node mobility and wireless channel fading.

3.4.1 Simulation Parameters

Each link is assigned one orthogonal channel among four sensed ones. The medium is accessed through a TDMA MAC with 40 slots per frame. As the specific spectrum assignment process is beyond the scope of this work, we simply use the following probabilistic model to select the assigned channel for each communication link: $\mathrm{P}[channel1] = 0.80$, $\mathrm{P}[channel2] = 0.10$, $\mathrm{P}[channel3] = 0.05$ and $\mathrm{P}[channel4] = 0.05$. The probabilities are purposely selected to achieve a high self-interference on the path and thus, to show the worst-case performance of the algorithm. The scheduling algorithm in the simulation is the same with the one used by BRAND, i.e., the necessary allocations of slots are performed at random among those available.

1. We are deliberately being imprecise regarding the definition of time because the execution of Algorithm 3 is event-driven. We could use the physical time, as observed by a single entity, or use Lamport's logical clock.

In practice, link rates depend on the local environment and may fluctuate with time. To approximate this behavior, we sample every link rate ϕ_i according to a normal distribution with mean μ_ϕ and standard deviation σ_ϕ where μ_ϕ is the mean link transmission rate on the corresponding assigned channel and σ_ϕ is taken proportional to μ_ϕ. In all the simulation results presented here $\sigma_\phi = \mu_\phi \times 0.10$. For the channels considered, we choose: $\mu_\phi(channel1) = 2$Mbps, $\mu_\phi(channel2) = 1.5$Mbps, $\mu_\phi(channel3) = 800$kbps and $\mu_\phi(channel4) = 250$kbps.

As described in Section 3.1.3, there are two sources of interference in a cognitive radio networks: Secondary-to-secondary and primary-to-secondary. To model the interference from other secondary sessions we use a probability, p_a, for denoting the chances of a given slot being free of other secondary communications, that is, is not reserved. The primary interference is modeled by the probability u introduced in Section 3.1.3.2.

3.4.2 BRAND parameters

BRAND calculates the available end-to-end bandwidth by computing the average end-to-end throughput realized over all possible demands and returning the highest value. In this evaluation, the demand values are taken in the range $I_d = [0, \min(\phi_1, \phi_2, ..., \phi_{N_H})]$ (kbps) with step $\Delta_\phi = 10$kbps.

3.4.3 Basis for comparison

To the best of our knowledge, there is no other work that tackles the problem of computing the available end-to-end bandwidth for a cognitive radio network. The closest to our work is that by Zhu et al [77], which computes the available bandwidth for legacy TDMA multi-hop network. We use such heuristic as basis for parts of our evaluation while taking into account the fact that it is not designed for cognitive radio networks. Furthermore, we use the algorithm proposed by Kodialam et al [68] and the optimal solution obtained by solving an integer linear program with *lpsolve* [104] for the case of 4-hop paths, provided in Appendix A.3.

As evident by the problem formulation, the integer linear program gets highly complicated for high number of hops so we were able to compute optimal values for paths of up to four hops. Therefore, for a longer path, an upper bound of the residual capacity can be obtained by simply taking the minimum of the computed capacity on every portion of four consequent links along the path.

3.4.4 End-to-End Bandwidth with BRAND

We perform the following experiment for comparing BRAND to Kodialam and the optimal. We choose the values of no secondary interference, p_a, for every link uniformly at random in $(0, 1)$ and generate PSAT tables for various path sizes. We run BRAND and Kodialam in MATLAB using these PSAT tables and compute the available end-to-end bandwidth for different values of primary user interference.

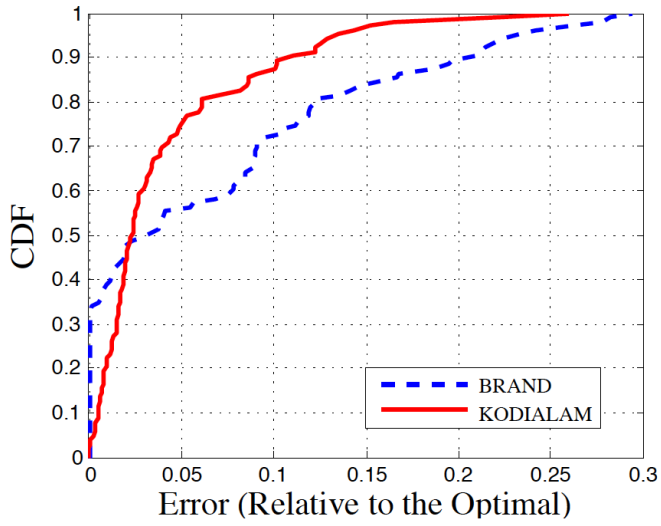
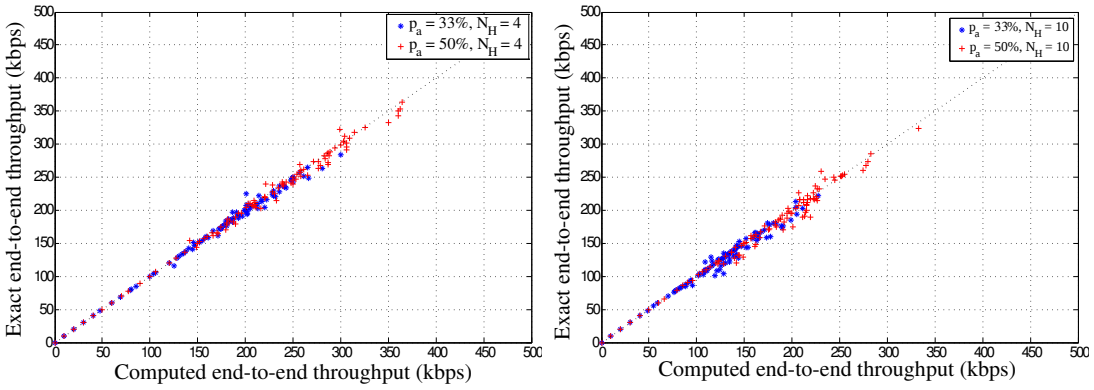


Figure 3.9 – BRAND vs Kodialam

In Figure 3.9, the available end-to-end bandwidth values are computed by BRAND and Kodialam and compared to the optimal values, obtained by using *lpsolve* to solve the integer linear program in Appendix A.3. Despite using a simple scheduling heuristic, BRAND's computed values are very competitive when compared to Kodialam, an algorithm introduced as a benchmarking tool, as well as the optimal scheduler.

3.4.5 End-to-End Throughput with Randomized Scheduling

The most novel and challenging part of BRAND is its algorithm for computing the average throughput of randomized scheduling, introduced in Section 3.2.3. Given the involved analysis of the algorithm, here we perform a simple experiment for verifying its correctness. For a specific value of hop-count and p_a we generate a PSAT table using MATLAB. The algorithm is applied using the PSAT table and the average end-to-end throughput is computed for all demands described in Section 3.4.1. With the same PSAT and traffic demands, we run simulations in MATLAB in which the slots are selected at random on every hop. The simulation is run multiple times and the seed for the random generator is changed every time. The throughput values measured at the end of each simulation are averaged over all runs. In Figure 3.10, for every demand, the computed and the measured throughput values for $p_a = 33\%$ and $p_a = 50\%$ are plotted on the x and y-axis, respectively. For these graphs, the values are averages over 5 runs. The probability of PU interference is set to 10%. The data shown are for path lengths of 4 hops, a reasonably sized multi-hop networks, and 10 hops, a large network. After 5 runs the measurements are already converging to the computed values.



(a) Numerical verification for a 4-hop path.

(b) Numerical verification for a 10-hop path.

Figure 3.10 – BRAND accuracy

3.4.6 Admission Control Performance

BRAND is designed for enabling admission control. As such, we expect any flow with demand less than or equal to the available bandwidth computed by BRAND to be admitted end-to-end. To verify that this is the case, we perform the following two-step experiment. In the first step, we apply BRAND on a 4-hop path and compute the available end-to-end bandwidth for various values of p_a. Specifically, we perform the computation on PSAT tables generated by taking the probabilities of no secondary interference, p_a, in $(0, 1)$ while the primary user interference is set to 10%. One PSAT table is generated per value of p_a and one bandwidth value per PSAT is computed by BRAND.

The available bandwidth values computed by BRAND in the first step are used as input in the second step of the experiment. Specifically, for every value of p_a and respective PSAT used in the first step, we run a simulation during which a single session with traffic demand equal to the available bandwidth computed by BRAND for this value of p_a is initiated end-to-end. We measure the end-to-end throughput realized during the simulation and plot it as function of the computed demand. To provide more data about the behavior of BRAND we repeat the same experiment for a 10-hop path, which is as long a path as one can be expected to encounter in deployed multi-hop cognitive radio networks. In Figure 3.11, the x-axis represents the available end-to-end bandwidth computed by BRAND for various probabilities of no secondary interference. The y-axis represents the end-to-end throughput measured in simulations – utilizing the same PSAT as the computation – when the traffic demand is equal to the respective end-to-end bandwidth computed by BRAND. The data shown here is for paths of four ($N_H = 4$) and ten hops ($N_H = 10$). As shown in this figure, the measured throughput is practically identical to the computed values of the available bandwidth. This demonstrates that BRAND delivers the bandwidth it promises and provides nearly 100% admission.

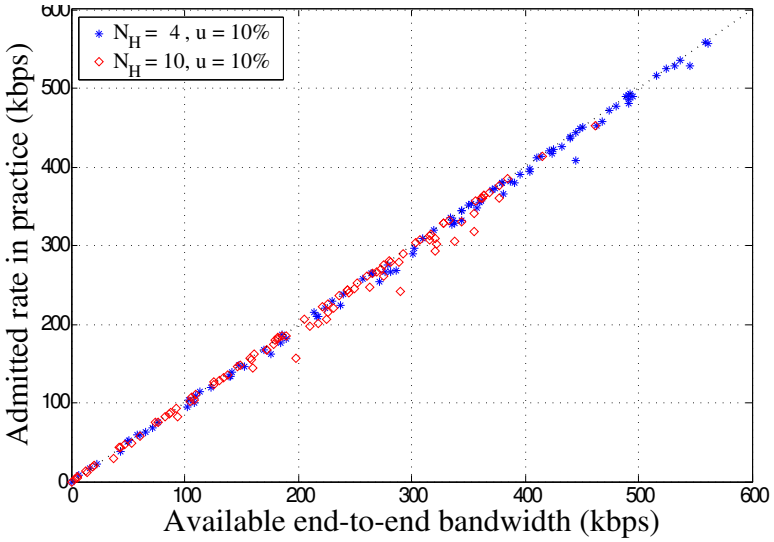
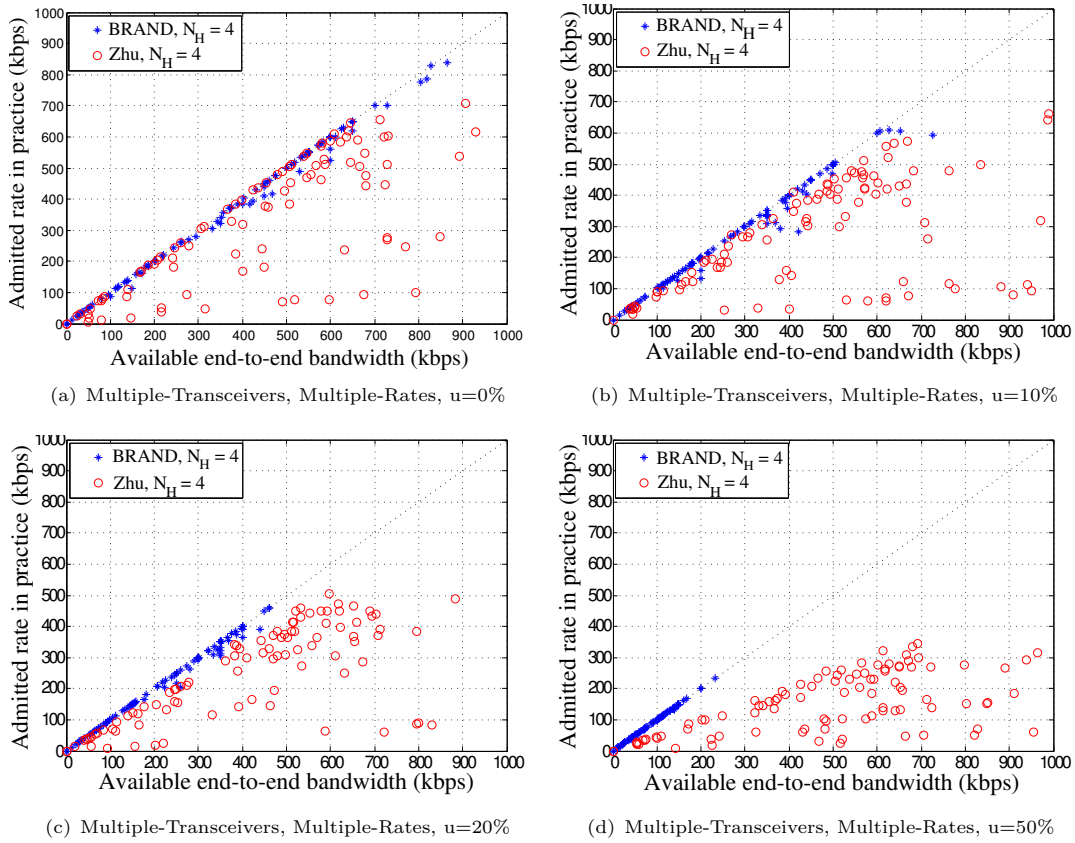


Figure 3.11 – BRAND admission performance

3.4.7 Cognitive Effect: Primary Users and Multi-Rate Links

We now evaluate the performance of BRAND for a cognitive network architecture in which nodes access the spectrum as secondary users. In addition to BRAND, we also use the Zhu heuristic in this part of the evaluation. The latter was designed for a single-rate, single-transceiver legacy architecture so clearly it would be unfair to expect it to perform as well as BRAND. Instead, the reason for which we include it in this evaluation is for quantifying the consequences of ignoring the primary user and multi-rate links when computing the available bandwidth.

To evaluate the "cognitive" effect, we repeat the experiment of the previous section using multi-transceivers, multiple-rates and varying probabilities of the Primary User occurrence. As shown in Figure 3.12 BRAND almost always estimates the correct value for the available capacity. On the other hand, ignoring the operating specifics of the cognitive radio networks leads to significant errors, as high as 900%, when calculating the available end-to-end bandwidth. The main reason for the Zhu heuristic underperforming in this experiment is that it allocates a constant number of slots end-to-end, assuming that all links on the path have the same rate as the first one. The assumption can lead it to miscalculate the path's available bandwidth and, most importantly, makes it very hard for adapting the heuristic to take multi-rate links into account. Finally, a careful analysis of the data, especially that of Figure 3.12(d), shows that, at times, Zhu realizes a higher available bandwidth in simulations. As predicted by Equation 3.1 in Section 3.1.3.2, protecting the Primary User comes at a cost in terms of bandwidth. This cost, which increases with the probability of Primary User occurrence, u, explains the smaller bandwidth values computed by BRAND when compared to a heuristic that completely ignores the Primary User.



(a) Multiple-Transceivers, Multiple-Rates, u=0%

(b) Multiple-Transceivers, Multiple-Rates, u=10%

(c) Multiple-Transceivers, Multiple-Rates, u=20%

(d) Multiple-Transceivers, Multiple-Rates, u=50%

Figure 3.12 – Effect of Primary Users and Multi-rate links

In conclusion, the data from this and the previous experiments shows the importance of designing a heuristic for computing the available bandwidth around the particularities of the cognitive radio.

3.4.8 Distributed BRAND: Mobility, Channel Fading and Multiple Flows

Finally, we consider a more realistic scenario that includes mobility and channel fading for evaluating the performance of distributed BRAND. We set up a multi-hop cognitive radio network of 30 nodes placed at random on a 1000mx1000m area, as illustrated in Figure 3.14. To make the simulation more closely model a real-life situation we do not start the experiment on an uninitialized network but instead on every node we set the status of a certain percentage of slots, selected at random between 0-100%, as allocated. The four available channels (red, blue, green, magenta) are randomly distributed. Then we select at random four source-destination pairs and start generating unicast sessions. Whenever a new traffic session arrives we use distributed

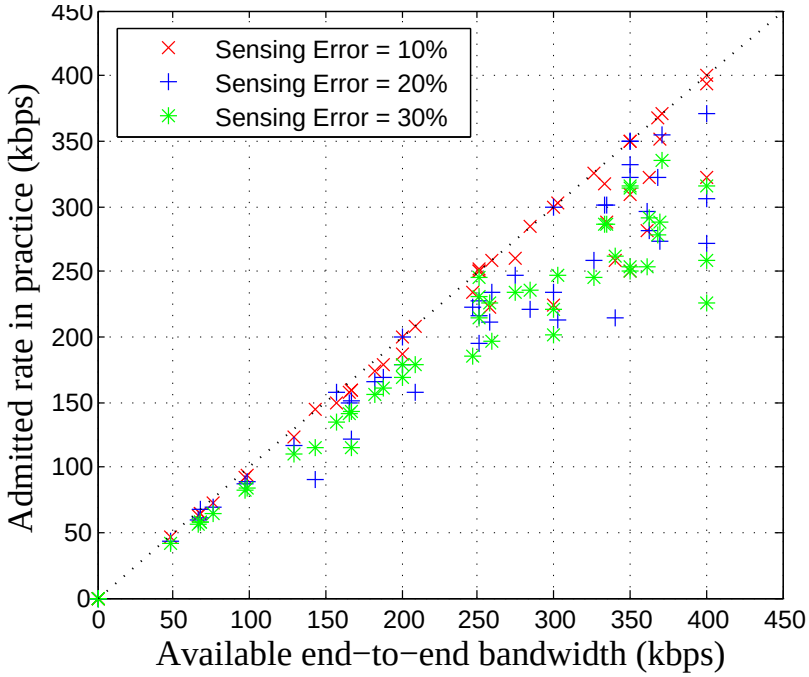


Figure 3.13 – BRAND is resilient to sensing errors

implementations of BRAND and Zhu for performing admission control. We carry out two sets of experiments, one with no mobility for isolating the effect of channel fading and a second where both fading and node mobility are present. The routing is simply performed according to the hop-count metric.

Figure 3.13 shows the data for static topologies. Every point is an average among all four sessions in the network. As the data shows, distributed BRAND, which takes into account fading, performs very well. Figure 3.15 shows the data for mobile scenario. For this set of experiments we use the random waypoint mobility model and consider two levels of maximum speed, 5m/s and 15 m/s, corresponding to pedestrian and vehicle movement, respectively. As the data in Figure 3.15(a) shows, both schemes are affected by mobility, in particular at high speed. Nevertheless, distributed BRAND is shown to be more resilient thanks to the simplicity of the randomized scheduling. For the static scenario channel fading is the only source of errors and BRAND does a good job at estimating the end-to-end bandwidth. Mobility is not modeled by BRAND but it shows moderate effect at pedestrian speed. As is to be expected, at vehicular speed the errors start becoming significant. A detailed look at the data for BRAND in Figure 3.15(b), shows that for the most part its estimate of the available capacity is good.

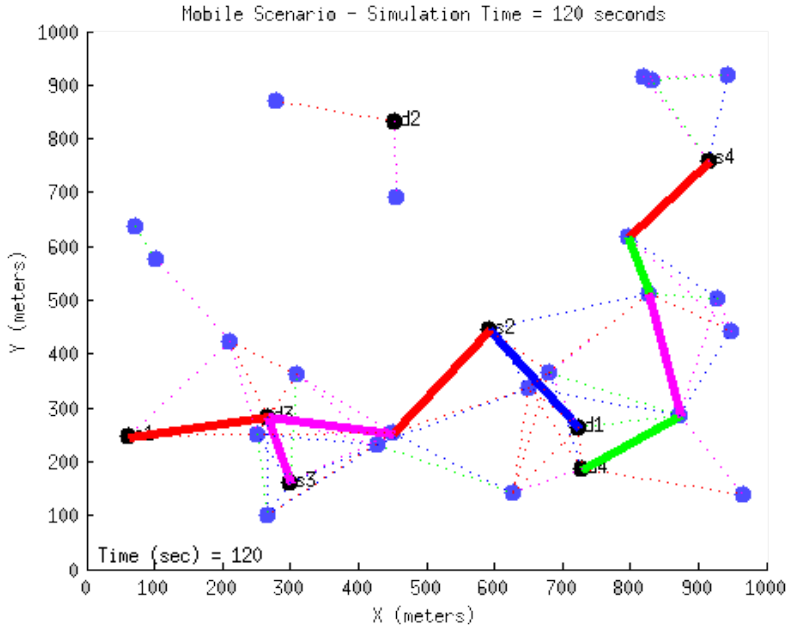
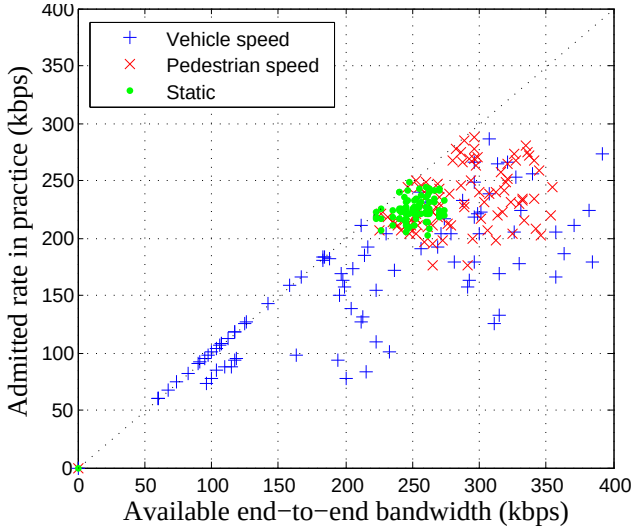


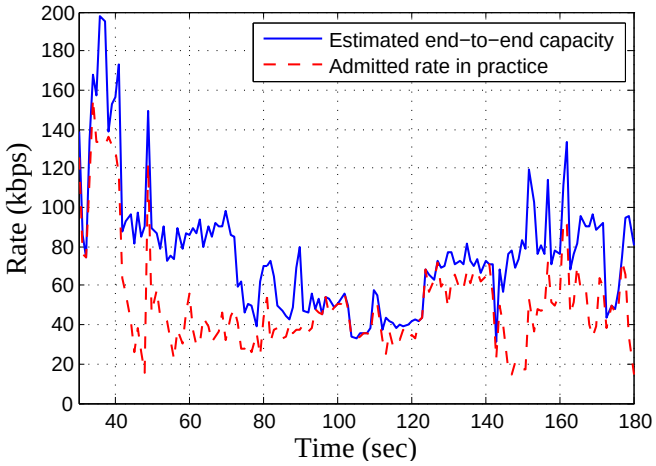
Figure 3.14 – Mobile scenario simulation

3.5 Conclusions

In this chapter, we have revisited the problem of admission control for the cognitive radio context. Our solution, BRAND, uses a polynomial time algorithm for estimating the available end-to-end throughput in TDMA-based multi-hop cognitive radio networks wherein each node is equipped with multiple transceivers. We have addressed the particular case of random slots selection at the MAC layer and provided an admission control scheme for end-to-end flows. Our method is based on the introduced *l-link available slot set decomposition* and the *interference-clique sliding approach*, an approximation scheme that reduces the calculation complexity from exponential to polynomial while still returning accurate and reliable results. We have shown that BRAND can run in a distributed fashion. This, coupled with our choice of computing the available bandwidth of a path, makes BRAND a feasible proposition for any routing approach. Using a thorough numerical analysis we have demonstrated the correctness of BRAND as well as its capability in accurately taking into account the cognitive radio context, in particular the Primary User and multi-rate links, and thus delivering almost 100% admission rate in a variety of conditions.



(a)



(b)

Figure 3.15 – Performance with mobility, channel fading and multiple concurrent flows.

4 Revisiting Transmission Count for cognitive radio networks

Transmission count that refers to the number of transmissions required for delivering a data packet over a link, is part of almost all state-of-the-art routing metrics for wireless networks. In traditional networks, peer-to-peer interference and channel errors are what define its value for the most part. In cognitive radio networks, however, there is a third culprit that can impact the transmission count: the primary user interference.

Motivated by this observation, in Section 4.2, we use a USRP N210 testbed to carry out an empirical study of the primary user impact on transmission count. We demonstrate that primary users have a distinct effect on transmission count over a wireless link, something that ETX [26], designed for traditional networks, fails to capture.

To resolve this, in Section 4.3, we present COExiST (for COgnitive radio EXpected transmISsion counT): a link metric that accurately captures the expected transmission count over a wireless link subject to primary user interference. Despite the involved computation, COExiST is shown to have a simple and elegant closed-form expression. We show that COExiST coupled with hop-by-hop Dijkstra-based routing satisfies the optimality, consistency and loop-freeness properties.

COExiST can be used as a stand-alone metric to compute paths with optimal transmission count, or be combined with other factors, e.g. the physical bit-rate, for optimizing other metrics, such as delay. To evaluate its performance, we implement in Section 4.4 a stand-alone version of COExiST as part of the OLSR routing protocol. We achive this purpose by building a prototype testbed using the USRP N210 radio platform and IRIS [105]. Then, we add our own CSMA/CA implementation and an implementation of COExiST for OLSR.

In Section 4.5, extensive experiments on this five-node USRP testbed demonstrate that COExiST accurately captures the actual transmission count in the presence of primary users and leads OLSR into selecting higher throughput paths.

4.1 Questioning the basic concepts

Despite the diversity of the contexts on which ETX has been applied so far, one thing has always been the same: the transmission count was mostly a function of channel errors and peer-to-peer [1] interference.

In cognitive radio networks, however, there is a third culprit that can impact the transmission count: primary user interference. Spectrum sensing [4, 42] and/or querying geo-location web services [52, 106] for identifying available spectrum are essential to cognitive networks, but they cannot guarantee zero interference from primary users [107]. It may be tempting to think of primary user interference as no different than interference caused by other peers. However, there are reasons to believe this may not be the case. For example, with 802.11, a successful packet transmission is always followed by a back-off, which can limit the damage a particular peer interferer can cause and shape how the pattern of losses is perceived by the transmitter. Obviously, primary users do not necessarily use 802.11.

Motivated by this observation, we carry out an empirical study on a five-node USRP testbed network. Our measurements show that when the interferer is a peer node, ETX is pretty adept at approximating the actual transmission count on a particular link. However, keeping everything the same and simply replacing the peer interferer with a primary user leads to significant gaps between ETX and the actual transmission count. Clearly, the primary user is having an adverse and distinct effect on the capability of two secondary nodes to communicate, which the traditional way of estimating transmission count is ill-equipped to cope with.

It has been clear from the early days of cognitive networks research that routing will need to be revisited[30], and several solutions have been proposed to this effect [39, 47, 62, 63, 82]. The proposed solutions run the gamut from complete system solutions, not backward compatible, to straightforward extensions of traditional solutions. None of these solutions is evaluated on real hardware. This is in part due to the fast-moving nature of cognitive networks research. Even fundamental things like hardware, policy, channel access protocols, relation between primary and secondary networks are not yet settled[51, 108]. Against this backdrop, we focus on addressing a problem essential to any wireless technology: estimating the link quality.

4.2 Primary user impact on transmission count

We present an empirical study on the impact of primary user interference on transmission count. Our measurements show that primary users present a distinct challenge when it comes to estimating the transmission count – something ETX, the *de facto* standard approach of estimating transmission count, fails to capture. We explore the reasons why, and provide pointers to potential solutions.

1. We use the terms peer and secondary user interchangeably.

Figure 4.1 – Experimental setup

4.2.1 Experimental setup

Hardware Our testbed consists of five USRP N210[109] software defined radios coupled with SBX daughterboards providing a 400-4400MHz frequency range. The SBX daughterboad is provided with two front-ends: one TX/RX used for secondary user communications, and one RX2 dedicated to spectrum sensing. Each USRP is connected to 64-bit host computer running the Ubuntu 12.04 LTS system (Fig. 4.1).

Software We use IRIS[105], an open source LGPLv3 software defined radio architecture. Unlike the GNURadio, Iris is designed specifically to support maximum reconfigurability while the radio is running, a capability that better fits our needs for a cognitive radio testbed. However, IRIS does not come with a MAC protocol implementation so we augmented its architecture to allow for carrier sensing and implemented a CSMA protocol similar to the 802.11 MAC. At the routing layer we use OLSR with an ETX implementation. The complete details of the software architecture can be found in Section 4.4.

Emulating primary users We model the primary user activity by transmitting packets at high power levels, thus interfering with SU communications. To control the ON/OFF period duration we vary the burst duration with typical continuous time distributions such as exponential or uniform[42]. Figure 4.2 shows a typical primary user behavior as utilized in our experiments.

4.2.2 Impact of primary users on transmission count

To study the potential impact of primary users, we set up a simple 2-node link – a third node acts as interferer and switches roles between being a peer and a primary user. UDP packets are

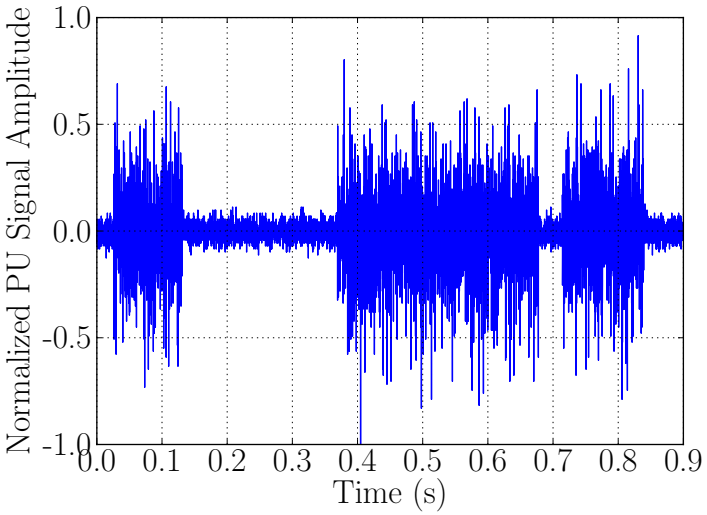


Figure 4.2 – Sampling of the primary user activity

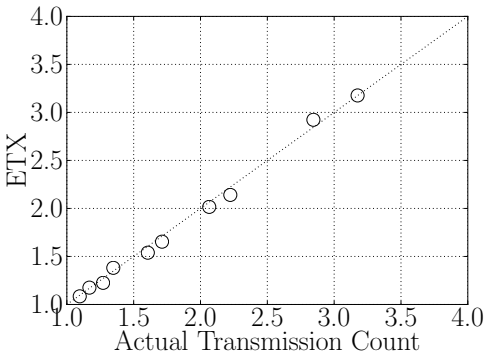
sent as fast as possible over the link and we collect the actual transmission count as well as ETX, as reported by the OLSR implementation. We repeat the experiment for a variety of PU levels, 0.2 to 0.7, and link probability of success when counting only channel errors, 0.5 to 1.

Figure 4.3(a) shows the ETX and actual transmission count when the interferer is a peer. Here, the peer is running the same CSMA protocol as the other two nodes and, thus, the transmission count is mostly due to channel errors, something ETX estimates fairly well. When we replace the peer interferer with a primary user things change: Figure 4.3(b) shows ETX performing poorly. This is due to the fact that ETX estimates the channel quality by sending (broadcast) packets on a regular basis and assumes the transmission failures are independent. However, when failures are due to primary user activity there is a high correlation between them, and the transmission count will highly depend on things like how active the primary users are, the pattern of their activity, etc.

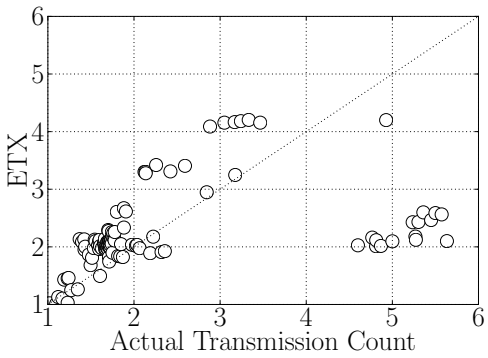
4.2.3 Capturing the impact of primary users

Having shown that ETX fails to capture the full impact of primary users on transmission count, we turn our attention to exploring alternative ways that will. Intuition says that, while the probe packets ETX uses may miss a good chunk of the primary user activity, the time a particular node spends to successfully transmit unicast packets should provide a good clue as to its activity. Besides, if a primary user is not affecting the transmission time, there is no need to adjust.

To verify our intuition, we perform the following experiment. We use the 3-node topology like before, with two USRP nodes functioning as the secondary network while the third as a



(a) **Interferer is a Peer.** There is a data point for every value of channel attenuation considered.



(b) **Interferer is a Primary User.** There is a data point for every combination of channel attenuation and primary user activity.

Figure 4.3 – Performance of ETX relative to PU/SU interference

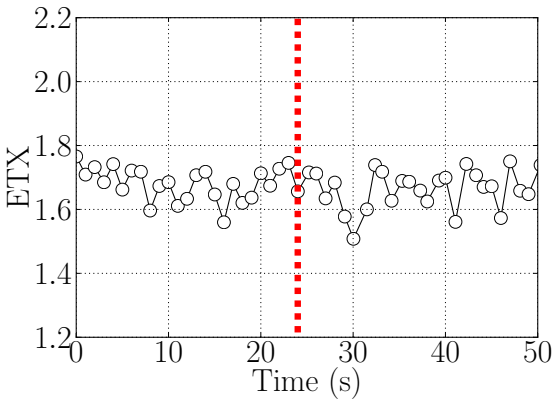
primary user. One of the secondary nodes transmits UDP packets to its peer as fast as possible for 50 seconds. The primary node is silent for the first 25 seconds and is activated for the last 25. During the experiment, we collect the ETX values computed by OLSR and at the MAC layer, the time between a successful transmission and the next attempt, T_t, and that between two retransmissions for unicast packets, T_r.

Figure 4.4 shows the observed values for ETX (Fig. 4.4(a)) and the normalized values of T_t, T_r (Fig. 4.4(a)), as function of the experiment time. ETX shows a poor correlation to PU activity. However,T_t and T_r, show a strong correlation to it.

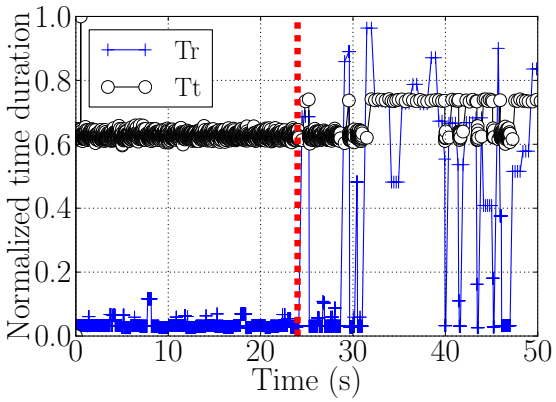
4.2.4 Would a straightforward solution work?

When presented to the challenge of estimating the impact of a primary user on transmission count, the straightforward solution that may cross one's mind is to treat the primary user as yet another probabilistic source of error, use history to estimate the ratio of time a primary user is active, and simply multiply ETX by this value. It is basically the strategy carried out in SAMER [82]. Figure 4.5 shows that this would not work. The data collected using the 3-node topology, with one node acting as a primary user, shows that for the same ratio of PU activity, different values of ON periods have a different impact on the packet delivery ratio. In particular, all points represent experiments with the same ratio of PU activity; what changes are the absolute values of the ON and OFF period durations. For the same ratio of ON/OFF durations, higher values of ON and OFF durations lead to lower packet reception ratios, and consequently, higher transmission count. Therefore, the straightforward solution of multiplying ETX by the primary user availability can lead to highly inaccurate estimates of transmission count.

The same conclusion can be drawn from the results presented in Figure 4.6. We set up an



(a) PU activity starts at time 25.



(b) PU activity starts at time 25. T_t and T_r are averages over 10 samples.

Figure 4.4 – Impact of primary users on T_r and T_t.

experiment with a fixed level of primary user activity ($u = 0.25$, $\bar{T}_{on} = 50ms$). At time 45s we multiplied the average duration of the ON periods by 5. We observe that ETX as well as the straightforward approach fail in capturing such a modification, while it has a real impact on the transmission count.

4.2.5 Summary

The above empirical study shows the following:

- Primary users present a distinct factor impacting the transmission count.
- The traditional way of computing the transmission count, ETX, is well adept at capturing peer interference and channel errors but fails when it comes to Primary User interference.

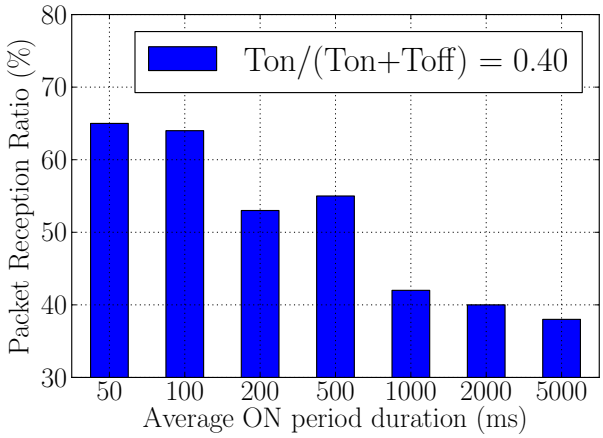


Figure 4.5 – Impact of $\bar{T}_{on}$ on the packet delivery ratio

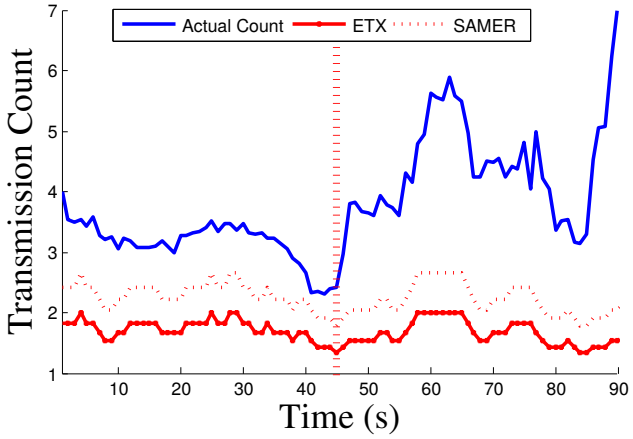


Figure 4.6 – Impact of $\bar{T}_{on}$ on the transmission count

- The straightforward solution of multiplying ETX by the primary user availability can perform poorly.
- A cross-layer approach of using MAC layer information can significantly improve our capability at capturing the impact of primary users.

4.3 COExiST: Revisiting the Expected Transmission Count

In this section, we present the design and computation of COExiST. As suggested by the empirical study in Section 4.2, to account for the impact of primary users on transmission count, COExiST utilizes the perceived primary user activity, the time between a successful

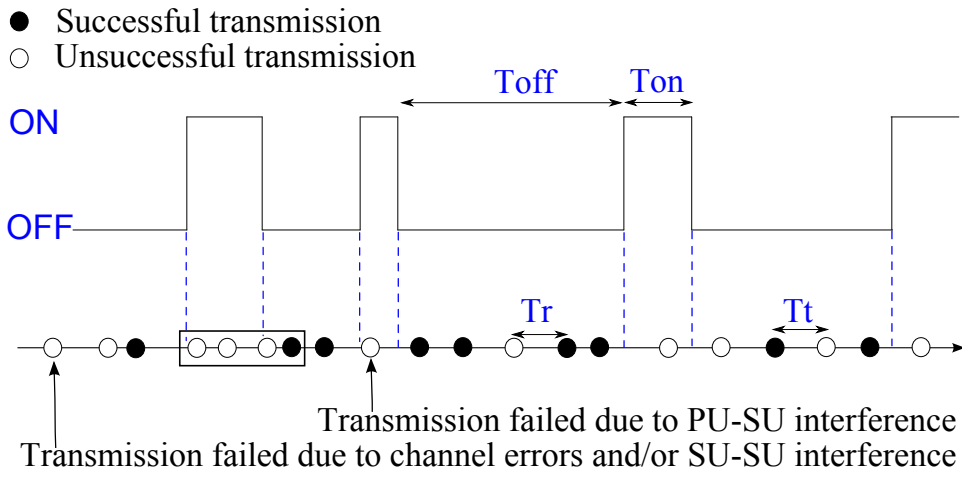


Figure 4.7 – PU and SU networks operating in parallel.

transmission and the next attempt, T_t, and the time between two retransmissions, T_r. COExiST is measurement driven – only run-time information as to primary user activity, the T_r, and the T_t is utilized for its computation. The COExiST of a path is the sum of the COExiST for each link on the path.

4.3.1 Model and preliminaries

We derive the COExiST of a given link according to the network model defined in Figure 4.7. The Primary and Secondary networks are not assumed synchronized, that is, the cognitive radios do not have perfect knowledge of the primary user activity pattern. Rather, the radios have basic information about the level of activity and the average activity period duration. A primary user is considered active if it interferes with communications performed on secondary user links.

For the rest of the analysis, we use u to denote the Primary User duty cycle [45] and $\bar{T}_{on}$, $\bar{T}_{off}$ to denote primary user ON/OFF period durations, respectively. These quantities are related by the formula $u = \bar{T}_{on}/(\bar{T}_{on} + \bar{T}_{off})$ [42, 44]. We model the distributions of the primary activity/non-activity periods using exponential distributions with parameters $\bar{T}_{on}$ and $\bar{T}_{off}$. In practice, such period durations are not always exponentially distributed and depend on the primary network characteristics. Nevertheless, as pointed out in [45], the exponential distribution is shown to suitably fit the empirical distributions observed for commercial systems, such as cellular networks. To keep the computation tractable, the same approximation is also used for T_r and T_t. COExiST, utilizes the probability of successfully transmitting a packet during the OFF period, p_s^{off}, to account for the effect of channel errors and SU-SU interference. Finally, our model assumes an unlimited number of transmission attempts at the MAC layer.

4.3.2 Computing COExiST using an absorbing discrete-time Markov chain

Let $N \in \mathbb{N}^*$ be the total number of MAC layer attempts required for successfully transmitting a packet over the link. COExiST estimates $\mathbb{E}[N]$ by resolving an absorbing discrete-time Markov Chain that takes into account $\bar{T}_r$ and $\bar{T}_t$ which represent the time correlation between consequent packet transmissions.

Definition 2 (Absorbing Discrete-time Markov chain for computing COExiST)**.** We model the Cognitive Radio MAC layer retransmission scheme with an absorbing discrete-time absorbing Markov chain wherein the states are defined as follows:

- I_0: The last layer-3 packet has been successfully transmitted during the OFF state in the Primary channel. The first transmission attempt of the current packet is pending during the OFF state for which the Primary channel is considered Idle. ***(initial state)***
- B_0: The last layer-3 packet has been successfully transmitted during the OFF state in the Primary channel. The first transmission attempt of the current packet is pending during the ON state for which the Primary channel is considered Busy. ***(transient state)***
- I_k, $k \in \mathbb{N}^*$: The packet has been transmitted k times without any success. The retransmission is pending during the OFF period for which the Primary channel is considered Idle. ***(transient state)***
- B_k, $k \in \mathbb{N}^*$: The packet has been transmitted k times without any success. The retransmission is pending during the ON period for which the Primary channel is considered Busy. ***(transient state)***
- R_k, $k \in \mathbb{N}$: The packet has been successfully transmitted with a total of k retransmissions ***(absorbing state)***

The corresponding Markov Chain, illustrated in Figure 4.8, has an infinite number of states. It converges probabilistically to one of the absorbing state, R_k, where k represents the number of retransmission attempts performed for successfully transmitting a particular packet over the link. The transition probabilities are defined $\forall k \in \mathbb{N}$ as follows:

- $p_{(I_k,R_k)}$ = Probability that the $(k{+}1)^{th}$ transmission attempt is successful **and** takes place before the end of the current OFF period,
- $p_{(I_k,I_{k+1})}$ = Probability that the $(k{+}1)^{th}$ transmission attempt is unsuccessful **and** takes place before the end of the current OFF period,
- $p_{(I_k,B_k)}$ = Probability that the $(k{+}1)^{th}$ transmission attempt takes place after the end of the current OFF period,
- $p_{(B_k,I_k)}$ = Probability that the $(k{+}1)^{th}$ transmission attempt takes place after the end of the current ON period,
- $p_{(B_k,B_{k+1})}$ = Probability that the $(k{+}1)^{th}$ transmission attempt takes place before the end of the current ON period.

The above transition probabilities depend on the modeling parameters: the success probabil-

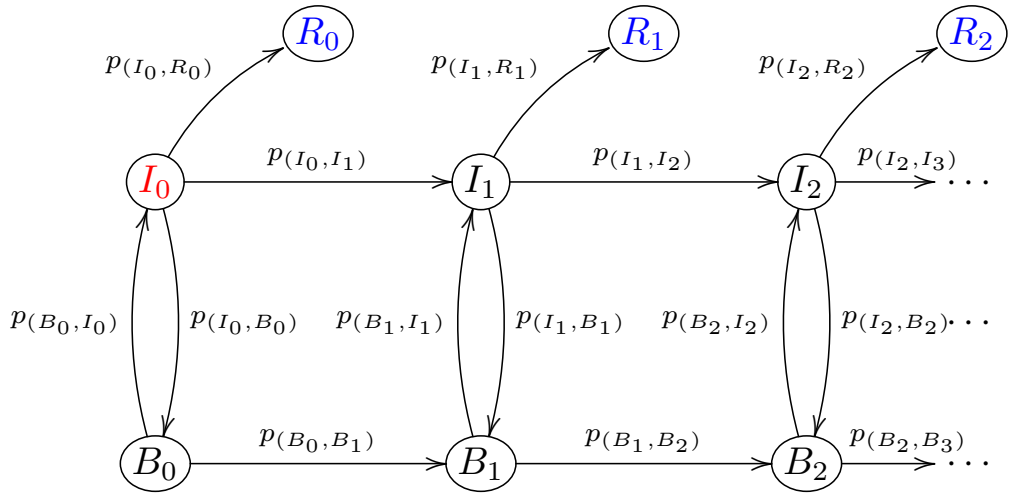


Figure 4.8 – Absorbing Discrete-Time Markov Chain for computing COExiST.

ity p_s^{off}, the duty cycle u, the average ON/OFF period durations $\bar{T}_{on}$ and $\bar{T}_{off}$, as well as the average MAC layer durations $\bar{T}_r$ and $\bar{T}_t$. These transition probabilities, for the most part, do not depend on the rank of the transmission attempt. For the case of $k = 0$, they depend on $\bar{T}_t$ but not $\bar{T}_r$. For the case of $k \neq 0$, they are identically expressed except that $\bar{T}_t$ is replaced with $\bar{T}_r$. Therefore, the Markov chain is composed of two homogeneous regions. One is composed of the states I_0 and B_0 while the other of all the remaining states. As a result, the Markov chain can be partially solved on both regions for computing COExiST.

Lemma 2 (Relations between the transition probabilities)**.** *The Markov chain transition probabilities satisfy the following relations:*

- $p_{(B_k,B_{k+1})} = 1 - p_{(B_k,I_k)}$
- $p_{(I_k,I_{k+1})} = (1 - p_s^{off}) \times (1 - p_{(I_k,B_k)})$
- $p_{(I_k,R_k)} = p_s^{off} \times (1 - p_{(I_k,B_k)})$

Denoting by $f_{(I_0,R_k)}$ the probabilities of reaching the absorbing state R_k, $k \in \mathbb{N}$ when starting from the initial state I_0, the expected transmission count equals

$$\mathbb{E}[N] = \sum_{k=0}^{+\infty} (k+1) f_{(I_0,R_k)} \tag{4.1}$$

This requires the calculation of the Markov chain transition probabilities as well as the reaching probabilities $f_{(I_0,R_k)}$.

1) Transition probabilities: Denoting by $\hat{T}_{on}$ the residual time in the ON period and applying the memoryless property of the exponential distribution, we have $\hat{T}_{on}$ distributed identically with T_{on}, that is, $\hat{T}_{on} \sim \mathcal{E}xp(1/\bar{T}_{on})$. Exactly the same analysis can be done with T_r, for which

$\hat{T}_r$ represents the residual time before the next retransmission takes place. With these, the computation of the transition probability $p_{(B_k,B_{k+1})}$ is as follows:

$$\begin{aligned} p_{(B_k,B_{k+1})} &= \int_{t=0}^{+\infty} \mathbb{P}[\hat{T}_r < t]\ f_{\hat{T}_{on}}(t)\mathrm{dt} \\ &= \int_{t=0}^{+\infty} \mathbb{P}[T_r < t]\ f_{T_{on}}(t)\mathrm{dt} = \frac{1}{1+\bar{T}_r/\bar{T}_{on}} \end{aligned}$$

Using the relations from Lemma 2 and introducing the variable ρ_r such that $\rho_r = \bar{T}_r/(\bar{T}_{on}+\bar{T}_{off})$ we have:

$$p_{(B_k,B_{k+1})} = \frac{u}{u+\rho_r} \qquad \text{and} \qquad p_{(B_k,I_k)} = \frac{\rho_r}{u+\rho_r}$$

Similarly, for the three remaining transition probabilities:

$$p_{(I_k,B_k)} = \frac{\rho_r}{1-u+\rho_r} \qquad p_{(I_k,R_k)} = \frac{p_s^{off}(1-u)}{1-u+\rho_r}$$

$$p_{(I_k,I_{k+1})} = \frac{(1-p_s^{off})(1-u)}{1-u+\rho_r}$$

For the expressions of the transition probabilities involving states I_0 and B_0, ρ_r is replaced with $\rho_t = \bar{T}_t/(\bar{T}_{on}+\bar{T}_{off})$.

2) Reaching probabilities: Computing the reaching probabilities in an absorbing discrete-time Markov chain can be done by applying the reachability equation, whose definition is repeated below:

Definition 3 (Reachability equation)**.** *The probability of reaching state j starting from state i can be computed as:*

$$f_{(i,j)} = p_{(i,j)} + \sum_{\forall k \neq j} p_{(i,k)} \times f_{(k,j)} \tag{4.2}$$

Theorem 3 (COExiST)**.** *The expected transmission count over a link subject to primary user interference is:*

$$\mathbb{E}[N] = \frac{1}{p_s^{off}(1-u)} + \frac{u}{\bar{T}_r} \times \frac{\bar{T}_t - \bar{T}_r}{\bar{T}_t/\bar{T}_{on} + 1 - u} \tag{4.3}$$

Proof. Applying the reachability equation on the first region of the Markov chain leads to:

$$f_{(I_0,R_0)} = \frac{p_{(I_0,R_0)}}{1 - p_{(I_0,B_0)}p_{(B_0,I_0)}}$$

and for $k \in \mathbb{N}^*$:

$$f_{(I_0,R_k)} = \frac{p_{(I_0,I_1)}f_{(I_1,R_k)} + p_{(I_0,B_0)}p_{(B_0,B_1)}f_{(B_1,R_k)}}{1 - p_{(I_0,B_0)}p_{(B_0,I_0)}}$$

$$f_{(I_1,R_{k+2})} = \left(\frac{p_{(I_k,I_{k+1})} + p_{(B_k,B_{k+1})}}{1 - p_{(I_k,B_k)}p_{(B_k,I_k)}}\right) f_{(I_1,R_{k+1})} - \left(\frac{p_{(I_k,I_{k+1})}p_{(B_k,B_{k+1})}}{1 - p_{(I_k,B_k)}p_{(B_k,I_k)}}\right) f_{(I_1,R_k)} \tag{4.6}$$

$$f_{(B_1,R_{k+2})} = \left(\frac{p_{(I_k,I_{k+1})} + p_{(B_k,B_{k+1})}}{1 - p_{(I_k,B_k)}p_{(B_k,I_k)}}\right) f_{(B_1,R_{k+1})} - \left(\frac{p_{(I_k,I_{k+1})}p_{(B_k,B_{k+1})}}{1 - p_{(I_k,B_k)}p_{(B_k,I_k)}}\right) f_{(B_1,R_k)} \tag{4.7}$$

$$Q(X) = X^2 - \left(\frac{p_{(I_k,I_{k+1})} + p_{(B_k,B_{k+1})}}{1 - p_{(I_k,B_k)}p_{(B_k,I_k)}}\right) X + \left(\frac{p_{(I_k,I_{k+1})}p_{(B_k,B_{k+1})}}{1 - p_{(I_k,B_k)}p_{(B_k,I_k)}}\right) \tag{4.8}$$

Since:

$$\mathbb{E}[N] = \sum_{k=0}^{+\infty}(k+1)f_{(I_0,R_k)} = \underbrace{\sum_{k=0}^{+\infty} f_{(I_0,R_k)}}_{=1} + \sum_{k=0}^{+\infty} k f_{(I_0,R_k)}$$

the desired results follows from applying Lemma 3. □

Lemma 3. *Computing the analytical expressions of the reaching probabilities leads to the following equations:*

$$\sum_{k=1}^{+\infty} k f_{(I_1,R_k)} = \frac{1}{p_s^{off}(1-u)} \tag{4.4}$$

$$\sum_{k=1}^{+\infty} k f_{(B_1,R_k)} = \frac{1}{p_s^{off}(1-u)} + \frac{u}{\rho_r} \tag{4.5}$$

Proof. Applying recursively the reachability equation, starting from states I_1 and B_1, and performing some linear combinations on the resulting equations leads to recursive expressions of the desired reaching probabilities, as given for $k \in \mathbb{N}^*$ by Equations 4.6 and 4.7. Therefore, these reaching probabilities satisfy the same linear second-order recurrence equations. However, they differ on their first terms, making the obtained probability values entirely different for the remaining terms of both sequences.

Each linear second-order recurrence equation can be solved for $k > 1$ using the following well-known method:

(a) Compute the roots r_1 and r_2 of the characteristic polynomial Q given in Equation 4.8.

(b) Compute $f_{(I_1,R_1)}$ and $f_{(I_1,R_2)}$ (respectively $f_{(B_1,R_1)}$ and $f_{(B_1,R_2)}$ for the second equation)

(c) Compute λ_I and μ_I (respectively λ_B and μ_B for the second equation) such that

$$\begin{cases} \lambda_I + \mu_I & = & f_{(I_1,R_1)} \\ \lambda_I r_1 + \mu_I r_2 & = & f_{(I_1,R_2)} \end{cases}$$

(d) Finally, combine the results: $f_{(I_1,R_k)} = \lambda_I r_1^{k-1} + \mu_I r_2^{k-1}$

Applying this method is straightforward in principle but it presents challenging calculations due to the dependence of the variables on three different parameters: u, p_s^{off} and ρ_r. The

algebraic expressions for r_1, r_2, λ_I (respectively λ_B) and μ_I (respectively μ_B) are long and complex. They can be computed, however, with the help of a mathematical tool, such as the open source calculation software *Maxima* [110]. For instance, by setting:

$$\sigma = {p_s^{off}}^2 u^4 + 2p_s^{off}(\rho_r p_s^{off} - p_s^{off} - 2\rho_r)u^3 + p_s^{off}(\rho_r^2 p_s^{off} - 4\rho_r p_s^{off} + p_s^{off} + 6\rho_r)u^2 \\ - 2(\rho_r - 1)\rho_r(p_s^{off} - 1)p_s^{off} u + \rho_r^2 {p_s^{off}}^2 - 2\rho_r^2 p_s^{off} + \rho_r^2$$

we get:

$$r_1 = \frac{\rho_r p_s^{off} - \rho_r - (p_s^{off} - 2)u^2 - [(\rho_r - 1)p_s^{off} + 2]u - \sqrt{\sigma}}{2(u^2 - u - \rho_r)}$$

and

$$r_2 = \frac{\rho_r p_s^{off} - \rho_r - (p_s^{off} - 2)u^2 - [(\rho_r - 1)p_s^{off} + 2]u + \sqrt{\sigma}}{2(u^2 - u - \rho_r)}$$

and then obtain:

$$\lambda_B = \frac{\rho_r p_s^{off}(u-1)\left[\sqrt{\sigma} + ({p_s^{off}}u^2 - 2u^2 + \rho_r p_s^{off} u - p_s^{off} u + 2u - \rho_r p_s^{off} + \rho_r)\right]}{2\sqrt{\sigma}(u^2 - u - \rho_r)}$$

as well as

$$\mu_B = \frac{\rho_r p_s^{off}(u-1)\left[\sqrt{\sigma} - ({p_s^{off}}u^2 - 2u^2 + \rho_r p_s^{off} u - p_s^{off} u + 2u - \rho_r p_s^{off} + \rho_r)\right]}{2\sqrt{\sigma}(u^2 - u - \rho_r)}$$

After checking the convergence requirements in Lemma 4, we can go straight away on the computation of the sought expectation terms, as:

$$\sum_{k=1}^{+\infty} k \times f_{(I_1,R_k)} = \lambda_I \sum_{k=1}^{+\infty} k \times r_1^{k-1} + \mu_I \sum_{k=1}^{+\infty} k \times r_2^{k-1} \\ = \frac{\lambda_I}{(1-r_1)^2} + \frac{\mu_I}{(1-r_2)^2}$$

which finally simplifies to $1/[p_s^{off}(1-u)]$. Similarly, for the second equation, we get:

$$\sum_{k=1}^{+\infty} k \times f_{(B_1,R_k)} = \frac{\lambda_B}{(1-r_1)^2} + \frac{\mu_B}{(1-r_2)^2}$$

which simplifies to $1/[p_s^{off}(1-u)] + u/\rho_r$. □

Lemma 4 (Convergence requirements)**.** *The roots of the characteristic polynomial given in Equation 4.8 strictly lie inside the unit disk, that is* $|r_1| < 1$ *and* $|r_2| < 1$.

Proof. Using well-know mathematical relations between the coefficients of the second degree

polynomial $Q(X)$, we obtain:

$$r_1 + r_2 = \frac{p_{(I_k,I_{k+1})} + p_{(B_k,B_{k+1})}}{1 - p_{(I_k,B_k)}p_{(B_k,I_k)}} \geq 0$$

and

$$r_1 r_2 = \frac{p_{(I_k,I_{k+1})}p_{(B_k,B_{k+1})}}{1 - p_{(I_k,B_k)}p_{(B_k,I_k)}} \geq 0$$

that leads to r_1 and r_2 being with positive real values. Suppose now $r_1 r_2 \geq 1$, we should have:

$$\frac{p_{(I_k,I_{k+1})}p_{(B_k,B_{k+1})}}{1 - p_{(I_k,B_k)}p_{(B_k,I_k)}} \geq 1$$

that is

$$p_{(I_k,I_{k+1})}p_{(B_k,B_{k+1})} \geq 1 - p_{(I_k,B_k)}p_{(B_k,I_k)}$$

However, using the relations depicted in Lemma 2 leads to:

$$p_{(I_k,I_{k+1})}p_{(B_k,B_{k+1})} = (1 - p_s^{off})(1 - p_{(I_k,B_k)})(1 - p_{(B_k,I_k)})$$

and thus

$$p_{(I_k,I_{k+1})}p_{(B_k,B_{k+1})} < 1 - p_{(I_k,B_k)}p_{(B_k,I_k)}$$

that contradicts the assumption that $r_1 r_2 \geq 1$.
Therefore, as $r_1 > r_2$, we obtain $|r_2| < 1$. We then introduce the following three sequences $(u_n)_{n\in\mathbb{N}}$, $(v_n)_{n\in\mathbb{N}}$ and $(w_n)_{n\in\mathbb{N}*}$ with general terms

$$u_n = \sum_{k=0}^{n} r_1^k \quad , \quad v_n = \sum_{k=0}^{n} r_2^k \quad \text{and} \quad w_n = \sum_{k=1}^{n} f_{(B_1,R_k)}$$

leading to $w_n = \lambda_B u_{n-1} + \mu_B v_{n-1}$, for $n \in \mathbb{N}^*$ with λ_B and μ_B non zero. Because $f_{(B_1,R_k)}$ are probability terms defined for $k \in \mathbb{N}^*$, we have $\lim_{n\to\infty} w_n = 1$, $\lim_{n\to\infty} v_n = 1/(1-r_2)$ and thus:

$$\lim_{n\to\infty} u_n = \frac{1 - \mu_B/(1 - r_2)}{\lambda_B}$$

In other words, as the serie $(u_n)_{n\in\mathbb{N}}$ converges, and $\lambda_B > 0$, $|r_1| < 1$. □

The method applied so far for computing COExiST is twofold. At the same time of computing an average of the transmission count, it provides an analytical expression of the associated probability distribution function. However, such a distribution is complex and cannot be obtained straightforwardly. For those who might only be interested in computing the average value, we present now a more straightforward calculation method.

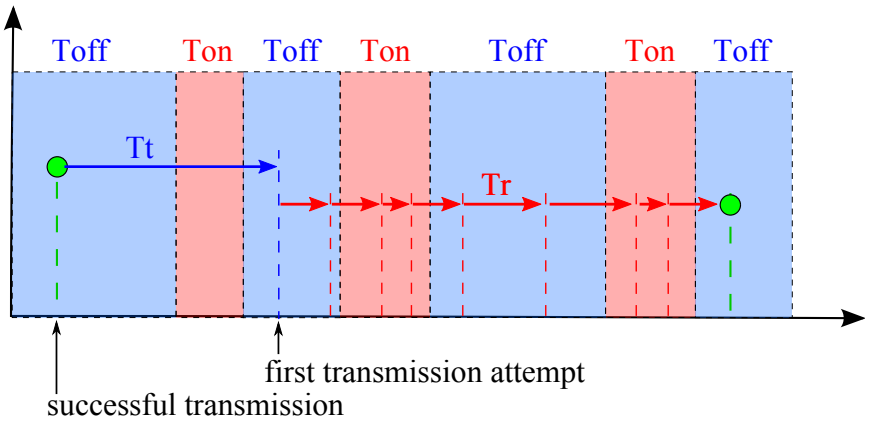


Figure 4.9 – Detailed transmission/retransmission scheme

4.3.3 Computing COExiST using a decomposition approach

A detailed vision of the transmission/retransmission scheme is depicted in Figure 4.9 and used as a basis for conducting the alternative computation strategy. The basic idea of this new method consists in decomposing the computation of COExiST by considering whether the first transmission attempt has occured during an OFF or ON period. From now on, we rather work with the transition rates of the exponential distributions, that is $\lambda_r = 1/\bar{T_r}$, $\lambda_t = 1/\bar{T_t}$, $\lambda_{off} = 1/\bar{T_{off}}$ and $\lambda_{on} = 1/\bar{T_{on}}$.

Theorem 4 (Decomposition)**.** *By denoting π_{off} the probability that the first transmission attempt occurs during an OFF period and α_{off} the expected transmission count given that the first transmission attempt occurs during an OFF period, COExiST can be computed through the following formula:*

$$\mathbb{E}[N] = \alpha_{off} \times \pi_{off} + \left[1 + \frac{\alpha_{off} - 1}{1 - p_s^{off}} + \frac{\lambda_r}{\lambda_{on}}\right] \times (1 - \pi_{off}) \tag{4.9}$$

Proof. Defining $\alpha_{on} = \mathbb{E}[N|1^{st}\text{tx attempt during an ON period}]$ and conditioning the expected transmission count whether the first transmission attempt has occured during an OFF or an ON period leads to:

$$\mathbb{E}[N] = \alpha_{off} \times \pi_{off} + \alpha_{on} \times (1 - \pi_{off})$$

The proof consists thus in showing that:

$$\alpha_{on} = 1 + \frac{\alpha_{off} - 1}{1 - p_s^{off}} + \frac{\lambda_r}{\lambda_{on}}$$

When the first transmission attempt occurs during the OFF period, it is likely to be successful

with a probability p_s^{off} and thus lead to an average of γ additional retransmission attempts [1] with probability $(1 - p_s^{off})$. However, if this first transmission attempt takes place during the ON period, it will be necessarily followed by an average of λ_r/λ_{on} retransmissions in the ON period before reaching the next OFF period. Note that, at the time a retransmission starts pending during the OFF period, the residual time before this retransmission occurs still follows an exponential distribution with rate λ_r. Similarly, the residual time before the current OFF period ends follows an exponential distribution with rate λ_{off}. Therefore, putting this into equations leads to the following system:

$$\begin{cases} \alpha_{off} &= 1 + (1 - p_s^{off})\gamma \\ \alpha_{on} &= 1 + \lambda_r/\lambda_{on} + \gamma \end{cases}$$

that, when substituting the value of γ, leads to:

$$\alpha_{on} = 1 + \frac{\alpha_{off} - 1}{1 - p_s^{off}} + \frac{\lambda_r}{\lambda_{on}}$$

□

Theorem 5 (π_{off} computation)**.** *The probability that the first transmission attempt occurs during an OFF period is given by:*

$$\pi_{off} = \frac{\lambda_t + \lambda_{on}}{\lambda_t + \lambda_{on} + \lambda_{off}} \tag{4.10}$$

Proof. At the time the first transmission is pending during an OFF or ON period, the residual time before such a transmission to occur as well as the current OFF or ON period finish respectively follow an exponential distribution with parameter λ_t, λ_{off} and λ_{on}. This statement holds whatever the observation interval is constant or exponentially distributed. Hereafter, by denoting τ_{off} and τ_{on} the probabilities that the first transmission occurs before the end of the current OFF or ON period, we get:

$$\tau_{off} = \frac{\lambda_t}{\lambda_t + \lambda_{off}} \tag{4.11}$$

and

$$\tau_{on} = \frac{\lambda_t}{\lambda_t + \lambda_{on}} \tag{4.12}$$

Considering that the last successful transmission took place during the OFF period indexed

1. We do not need to compute γ at this stage.

first, we get:

$$\begin{aligned}\pi_{off} &= \sum_{k=1}^{+\infty} \left[(1-\tau_{off})(1-\tau_{on})\right]^{k-1} \tau_{off} \\ &= \frac{\tau_{off}}{1-(1-\tau_{off})(1-\tau_{on})} = \frac{\tau_{off}}{\tau_{on}+\tau_{off}-\tau_{on}\tau_{off}} \\ &= \frac{\lambda_t+\lambda_{on}}{\lambda_t+\lambda_{on}+\lambda_{off}}\end{aligned}$$

that concludes the proof. □

Theorem 6 (α_{off} computation)**.** *The expected transmission count given that the first transmission attempt occurred during an OFF period is given by:*

$$\alpha_{off} = \frac{1}{p_s^{off}(1-u)} - \frac{u}{1-u} \tag{4.13}$$

Proof. Establishing the proof is equivalent to computing the γ term used in Theorem 4, as $\alpha_{off} = 1 + (1-p_s^{off})\gamma$. For this typical purpose we first compute the probability of an OFF period seeing a successful transmission, referred to as μ_{off}. Using the transition rates defined so far, μ_{off} satisfies the following equation:

$$\mu_{off} = \frac{\lambda_r p_s^{off}}{\lambda_{off}+\lambda_r} + \mu_{off} \times \frac{\lambda_r(1-p_s^{off})}{\lambda_{off}+\lambda_r}$$

and therefore

$$\mu_{off} = \frac{\lambda_r p_s^{off}}{\lambda_{off}+\lambda_r p_s^{off}}$$

We remind that the duty cycle can also be expressed as:

$$u = \frac{\lambda_{off}}{\lambda_{off}+\lambda_{on}} \tag{4.14}$$

Consequently, γ can be computed as follows:

$$\begin{aligned}\gamma &= \sum_{k=1}^{+\infty}(k-1)\left[\frac{\lambda_r}{\lambda_{on}}+\frac{\lambda_r}{\lambda_{off}}\right](1-\mu_{off})^{k-1}\mu_{off}\\ &= \lambda_r\frac{\lambda_{on}+\lambda_{off}}{\lambda_{on}\lambda_{off}}\times\mu_{off}\sum_{k=0}^{+\infty}k(1-\mu_{off})^k\\ &= \lambda_r\frac{\lambda_{on}+\lambda_{off}}{\lambda_{on}\lambda_{off}}\times\frac{1-\mu_{off}}{\mu_{off}} = \lambda_r\frac{\lambda_{on}+\lambda_{off}}{\lambda_{on}\lambda_{off}}\times\frac{\lambda_{off}}{\lambda_r p_s^{off}}\\ &= \frac{1}{(1-u)p_s^{off}}\end{aligned}$$

In fact, γ is equivalent to the sum $\sum_{k=1}^{+\infty}kf_{(I_1,R_k)}$ provided in Equation 4.4. Hereafter,

$$\alpha_{off} = 1+(1-p_s^{off})\gamma = \frac{1}{p_s^{off}(1-u)}+1-\frac{1}{1-u} = \frac{1}{p_s^{off}(1-u)}-\frac{u}{1-u}$$

that concludes the proof. □

Theorem 7 (COExiST)**.** *The expected transmission count over a link subject to primary user interference is given by:*

$$\mathbb{E}[N] = \frac{1}{p_s^{off}(1-u)}+\frac{\lambda_{off}}{\lambda_{on}}\times\frac{\lambda_t-\lambda_r}{\lambda_t+\lambda_{on}+\lambda_{off}} \tag{4.15}$$

that is equivalent to Equation 4.3.

Proof. Applying Equations 4.9, 4.10 and 4.13 we obtain:

$$\begin{aligned}\mathbb{E}[N] &= \alpha_{off}\times\pi_{off}+\left(\alpha_{off}+\frac{1}{1-u}+\frac{\lambda_r}{\lambda_{on}}\right)\times(1-\pi_{off})\\ &= \alpha_{off}+\frac{\lambda_{off}}{\lambda_t+\lambda_{on}+\lambda_{off}}\left(\frac{1}{1-u}+\frac{\lambda_r}{\lambda_{on}}\right)\\ &= \frac{1}{p_s^{off}(1-u)}-\frac{\lambda_{off}}{\lambda_{on}}+\frac{\lambda_{off}}{\lambda_t+\lambda_{on}+\lambda_{off}}\left(\frac{1}{1-u}+\frac{\lambda_r}{\lambda_{on}}\right)\end{aligned}$$

that simplifies to

$$\mathbb{E}[N] = \frac{1}{p_s^{off}(1-u)}+\frac{\lambda_{off}}{\lambda_{on}}\times\frac{\lambda_t-\lambda_r}{\lambda_t+\lambda_{on}+\lambda_{off}}$$

and then

$$\mathbb{E}[N] = \frac{1}{p_s^{off}(1-u)}+\left(\frac{1}{\lambda_t}-\frac{1}{\lambda_r}\right)\times\frac{u\lambda_r}{\lambda_{on}/\lambda_t+1-u}$$

by using some algebraïc operations. □

4.3.4 Practical application of COExiST

Theorem 8 (COExiST as a function of ETX)**.** *When the probing packets used for computing ETX are sent independently of the primary users activity pattern, COExiST can be expressed as the following function of ETX:*

$$\mathbb{E}[N] = ETX + \frac{u}{\bar{T}_r} \times \frac{\bar{T}_t - \bar{T}_r}{\bar{T}_t/\bar{T}_{on} + 1 - u} \tag{4.16}$$

$$\mathbb{E}[N] = ETX + \left(\frac{1}{\lambda_t} - \frac{1}{\lambda_r}\right) \times \frac{u\lambda_r}{\lambda_{on}/\lambda_t + 1 - u} \tag{4.17}$$

Proof. The probing packets are sent in broadcast mode with a higher priority than unicast packets. As per 802.11, the probing packets are neither acknowledged nor retransmitted in case of errors. In addition, we consider the case of the probing packets being discarded if primary user activity is detected by a preceding in-band sensing mechanism. Therefore, the probability for such a probe to be successfully received is:

$$\begin{aligned} PRR &= \mathbb{P}[\text{receive probe}|\text{tx probe during OFF period}]\mathbb{P}[\text{tx probe during OFF period}] \\ &\quad + \underbrace{\mathbb{P}[\text{receive probe}|\text{tx probe during ON period}]}_{=0}\mathbb{P}[\text{tx probe during ON period}] \\ &= \underbrace{\mathbb{P}[\text{receive probe}|\text{tx probe during OFF period}]}_{=p_s^{off}}\underbrace{\mathbb{P}[\text{tx probe during OFF period}]}_{=1-u \text{ by indep.}} \end{aligned}$$

which is equivalent to PRR $= p_s^{off}(1-u)$. As 1/PRR = ETX, that concludes the proof. □

The value of Theorem 8 is twofold. It shows that ETX is a special case of COExiST for $u = 0$ and/or $\bar{T}_t = \bar{T}_r$. And more important, in conjunction with Theorem 9 below, it paves the way for leveraging popular ETX implementations to quickly deploy COExiST. It is the approach we use in Section 4.4.

Theorem 9 (Routing with COExiST)**.** *COExiST coupled with hop-by-hop Dijkstra-based routing satisfies the optimality, consistency and loop-freeness properties.*

Proof. According to [28], establishing the proof is equivalent to demonstrating that the path weight metric is right-monotonic and right-isotonic. As COExiST is additive, it suffices to show that the metric is non-negative – this is straightforward from Eq. (4.1). □

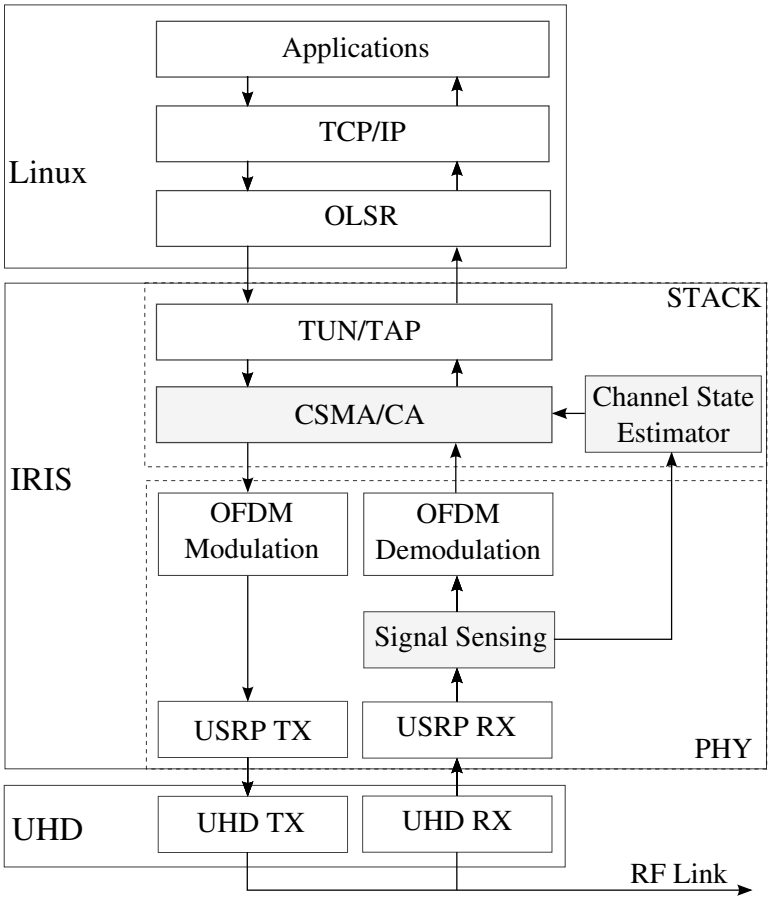


Figure 4.10 – Software Architecture

4.4 Testbed implementation

To evaluate COExiST, we have used the USRP N210 radio platform and the IRIS software package. IRIS is a component-based application whose architecture and parameters are fully reconfigurable. Such reconfiguration, used for instance for tuning the transmission frequency to another vacant channel, can be performed in real-time, something currently not possible with the GNURadio. We have made significant additions to IRIS that were necessary to run our experiments. Where possible, we used open-source libraries, such as OLSR with ETX, and when not, we added our own implementations, including a CSMA/CA MAC, a primary user model, described in Section 4.2, and COExiST. Figure 4.10 shows the architecture of the software running on our testbed.

4.4.1 CSMA/CA implementation

IRIS does not yet include a MAC layer component nor any mechanisms that would easily allow running IP applications over USRP radios. To rectify this, we have implemented our own MAC layer following the specifications of the IEEE 802.11 standard. In particular, we have realized the DCF (CSMA/CA) part of 802.11.

A main challenge in implementing a CSMA MAC on USRP radios is implementing carrier sensing. We developed our own solution consisting of a *Signal Sensing* component that computes the complex signal recovered by the UHD driver and estimates the power of the received signal or RSSI. The value, coded in 16 bits, is then passed onto the *Channel State Estimator* component at the frequency of once per physical frame received. The *Channel State Estimator* module estimates the current channel state by comparing to a threshold value. For the simple case of a single threshold mechanism, the activity threshold must be calibrated by calculating the noise-floor and adding 10 dB, as recommended by the IEEE 802.11 standard. If the channel state changes, it sends a *Sensing Change* message to the main CSMA/CA component, the equivalent of the Clear Channel Assessment (CCA) in 802.11.

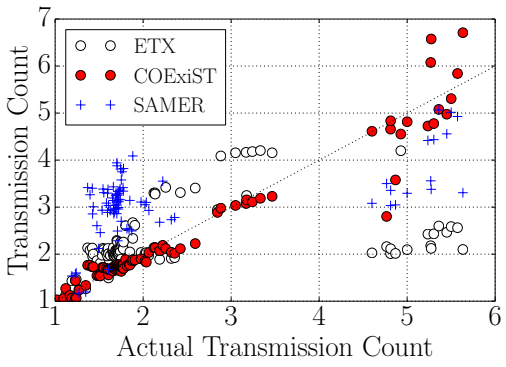
We use the Google protocol buffers to define the structure of the CSMA/CA and leverage the boost library to synchronize the transmission and reception threads inside the main CSMA/CA component. Finally, we interface our MAC layer to the linux IP stack using the Tun/Tap component provided in IRIS.

4.4.2 COExiST implementation

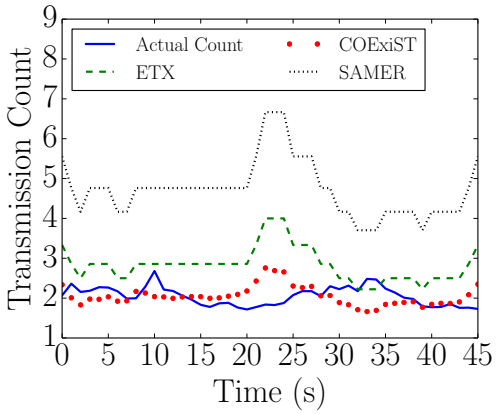
We use Equation 4.16 from Section 4.3 to implement COExiST in OLSR. We use OLSRd, an open-source implementation that also includes an implementation of ETX. As Equation 4.16 shows, we can leverage the ETX value and add the second term, which is solely a function of T_{on}, T_r, T_t and u. Our MAC implementation collects these values and passes them on to OLSR, where a simple modification allows replacing ETX with COExiST. As it does with ETX, OLSR updates COExiST at the default rate of 1/sec. Note that T_r and T_t are based on unicast traffic. For bootstrapping the computation and for the cases where there is no unicast traffic, we use the minimum possible values based on the channel access parameters. Also note that we took a particular care in implementing COExiST in conjunction with ETX. That is, each COExiST sample corresponds to a moving average over the last 10 or 20 seconds window, this time being set as a parameter.

4.5 Performance evaluation

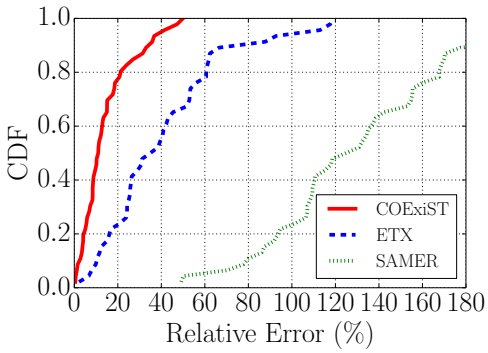
In this section we evaluate the performance of COExiST and compare it with ETX and the actual transmission count. In summary, we make the following main observations: In Section 4.5.2, we demonstrate that COExiST is a very good approximation of the actual transmission count



(a) Every data point is an average over three 5-minute experiments.



(b) A 45 seconds sample from a 5-minute experiment is shown here. Both COEXisT and ETX are computed once a second but COExiST follows the actual transmission count far more closely.



(c) The CDF of errors for the data on the left (b). For COExiST, 80% of the time the relative error is less than 20%.

Figure 4.11 – Accuracy of COExiST.

– 80% of the time the error is less than 20%. In constrast, the 80^{th} percentile of error for ETX and SAMER are found to be 60% and 160%, respectively. In Section 4.5.3, we demonstrate that COExiST continues to provide a very good approximation to the transmission count even in face of imperfect estimates of the primary user activity. In Section 4.5.4, we show that when OLSR uses COExiST it computes higher throughput multi-hop paths than when it uses ETX, SAMER or the STOP-RP metrics.

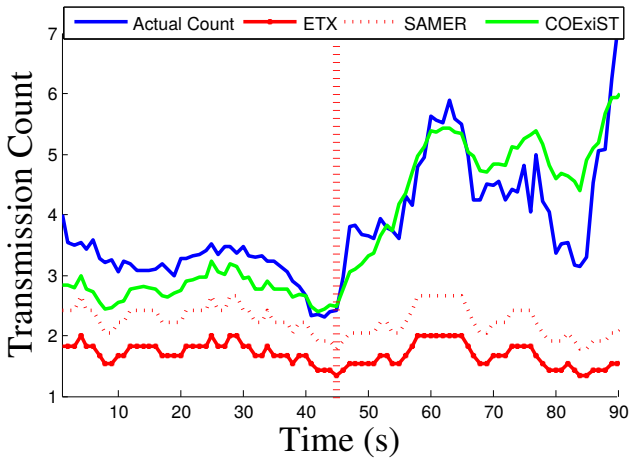


Figure 4.12 – COExiST captures the variation of $\bar{T}_{on}$

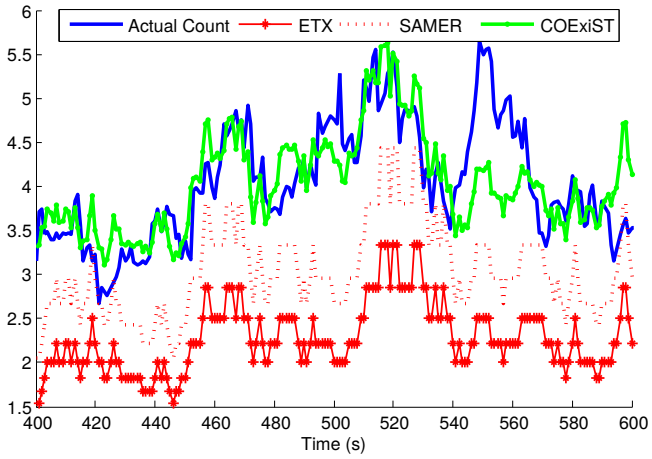


Figure 4.13 – COExiST vs ETX and SAMER

4.5.1 Experimental setup

Unless otherwise specified, the experimental setup is as follows. The testbed and primary user activity are as described in Section 4.2.1, and the software architecture as described in Section 4.4. We carry out two groups of experiments. The first group (Sections 4.5.2 and 4.5.3) is targeted at evaluating the accuracy of COExiST at estimating transmission count over a link. For this, three USRP radios are deployed, with two of them representing the secondary network and the third, the primary user.

The second group (Sections 4.5.4) is targeted at evaluation the overall performance of CO-ExiST. For this, we use all five USRP radios, with four of them creating a multi-hop secondary

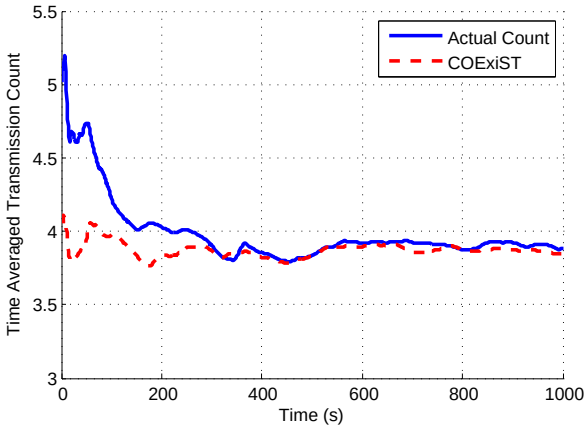


Figure 4.14 – COExiST is close to the actual count time-average

network, and the fifth utilized to create up to two primary users. For this group of experiments we connect the USRP radios via RF cables to an RF switch matrix. This enables us to create a multi-hop topology using otherwise licensed frequencies and create two primary users using a single USRP.

In all experiments we use Iperf [111] to generate UDP traffic. The radios are configured to send packets at 1Mbps data rate and the data packet size is set to 1500 Bytes. A single experiment is run for 5 minutes and the data presented is an average over 3 runs.

Basis for Comparison: We compare COExiST with ETX, the actual transmission count as well as two metrics proposed as part of two routing protocols designed for cognitive radio networks, namely SAMER [82] and STOD-RP [46]. SAMER essentially multiplies the packet reception ratio by the fraction of time with no primary users activity, while STOD-RP combines link quality with spectrum availability by dividing ETT [27] by the time duration of the link.

4.5.2 Accuracy of COExiST

To measure the accuracy of COExiST and compare it to the other metrics we carry a series of experiments using the three node topology, with two nodes representing the secondary network and the third the primary user. Between every experiment we change the placement of all nodes as well as the pattern of the PU activity. In every experiment we measure the actual transmission count and collect transmission counts collected by COExiST, ETX and SAMER. STOD-RP is not included in this experiment as it does not compute the transmission count but rather the transmission time for a successful packet transmission. The data collected, shown in Figure 4.11(a), shows that COExiST matches the actual transmission count fairly closely while ETX and SAMER end up either overestimating or underestimating it for a significant number of measurement points. To quantify the error for all metrics, Figure 4.11(c) shows the CDFs of

the errors when compared to the actual transmission count. COExiST is shown to be a very good approximation of the actual transmission count – 80% of the time the estimation error is less than 20%. At the same time, the data shows ETX and SAMER performing poorly – the 80th percentile for the error rate is 60% and 160%, respectively.

COExiST also adapts quickly to the variations of the involved parameters. Indeed, as depicted in Figure 4.12, that completes Figure 4.6, COExisT still accurately estimates the transmission count when $\bar{T}_{on}$ varies from 50ms to 250ms.

Note that, due to values that can take T_r and T_t, COExiST can be either greater or lower than ETX. Therefore, the straightforward adaption of ETX considered in SAMER that consists in multiplying it by $1/(1-u)$ can perform better and not necessarily worse than ETX. This is typically illustrated in Figure 4.13.

Finally we noticed that for the cases of COExiST inaccuracy, the instantaneous value of COExiST is often closer to the time average of the transmission count than the instantaneous transmission count itself. Such a phenomenon is illustrated in Figure 4.14.

4.5.3 Sensitivity of COExiST to input errors

Figure 4.15 shows the performance of COExiST in the presence of errors in the estimate of the primary user activity, as well as when the UDP packets are of different sizes. To induce a specific amount of errors, we simply modify the OLSR-COExiST implementation to artificially add errors to the parameters of primary user activity coming from the lower layers. We did this to simulate a real-life scenario where estimation errors are to be expected. Despite the significant errors, COExiST is shown to maintain its accuracy. Furthermore, Figure 4.15(c) shows that the accuracy of COExiST presents little sensitivity to packet size, except for the very small. This is due to the fact that very small packet can "get lucky", that is, get through the link even though the primary user is active since, as we show in Figure 4.2 in Section 4.2, its activity is not a perfect step function.

4.5.4 Transmission Count Accuracy & Throughput

Finally, we evaluate the impact of an accurate transmission count on throughput. For this we carry out two experiments.

In the **first experiment** we use the three node topology and, to overcome the limitation due to the limited number of USRP radios we posses, we try to create in time the equivalent of several links on a multi-hop topology. To do this, we carry multiple experiment where we have a single source transmitting as fast as possible to a single destination while a primary user is interfering and vary the node placement and the PU level of activity from one experiment to another. During each experiment we collect the COExiST and ETX values as well as the realized UDP throughput. Figure 4.16 shows the collected values for COExiST (y-axis) and ETX (x-axis) for every experiment. For two experiments we show the respective UDP throughput ranges observed

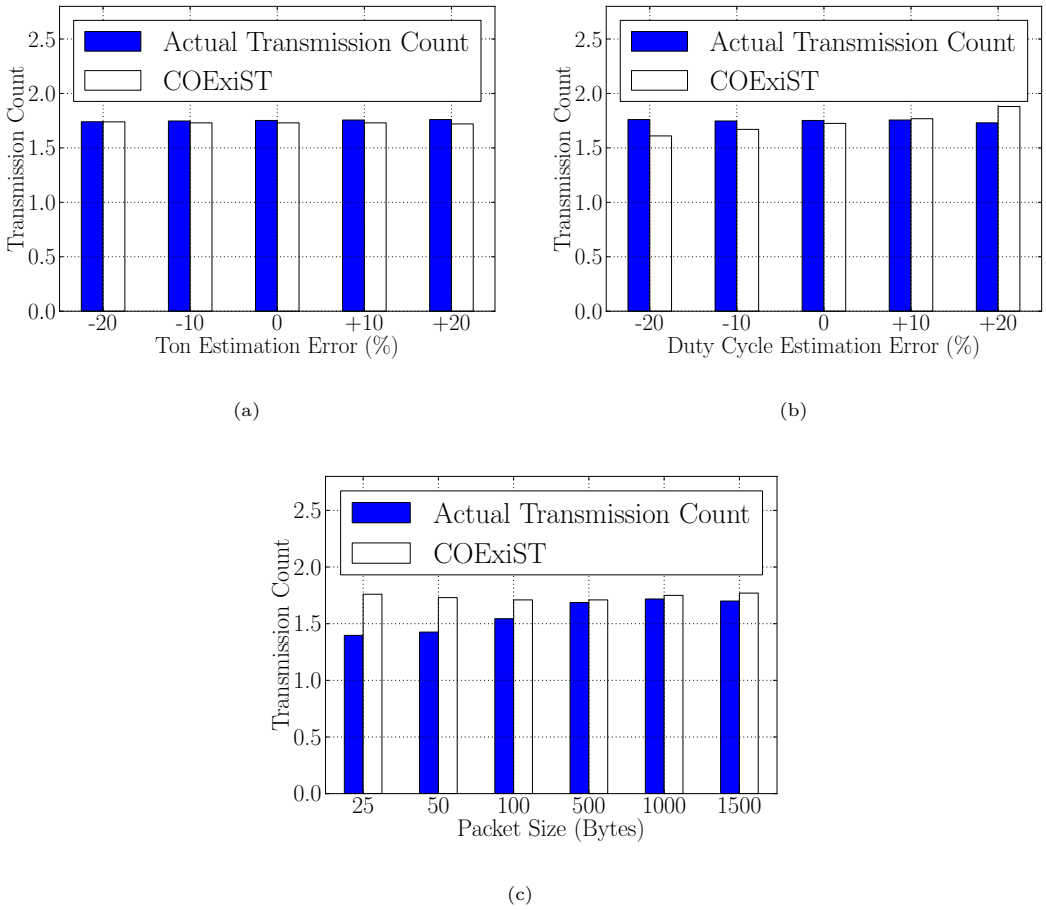


Figure 4.15 – Sensitivity of COExiST to input errors

(208 to 221 Kbps for one, 239 to 257 Kbps for the other). ETX is smallest for the experiment where the smaller throughput was realized – 2.0 for 208 to 221 Kbps, 2.2 for 239 to 257 Kbps – while the opposite is observed with COExiST. The difference observed is obviously due to the time dimension – in a larger network the difference would be due to the space dimension. Either or, a routing protocol minimizing COExiST would select higher throughput links.

In the **second experiment**, we evaluate the performance of all metrics on the multi-hop topology. In addition to the performance measurements, we also show in Figure 4.17 the state of the network at the time of the experiment, including the level of channel errors and primary user activity. Note that, as mentioned above, for this experiment, we use an RF switch matrix which allows us to control the channel errors and the level of primary user activity on every link. In the deployed topology, the primary users are hidden to USRP 1.

Figure 4.17 shows that COExiST is the only metric that identifies the highest throughput path, 1-2-4. This is due to the fact that SAMER considers primary users as a new source of

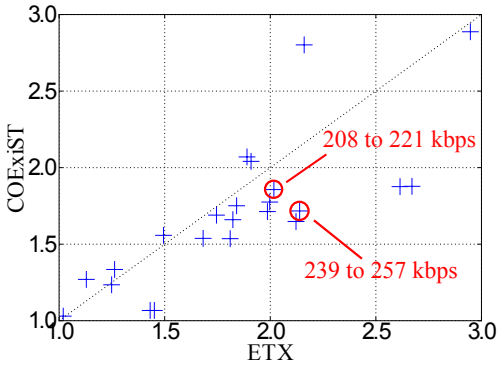


Figure 4.16 – Better accuracy on transmission count leads to better throughput

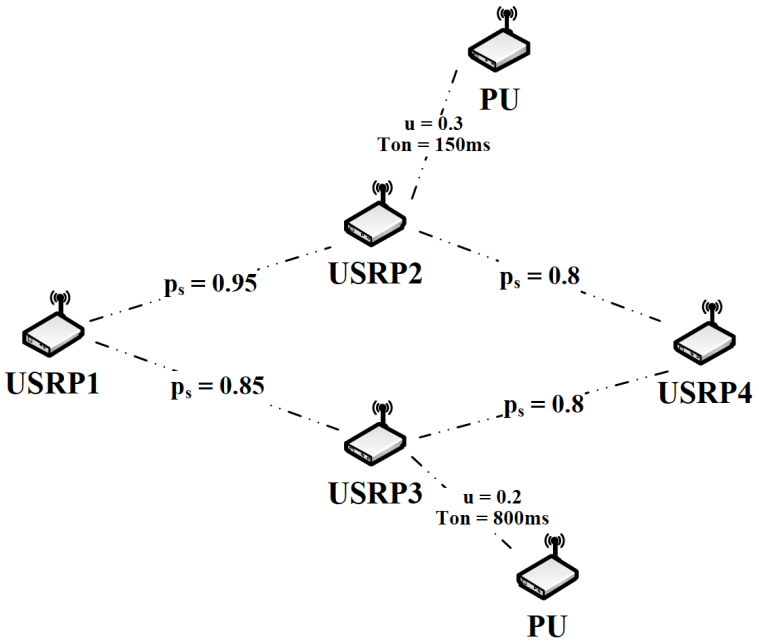
independent channel errors. However, as we have shown in Theorem 8, the packet reception ratio computed by sending broadcast probes is already impacted by the primary user activity. Therefore, multiplying the packet reception ratio by the fraction of time with no primary user activity cannot suffice to capture the actual effect of primary users activity on the transmission count and, ultimately, the realized throughput. On the other hand, STOD-RP adopts a different strategy by considering the absolute time a link is available so as to favor links with less PU activity. However, the absolute time a link is free of PU activity does not tell the whole story – a link can be free of PU activity for a while only with the PU becoming suddenly active. STOD-RP is slow in penalizing such link.

4.6 Discussion

In cognitive radio networks, there have been several proposed approaches to routing. Some have advocated for complete system solutions that address joint route-spectrum selection, protection to primary users [47], [62], [79], QoS provisioning [39]. We believe COExiST is complementary to these approaches. No matter how good the sensing and spectrum assignment are, they cannot guarantee PU free networking. COExiST can be leveraged for improving routing once the spectrum assignment converges, and it can be used as part of the spectrum assignment decision by quantifying the impact of primary users on performance. Furthermore, combining COExiST with traditional routing approaches, as we did with OLSR in this work, can allow for backward compatible solutions that can help the market penetration of cognitive radio networks.

4.7 Conclusions

This chapter has presented COExiST, an approach for estimating the transmission count in multi-hop cognitive radio networks. COExiST can be used as a stand-alone metric for quantify-



Path	ETX	SAMER	STOD-RP	COExiST	Throughput (kbps)
1-2-4	3.51	5.01	0.12	**3.48**	**165.24**
1-3-4	**3.21**	**4.01**	**0.01**	3.67	141.12

Figure 4.17 – COExiST correctly estimates that 1-2-4 is the best path.

ing link qualities and computing transmission-count optimal paths, or be combined with other parameters for creating more sophisticated routing metrics, depending on the particular needs. COExiST is measurement-driven, in that, all its inputs are collected at run-time. Using measurements on a five-node USRP N210 testbed, we show that COExiST accurately captures the transmission count for a variety of primary user activity levels and channel errors.

There are several interesting future directions that could be explored. First, it would be very interesting to evaluate COExiST on a larger scale testbed with more hops. Second, it would be interesting to explore how COExiST can be used as a building block for creating more sophisticated routing metrics customized to multi-hop cognitive networks.

5 Greedy routing: a promising solution for Quality of Service routing

In this chapter, we study Quality of Service routing based on multiple additive metrics such as delay and jitter. The objective is to find paths that respect a set of end-to-end constraints for each QoS metric. In Section 5.1, we define the associated multi-constrained path problem and show that it cannot be solved using traditional path calculation algorithms. We review the proposed approximation algorithms in Section 5.2 and propose to implement two of those in a real network environment. For this typical purpose, we select a greedy routing strategy that simply extends the Dijkstra's shortest path algorithm. In Section 5.3, we conduct a numerical analysis and show that such a scheme performs well in practice. However, we show that for the routing to be efficient, packets have to be source routed.

Finally, in Section 5.5, we introduce new concepts for evaluating greedy routing strategies and investigating better performance bounds than the best currently established.

5.1 Preliminaries

QoS requirements can necessitate that several metrics not exceed some specific values. For example, it can be required for a given flow that the total transmission delay does not exceed $15ms$, jitter be lower that $2ms$ and packets do not pass through more than 4 routers. Therefore, packets must be routed along a path that satisfies the given constraints. Such a routing problem has been referred to as QoS routing [112] or the multi-constrained path problem. In the remaining of this section we introduce a set of notations and definitions that serve to formalize such a problem and understand the relative issues.

5.1.1 Routing algebra

We model a multi-hop network with a directed graph $G(V, E)$ where V is the set of nodes and E the set of edges. An edge (u, v) represents the link connectivity between nodes u and v for packets being transmitted from u to v. The graph is provided with a *path weight structure* represented by the quadruplet $(S, \oplus, w, \preceq)$ where S is the set of paths, $\oplus$ is the path concatenation

operation, w is a function mapping a path to a weight and providing a totally ordered structure of S with the relation $\preceq$. With respect to the routing algebra defined in [28, 113] a path p is considered better than a path q if and only if $w(p) \preceq w(q)$. In a shortest path routing algebra, this is equivalent to $w(p) \leq w(q)$, with w being interpreted as a path length.

5.1.2 Multi-constrained QoS routing

In multi-constrained QoS routing, every edge is associated with K additive weights. For each of these K weights, the associated path weight simply equals the sum of the corresponding edge weights along the path. In other words, for any path p, $w_k(p) = \sum_{e \in p} w_k(e)$. The Multi-Constrained Path (MCP) problem consists thus of finding a $s - t$ path that satisfies a set of K QoS constraints. Such a path is referred to as a *feasible* path.

Definition 4 (Multi-Constrained Path (MCP) problem)**.** *INSTANCE: A directed graph $G(V, E)$ with K non-negative real-valued edge weights and a set $\{W_1, ..., W_K\}$ of K path constraints. PROBLEM: Find an s-t path p such that $\forall k \in \{1, ..., K\},\ w_k(p) \leq W_k$.*

There can be several feasible paths in a graph G. Unfortunately, for $K > 1$, the MCP problem has been demonstrated to be NP-complete [16, 112]. However, to differentiate feasible paths, one might be interested in finding the path with minimal length. The problem that additionally optimizes the path length as been referred to as the *Multi-Constrained Optimal Path (MCOP) problem*. It depends on the definition of the path length. They are multiple candidates for such a length. Straightforward choices consisting of computing a length based on linear combinations of the K metric components have been shown to fail in finding feasible paths [16]. The best known length for determining feasible paths is defined by the norm $w(p) = \max_{1 \leq k \leq K} \frac{w_k(p)}{W_k}$.

Definition 5 (Multi-Constrained Optimal Path (MCOP) problem)**.** *INSTANCE: A directed graph $G(V, E)$ with K non-negative real-valued edge weights, a set $\{W_1, ..., W_K\}$ of K path constraints and provided with a length w defined for any path p by the norm $w(p) = \max_{1 \leq k \leq K} \frac{w_k(p)}{W_k}$. PROBLEM: Find an s-t path p with minimal length.*

The MCOP problem is thus equivalent to a shortest path problem in a K-dimensional graph provided with the non-linear length w. If the solution p' of the MCOP problem have length $w(p') \leq 1$, it guarantees that p' is a feasible path. When the shortest path does not satisfy this property, the length quantifies the distance to the closest feasible path.

It is well known that for 1-dimensional additive metrics, the shortest path problem can be solved greedily in $\mathcal{O}(|V| \log |V| + |E|)$ using the Dijkstra's shortest path algorithm with the Fibonacci heap structure. However, as demonstrated in [28], for the Dijkstra's algorithm to be optimal, the underlying path weight structure needs to satisfy both right-monotonicity and right-isotonicity properties, defined hereunder.

Definition 6 (Right Monotonicity)**.** *A path weight structure $(S, \oplus, w, \preceq)$ is right-monotonic if and only if $\forall (p, q) \in S^2,\ w(p) \preceq w(p \oplus q)$.*

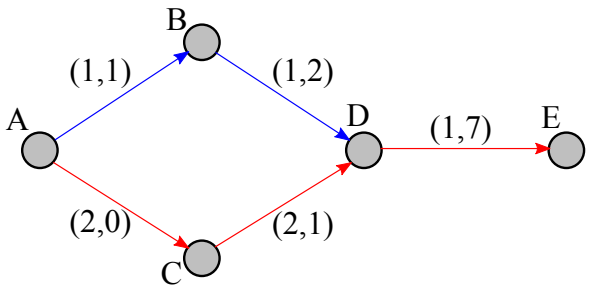


Figure 5.1 – In MCOP routing, subpaths of optimal paths are not necessarily optimal.

Definition 7 (Right Isotonicity)**.** *A path weight structure* $(S, \oplus, w, \preceq)$ *is right-isotonic if and only if* $\forall (p,q,r) \in S^3,\ w(p) \preceq w(q) \ \Rightarrow\ w(p \oplus r) \preceq w(q \oplus r)$.

Clearly, the norm defined so far in the MCOP problem results in a path weight structure that only satisfies the right-monotonicity property, as each QoS metric is additive. However, it is not right-isotonic. To elucidate this phenomenon we can have a look to the scenario described in Figure 5.1. We consider a five nodes network with edges associated with two additive weights. We seek an optimal feasible path from A to E with respect to the constraints $W_1 = W_2 = 8$. Using Dijkstra's algorithm with the MCOP path weight structure, we obtain $A \to B \to D$ to be the shortest $A - D$ path with weight $w(A \to B \to D) = 3/8 = 0.375$. As a consequence path $A \to C \to D$ will not be further examined during the Dijkstra's algorithm execution. In particular, node E will never be relaxed according to such a path. The algorithm finally returns $A \to B \to D \to E$ as the shortest $A - E$ path. As $w(A \to B \to D \to E) = 10/8 = 1.25 > 1$, such a path is not feasible. However, one can remark that path $A \to C \to D \to E$ has weight 1 and is thus a feasible path. In the following, such a path $A \to C \to D$ is said to be *non-dominated* as it could potentially lead to better super-paths. Therefore, a simple Dijkstra's based algorithm cannot be guaranteed to find the optimal MCOP path. This is due to the fact that the sub-structure optimality property [114] does not hold for the case of such a path weight structure. Such a greedy routing strategy actually belongs to the set of heuristics/approximation algorithms designed for the MCOP routing problem.

Definition 8 (β-approximation algorithm)**.** *Denoting by* p' *the MCOP shortest path,* $\mathcal{A}_\beta$ *is called a* β*-approximation algorithm of the MCOP problem if and only if for any instance of MCOP,* $\mathcal{A}_\beta$ *finds a path* p *such that* $w(p) \leq \beta w(p')$.

5.2 Motivation

Many heuristics and approximation algorithms have been proposed in the literature for addressing the MCOP problem. In particular, Xue *et al.* have mainly focused on the design of

Fully Polynomial Time Approximation Schemes (FPTAS) for the MCOP problem [13, 14, 15, 115] while Van Mieghem and Kuipers have investigated the complexity of exact routing solutions for the MCP problem [16, 17]. Finding feasible paths in general and the optimal path in particular requires the design of algorithms with very high complexity where each non-dominated path needs to be further examined as it can be part of the optimal path. This is due to the fact that sub-paths of an optimal path might not be optimal. Processing non-dominated path is not carried out in the Dijkstra's algorithm as illustrated in Figure 5.1. Indeed, for ensuring a polynomial time complexity, non shortest paths found at an intermediate node are no longer considered for future relaxation phases.

5.2.1 Strategies for finding feasible paths

The work in [16, 17] relies on the concept of path dominance. The proposed heuristic, SAMCRA, adopts a self-adaptive k-shortest path strategy to seek and store any non-dominated path to be further examined. The search space can then be reduced by testing if a given path is dominated by another or violate one of the path constraints. For instance, such a simple dominance condition can be considered: a path p is dominated by a path q if $\forall k \in 1, ..., K, w_k(p) \geq w_k(q)$. Such a dominance condition is not unique and others could be designed according to the edge weight distributions. The search space reduction level depends thus on the topology as well as the weight and constraint distributions. SAMCRA is thus sure to find a feasible path to the MCP problem. It can be complemented with another search space reduction technique called the *look-ahead* method [16]. This technique consists of computing at each node the shortest path trees relative to each weight component k. Doing so provides lower bound values that can be leveraged to reduce the SAMCRA execution time. Van Mieghem and Kuipers have proven SAMCRA to have a $\mathcal{O}(k_{max}|V|\log(k_{max}|V|)+k_{max}^2K|E|)$ worst-case computational complexity, with k_{max} being the maximum number of non-dominated paths to be stored at each node. Depending on the weights granularity, it is possible that k_{max} grows as $\mathcal{O}(|V|!) = \mathcal{O}(\exp(|V|\ln|V|)$ and SAMCRA algorithm ends up being not tractable. For SAMCRA to have a reasonable execution time, one can decide to limit the value of k_{max} at the cost of loss in the algorithm exactness. However, this prevents SAMCRA from having a theoretical performance guarantee. In [17], Van Mieghem and Kuipers claim that the MCP QoS routing problem is not NP-Complete in the strong sense, as the NP-hard behavior might appear only for a set of conditions to be jointly satisfied. For example, if the MCP constraints are set very high with respect to the edge weight values, it is likely that a feasible path will be found easily.

5.2.2 Strategies for approximating optimal MCOP paths

Finding a solution to the MCOP problem is even more difficult than for the MCP problem. The research community has investigated a lot the design of FPTAS algorithms whose performance is bounded by $1+\epsilon$, $\epsilon > 0$ and the associated worst-case computational complexity is

	K	Enforcing	Approximating	Guarantee	Worst-case time complexity
Lorenz and Raz [117]	$= 2$	Delay	Cost	$1+\epsilon$	$\mathcal{O}(\lvert V\rvert\lvert E\rvert(\log\log\lvert V\rvert + 1/\epsilon))$
Goel *et al.* [118]	$= 2$	Cost	Delay	$1+\epsilon$	$\mathcal{O}((\lvert E\rvert + \lvert V\rvert\log\lvert V\rvert)\mathcal{H}/\epsilon)$
K-Approx [14]	≥ 2	None	K constraints	K	$\mathcal{O}(K\lvert E\rvert + \lvert V\rvert\log\lvert V\rvert)$
Xue *et al.* [14]	≥ 2	None	K constraints	$1+\epsilon$	$\mathcal{O}(\lvert E\rvert(\lvert V\rvert/\epsilon)^{K-1})$
Xue *et al.* [14]	≥ 2	None	K constraints	$1+\epsilon$	$\mathcal{O}(\lvert E\rvert(\mathcal{H}/\epsilon)^{K-1} + \lvert V\rvert\log\lvert V\rvert)$
Greedy [15]	≥ 2	None	K constraints	K	$\mathcal{O}(K\lvert E\rvert + \lvert V\rvert\log\lvert V\rvert)$
Xue *et al.* [13]	≥ 2	1 constraint	K-1 constraints	$1+\epsilon$	$\mathcal{O}(\lvert E\rvert\lvert V\rvert\log\log\log\lvert V\rvert + \lvert E\rvert(\lvert V\rvert/\epsilon)^{K-1})$

Table 5.1 – Worst-case computational complexity of the most relevant multi-constrained approximation algorithms (borrowed and completed from [14])

polynomial in $1/\epsilon$. Most of the underlying optimization techniques have been inherited from historical 2-constrained MCP problems such as the *Delay Constrained Least Cost* problem (DCLC). In DCLC, one seeks a path with minimum cost and subjected to a given delay constraint. The approximation technique consists of executing a sequence of scaling, rounding and polynomial time approximate testing procedures introduced by Hassin in [116] and further improved by Lorenz and Raz in [117]. At each iteration step, an auxiliary graph is constructed with edge weights being scaled and rounded to some integer values. The solution is then improved iteratively by computing finer upper and lower bounds on the optimal path weight. Most of the solutions presented in Table 5.1 are basically built on adapting and/or improving the Hassin technique. In this table, $\mathcal{H}$ refers to the hop count of the longest computed path. The depicted K-approximation algorithms, *K-Approx* and *Greedy*, are simple adaptations of the Dijkstra's shortest path algorithm [114].

5.2.2.1 K-Approx

As depicted in Algorithm 4, K-approx simply applies the single-metric Dijkstra's shortest path algorithm using an auxiliary edge weight. For each edge e, the auxiliary weight is set to $w'(e) = \max_{1\leq k\leq K}[w_k(e)/W_k]$ and for any path p, $w'(p) = \sum_{\forall e\in p} w'(e)$. Therefore, K-approx presents the advantage of routing according to a right-monotonic and right-isotonic path weight metric in the auxiliary graph. In other words, it guarantees to find the actual shortest path in the auxiliary graph. It can thus be combined with the OLSR routing protocol.

5.2.2.2 Greedy

Unlike K-approx, Greedy does not use an auxiliary metric but the true MCOP metric. Greedy approximates the optimal MCOP path following the Dijkstra's shortest path algorithm as described in Algorithm 5. The worst-case computational complexity of Greedy is demonstrated in Theorem 10. Contrary to K-approx, Greedy is not guaranteed to return the shortest $s-t$ path in G with respect to the path weight structure defined so far for the MCOP problem.

Theorem 10 (Complexity of Greedy [15])**.** *Assuming a Fibonacci heap structure implementing the min-priority queue, Greedy has worst case time complexity* $\mathcal{O}(\lvert V\rvert\log\lvert V\rvert + K\lvert E\rvert)$.

Proof. The initialization phase (lines 2-8) takes $\mathcal{O}(K|V|)$. Using the Fibonacci heap structure, selecting each white node u with minimum length $\lambda[u]$ has amortized cost $\mathcal{O}(\log|V|)$. At each iteration of the while loop, only one node u is extracted from the set of white nodes, that results in every node u being selected at once during the algorithm. This also results in every edge considered in the *foreach* loop (line 16) being examined at once. Therefore, the algorithm executes at most $|E|$ relaxation phases. A relaxation phase (lines 17-21) consists of path weight calculations that take at most K and one decrease-key operation that has amortized cost $\mathcal{O}(1)$ when using the Fibonacci heap structure. Consequently, the worst case time complexity of Greedy is $\mathcal{O}(K|V| + |V|\log|V| + |E| \times (1 + K))$, that is $\mathcal{O}(|V|\log|V| + K|E|)$. □

Algorithm 4: K-Approx(G, s)

```
   Input  : A graph G(V, E) with non negative real-valued weights
   Input  : The estimated s-t shortest paths p[t] ∀t ∈ V\{s}
 1: // Initialization
 2: foreach node v ∈ V do
 3:     color[v] := White;
 4:     p[v] := null;
 5:     λ[v] := ∞;
 6: λ[s] := 0;
 7: // Main Loop
 8: while there are still white nodes in V do
 9:     // Get white node u with minimum length λ[u]
10:     u := Extract_Min_White(V);
11:     // Color Black the extracted node
12:     color[u] := Black;
13:     // Relax remaining white nodes from u
14:     foreach white node in V such that (u, v) ∈ E do
15:         λ_relax := λ[u] + max_{1≤k≤K} w_k(u,v)/W_k;
16:         if λ_relax < λ[v] then
17:             λ[v] := λ_relax;
18:             p[v] := p[u] ⊕ (u, v);
```

5.2.3 Application in practice

The approximation schemes presented in Table 5.1 have all been designed and evaluated in terms of worst-case computational complexity and evaluated numerically with weights randomly distributed over practical topologies such as NSFNET (14 nodes and 21 edges) [119], ARPANET (20 nodes and 32 edges) [120] and Italian National Network (33 nodes and 67 edges) [120]. However, no real implementation has been carried out for the general K-constrained routing. Only few works such as [18] have addressed protocol aspects of multi-constraint QoS routing in a networking environment. In [18], a MCP routing scheme has been implemented in the ns-2 simulator. The embedded MCP algorithm seeks feasible paths by using a k-shortest path

Algorithm 5: Greedy(G, s)

Input : A graph $G(V, E)$ with non negative real-valued weights
Input : The estimated s-t shortest paths $p[t]$ $\forall t \in V \backslash \{s\}$

// Initialization
foreach *node* $v \in V$ **do**
    $color[v] := White$;
    $p[v] := null$;
    $\lambda[v] := \infty$;
    $\forall k,\ 1 \leq k \leq K, d_k[v] := \infty$;
$\lambda[s] := 0$;
$\forall k,\ 1 \leq k \leq K, d_k[s] := 0$;
// Main Loop
while *there are still white nodes in* V **do**
    // Get white node u with minimum length $\lambda[u]$
    $u :=$ Extract_Min_White(V);
    // Color Black the extracted node
    $color[u] := Black$;
    // Relax remaining white nodes from u
    foreach *white node in* V *such that* $(u, v) \in E$ **do**
        $\lambda_{relax} := w(p[u] \oplus (u, v)) = \max_{1 \leq k \leq K} \frac{d_k[u] + w_k(u,v)}{W_k}$;
        if $\lambda_{relax} < \lambda[v]$ **then**
            $\lambda[v] := \lambda_{relax}$;
            $\forall k,\ 1 \leq k \leq K, d_k[v] := d_k[u] + w_k(u, v)$;
            $p[v] := p[u] \oplus (u, v)$;

technique. However, it does not try to approximate the optimal MCOP path. The study shows that the routing efficiency depends on the multiplicity of the QoS constraints as well as the routing protocol dynamics (link-state update interval).

As a result the approximation algorithms presented in Table 5.1 have not been implemented in practical systems and have been evaluated only in terms of algorithmic performance. As quality of service requirements can be expressed through K multiple constrained, we study in the following section possible implementations of the approximation algorithms presented so far.

5.3 Evaluation of Greedy for practical QoS routing

5.3.1 OLSR and K-constrained routing

OLSR [101] has been successfully adopted as a reference routing protocol for mobile ad hoc networks. It is an optimized proactive link-state protocol that uses multipoint relays (MPR) to reduce the amount of controlling message transmissions within the network. The MPR set relative to a given node is selected so that any 2-hop neighbour can be reached by forwarding the packets through one of the MPR. The MPR set calculation is done using the *HELLO* messages exchanged locally. They are necessary for discovering the local 2-hop neighborhood and

the associated link states (bidirectional or not). In particular, the MPRs broadcast, without repetition, Topology Control (TC) messages in the network. These messages provide sufficient information for computing shortest paths in a partial topology at least comprising the links betweens nodes and their MPR selectors. The shortest path is commonly computed using the Dijkstra's algorithm [114]. The resulting path is then used to update the node routing tables.

Both MPR selection and path calculation processes can be subjected to quality of service optimization [121]. Most works dedicated to augmenting OLSR with quality of service capabilities have addressed both issues. In these works, the underlying QoS path calculation was carried out by computing the widest shortest path (path with bandwidth guarantee and minimum delay). Such a path, however, differs from the general K-constrained MCOP path. In the following, instead, we study the applicability of general K-constrained routing schemes.

5.3.2 Selecting Greedy for computing the MCOP paths

The review of approximation algorithms provided in Table 5.1 for the general K-constrained problem has permitted to classify the solutions in the following two categories: K-approximations with low performance guarantees, but worst-case low computational cost, and $(1+\epsilon)$-approximations with high performance guarantees, but worst-case high computational cost.

Despite the large theoretical performance bounds established in [14, 15] , both *Greedy* and *K-Approx* are the most efficient algorithms in terms of worst-case computational complexity. This is due to the fact that these algorithms are simple adaptations of the Dijkstra's shortest path algorithm [114]. An interesting study carried out in [15] revealed that for small topologies with no more than 15 nodes, the best MCOP approximation algorithms with $\epsilon = 0.1$ presents execution time that is significantly higher than those of *Greedy* and *K-Approx*. They further showed that this execution time grows with the number of nodes. At the same time, they showed that K-Approx and Greedy perform very well compared to the $(1-\epsilon)$-approximation.

To evaluate the impact of weight distributions on the relative performance of Greedy and K-Approx, we performed large-scale simulations wherein we compared these algorithms to an exact MCOP algorithm based on SAMCRA. We implemented these three algorithms by extending the shortest path calculation modules provided in the Java JGraphT library [122], a free Java class library that provides mathematical graph-theory objects and algorithms. We generated random topologies composed of 10 to 100 nodes with multiple values of K and different weight distributions. We selected at random a set of 20 source-destination pairs and executed the path calculations. For the sake of clarity, as we focused on the shortest path weight errors rather than the path feasibility, we set $W_k = W = 1$ for all k. Figure 5.2 shows that Greedy, K-Approx and SAMCRA compute different paths for a given source-pair destination. SAMCRA computes the optimal MCOP path, colored red, as it takes into consideration any possible non-dominated path. For the typical experiment represented in this figure, where K was set to 3 and the weights uniformly distributed over (10,100), the relative errors in path weight for Greedy and K-Approx was 6.95% and 21.12% respectively.

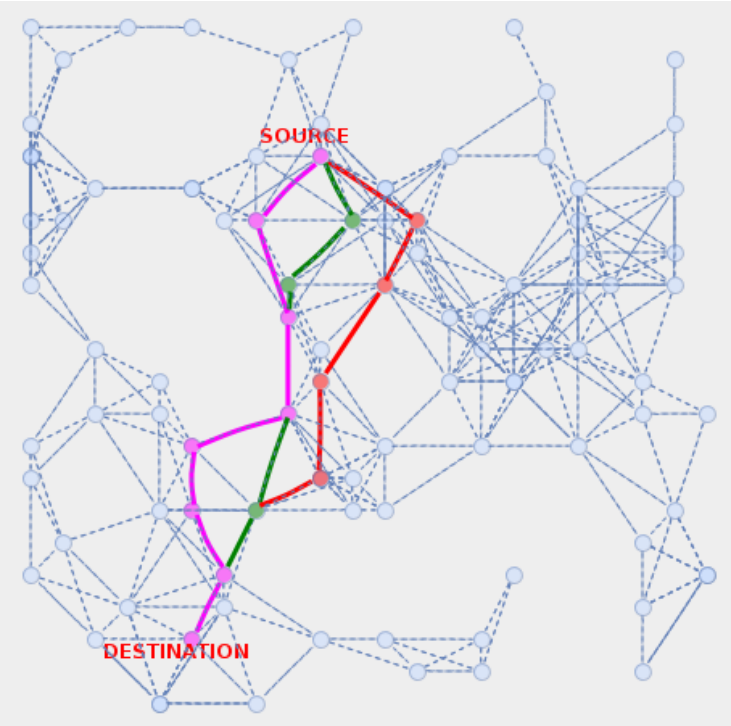


Figure 5.2 – Greedy and K-Approx do not necessarily find the optimal MCOP path

Our numerical analysis led to the same conclusions as in [15], in the sense that Greedy outperforms K-Approx and finds better paths. This results from K-Approx using a path weight that is an upper bound on the true MCOP path weight. Then, Greedy finds paths whose length errors relative to the optimal path are no more than 10% in average. We extend the analysis provided in [15] to show how the performance of Greedy is impacted by the weight distribution as well as the number of constraints. Figure 5.3 presents the error repartition in weight, computed over 100 runs, for the path computed by Greedy relative to the one computed by SAMCRA when $K = 3$ with different weight distributions. Our analysis shows that, when the weights are uniformly distributed over (0,100), the average error is of 7.22%. Over (10,100) we obtain 3.52%. Then 2.24% over (25,100) and finally 0.89% over (50,100). This analysis shows that Greedy can perform very well as it often finds the optimal path (at least 28% of the time in our experiments). Such a performance does not depend only on the value of K but also the max/min edge weight ratio. We performed the same analysis with weights distributed uniformly over (25,100) and used a different number of constraints. Figure 5.4 shows that the average performance of Greedy is not impacted a lot by the value of K. Furthermore, Greedy presents significantly lower execution time than the SAMCRA-based optimal algorithm and the practical results are far from the worst-case K-bound derived in [15]. This might be due to unexplored conditions that limit the effects of the path weight metric non-isotonicity. Therefore, due to its simplicity and good performance

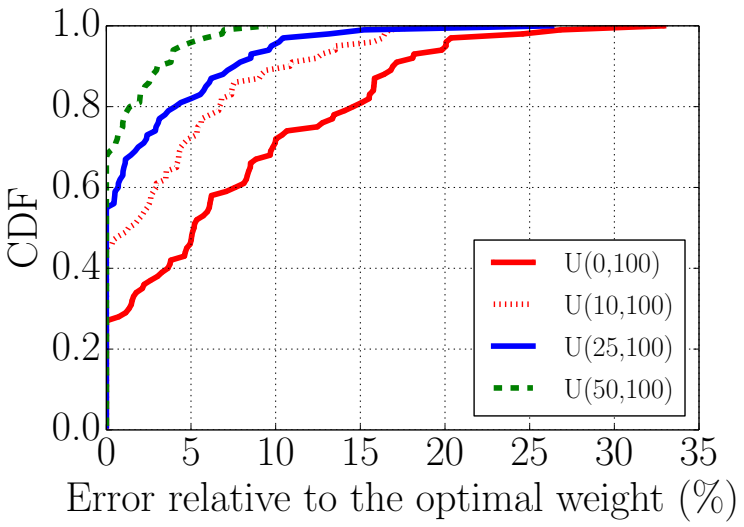


Figure 5.3 – Impact of weight distributions on the performance of Greedy

achieved in practice, we have integrated Greedy in the OLSR implementation deployed on our USRP testbed. Then, we integrated K-approx straightforwardly.

5.3.3 Greedy requires source routing

As the path weight structure defined by the MCOP metric is not right-isotonic, Greedy cannot be coupled with hop-by-hop routing [28]. Such a phenomenon is illustrated in Figure 5.5 wherein the QoS constraints are set to $W_1 = W_2 = 8$. While Greedy is able to compute the actual $A - E$ shortest path, that is $A \rightarrow B \rightarrow D \rightarrow E$, the shortest $B - E$ path is $B \rightarrow C \rightarrow E$. Therefore, with hop-by-hop routing, a packet is first transmitted from A to B and then relayed to C before reaching E. The packet is thus forwarded along the wrong path. This phenomenon has been identified as *inconsistency* problem [28] and could eventually lead to forwarding-loops. Consequently, Greedy must be coupled with a source routing scheme to avoid such issues.

5.3.4 Summary

We selected Greedy, a simple extension of the Dijkstra's shortest path algorithm, as the routing algorithm to be incorporated into OLSR because it presents satisfactory results in practice. While the best known theoretical bound of such scheme is a unique function of the number of constraints K, we showed that the achieved performance also depends on the edge weight distribution.

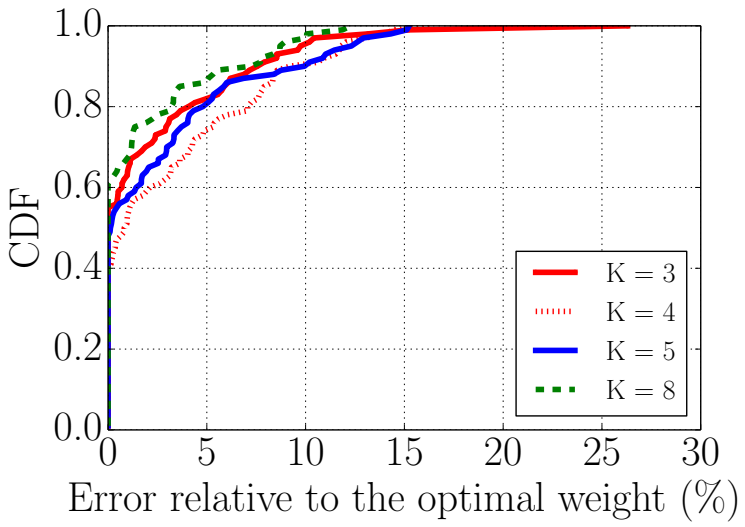


Figure 5.4 – Impact of the number of constraints on the performance of Greedy

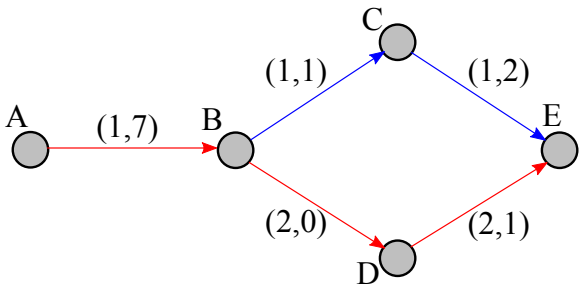


Figure 5.5 – Greedy must be coupled with source routing.

5.4 Testbed implementation

Implementing Greedy on our USRP testbed presented a double challenge. First, it required extending the open source OLSRd implementation [123] to handle vectorial link quality metrics. These metrics must be collected locally using cross-layer mechanisms. Second, OLSRd does not implement source routing. We thus implemented our own source routing and QoS metrics measurement modules which we then added to the IRIS architecture, as illustrated in Figure 5.6.

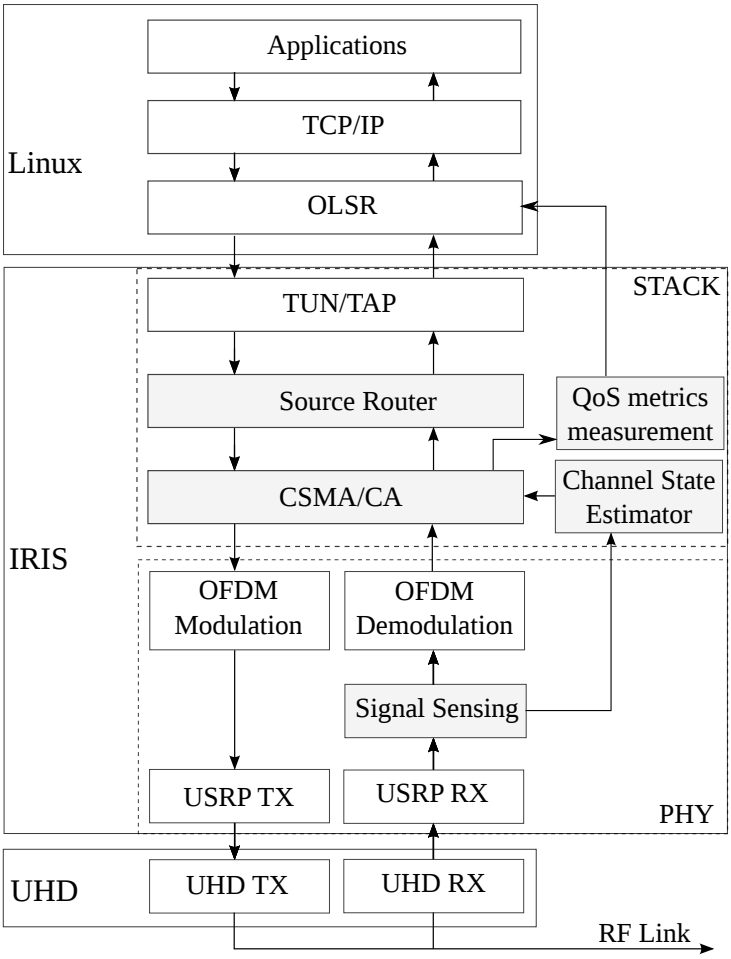


Figure 5.6 – Software Architecture for QoS routing

5.4.1 QoS metric measurements

We implemented a set of monitoring functions to capture multiple additive QoS routing metrics such as packet transmission delay, jitter, packet loss ratio or expected transmission count. Delay and jitter can be estimated by repeatedly executing a *ping* command and collecting the available statistics. Packet loss ratio and ETX are already implemented in OLSRd. However, as described in Chapter 4, ETX fails to capture the actual transmission count in cognitive radio networks. Therefore, OLSRd was extended to compute COExiST (Chapter 3) and use it as one of the QoS routing metrics. Finally, we developed an independent IRIS module that captures specific MAC layer metrics such as the average access time.

5.4.2 Multi-metric consideration

The standard OLSRd implementation comes with a *Link Quality* extension. It serves for computing ETX based on bidirectional packet reception ratio (PRR) information. As a result a single metric is exchanged between the nodes. However, as shown in algorithm 5, Greedy must consider every k components of the links reported in the topology. Therefore, we extended OLSRd so that it can exchange controlling messages embedding all necessary weight components, regardless of the value of K. Every weight is represented as a float and stored on 4 bytes. As a consequence, for any value of K, the amount of data relative to the vectorial metrics to be exchanged is $4 \times K$ bytes.

We extended the Dijkstra's shortest path calculation procedure to reflect the specification given in Algorithm 5. OLSRd is thus capable of computing, at each node, all source-destination paths corresponding to one or multiple sets of QoS constraints. However, it does not keep the whole calculated path but only stores the next hop information to construct the routing table and route packets on a hop-by-hop fashion. Up to this point, OLSRd is capable of computing the path relative to the operation given in Algorithm 5 at line 21. However, it does not modify the hop-by-hop forwarding scheme and this leads to two major issues. First, as demonstrated in Section 5.3.3, Greedy may lead to routing inconsistency with packets forwarded along a different path than the one computed at the source. Second, doing so does not suffice to describe the forwarding decision for flows associated with different QoS requirements and equivalently different end-to-end constraints.

Different strategies can be adopted to overcome these issues. One could mark every end-to-end flow with a specific id and complete the routing tables with entries depending on the flows id. While this approach might work in practice, it might suffer from large forwarding time when lots of flows have to be routed across the same node. In addition, it requires additional signalling procedures that might affect the overall protocol overhead. Instead we adopted a simpler strategy that consists of source routing the packets. Indeed, we can store the routes as a sequence of IP addresses and include it in the source routed packet header.

5.4.3 Source routing implementation

As mentioned in Section 5.3.3, source routing is necessary for avoiding routing inconsistency. Therefore, we looked for the best current implementations running on Linux systems. While source routing is considered in the specification of the IP protocol, it is not straightforward to enable such a feature in Linux. This is mainly due to the associated security issues. An attacker could easily modify the route inserted in the source routed packet and exploit routing threats. Only few source routing implementations are available. One has been built on the *Click Modular Router* [124] and used for the Dynamic Source Routing (DSR) protocol.

The Click implementation does not call specific system functions to do source routing. It defines and processes its own *DSR_header* structure. Exactly the same approach could be

implemented as an IRIS module. We adopted the last strategy for not overloading our testbed architecture. As illustrated in Figure 5.6, the resulting source routing architecture processes IP packets just before passing them onto the CSMA/CA block. While the source router is not located inside OLSR, we managed to exchange the route information between these two processes using some cross-layer mechanisms. The simplest consists of writing system-like routing tables that can be read from the source routing module and are composed of the computed paths for each constraint sets. According to the flow QoS requirements, the source router positions the route to the destination in a source routing packet header. Then, when the packet is forwarded across the network, the visited nodes simply modify the value of a pointer to the current node in the route.

5.4.4 QoS flow differentiation & proof of concept

Packets related to flows with different QoS requirements can be distinguished by the value of their *Type of Service* (TOS) field. The Source Router mentioned above is able to read this field and places the associated route in the packet source routing header. However, to get the corresponding route, it must read a routing table associated with the flow QoS requirements. This can be done in the Linux system using the *ip* command for defining policy routing tables according to the TOS value.

By doing so we are capable of sending two different IP traffic flows between a given source and destination pair. Each flow has been assigned a specific set of QoS constraints so that the related packets can be routed through two distinguished paths. Then, by modifying the level of Primary User activity or channel error on one topology link, we degrade the QoS metrics so that one of the traffic flow is re-routed through a different path, which is computed by Greedy according to the new values of the QoS metrics.

We then carried out exactly the same experiment by bypassing the source routing process and routing packets on a hop-by-hop fashion. Due to routing inconsistency, packets were not routed along the right path computed by Greedy and the performance expected at the end of the Greedy path calculation were not satisfied. This shows again the importance of source routing in QoS routing.

5.4.5 Summary

The Greedy algorithms represents a good approach for a practical implementation of multi-constrained QoS routing schemes. Paths computed by Greedy present satisfactory performance results compared to the optimal path computed by an exact MCOP algorithm. Moreover, as a simple extension of the Dijkstra's shortest path algorithm, Greedy can be integrated to any link state routing protocols, such as OSPF and OLSR. Nevertheless, doing so requires packets to be source routed along the computed paths to avoid routing inconsistencies and implement stable routing policies.

Despite its virtues in practice, we believe the best known upper bound on Greedy is far from tight. In particular, the best known bound depends on the number of weight constraints but not on the edge weight distributions or the constraints themselves. In the remaining of this chapter we aim at giving insights for investigating lower theoretical bounds.

5.5 Perspectives for evaluating the performance of greedy strategies

As stated in Section 5.2, the best theoretical performance bound of Greedy is a unique function of K. However, in Section 5.3.2, the performance of Greedy has been shown (by simulation) to depend on the weight distributions as well. This chapter provides some insights that could be used to investigate bounds computed as a function of the weight distributions.

5.5.1 Conditions that impact the performance of Greedy

As stated in Section 5.2, the Greedy algorithm fails in computing the optimal MCOP paths because it does not take into account potential non-dominated paths encountered along its execution. One may wonder which conditions might lead to the appearance of non-dominated paths.

Using the routing algebra defined in Section 5.1.1, we show in Lemma 5 that the non-dominance can be a function of the edges/paths to be further aggregated.

Lemma 5. *Considering two concurrent paths $(f, g) \in S^2$ such that $w(f) \leq w(g)$ and a path h such that $\forall k \in \{1, ..., K\}$, $w_k(h) = \sum_{e \in h} w_k(e) = W_k \mathcal{W}$, $\mathcal{W}$ being an arbitrary positive constant, we guarantee that $w(f \oplus h) \leq w(g \oplus h)$. In other words, f dominates g with respect to h.*

Proof. With such a scenario we have:

$$w(f \oplus h) = \max_{1 \leq k \leq K} \frac{\sum_{e \in f} w_k(e) + \sum_{e' \in h} w_k(e')}{W_k} = \max_{1 \leq k \leq K} \left[\frac{\sum_{e \in f} w_k(e)}{W_k} + \frac{\sum_{e' \in h} w_k(e')}{W_k} \right]$$

that is

$$w(f \oplus h) = \max_{1 \leq k \leq K} \left[\frac{\sum_{e \in f} w_k(e)}{W_k} \right] + \mathcal{W} = w(f) + \mathcal{W}$$

and similarly $w(g \oplus h) = w(g) + \mathcal{W}$. Therefore, as $w(f) \leq w(g)$ then $w(f \oplus g) \leq w(g \oplus h)$. □

Therefore, the weight distribution has an effect on the path weight structure isotonicity and the resulting performance of the Greedy routing scheme.

5.5.2 Obtaining a simple approximation algorithm based on Greedy

Applying the concept mentioned above, the simplest way to get a lower bound on the optimal MCOP path in G is to run Greedy in an auxiliary graph $G'(V, E')$ where $E' = E$ and $\forall e \in$

E', $w_k(e) := W_k \mathcal{W}_e$ with $\mathcal{W}_e = \min_{1 \leq k \leq K} [w_k(e)/W_k]$. Doing so ensures the path weight structure to be right-monotic and right-isotonic in G'. This right-isotonicity enables Greedy to find the MCOP shortest path in the modified graph G' in polynomial time.

Theorem 11. *The path weight structure in the modified graph G' is right-monotonic and right-isotonic.*

Proof. Every QoS metric in G' is additive. Therefore, the path length in G' will increase when aggregating a sub-path. In other words, the path weight structure is right-monotonic. Then, for any path h whose length is evaluated in G', we have:

$$w_k^{G'}(h) = \sum_{e \in h} w_k^{G'}(e) = \sum_{e \in h} W_k \min_{1 \leq k \leq K} [w_k(e)/W_k]$$

By denoting $\mathcal{W} = \sum_{e \in h} \min_{1 \leq k \leq K} [w_k(e)/W_k]$, we obtain $w_k^{G'}(h) = W_k \mathcal{W}$. From now, for any paths f, g such that $w^{G'}(f) \leq w^{G'}(g)$ we obtain $w^{G'}(f \oplus h) \leq w^{G'}(g \oplus h)$ by applying Lemma 5 and thus the path weight structure is right-isotonic in G'. □

Theorem 12. *The optimal path computed by Greedy in G' is a β-approximation of the optimal path p_{OPT} in G with β function of weights and constraints as follows:*

$$\beta = \frac{\max\limits_{\forall (e,k)} [w_k(e)/W_k]}{\min\limits_{\forall (e,k)} [w_k(e)/W_k]}$$

Proof. The proof simply consists of deriving an upper-bound on $w(p')$ in G and a lower bound on $w(p')$ in the auxiliary graph G'. Let H denote the hop count of path p', we have:

$$w^G(p') = \max_{1 \leq k \leq K} \frac{w_k^G(p')}{W_k} = \max_{1 \leq k \leq K} \frac{\sum_{\forall e \in p'} w_k^G(e)}{W_k} = \max_{1 \leq k \leq K} \sum_{\forall e \in p'} \frac{w_k^G(e)}{W_k} \leq H \max_{\forall (e,k)} \frac{w_k^G(e)}{W_k}$$

Similarly we obtain in the auxiliary graph G':

$$w^{G'}(p') = \max_{1 \leq k \leq K} \frac{w_k^{G'}(p')}{W_k} = \max_{1 \leq k \leq K} \sum_{\forall e \in p'} \frac{W_k \mathcal{W}_e}{W_k} \geq H \min_{\forall (e,k)} \frac{w_k^G(e)}{W_k}$$

As $\forall (e,k),\ w_k^{G'}(e) \leq w_k^G(e)$, we have $w^{G'}(p') \leq w^G(p_{OPT})$. Hereafter, we obtain:

$$w^G(p') H \min_{\forall (e,k)} \frac{w_k^G(e)}{W_k} \leq w^G(p') w^{G'}(p') \leq H \max_{\forall (e,k)} \frac{w_k^G(e)}{W_k} w^G(p_{OPT})$$

and thus in the initial graph G:

$$w^G(p') \leq \left(\frac{\max\limits_{\forall(e,k)} [w_k(e)/W_k]}{\min\limits_{\forall(e,k)} [w_k(e)/W_k]} \right) w^G(p_{OPT})$$

that concludes the proof. □

5.5.3 Insight for reducing the bound

Without loss of generality, we now consider the case of multi-constrained routing scenarios with the K constraints set to 1. This assumption does not change the difficulty of the MCOP problem and can be simply adapted to the more general case $W_k \neq 1$ by scaling the k^{th} edge weights using the following operation: $w_k(e) := w_k(e)/W_k$.

In the following we simply call an *update* operation the procedure consisting of changing the weights of a given edge in the initial graph $G(V, E)$ and/or the generated auxiliary graphs $G'(V, E')$. The update operation performed for the design of the simple approximation algorithm presented in Section 5.5.2 is relatively raw and leads to a bound that can be large in comparison to the K-bound derived so far for Greedy. In the following lemma, we propose a smoother update operation that locally breaks non-isotonicity.

Lemma 6. *Consider a set of n concurrent paths $f_1, f_2, ..., f_n$ such that $w(f_1) \leq w(f_2) \leq ... \leq w(f_n)$ and an edge l with K multiple weights $w_k(l),\ 1 \leq k \leq K$. By updating edge l so that*

$$\forall k \in \{1, ..., K\},\ w_k(l) := \min \left(w_k(l), w(f_2) - w_k(f_1) + \min_{1 \leq k \leq K} w_k(l) \right)$$

we ensure that $\forall i > 1,\ w(f_1 \oplus l) \leq w(f_i \oplus l)$.

Proof. By updating the edge weights this way, we obtain:

$$w(f_1 \oplus l) \leq \max_{1 \leq k \leq K} \left[w_k(f_1) + w(f_2) - w_k(f_1) + \min_{1 \leq k \leq K} w_k(l) \right] \leq \max_{1 \leq k \leq K} \left[w(f_j) + \min_{1 \leq k \leq K} w_k(l) \right]$$

leading to

$$w(f_1 \oplus l) \leq w(f_j) + \min_{1 \leq k \leq K} w_k(l) = \max_{1 \leq k \leq K} w_k(f_j) + \min_{1 \leq k \leq K} w_k(l) \leq \max_{1 \leq k \leq K} [w_k(f_j) + w_k(l)] = w(f_j \oplus l)$$

which concludes the proof. □

Iteratively updating edges during the relaxation phases of the Greedy routing algorithm could eventually lead to an approximation algorithm whose bound is better than the β derived in Section 5.5.2. A thorough analysis is required to demonstrate that doing so breaks the path weight structure non-isotonicity and guarantees finding the optimal path in G' or at least a lower bound on the optimal path length in G.

5.6 Conclusions

The Greedy routing algorithm, that extends the Dijkstra's shortest path algorithm, can be considered as a promising solution for multi-constrained quality of service routing. It achieves satisfactory results in practice and, due to its polynomial time complexity, can be integrated in any link state routing protocol, such as OLSR. The introduced control message overhead is linear with the number of constraints and is limited to few bytes. However, for the routing to be efficient, packets must be source routed. While such a forwarding scheme might present security risks, it is very convenient for the implementation of policy routing strategies. The latter can use the TOS field of the IP packets to define a class of Quality of Service requirements that is further associated with a set of end-to-end constraints.

The Greedy routing algorithm could be even more appreciated with better performance guarantees. An approach that could establish a tighter bound consists of updating some of the edge weights to break the path weight metric non-isotonicity. However, establishing a better bound than the currently best known, requires a thorough analysis to proove the optimality property of the paths computed in the processed auxiliary graphs, and is left for future work.

6 Conclusions and Perspectives

6.1 Conclusions

Cognitive Radio today is considered the most promising technology for addressing the spectrum scarcity challenge. One of its objectives is to exploit the best vacant frequency channels while minimizing the interference to the licensed users. While the standardization efforts have been focused on infrastructure-based solutions, we have mostly focused our attention on the distributed ad hoc architectures.

In cognitive radio ad hoc networks, a prerequisite for ensuring primary user protection is spectrum sensing. It can consist of measuring the power of the received signal and comparing it to a given threshold. Unfortunately, sensing is not perfect as Primary and Secondary networks are not synchronized. The resulting mutual Primary-Secondary interference, which does not exist for the case of legacy ad hoc networks, must be considered in the design of cross-layer quality of service schemes. As outlined in this thesis, not considering such effects can lead to poor estimations of quality of service metrics, such as the available bandwidth or the expected transmission count.

The available bandwidth is related to the way the communication resources are exploited and, therefore, it highly depends on the implemented medium access scheme. As there is no clear winner among the MAC strategies proposed for cognitive radio ad hoc networks, we have proposed to use a TDMA access scheme. The reason was two-fold: first, it is the most used MAC scheme in tactical networks and second, it is inherently more suited to the sensing synchronization requirements. However, after surveying the literature we found no algorithm that distributively computes the residual bandwidth on a given path. The solutions proposed are either specific to random access schemes, or based on centralized heuristics that must be run offline.

We proved that computing the maximum available bandwidth on a given path in TDMA-based cognitive radio networks is NP-complete. This is due to the maximal bandwidth being dependent on an optimal slot schedule. Consequently, we relaxed the calculation problem by considering a random slot allocation strategy. We showed that, given the simple nature of this allocation heuristic, the actual path available bandwidth calculation remains complex. Therefore, we proposed BRAND, a polynomial time algorithm that can accurately compute the available end-to-end bandwidth in a distributed fashion. We demonstrated the robustness of distributed

BRAND and showed its efficiency in doing admission control. We performed a thorough numerical analysis demonstrating the surprisingly good performance of randomized slot scheduling when compared to the best known offline allocation schemes.

We then proposed a novel cognitive radio routing metric by revisiting the expected transmission count calculation problem. In legacy networks, such a quantity is commonly computed using the ETX metric. As outlined in this thesis, ETX was shown to fail at capturing the actual transmission count in cognitive radio network. We also showed that simply considering primary users as a new source of independent packet errors does not perform well either. Therefore, using mathematical analysis, we derived COExiST, an approach capable of accurately estimating the actual transmission count in cognitive radio networks. Despite its complexity, COExiST was shown to have an elegant closed-form expression. We demonstrated that COExiST extends ETX by adding a term that depends on the primary user activity levels and durations as well as the MAC dynamics.

COExiST is additive and can be coupled with a link state routing protocol for finding paths with minimal transmissions count per packet. It can be extended and/or integrated into more sophisticated metrics that take into account channel diversity. Our experimental study showed that while COExiST is aimed at measuring the transmission count, it also permits to identify high-throughput paths, that state-of-the art cognitive routing solutions often fail to capture.

A flow can be associated with a set of quality of service requirements defined by several metrics such as transmission delay, jitter or any other cost function. Optimizing simultaneously all these criteria must be done subject to a set of constraint values that the metrics should not exceed. We studied the related multi-constrained routing problem and decided to implement one of these strategies on a USRP testbed. We selected a greedy routing strategy that simply extends the Dijkstra's shortest path algorithm and incorporated it into the OLSR routing protocol. While the multi-constrained routing problem is NP-complete, such a greedy heuristic was shown to perform well in practice. Furthermore, we showed that for the routing to be consistent the packets have to be source routed.

Finally, our study showed that the performance of the greedy strategy can be significantly better than what the state of the art bounds on its performance seem to indicate. In particular, the best known bound depends on the number of constraints only but our data showed that the performance also depends on the weight distribution, which could point to a path for finding tighter bounds in the future.

6.2 Perspectives

The work realized in this thesis opens up several research directions. First of all, routing in the heterogeneous Tactical Internet could be unified through an architecture composed of technology-dependent (TD) and technology-independent (TI) blocks. The TD layer would focus on the estimation of QoS metrics relative to the virtual links in the overlay network while the

TI layer would compute, for a given QoS flow, the sequence of gateways to go through by using multi-constrained routing. It would also perform admission control using the available end-to-end bandwidth estimation.

It is likely that successive heterogeneous networks do not interfere with each others. As a result the available bandwidth along a path in the overlay network could simply equal the minimum bandwidth of the virtual links composing the path. Therefore, BRAND and the techniques developed so far for estimating the available end-to-end bandwidth in random access networks could be leveraged to estimate the end-to-end bandwidth inside the heterogeneous networks.

BRAND has been designed to perform admission control. However, it could also be used to find the path with maximal available end-to-end bandwidth. The main challenge is that of the available bandwidth is not an isotonic metric due to intra-path interference. Therefore, a simple greedy routing strategy based on BRAND cannot be guaranteed to find such a path.

Therefore, the approach introduced in this thesis for evaluating the performance of greedy routing algorithms could be extended to the general case of non-isotonic metrics. We could start designing an exact algorithm that processes any non-dominated path to analyze the conditions that impact the performance of such greedy schemes. Then, we could develop strategies to evaluate the impact of the metrics on non-isotonicity and derive theoretical bounds that guarantee a certain performance level.

In this thesis, we proposed an algorithm for estimating the available end-to-end bandwidth in cognitive radio networks. Such an algorithm is based on a random slot allocation process. While such a randomized scheduling has been shown to perform well in practice compared to an optimal slot allocation scheme, it is not yet provided with any performance guarantees. One research direction could thus be to precisely provide such guarantees.

Finally, the performance of COExiST has been evaluated on small network topologies comprising USRP radios. Our perspective is to extend the analysis of COExiST to large-scale topologies by using network simulation tools. At the same time, it would be interesting to study the impact of approximating the quantities $\overline{T}_r$ and $\overline{T}_t$ on the performance of COExiST. Indeed, doing so would permit to compute COExiST without measuring these quantities at the MAC layer.

A Complements for computing BRAND

This appendix provides additional materials helpful for the understanding of BRAND (Chapter 3) as well as the integer program formulation used in Section 3.4.

A.1 General expression of the update equation

When processing the third Interference-Clique, we observed that:

$$C^{(1)}_{3,4,5} = \underbrace{C^{(1)}_{1,2,3,4,5}}_{\substack{\text{impacted by allocations}\\ \text{on } l_1 \text{ and then } l_2}} + \underbrace{C^{(1)}_{1,\overline{2},3,4,5}}_{\substack{\text{impacted by allocations}\\ \text{on } l_1 \text{ but not } l_2}} + \underbrace{C^{(1)}_{\overline{1},2,3,4,5}}_{\substack{\text{impacted by allocations}\\ \text{on } l_2 \text{ but not } l_1}} + \underbrace{C^{(1)}_{\overline{1},\overline{2},3,4,5}}_{\text{not impacted}} \tag{A.1}$$

that leads to:

$$C^{(3)}_{3,4,5} = C^{(1)}_{3,4,5} - \left[p_1 + p_2(1-p_1)\right].C^{(1)}_{1,2,3,4,5} - p_1.C^{(1)}_{1,\overline{2},3,4,5} - p_2.C^{(1)}_{\overline{1},2,3,4,5} \tag{A.2}$$

Measuring the impact of allocations on l_1 and l_2 is equivalent to transferring slots from one set to another. Indeed, from the previous equations, we can conclude that on average $p_1.\left[C^{(1)}_{1,2,3,4,5} + C^{(1)}_{1,\overline{2},3,4,5}\right]$ slots and $\left[p_2(1-p_1).C^{(1)}_{1,2,3,4,5} + \ p_2.C^{(1)}_{\overline{1},2,3,4,5}\right]$ slots in the set $E^{(1)}_{3,4,5}$ are respectively reserved on links l_1 and l_2. Due to the 2-hop nature of the interference, when updating the sets resulting from the 3-link decomposition related to clique 3, on average $p_1.\left[C^{(1)}_{1,2,3,4,5} + C^{(1)}_{1,\overline{2},3,4,5}\right]$ slots from the set $E^{(1)}_{3,4,5}$ are transferred to the set $E^{(1)}_{\overline{3},4,5}$ and $\left[p_2(1-p_1).C^{(1)}_{1,2,3,4,5} + p_2.C^{(1)}_{\overline{1},2,3,4,5}\right]$ to the set $E^{(1)}_{\overline{3},\overline{4},5}$. Every set resulting from the 3-link available slot set decomposition related to clique 3 is then similarly updated.

More generally speaking, when processing the i^{th} clique, the influence of allocations on the two previous links can be correctly considered by updating the sets resulting from its 3-link available slot set decomposition as follows:

$$\mathbf{C}_i^{(i)} = \mathbf{C}_i^{(1)} - \mathbf{p}_i \times \mathbf{I}_i + \mathbf{u}_i + \mathbf{v}_i \tag{A.3}$$

where

$$\mathbf{C}_i^{(j)} = \begin{pmatrix} C^{(j)}_{\overline{i},\overline{i+1},\overline{i+2}} & C^{(j)}_{i,\overline{i+1},\overline{i+2}} & \cdots & C^{(j)}_{i,i+1,i+2} \end{pmatrix}$$
$$\mathbf{p}_i = \begin{pmatrix} p_{i-2} & p_{i-1} & [p_{i-2} + p_{i-1}.(1-p_{i-2})] \end{pmatrix}$$

and

$$\mathbf{I}_i = \begin{pmatrix} C^{(i-2)}_{i-2,\overline{i-1},\overline{i},\overline{i+1},\overline{i+2}} & \cdots & C^{(i-2)}_{i-2,\overline{i-1},i,i+1,i+2} \\ C^{(i-2)}_{\overline{i-2},i-1,\overline{i},\overline{i+1},\overline{i+2}} & \cdots & C^{(i-2)}_{\overline{i-2},i-1,i,i+1,i+2} \\ C^{(i-2)}_{i-2,i-1,\overline{i},\overline{i+1},\overline{i+2}} & \cdots & C^{(i-2)}_{i-2,i-1,i,i+1,i+2} \end{pmatrix} \tag{A.4}$$

The vector $\mathbf{u}_i$ serves to compensate a set that does not suffer from all of the interferences. The vector $\mathbf{v}_i$ is then used to update the sets receiving slots becoming unavailable for reservation on certain links. The non zero values of vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ depend on some variables used in $\mathbf{p}_i$ and $\mathbf{I}_i$ and are given in the following:

$$u_1 = p_{i-2}.C^{(i-2)}_{i-2,\overline{i-1},\overline{i},\overline{i+1},\overline{i+2}} + p_{i-1}.C^{(i-2)}_{\overline{i-2},i-1,\overline{i},\overline{i+1},\overline{i+2}} + [p_{i-2} + p_{i-1}.(1-p_{i-2})].C^{(i-2)}_{i-2,i-1,\overline{i},\overline{i+1},\overline{i+2}}$$
$$u_3 = p_{i-2}.\left[C^{(i-2)}_{i-2,\overline{i-1},\overline{i},i+1,\overline{i+2}} + C^{(i-2)}_{i-2,i-1,\overline{i},i+1,\overline{i+2}}\right]$$
$$u_4 = p_{i-2}.C^{(i-2)}_{i-2,\overline{i-1},\overline{i},\overline{i+1},i+2} + p_{i-1}.C^{(i-2)}_{\overline{i-2},i-1,\overline{i},\overline{i+1},i+2} + [p_{i-2} + p_{i-1}.(1-p_{i-2})].C^{(i-2)}_{i-2,i-1,\overline{i},\overline{i+1},i+2}$$
$$u_7 = p_{i-2}.\left[C^{(i-2)}_{i-2,\overline{i-1},\overline{i},i+1,i+2} + C^{(i-2)}_{i-2,i-1,\overline{i},i+1,i+2}\right]$$

and

$$v_1 = p_{i-1}.\left[C^{(i-2)}_{\overline{i-2},i-1,\overline{i},i+1,\overline{i+2}} + C^{(i-2)}_{\overline{i-2},i-1,i,i+1,\overline{i+2}} + C^{(i-2)}_{\overline{i-2},i-1,i,\overline{i+1},\overline{i+2}}\right] + p_{i-2}.C^{(i-2)}_{i-2,\overline{i-1},i,\overline{i+1},\overline{i+2}}$$
$$+ [p_{i-2} + p_{i-1}.(1-p_{i-2})].C^{(i-2)}_{i-2,i-1,i,\overline{i+1},\overline{i+2}} + p_{i-1}.(1-p_{i-2}).\left[C^{(i-2)}_{i-2,i-1,i,i+1,\overline{i+2}} + C^{(i-2)}_{i-2,i-1,\overline{i},i+1,\overline{i+2}}\right]$$
$$v_3 = p_{i-2}.\left[C^{(i-2)}_{i-2,\overline{i-1},i,i+1,\overline{i+2}} + C^{(i-2)}_{i-2,i-1,i,i+1,\overline{i+2}}\right]$$
$$v_4 = p_{i-2}.\left[C^{(i-2)}_{i-2,\overline{i-1},i,\overline{i+1},i+2} + C^{(i-2)}_{i-2,i-1,i,\overline{i+1},i+2}\right] + p_{i-1}.\left[C^{(i-2)}_{\overline{i-2},i-1,i,\overline{i+1},i+2} + C^{(i-2)}_{\overline{i-2},i-1,\overline{i},i+1,i+2}\right.$$
$$\left. + C^{(i-2)}_{\overline{i-2},i-1,i,i+1,i+2}\right] + p_{i-1}.(1-p_{i-2}) \times \left[C^{(i-2)}_{i-2,i-1,i,\overline{i+1},i+2} + C^{(i-2)}_{i-2,i-1,\overline{i},i+1,i+2} + C^{(i-2)}_{i-2,i-1,i,i+1,i+2}\right]$$
$$v_7 = p_{i-2}.\left[C^{(i-2)}_{i-2,\overline{i-1},i,i+1,i+2} + C^{(i-2)}_{i-2,i-1,i,i+1,i+2}\right]$$

A.2 Approximation in the update equation

In the following, we illustrate the approximation scheme to be used when processing any clique $i \geq 4$. We now show how to characterize a set used in the clique i set update equation, say $E^{(i-2)}_{i-2,\overline{i-1},i,i+1,i+2}$, as a function of its initial state. We start by doing the 2-link available slot set decomposition related to l_{i-2} and l_{i-1}. This decomposition leads to four disjointed sets: $E^{(j)}_{\overline{i-2},\overline{i-1}}$,$E^{(j)}_{\overline{i-2},i-1}$,$E^{(j)}_{i-2,\overline{i-1}}$ and $E^{(j)}_{i-2,i-1}$. Let us work on the third one, that is $E^{(j)}_{i-2,\overline{i-1}}$. This set can also be divided in eight disjointed subsets resulting from the 3-link available slot set decomposition of clique i. This time, the slot space equals $E^{(j)}_{i-2,\overline{i-1}}$ rather than $\{1,2,...,S.B\}$, leading to subsets of the following form $E^{(j)}_{i-2,\overline{i-1},i,i+1,i+2}$, taken as an example. Such a decomposition is also illustrated in Figure A.1 for the set $E^{(j)}_{i-2,i-1}$.

A property of the set $E^{(j)}_{i-2,\overline{i-1}}$ is that along the estimation process, it can only transfer slots to the set $E^{(j)}_{\overline{i-2},\overline{i-1}}$ and cannot receive slots from another. Therefore, the number of slots that initially belonged to the set $E^{(j)}_{i-2,\overline{i-1}}$ and had become unavailable just before node n_{i-2} did its reservations for communication on link l_{i-2} equals $C^{(1)}_{i-2,\overline{i-1}} - C^{(i-2)}_{i-2,\overline{i-1}}$. These slots had become unavailable due to allocations on l_{i-4} and l_{i-3}. Because of the random nature of the slot reservation process, these slots were taken uniformly at random among the subsets partitioning $E^{(j)}_{i-2,\overline{i-1}}$. We can thus represent the number of slots that had become unavailable in the set $E^{(j)}_{i-2,\overline{i-1},i,i+1,i+2}$ with the discrete random variable $X_{i-2,\overline{i-1},i,i+1,i+2}$ taking its values in the set $\{0,...,C^{(1)}_{i-2,\overline{i-1}} - C^{(i-2)}_{i-2,\overline{i-1}}\}$ and following an hypergeometric distribution with parameters $(C^{(1)}_{i-2,\overline{i-1}}, C^{(1)}_{i-2,\overline{i-1},i,i+1,i+2}, C^{(1)}_{i-2,\overline{i-1}} - C^{(i-2)}_{i-2,\overline{i-1}})$. From this identification we can deduce that, for $C^{(1)}_{i-2,\overline{i-1}}$ strictly positive, the average value of this random variable equals $[(C^{(1)}_{i-2,\overline{i-1}} - C^{(i-2)}_{i-2,\overline{i-1}})/C^{(1)}_{i-2,\overline{i-1}}] \times C^{(1)}_{i-2,\overline{i-1},i,i+1,i+2}$.

More generally, just before node n_{i-2} did its allocations, an average proportion $C^{(i-2)}_{i-2,\overline{i-1}}/C^{(1)}_{i-2,\overline{i-1}}$ of the initially available slots remained available in every subset partitioning $E^{(j)}_{i-2,\overline{i-1}}$. Hereafter, the quantity $C^{(i-2)}_{i-2,\overline{i-1},i,i+1,i+2}$ can be correctly evaluated as follows:

$$C^{(i-2)}_{i-2,\overline{i-1},i,i+1,i+2} = C^{(1)}_{i-2,\overline{i-1},i,i+1,i+2} \times \alpha_{i-2,\overline{i-1}} \tag{A.5}$$

where the reduction factor of the set $E^{(j)}_{i-2,\overline{i-1}}$ equals

$$\alpha_{i-2,\overline{i-1}} = \begin{cases} 0 \quad \text{, if } C^{(1)}_{i-2,\overline{i-1}} = 0 \\ \frac{C^{(i-2)}_{i-2,\overline{i-1}}}{C^{(1)}_{i-2,\overline{i-1}}} = \frac{C^{(i-2)}_{i-2,\overline{i-1},\bar{i}} + C^{(i-2)}_{i-2,\overline{i-1},i}}{C^{(1)}_{i-2,\overline{i-1},\bar{i}} + C^{(1)}_{i-2,\overline{i-1},i}} & \text{, else} \end{cases} \tag{A.6}$$

and is related to clique $(i-2)$ as it can be computed at the beginning of the process of this clique.

The same analysis can be carried out for the two other sets of interest $E^{(j)}_{\overline{i-2},i-1}$ and $E^{(j)}_{i-2,i-1}$. However, it differs a little for $E^{(j)}_{\overline{i-2},i-1}$ when $C^{(1)}_{\overline{i-2},i-1} = 0$ as this set can receive slots from $E^{(j)}_{i-2,i-1}$ due to allocations on previous links. For this case, to correctly update the resulting subsets, we compute the proportion of slots transferred from $E^{(j)}_{i-2,i-1}$ to $E^{(j)}_{\overline{i-2},i-1}$. The quantity $C^{(i-2)}_{\overline{i-2},i-1,i,i+1,i+2}$ can be correctly evaluated as follows:

$$C^{(i-2)}_{\overline{i-2},i-1,i,i+1,i+2} = C^{(1)}_{i-2,i-1,i,i+1,i+2} \times \tau_{i-2} \tag{A.7}$$

with

$$\tau_{i-2} = \frac{C^{(i-2)}_{\overline{i-2},i-1}}{C^{(1)}_{i-2,i-1}} = \frac{C^{(i-2)}_{\overline{i-2},i-1,\bar{i}} + C^{(i-2)}_{\overline{i-2},i-1,i}}{C^{(1)}_{i-2,i-1,\bar{i}} + C^{(1)}_{i-2,i-1,i}} \tag{A.8}$$

if $C^{(1)}_{i-2,i-1}$ is strictly positive, and zero otherwise.

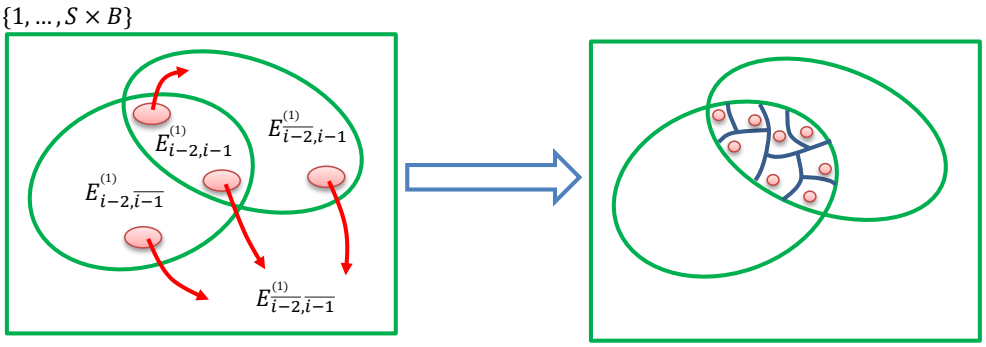


Figure A.1 – Clique reduction approximation

A.3 Integer program formulation

For the case of 4-hop paths with already allocated slots, we have used *lpsolve 5.5*[104] to valuate the maximum available bandwidth provided by an optimal slot allocation process. The optimal solution can be found by realizing the objective defined in the linear integer program described herebelow.

$$
\begin{aligned}
\underset{\mathbf{x}}{\text{maximize}} \quad & \min\left(\mathbf{P}.\left\{\mathbf{A}.\mathbf{x}+\mathbf{b}\right\}\right) \\
\text{subject to} \quad & x_1 + x_2 \leq C^{(1)}_{1,2,\overline{3},\overline{4}}, \\
& x_3 + x_4 \leq C^{(1)}_{1,\overline{2},3,\overline{4}}, \\
& x_5 + x_6 \leq C^{(1)}_{\overline{1},2,3,\overline{4}}, \\
& x_7 + x_8 \leq C^{(1)}_{\overline{1},2,\overline{3},4}, \\
& x_9 + x_{10} \leq C^{(1)}_{\overline{1},\overline{2},3,4}, \\
& x_{11} + x_{12} + x_{13} \leq C^{(1)}_{1,2,3,\overline{4}}, \\
& x_{14} + x_{15} \leq C^{(1)}_{1,2,\overline{3},4}, \\
& x_{16} + x_{17} \leq C^{(1)}_{1,\overline{2},3,4}, \\
& x_{18} + x_{19} + x_{20} \leq C^{(1)}_{\overline{1},2,3,4}, \\
& x_{21} + x_{22} + x_{23} \leq C^{(1)}_{1,2,3,4}, \\
& x_i \text{ integer and positive.}
\end{aligned}
$$

with the following variable definitions

$$\mathbf{x} = (x_1, x_2, \ldots, x_{23})^T$$

$$\mathbf{b} = \begin{pmatrix} C^{(1)}_{1,\overline{2},\overline{3},\overline{4}} + C^{(1)}_{1,\overline{2},\overline{3},4} \\ C^{(1)}_{\overline{1},2,\overline{3},\overline{4}} \\ C^{(1)}_{\overline{1},\overline{2},3,\overline{4}} \\ C^{(1)}_{\overline{1},\overline{2},\overline{3},4} + C^{(1)}_{1,\overline{2},\overline{3},4} \end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix} 1\,0\,1\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,1\,0\,1\,0\,0\,0\,0\,1\,0\,0 \\ 0\,1\,0\,0\,1\,0\,1\,0\,0\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0\,0\,1\,0 \\ 0\,0\,0\,1\,0\,1\,0\,0\,1\,0\,0\,0\,1\,0\,0\,0\,1\,0\,1\,0\,0\,0\,1 \\ 0\,0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,0\,0\,1\,0\,1\,0\,0\,0\,1\,1\,0\,0 \end{pmatrix}$$

and

$$\mathbf{P} = \begin{pmatrix} \phi_1 f_1/S & 0 & 0 & 0 \\ 0 & \phi_2 f_2/S & 0 & 0 \\ 0 & 0 & \phi_3 f_3/S & 0 \\ 0 & 0 & 0 & \phi_4 f_4/S \end{pmatrix}$$

Publications

International Conferences

[1] G. Artero Gallardo, G. Jakllari, L. Canourgues, and A.-L. Beylot. "On estimating the end-to-end bandwidth in multi-transceiver multi-hop cognitive radio networks (regular paper)," in *ACM Workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks - PM^2HW^2N*, pp.117-124, Barcelona, 03/11/2013-08/11/2013. http://dx.doi.org/10.1145/2512840.2512856

International Conferences (under review)

[1] G. Artero Gallardo, J.-G. Krieg, G. Jakllari, L. Canourgues, and A.-L. Beylot. "COExiST: Revisiting transmission count for cognitive radio networks".

International Journals (under revision)

[1] G. Artero Gallardo, G. Jakllari, L. Canourgues, and A.-L. Beylot. "On estimating the end-to-end bandwidth in multi-transceiver multi-hop cognitive radio networks (Extended version)," in *IEEE Transactions on Mobile Computing*.

Bibliography

[1] L. Cerdà-Alabern, A. Neumann, and P. Escrich, "Experimental evaluation of a wireless community mesh network," in *Proceedings of the 16th ACM International Conference on Modeling, Analysis & Simulation of Wireless and Mobile Systems*, ser. MSWiM '13. New York, NY, USA: ACM, 2013, pp. 23–30. [Online]. Available: http://doi.acm.org/10.1145/2507924.2507960

[2] B. Wang and K. Liu, "Advances in cognitive radio networks: A survey," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 5–23, feb. 2011.

[3] "Evaluation of the Performance of Prototype TV-Band White Space Devices," *FCC press release*, November 2008. [Online]. Available: http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-08-260A6.pdf

[4] P. Bahl, R. Chandra, T. Moscibroda, R. Murty, and M. Welsh, "White space networking with wi-fi like connectivity," in *ACM SIGCOMM 2009*. New York, NY, USA: ACM, 2009, pp. 27–38.

[5] I. Mitola, J. and J. Maguire, G.Q., "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, aug 1999.

[6] "Report to the president: Realizing the full potential of government-held spectrum to spur economic growth," *President's Council of Advisors on Science and Technology*, July 2012. [Online]. Available: http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast_spectrum_report_final_july_20_2012.pdf

[7] J. Chapin and V. Chan, "Architecture Concepts for a Future Heterogeneous, Survivable Tactical Internet," in *IEEE Military Communications Conference*, November 2013.

[8] L. Canourgues, "Routing Algorithms in Large Scale Tactical Mobile Ad Hoc Networks," PhD Thesis, Institut National Polytechnique de Toulouse, Toulouse, France, May 2008.

[9] "Workshop on Future Heterogeneous Networks," NASA Ames Research Center, Moutain View, California, Mar. March 2011. [Online]. Available: http://www.rle.mit.edu/futurehetnets/

[10] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris, "Resilient overlay networks," in *SOSP*, 2001, pp. 131–145. [Online]. Available: http://doi.acm.org/10.1145/502034.502048

[11] T. Clausen, C. Dearlove, P. Jacquet, and U. Herberg, "The Optimized Link State Routing Protocol Version 2," RFC 7181 (Proposed Standard), Internet Engineering Task Force, Apr. 2014. [Online]. Available: http://www.ietf.org/rfc/rfc7181.txt

[12] C. Dearlove, T. Clausen, and P. Jacquet, "Link Metrics for the Mobile Ad Hoc Network (MANET) Routing Protocol OLSRv2 - Rationale," RFC 7185 (Informational), Internet Engineering Task Force, Apr. 2014. [Online]. Available: http://www.ietf.org/rfc/rfc7185.txt

[13] G. Xue, W. Zhang, J. Tang, and K. Thulasiraman, "Polynomial time approximation algorithms for multi-constrained qos routing," *IEEE/ACM Trans. Netw.*, vol. 16, no. 3, pp. 656–669, 2008. [Online]. Available: http://doi.acm.org/10.1145/1399562.1399575

[14] G. Xue, A. Sen, W. Zhang, J. Tang, and K. Thulasiraman, "Finding a path subject to many additive qos constraints," *IEEE/ACM Trans. Netw.*, vol. 15, no. 1, pp. 201–211, 2007. [Online]. Available: http://doi.acm.org/10.1145/1241832.1241848

[15] G. Xue and W. Zhang, "Multiconstrained qos routing: Greedy is good," in *Proceedings of the Global Communications Conference, 2007. GLOBECOM '07, Washington, DC, USA, 26-30 November 2007*, 2007, pp. 1866–1871. [Online]. Available: http://dx.doi.org/10.1109/GLOCOM.2007.359

[16] P. V. Mieghem and F. A. Kuipers, "Concepts of exact qos routing algorithms," *IEEE/ACM Trans. Netw.*, vol. 12, no. 5, pp. 851–864, 2004. [Online]. Available: http://dx.doi.org/10.1109/TNET.2004.836112

[17] F. A. Kuipers and P. V. Mieghem, "Conditions that impact the complexity of qos routing," *IEEE/ACM Trans. Netw.*, vol. 13, no. 4, pp. 717–730, 2005. [Online]. Available: http://dx.doi.org/10.1109/TNET.2005.852882

[18] V. D. Nguyen, T. Begin, and I. G. Lassous, "Multi-constrained routing algorithm: A networking evaluation," in *IEEE 37th Annual Computer Software and Applications Conference, COMPSAC Workshops 2013, Kyoto, Japan, July 22-26, 2013*, 2013, pp. 719–723. [Online]. Available: http://dx.doi.org/10.1109/COMPSACW.2013.107

[19] C. Dovrolis, P. Ramanathan, and D. Moore, "Packet-dispersion techniques and a capacity-estimation methodology," *IEEE/ACM Trans. Netw.*, vol. 12, no. 6, pp. 963–977, 2004. [Online]. Available: http://dx.doi.org/10.1109/TNET.2004.838606

[20] M. Garetto, T. Salonidis, and E. W. Knightly, "Modeling per-flow throughput and capturing starvation in CSMA multi-hop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 4, pp. 864–877, 2008. [Online]. Available: http://doi.acm.org/10.1145/1453698.1453708

[21] T. Salonidis, M. Garetto, A. Saha, and E. W. Knightly, "Identifying high throughput paths in 802.11 mesh networks: a model-based approach," in *Proceedings of the IEEE International Conference on Network Protocols, ICNP 2007, October 16-19, 2007, Beijing, China*, 2007, pp. 21–30. [Online]. Available: http://dx.doi.org/10.1109/ICNP.2007.4375833

[22] N. V. Nguyen, I. G. Lassous, V. Moraru, and C. Sarr, "Retransmission-based available bandwidth estimation in IEEE 802.11-based multihop wireless networks," in *Proceedings of the 14th International Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems, MSWiM 2011, Miami, Florida, USA, October 31 - November 4, 2011*, 2011, pp. 377–384. [Online]. Available: http://doi.acm.org/10.1145/2068897.2068961

[23] C. Sarr, C. Chaudet, G. Chelius, and I. G. Lassous, "Bandwidth estimation for IEEE 802.11-based ad hoc networks," *IEEE Trans. Mob. Comput.*, vol. 7, no. 10, pp. 1228–1241, 2008. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TMC.2008.41

[24] T. Abreu, B. Baynat, T. Begin, and I. G. Lassous, "Hierarchical modeling of IEEE 802.11 multi-hop wireless networks," in *16th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, MSWiM '13, Barcelona, Spain, November 3-8, 2013*, 2013, pp. 143–150. [Online]. Available: http://doi.acm.org/10.1145/2507924.2507949

[25] T. Abreu, B. Baynat, T. Begin, I. G. Lassous, and N. Nguyen, "Modeling of IEEE 802.11 multi-hop wireless chains with hidden nodes," in *17th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, MSWiM'14, Montreal, QC, Canada, September 21-26, 2014*, 2014, pp. 159–162. [Online]. Available: http://doi.acm.org/10.1145/2641798.2641826

[26] D. S. J. D. Couto, D. Aguayo, J. C. Bicket, and R. Morris, "A high-throughput path metric for multi-hop wireless routing," in *Proceedings of the Ninth Annual International Conference on Mobile Computing and Networking, MOBICOM 2003, 2003, San Diego, CA, USA, September 14-19, 2003*, 2003, pp. 134–146. [Online]. Available: http://doi.acm.org/10.1145/938985.939000

[27] R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking, MOBICOM 2004, 2004, Philadelphia, PA, USA, September 26 - October 1, 2004*, 2004, pp. 114–128. [Online]. Available: http://doi.acm.org/10.1145/1023720.1023732

[28] Y. Yang and J. Wang, "Design guidelines for routing metrics in multihop wireless networks," in *INFOCOM 2008. 27th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 13-18 April 2008, Phoenix, AZ, USA*, 2008, pp. 1615–1623. [Online]. Available: http://dx.doi.org/10.1109/INFOCOM.2008.222

[29] "United States Frequency Allocations: The Radio Spectrum Chart," FCC, Oct. October 2011. [Online]. Available: http://www.ntia.doc.gov/files/ntia/publications/spectrum_wall_chart_aug2011.pdf

[30] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[31] I. F. Akyildiz, W.-Y. Lee, and K. R. Chowdhury, "CRAHNs: Cognitive radio ad hoc networks," *Ad Hoc Netw.*, vol. 7, no. 5, pp. 810–836, Jul. 2009.

[32] "IEEE Standard for Information Technology–Telecommunications and information exchange between systems Wireless Regional Area Networks (WRAN)–specific requirements part 22: Cognitive wireless RAN medium access control (MAC) and physical layer (PHY) specifications: Policies and procedures for operation in the TV bands," *IEEE Std 802.22-2011*, pp. 1–680, January 2011.

[33] H. Kim and K. G. Shin, "In-band spectrum sensing in IEEE 802.22 wrans for incumbent protection," *IEEE Trans. Mob. Comput.*, vol. 9, no. 12, pp. 1766–1779, 2010. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TMC.2010.169

[34] A. B. Flores, R. E. Guerra, E. W. Knightly, P. Ecclesine, and S. Pandey, "IEEE 802.11af: a standard for TV white space spectrum sharing," *IEEE Communications Magazine*, vol. 51, no. 10, pp. 92–100, 2013. [Online]. Available: http://dx.doi.org/10.1109/MCOM.2013.6619571

[35] "Rural Connect TV WhiteSpaces Radio," Carlson Wireless Technologies, Jan. 2014. [Online]. Available: http://www.carlsonwireless.com/ruralconnect/

[36] "Providing Broadband Wireless and Machine to Machine Solutions Using Whitespaces and Spectrum Sharing," Whitespace Alliance, 2012. [Online]. Available: http://www.whitespacealliance.org/

[37] "Une gestion dynamique du spectre pour l'innovation et la croissance," Mission Spectre, Mar. 2014. [Online]. Available: http://www.mission-spectre2014.fr/fileadmin/ressources/fichiers/Rapport_Mission_Spectre_J-Toledano.pdf

[38] "CREW project: Cognitive Radio Experimentation World," FP7-CREW, started in October 2010. [Online]. Available: http://www.crew-project.eu/

[39] K. Chowdhury and I. Akyildiz, "CRP: A Routing Protocol for Cognitive Radio Ad Hoc Networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 794–804, april 2011.

[40] W. Lee and I. F. Akyildiz, "Optimal spectrum sensing framework for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3845–3857, 2008. [Online]. Available: http://dx.doi.org/10.1109/T-WC.2008.070391

[41] M. López-Benítez and F. Casadevall, "Improved energy detection spectrum sensing for cognitive radio," *IET Communications*, vol. 6, no. 8, pp. 785–796, 2012. [Online]. Available: http://dx.doi.org/10.1049/iet-com.2010.0571

[42] H. Kim and K. G. Shin, "Efficient discovery of spectrum opportunities with mac-layer sensing in cognitive radio networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 5, pp. 533–545, May 2008.

[43] M. López-Benítez and F. Casadevall, "Signal uncertainty in spectrum sensing for cognitive radio," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1231–1241, 2013. [Online]. Available: http://dx.doi.org/10.1109/TCOMM.2013.021413.110807

[44] C. Jiang, Y. Chen, K. J. R. Liu, and Y. Ren, "Renewal-theoretical dynamic spectrum access in cognitive radio network with unknown primary behavior," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 3, pp. 406–416, 2013.

[45] M. López-Benítez and F. Casadevall, "Time-dimension models of spectrum usage for the analysis, design, and simulation of cognitive radio networks," *IEEE T. Vehicular Technology*, vol. 62, no. 5, pp. 2091–2104, 2013. [Online]. Available: http://dx.doi.org/10.1109/TVT.2013.2238960

[46] G.-M. Zhu, I. F. Akyildiz, and G.-S. Kuo, "STOD-RP: A Spectrum-Tree Based On-Demand Routing Protocol for Multi-Hop Cognitive Radio Networks," in *GLOBECOM*, 2008, pp. 3086–3090.

[47] L. Ding, T. Melodia, S. Batalama, and M. J. Medley, "ROSA: distributed joint routing and dynamic spectrum allocation in cognitive radio ad hoc networks," in *ACM MSWiM 2009*. New York, NY, USA: ACM, 2009, pp. 13–20.

[48] C. Jiang, Y. Chen, Y. Gao, and K. J. R. Liu, "Joint spectrum sensing and access evolutionary game in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2470–2483, 2013.

[49] S. Sorooshyari, C. W. Tan, and M. Chiang, "Power control for cognitive radio networks: axioms, algorithms, and analysis," *IEEE/ACM Trans. Netw.*, vol. 20, no. 3, pp. 878–891, Jun. 2012.

[50] C. Cormio and K. R. Chowdhury, "A survey on MAC protocols for cognitive radio networks," *Ad Hoc Netw.*, vol. 7, no. 7, pp. 1315–1329, Sep. 2009.

[51] A. De Domenico, E. Strinati, and M. Di Benedetto, "A survey on MAC strategies for cognitive radio networks," *Communications Surveys Tutorials, IEEE*, vol. 14, no. 1, pp. 21–44, 2012.

[52] R. Murty, R. Chandra, T. Moscibroda, and P. Bahl, "Senseless: A database-driven white spaces network," *Mobile Computing, IEEE Transactions on*, vol. 11, no. 2, pp. 189–203, 2012.

[53] C. Cordeiro and K. Challapali, "C-MAC: A Cognitive MAC Protocol for Multi-Channel Wireless Networks," in *DySPAN 2007*, april 2007, pp. 147–157.

[54] Y. Kondareddy and P. Agrawal, "Synchronized MAC Protocol For Multi-Hop Cognitive Radio Networks," in *IEEE ICC '08*, may 2008, pp. 3198–3202.

[55] T. Chen, H. Zhang, G. Maggio, and I. Chlamtac, "Cogmesh: A cluster-based cognitive radio network," in *New Frontiers in Dynamic Spectrum Access Networks, 2007. DySPAN 2007. 2nd IEEE International Symposium on*, 2007, pp. 168–178.

[56] B. Hamdaoui and K. Shin, "OS-MAC: An efficient MAC protocol for spectrum-agile wireless networks," *Mobile Computing, IEEE Transactions on*, vol. 7, no. 8, pp. 915–930, 2008.

[57] X. Zhang and H. Su, "CREAM-MAC: Cognitive radio-enabled multi-channel MAC protocol over dynamic spectrum access networks," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 1, pp. 110–123, 2011.

[58] J. Jia, Q. Zhang, and X. Shen, "Hc-mac: A hardware-constrained cognitive mac for efficient spectrum management," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 1, pp. 106–117, 2008.

[59] H. Bany Salameh, M. Krunz, and O. Younis, "Mac protocol for opportunistic cognitive radio networks with soft guarantees," *Mobile Computing, IEEE Transactions on*, vol. 8, no. 10, pp. 1339–1352, 2009.

[60] Y. Zhao, M. Song, and C. Xin, "FMAC: A fair MAC protocol for coexisting cognitive radio networks," in *INFOCOM, 2013 Proceedings IEEE*, 2013, pp. 1474–1482.

[61] H. Khalifé, N. Malouch, and S. Fdida, "Multihop cognitive radio networks: to route or not to route," *Netwrk. Mag. of Global Internetwkg.*, vol. 23, no. 4, pp. 20–25, Jul. 2009.

[62] K. R. Chowdhury and M. D. Felice, "Search: A routing protocol for mobile cognitive radio ad-hoc networks," *Computer Communications*, vol. 32, no. 18, pp. 1983–1997, 2009.

[63] M. Caleffi, I. F. Akyildiz, and L. Paura, "OPERA: Optimal Routing Metric for Cognitive Radio Ad Hoc Networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2884–2894, 2012.

[64] A. S. Cacciapuoti, C. Calcagno, M. Caleffi, and L. Paura, "CAODV: Routing in mobile ad-hoc cognitive radio networks," in *IFIP Wireless Days 2010, Venice, Italy, October 20-22, 2010*. IEEE, 2010, pp. 1–5.

[65] D. Xue and E. Ekici, "Guaranteed opportunistic scheduling in multi-hop cognitive radio networks," in *INFOCOM*, 2011, pp. 2984–2992.

[66] K. R. Chowdhury, M. D. Felice, and I. F. Akyildiz, "TCP CRAHN: A transport control protocol for cognitive radio ad hoc networks," *IEEE Trans. Mob. Comput.*, vol. 12, no. 4, pp. 790–803, 2013. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TMC.2012.59

[67] R. L. Carter and M. E. Crovella, "Measuring bottleneck link speed in packet-switched networks," *Perform. Eval.*, vol. 27-28, pp. 297–318, Oct. 1996. [Online]. Available: http://dx.doi.org/10.1016/0166-5316(96)00036-3

[68] M. S. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proceedings of the 11th Annual International Conference on Mobile Computing and Networking, MOBICOM 2005, Cologne, Germany, August 28 - September 2, 2005*, 2005, pp. 73–87. [Online]. Available: http://doi.acm.org/10.1145/1080829.1080837

[69] Y. Yang and R. Kravets, "Contention-aware admission control for ad hoc networks," *IEEE Trans. Mob. Comput.*, vol. 4, no. 4, pp. 363–377, 2005. [Online]. Available: http://dx.doi.org/10.1109/TMC.2005.52

[70] N. V. Nguyen, I. G. Lassous, V. Moraru, and T. Razafindralambo, "Characterisation and application of idle period durations in IEEE 802.11 dcf-based multihop wireless networks," in *The 15th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, MSWiM '12, Paphos, Cyprus, October 21-25, 2012*, 2012, pp. 277–284. [Online]. Available: http://doi.acm.org/10.1145/2387238.2387286

[71] B. Nardelli and E. W. Knightly, "Closed-form throughput expressions for CSMA networks with collisions and hidden terminals," in *Proceedings of the IEEE INFOCOM 2012, Orlando, FL, USA, March 25-30, 2012*, 2012, pp. 2309–2317. [Online]. Available: http://dx.doi.org/10.1109/INFCOM.2012.6195618

[72] R. P. Laufer and L. Kleinrock, "On the capacity of wireless CSMA/CA multihop networks," in *Proceedings of the IEEE INFOCOM 2013, Turin, Italy, April 14-19, 2013*, 2013, pp. 1312–1320. [Online]. Available: http://dx.doi.org/10.1109/INFCOM.2013.6566924

[73] M. Alicherry, R. Bhatia, and E. L. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in *Proceedings of the 11th Annual International Conference on Mobile Computing and Networking, MOBICOM 2005,*

Cologne, Germany, August 28 - September 2, 2005, 2005, pp. 58–72. [Online]. Available: http://doi.acm.org/10.1145/1080829.1080836

[74] M. S. Kodialam and T. Nandagopal, "Characterizing achievable rates in multi-hop wireless networks: the joint routing and scheduling problem," in *Proceedings of the Ninth Annual International Conference on Mobile Computing and Networking, MOBICOM 2003, 2003, San Diego, CA, USA, September 14-19, 2003*, 2003, pp. 42–54. [Online]. Available: http://doi.acm.org/10.1145/938985.938991

[75] H. Li, Y. Cheng, P. Wan, and J. Cao, "Local sufficient rate constraints for guaranteed capacity region in multi-radio multi-channel wireless networks," in *INFOCOM 2011. 30th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 10-15 April 2011, Shanghai, China*, 2011, pp. 990–998. [Online]. Available: http://dx.doi.org/10.1109/INFCOM.2011.5935328

[76] R. Gupta, J. Musacchio, and J. C. Walrand, "Sufficient rate constraints for qos flows in ad-hoc networks," *Ad Hoc Networks*, vol. 5, no. 4, pp. 429–443, 2007. [Online]. Available: http://dx.doi.org/10.1016/j.adhoc.2006.01.003

[77] C. Zhu and M. S. Corson, "Qos routing for mobile ad hoc networks," in *INFOCOM*, 2002. [Online]. Available: http://www.ieee-infocom.org/2002/papers/121.pdf

[78] X. Lin and S. Rasool, "Distributed and provably efficient algorithms for joint channel-assignment, scheduling, and routing in multichannel ad hoc wireless networks." in *IEEE/ACM Trans. Netw.*, 2009, pp. 1874–1887. [Online]. Available: http://doi.acm.org/10.1145/1721711.1721725

[79] A. Sampath, L. Yang, L. Cao, H. Zheng, and B. Y. Zhao, "High Throughput Spectrum-aware Routing for Cognitive Radio Networks," in *CROWNCOM*, 2007.

[80] F. Tang, L. Barolli, and J. Li, "A joint design for distributed stable routing and channel assignment over multihop and multiflow mobile ad hoc cognitive networks," *Industrial Informatics, IEEE Transactions on*, vol. 10, no. 2, pp. 1606–1615, May 2014.

[81] M. Youssef, M. Ibrahim, M. Abdelatif, L. Chen, and A. Vasilakos, "Routing metrics of cognitive radio networks: A survey," *Communications Surveys Tutorials, IEEE*, vol. PP, no. 99, pp. 1–18, 2013.

[82] I. Pefkianakis, S. H. Y. Wong, and S. Lu, "SAMER: Spectrum Aware Mesh Routing in Cognitive Radio Networks," *IEEE DySPAN*, pp. 1–5, 2008.

[83] W. Kim, S. Oh, M. Gerla, and K. Lee, "Coroute: A new cognitive anypath vehicular routing protocol," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International*, July 2011, pp. 766–771.

[84] X. Tang, Y. Chang, and K. Zhou, "Geographical opportunistic routing in dynamic multi-hop cognitive radio networks," in *Computing, Communications and Applications Conference (ComComAp), 2012*, Jan 2012, pp. 256–261.

[85] H.-P. Shiang and M. van der Schaar, "Distributed resource management in multihop cognitive radio networks for delay-sensitive transmission," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 2, pp. 941–953, Feb 2009.

[86] G. Cheng, W. Liu, Y. Li, and W. Cheng, "Spectrum aware on-demand routing in cognitive radio networks," in *New Frontiers in Dynamic Spectrum Access Networks, 2007. DySPAN 2007. 2nd IEEE International Symposium on*, April 2007, pp. 571–574.

[87] J. Kim and M. Krunz, "Spectrum-aware beaconless geographical routing protocol for mobile cognitive radio networks," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, Dec 2011, pp. 1–5.

[88] M. Xie, W. Zhang, and K.-K. Wong, "A geometric approach to improve spectrum efficiency for cognitive relay networks," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 1, pp. 268–281, January 2010.

[89] S. Singh, M. Woo, and C. S. Raghavendra, "Power-aware routing in mobile ad hoc networks," in *Proceedings of the 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, ser. MobiCom '98. New York, NY, USA: ACM, 1998, pp. 181–190. [Online]. Available: http://doi.acm.org/10.1145/288235.288286

[90] Y. Yang, J. Wang, and R. Kravets, "Load-balanced routing for mesh networks," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 10, no. 4, pp. 3–5, Oct. 2006. [Online]. Available: http://doi.acm.org/10.1145/1215976.1215979

[91] ——, "Interference-aware load balancing routing for multihop wireless networks," *Technical Report UIUCDCS-R-2005-2526, Dept. of Computer Science, Univ. of Illinois at Urbana-Champaign*, 2005. [Online]. Available: http://hdl.handle.net/2142/10974

[92] O. Gnawali, R. Fonseca, K. Jamieson, D. Moss, and P. Levis, "Collection tree protocol," in *ACM SenSys*, 2009, pp. 1–14.

[93] S. Moeller, A. Sridharan, B. Krishnamachari, and O. Gnawali, "Routing without routes: the backpressure collection protocol," in *ACM/IEEE IPSN*, 2010, pp. 279–290.

[94] H. Dubois-Ferrierre, M. Grossglauser, and M. Vetterli, "Valuable detours: Least-cost anypath routing," *IEEE/ACM Transactions on Networking*, vol. 19, no. 2, pp. 333–346, April 2011.

[95] R. P. Laufer, H. Dubois-Ferriere, and L. Kleinrock, "Polynomial-time algorithms for multirate anypath routing in wireless multihop networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 3, pp. 742–755, 2012.

[96] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "Xors in the air: practical wireless network coding," *IEEE/ACM Trans. Netw.*, vol. 16, no. 3, pp. 497–510, 2008.

[97] R. Hou, K. Lui, F. Baker, and J. Li, "Hop-by-hop routing in wireless mesh networks with bandwidth guarantees," *IEEE Trans. Mob. Comput.*, vol. 11, no. 2, pp. 264–277, 2012. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TMC.2011.25

[98] J. Redi and R. Ramanathan, "The DARPA WNaN network architecture," in *IEEE MILCOM 2011*, nov. 2011, pp. 2258–2263.

[99] I. Rhee, A. Warrier, J. Min, and L. Xu, "DRAND: distributed randomized TDMA scheduling for wireless ad-hoc networks," in *ACM MobiHoc '06*. New York, NY, USA: ACM, 2006, pp. 190–201.

[100] D. Maltz, J. Broch, J. Jetcheva, and D. Johnson, "The effects of on-demand behavior in routing protocols for multihop wireless ad hoc networks," *Selected Areas in Communications, IEEE Journal on*, vol. 17, no. 8, pp. 1439–1453, 1999.

[101] T. Clausen and P. Jacquet, "Optimized Link State Routing Protocol (OLSR)," RFC 3626 (Experimental), Internet Engineering Task Force, Oct. 2003. [Online]. Available: http://www.ietf.org/rfc/rfc3626.txt

[102] P. Gupta and P. Kumar, "The capacity of wireless networks," *Information Theory, IEEE Transactions on*, vol. 46, no. 2, pp. 388–404, Mar 2000.

[103] L. Lamport, "Time, Clocks, and the Ordering of Events in a Distributed System," *Commun. ACM*, vol. 21, no. 7, pp. 558–565, Jul. 1978. [Online]. Available: http://doi.acm.org/10.1145/359545.359563

[104] M. Berkelaar, K. Eikland, and P. Notebaert, "lp_solve 5.5, open source (mixed-integer) linear programming system," Software, May 1 2004. [Online]. Available: http://lpsolve.sourceforge.net/5.5/

[105] P. D. Sutton, J. Lotze, H. Lahlou, S. A. Fahmy, K. E. Nolan, B. Özgül, T. W. Rondeau, J. Noguera, and L. Doyle, "Iris: an architecture for cognitive radio networking testbeds," *IEEE Communications Magazine*, vol. 48, no. 9, pp. 114–122, 2010.

[106] D. Makris, G. Gardikis, and A. Kourtis, "Quantifying tv white space capacity: A geolocation-based approach." *IEEE Communications Magazine*, vol. 50, no. 9, pp. 145–152, 2012.

[107] X. Ying, J. Zhang, L. Yan, G. Zhang, M. Chen, and R. Chandra, "Exploring indoor white spaces in metropolises," in *ACM MOBICOM*, 2013, pp. 255–266.

[108] J. Marinho and E. Monteiro, "Cognitive radio: survey on communication protocols, spectrum decision issues, and future research directions," *Wirel. Netw.*, vol. 18, no. 2, pp. 147–164, Feb. 2012.

[109] Ettus Research. (2014) Ettus Research, A National Instruments Company. http://www.ettus.com/. [Online]. Available: http://www.ettus.com/

[110] Maxima. (2013) Maxima, a computer algebra system. version 5.30.0. http://maxima.sourceforge.net/. [Online]. Available: http://maxima.sourceforge.net/

[111] "Iperf-tool, http://dast.nlanr.net/projects/iperf/."

[112] Z. Wang and J. Crowcroft, "Quality-of-service routing for supporting multimedia applications," *Selected Areas in Communications, IEEE Journal on*, vol. 14, no. 7, pp. 1228–1234, Sep 1996.

[113] J. L. Sobrinho, "Algebra and algorithms for qos path computation and hop-by-hop routing in the internet," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 541–550, Aug. 2002. [Online]. Available: http://dx.doi.org/10.1109/TNET.2002.801397

[114] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson, *Introduction to Algorithms*, 2nd ed. McGraw-Hill Higher Education, 2001.

[115] G. Xue and S. K. Makki, "Multiconstrained qos routing: A norm approach," *IEEE Trans. Computers*, vol. 56, no. 6, pp. 859–863, 2007. [Online]. Available: http://dx.doi.org/10.1109/TC.2007.1016

[116] R. Hassin, "Approximation schemes for the restricted shortest path problem," *Math. Oper. Res.*, vol. 17, no. 1, pp. 36–42, Feb. 1992. [Online]. Available: http://dx.doi.org/10.1287/moor.17.1.36

[117] D. H. Lorenz and D. Raz, "A simple efficient approximation scheme for the restricted shortest path problem," *Oper. Res. Lett.*, vol. 28, no. 5, pp. 213–219, Jun. 2001. [Online]. Available: http://dx.doi.org/10.1016/S0167-6377(01)00069-4

[118] A. Goel, K. Ramakrishnan, D. Kataria, and D. Logothetis, "Efficient computation of delay-sensitive routes from one source to all destinations," in *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, 2001, pp. 854–858 vol.2.

[119] X. Chu and B. Li, "Dynamic routing and wavelength assignment in the presence of wavelength conversion for all-optical networks," *Networking, IEEE/ACM Transactions on*, vol. 13, no. 3, pp. 704–715, June 2005.

[120] R. Andersen, F. Chung, A. Sen, and G. Xue, "On disjoint path pairs with wavelength continuity constraint in wdm networks," in *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 1, March 2004, pp. –535.

[121] H. Badis, A. Munaretto, K. Al Agha, and G. Pujolle, "Optimal path selection in a link state qos routing protocol," in *Vehicular Technology Conference, 2004. VTC 2004-Spring. 2004 IEEE 59th*, vol. 5, May 2004, pp. 2570–2574 Vol.5.

[122] "Java JGraphT library," 2014. [Online]. Available: http://jgrapht.org/

[123] "olsrd, an ad hoc wireless mesh routing daemon." [Online]. Available: http://www.olsr.org/

[124] E. Kohler, R. Morris, B. Chen, J. Jannotti, and M. F. Kaashoek, "The click modular router," *ACM Trans. Comput. Syst.*, vol. 18, no. 3, pp. 263–297, Aug. 2000. [Online]. Available: http://doi.acm.org/10.1145/354871.354874

Résumé

Les télécommunications sans fil ont connu ces dernières années un immense succès à tel point que le spectre des fréquences est désormais surchargé et nécessite la disponibilité de nouvelles ressources. Pour répondre à ce besoin, des techniques de réutilisation dynamique du spectre ont alors vu le jour sous la dénomination de radio cognitive. Elles consistent à partager de manière opportuniste et efficace certaines fréquences ayant été initialement allouées à d'autres systèmes. Cette thèse se place dans le contexte de réseaux sans fil tactiques hétérogènes comportant des segments de radios cognitives. La difficulté provient alors de la garantie de qualité de service de bout en bout : respect du débit négocié, du délai et de la gigue. Nous nous sommes tout d'abord intéressés au contrôle d'admission dans ce type de réseaux en proposant une méthode de calcul de bande passante résiduelle de bout en bout s'appuyant sur un algorithme de complexité polynomiale et pouvant être implanté de manière distribuée. Nous nous sommes ensuite concentrés sur le routage en proposant une nouvelle métrique tenant compte des particularités de ce type de réseaux. Enfin, nous nous focalisons sur la thématique du routage à contraintes multiples en étudiant et implantant en environnement réel des algorithmes d'approximation proposés dans la littérature.

Mots clés : Réseaux ; Qualité de Service ; Radio cognitive ; Bande passante ; Routage

Abstract

The unprecedented success of wireless telecommunication systems has resulted in the wireless spectrum becoming a scarce resource. Cognitive Radio systems have been proposed as the enabling technology allowing unlicensed equipments to opportunistically access the licensed spectrum when not in use by the licensed users. The focus of this thesis is on heterogeneous tactical networks deploying cognitive radios in parts or in their entirety. Such networks can be organized in multiple sub-networks, each characterized by a specific topology, medium access scheme and spectrum access policy. As a result, providing end-to-end Quality of Service guarantees in terms of bandwidth, delay and jitter, emerges as a key challenge. We first address the admission control in multi-hop cognitive radio networks and propose a polynomial time algorithm that can be implemented in a distributed fashion for estimating the end-to-end bandwidth. Then, we focus on routing and propose a new metric that takes into account the specifics of such networks. Finally, as quality of service requirements can be expressed using multiple metrics, we turn our attention to multi-constrained routing and implement on a real testbed low complexity approximation algorithms.

Keywords: Networks; Quality of Service; Cognitive radio; Bandwidth; Routing