Channel Access and Reliability Performance in Cognitive Radio Networks: Modeling and Performance Analysis

Indika Anuradha Mendis Balapuwaduge

Channel Access and Reliability Performance in Cognitive Radio Networks: Modeling and Performance Analysis

Doctoral Dissertation for the Degree *Philosophiae Doctor (PhD)* at the Faculty of Engineering and Science, Specialization in Information and Communication Technology

University of Agder
Faculty of Engineering and Science
2016

Doctoral Dissertation at the University of Agder 139
ISBN: 978-82-7117-832-1
ISSN: 1504-9272

Printed by Wittusen & Jensen
Oslo

❝*We are what we think. All that we are arises with our thoughts. With our thoughts, we make the world.*❞*....*

— Gautama Supreme Buddha

Dedicated to my parents

Kingsley Mendis and Samanthi Chandrika

Abstract

During the past one and a half decade, cognitive radio networks (CRNs) have received tremendous attention from both academia and spectrum regulatory bodies in order to promote efficient spectrum utilization by allowing unlicensed users exploit the vacant frequency bands of the licensed spectrum. The concept of CRN has inspired innumerable research ideas and efforts to explore dynamic spectrum access (DSA) strategies to provide satisfactory performance for both primary and secondary networks. The objective of this doctoral dissertation is two folds, as 1) to design and analyze DSA schemes under different network scenarios; and 2) to assess the reliability aspect of DSA from a dependability perspective.

To design and analyze DSA schemes, this dissertation focuses on providing more channel access opportunities for secondary networks (SNs) by employing four techniques: channel aggregation; priority-based queuing; spectrum leasing; and channel reservation. In particular, two priority-based queuing schemes are proposed for DSA strategies supported by channel aggregation. It is shown that, by introducing queues with appropriate sizes, the overall performance of the SN is improved. Next, considering the willingness of primary networks (PNs) to lease a part of their licensed spectrum to the SN, both static and dynamic spectrum leasing schemes are proposed in CRNs. Associated with them, a pricing scheme is also introduced to estimate the leasing cost in order to provide incentives to the PN. The proposed leasing cost function is monotonically increasing with the number of leased channels as well as the achieved QoS level in the SN. Furthermore, an optimal number of leased channels is identified under the QoS constraints of the PN.

Due to various interruptions such as priority users' arrivals and random link failures, ongoing services may be terminated. In this thesis work, channel reservation is studied as an effective technique to ensure service completion of ongoing transmissions in CRNs. Both static and dynamic channel reservation schemes are proposed and analyzed under several network scenarios, including both homogeneous and heterogeneous failures. It is demonstrated that when an optimal number of channels are reserved, a significant performance improvement of the SN can be achieved without sacrificing PN's QoS level.

So far, the analysis of the reliability aspect of CRNs from a dependability perspective still remains as a nearly un-chartered topic. To fill this gap, we first introduce a set of reliability and availability metrics including system times, retainability and network unserviceable probability in accordance with the ITU-T recommendations and dependability theory. Based on the developed continuous time Markov chain (CTMC) analytical models, the thesis examines channel availability, mean channel available and unavailable times and a few other reliability performance metrics for existing DSA schemes. In addition, an analytical approach is proposed to calculate the probability of not meeting a given availability requirement, by applying a uniformization based method in a CTMC.

The heterogeneity of channel bandwidth, multiple types of services with distinct QoS requirements and different failure types are also considered in the mathematical models developed in this thesis, considering potential applicability of those models in real-life networks. Moreover, the validity of the proposed analytical models is verified through extensive simulations carried out by employing various probabilistic distributions for traffic arrivals and service departures. Overall, we believe that the proposed DSA schemes, techniques for CRN performance enhancement and the reliability assessment framework in this dissertation contribute positively towards future deployment of CRNs.

Preface

This dissertation is a result of the research work carried out at the Department of Information and Communication Technology (ICT), University of Agder (UiA), in Grimstad, Norway, from September 2012 to June 2016. Professor Frank Y. Li, UiA, has been the main supervisor of this Ph.D. work and Professor Vicent Pla, Universitat Politècnica de València (UPV), has been the co-supervisor. From February 2014 to June 2014, the author was a visiting researcher at the University of Minnesota (UMN), USA. His host professor at UMN was Professor Mostafa Kaveh.

The research work presented in this dissertation has been funded by UiA and the visit to UMN was supported by the European Commission under the 7th Framework Program (FP7) through the Security, Services, Networking, and Performance of Next Generation IP-Based Multimedia Wireless Networks Project under Grant Agreement Number 247083.

Production note: LaTeX has been adopted as the tool for writing this dissertation, as well as the papers produced during the author's Ph.D. study. The mathematical calculations and simulation results are obtained by using MATLAB.

Acknowledgments

After an endeavor lasting for three years and nine months, my Ph.D. carrier has finally come to an end. Hereby, I would like to take this opportunity to show my gratitude to my supervisors, colleagues, friends and family who gave me supports through all these years.

First and foremost, I am grateful to my supervisor, Professor Frank Yong Li, from the bottom of my heart for his constant support and guidance during my journey towards Ph.D.. I was lucky to continue my study towards the Ph.D. degree with him after the successful completion of my two-year M.Sc. degree. During the past six years, I have benefited tremendously from his guidance, encouraging thoughts and great enthusiasm for high quality research. It has been a pleasure to make this long journey under his tutoring for both research and personal life guidance. I am indebted to him for checking meticulously my manuscripts and giving me invaluable insights into my research. Without his deep and broad knowledge, as well as the valuable comments, I would not have gone so far.

I express my sincere gratitude to my co-supervisor Professor Vicent Pla from UPV who monitored my work remotely and took efforts in reading various versions of my manuscripts during numerous evenings and weekends. His insightful comments have significantly improved the quality of my research work. It has been such a pleasure and inspiration to work with him. Look forward to having more opportunities to collaborate with him in the future. I would also like to extend my thanks to Professor Mostafa Kaveh, the Past-President of the IEEE Signal Processing Society, who has given me a valuable opportunity to work with his research group during my four month research stay at UMN.

Special thanks to Professor Frank Reichert, the Rector of UiA, for his generous support with many useful discussions on my Ph.D. work. I wish to thank Professors Andreas Prinz and Folke Haugland, the former and current Head of the Department of ICT, for their administrative support. I appreciate especially the assistance provided by the coordinators of the Ph.D. program at the Department of ICT, UiA, Mrs. Emma Elisabeth Horneman and Mrs. Tonje Sti for making all the practicalities run smoothly.

I am grateful to Dr. Chandana Perera, Dr. Keerthi Gunawickrama, Dr. Nisabha Jayasundere and other senior lecturers from the Faculty of Engineering, University of Ruhuna for giving me the opportunity to study at the UiA. I would also like to thank Dr. Lei Jiao who has proven to be a great fellow researcher. He provided numerous ideas during our joint work. My appreciation goes also to the colleagues of the Department of ICT and all the friends I met here in Grimstad, including Lakshmikanth Guntupalli, Huihui Yang, Ali Saleh Abomhara, Debasish Ghose, Vimala Srinivas, Yulin Si, Martin Gerdes, Renée Patrizia, Alireza Borhani, Meisam Naderi, Harsha Sandaruwan, Sasanka Ranasinghe, Ranga Wijesekara, Thilina Weerasinghe and Jagath Sri Lal. They provided a sense of community and mutual support to me and for each other. In particular, I thank Geeth Wijesiri and Sandun Konara who arranged delicious dinners for me and threw nice talks during weekends.

My deepest gratitude, love and affection belong to my parents, Kingsley Mendis and Samanthi Chandrika, and my brother, Mahesh Chathuranga, who are always encouraging me in my quest for the highest degree. Last but not least, I wish to thank my best friend for many years, Priyantha Godakumburage, who always took his time to listen to me and helped me when nobody else would.

Indika Balapuwaduge
June 2016
Grimstad, Norway

List of Publications

All the papers listed below are an outcome of the research work carried out by the author of this dissertation, including one submitted and nine published papers. In the first part of this list, Papers A-D are reproduced as Part II of this dissertation. Papers 5-10 which are published within the framework of this doctoral thesis are listed as the second part of this list, but they are not included in this dissertation. While the four included papers are sorted according to the topics in this dissertation, the other six papers are sorted in a reverse chronological order according to the dates of the latest publication.

Papers Included in the Dissertation

Paper A: I. A. M. Balapuwaduge, F. Y. Li, A. Rajanna, and M. Kaveh, "Channel Occupancy-based Dynamic Spectrum Leasing in Multi-channel CRNs: Strategies and Performance Evaluation," *IEEE Transactions on Communications*, vol. 64, no. 3, pp. 1313-1328, March 2016.

Paper B: I. A. M. Balapuwaduge, F. Y. Li, and V. Pla, "System Times and Channel Availability for Secondary Transmissions in CRNs: A Dependability Theory based Analysis," *IEEE Transactions on Vehicular Technology*, Early access available in IEEE Xplore, June 2016, DOI:10.1109/TVT.2016.2585200.

Paper C: I. A. M. Balapuwaduge, F. Y. Li, and V. Pla, "Significance of Channel Failures on Network Performance in CRNs with Reserved Spectrum," in *Proc. IEEE International Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, 23-27 May 2016.

Paper D: I. A. M. Balapuwaduge, F. Y. Li, and V. Pla, "Reliability Analyses of Dynamic Spectrum Reservation Strategies for CR Networks under Random Channel Failure and Recovery," submitted to *IEEE Transactions on Wireless Communications*, March 2016 (first round); November 2016 (second round).

Other Publications Not Included in the Dissertation

Paper 5: I. A. M. Balapuwaduge, A. Rajanna, M. Kaveh, and F. Y. Li, "Performance Evaluation of Three Dynamic Channel Access Strategies for Spectrum Leasing in CRNs," in *Proc. IEEE International Conference on Communications (ICC)*, London, UK, 8-12 June 2015, pp. 7570-7575.

Paper 6: E. Jalali, I. A. M. Balapuwaduge, F. Y. Li, and V. Pla, "A Dynamic Channel Access Strategy for Underlay CRNs: Markov Modeling and Performance Evaluation," *Transactions on Emerging Telecommunications Technologies*, February 2015, DOI: 10.1002/ett.2928.

Paper 7: L. Jiao, I. A. M. Balapuwaduge, F. Y. Li, and V. Pla, "On the Performance of Channel Assembling and Fragmentation in Cognitive Radio Networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5661-5675, October 2014.

Paper 8: I. A. M. Balapuwaduge and F. Y. Li, "System Times and Channel Availability Analyses in Multi-channel Cognitive Radio Networks," in *Proc. IEEE International Conference on Communications (ICC)*, Sydney, Australia, 10-14 June 2014, pp. 1320-1325.

Paper 9: I. A. M. Balapuwaduge, L. Jiao, V. Pla, and F. Y. Li, "Channel Assembling with Priority-based Queues in Cognitive Radio Networks: Strategies and Performance Evaluation," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 630-645, February 2014.

Paper 10: I. A. M. Balapuwaduge, L. Jiao, and F. Y. Li, "Complexity Analysis of Spectrum Access Strategies with Channel Aggregation in Cognitive Radio Networks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Anaheim, CA, USA, 3-7 December 2012, pp. 1295-1301.

Contents

List of Figures

List of Tables

List of Abbreviations

5G	Fifth Generation
ADC	Analog to Digital Conversion
BE	Best-Effort
CA	Channel Aggregation
CB	Channel Bonding
CCDF	Complementary Cumulative Distribution Function
CEPT	European Conference of Postal and Telecommunications Administrations
CF	Channel Fragmentation
CRAHN	Cognitive Radio Ad Hoc Network
CRN	Cognitive Radio Networks
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance
CTMC	Continuous Time Markov Chain
CQI	Channel Quality Information
DCA	Dynamic Channel Access
DCR	Dynamic Channel Reservation
DES	Discrete-Event Simulation
DFA	Dynamic Fully Adjustable
DSA	Dynamic Spectrum Access
DSL	Dynamic Spectrum Leasing
DTMC	Discrete Time Markov Chain
ECC	Electronic Communications Committee
ESU	Elastic Secondary User
E-UTRAN	Evolved UMTS Terrestrial Radio Access
FCC	Federal Communications Commission
FEC	Forward Error Correction
HTTP	Hypertext Transfer Protocol
ICT	Information and Communication Technology
IP	Internet Protocol
ISM	Industrial, Scientific and Medical
ITU	International Telecommunication Union
L-CRN	Leased Cognitive Radio Network

LSA	Licensed Shared Access
LTE-A	Long Term Evolution-Advanced
MAC	Medium Access Control
MCT	Mean Cycle Time
MDT	Mean Down Time
MGF	Moment Generating Function
MTBF	Mean Time Between Failures
MTFF	Mean Time to First Failure
MTTF	Mean Time To Failure
MUT	Mean Up Time
N-CRN	Unleased Cognitive Radio Network
NBC	Narrowband Channel
NUP	Network Unserviceable Probability
PDF	Probability Density Function
PH	Phase-type
PN	Primary Network
POL	Primary user Opportunistically access Leased CRN
PU	Primary User
QoS	Quality of Service
RF	Radio Frequency
RSU	Real-time Secondary User
RV	Random Variable
SCR	Static Channel Reservation
SEL	Secondary user Exclusively access Leased CRN
SN	Secondary Network
SPTF	Spectrum Policy Task Force
SSL	Static Spectrum Leasing
SU	Secondary User
UMTS	Universal Mobile Telecommunications System
URC	Ultra Reliable Communication
UWB	Ultra-Wideband
VBR	Variable Bit Rate
VoIP	Voice over IP
WBC	Wideband Channel
WG	Working Group
WiFi	Wireless Fidelity
WLAN	Wireless Local Area Network

PART I

Chapter 1

Introduction

Currently large chunks of spectrum are used inefficiently due to static spectrum allocation policies in which frequency bands are allocated exclusively to licensed users. Cognitive radio (CR) appears as a key technology to improve spectrum utilization by enabling unlicensed users to opportunistically exploit the spectrum of the licensed owners. Dynamic spectrum allocation and management are two key challenges in the design of cognitive radio networks (CRNs). In addition, the reliability aspect of CRNs has recently received an increasing amount of attention owing to the requirements of the next generation communication systems, such as ultra reliable communication (URC) and higher service retainability. This dissertation proposes dynamic spectrum allocation schemes together with techniques to improve the performance of both primary and secondary networks. The reliability aspect of CRNs is also explored. Furthermore, continuous time Markov chain (CTMC) models are developed to analyze the performance of these proposed schemes. In this chapter, we provide the basics of the CR/CRN concept and the related technological background information together with the identified research questions, thesis outline and contributions.

According to the facts and figures published by the international telecommunication union (ITU) regarding information and communication technology (ICT) industry, it is estimated that over 3.2 billion people have access to the Internet in 2015 [1]. Since 2000, this number has been octupled. Meanwhile, by the end of 2015, there were more than 7 billion mobile cellular subscriptions in the world, corresponding to a penetration rate of 97%. As the most dynamic segment in ICT, mobile communication is providing Internet services and consequently the mobile broadband penetration rate has reached 47% globally. Accordingly, capacity, throughput, reliability, service quality and resource availability of wireless services become essential factors for future mobile and wireless communications. Essentially, all these wireless technologies, standards, services and allocation policies rely on one common natural resource, i.e., radio spectrum.

Radio spectrum spans over the electromagnetic frequencies between 3 kHz and 300 GHz. Existing radio spectrum access techniques are based on the fixed allocation of radio resources. These methods with fixed assigned bandwidth for exclusive usage of licensed users are often not efficient since most of the spectrum bands are under-utilized, either/both in the space domain or/and in the time domain. In reality, it is observed that many spectrum bands are largely un-occupied in many places [2], [3]. For instance, the spectrum bands which are exclusively allocated for TV broadcasting services in USA remain un-occupied from midnight to early morning according to the real-life measurement performed in [4]. In addition to the wastage of radio resources, spectrum under-utilization constraints spectrum availability for other intended users. Furthermore, legacy fixed spectrum allocation techniques are not capable of adapting to the changes and interactions in the system, leading to degraded network performance.

Unlike in the static spectrum allocation, a fraction of the radio spectrum is allocated for open access as license-free bands, e.g., the industrial, scientific and medical (ISM) bands (902-928, 2400-2483.5, 5725-5850 MHz). In 1985, the federal communications commission (FCC) permitted to use the ISM bands for private and unlicensed occupancy, however, under certain restrictions on transmission power [5]. Consequently, standards like IEEE 802.11 for wireless local area networks (WLANs) and IEEE 802.15 for wireless personal area networks (WPAN) have grown rapidly with open access spectrum policies in the 2.4 GHz and 5 GHz ISM bands. With the co-existence of both similar and dissimilar radio technologies, 802.11 networks face challenges for providing satisfactory quality of service (QoS). This and the above mentioned spectrum under-utilization issues motivate the spectrum regulatory bodies to rethink about more flexible spectrum access for license-exempt users or more efficient radio spectrum management. Cognitive radio (CR) is probably the most promising technology for achieving efficient spectrum utilization in future wireless networks.

1.1 Software Defined Radio and Cognitive Radio

CR is regarded as an evolutionary concept from software defined radio (SDR). Herein, we briefly revisit the SDR technology prior to discussing CR and CRNs.

1.1.1 Software Defined Radio

A software defined radio is a reconfigurable wireless communication system in which the transmission parameters such as operating frequency band, modulation mode and protocol can be dynamically controlled through software. The dynamic control of those functions is performed by software-controlled signal processing algorithms and programmable hardware [6]. The physical layer functions such as analog to digital conversion (ADC), forward error correction (FEC), radio frequency (RF) filtering and transmission power amplification are software defined in SDR by providing reconfigurable processors instead of analog components with static hardware. According to the main functions of SDR, it is a multi-band and multi-standard system which can support diverse services such as telephony, data and video streaming [7]. Moreover, simultaneous operation over multiple frequency bands is also a significant feature of SDRs.

As an enabling technology, SDR forms the basis for the implementation of many communication systems especially for CR. One of the most important application areas of SDR is military communication. The abilities of changing encryption codes, modulation scheme and type of voice codec are some of the essential features in tactical communications. Furthermore, in emergency situations, SDR can assist to bridge different types of incompatible radio devices since SDR supports different communication standards.

1.1.2 Cognitive Radio and Cognitive Radio Networks

The concept of CR was coined by Joseph Mitola III in 1999 in his Ph.D. dissertation as an implementation based on SDR. In that dissertation, the level of autonomy of SDR is further enhanced by integrating artificial intelligence and machine learning features into CR nodes [8]. CR provides mechanisms for spectrum sensing, spectrum management and spectrum access for unlicensed users. A CR node is able to sense spectrum for channel availability measurements and modify its transmitting parameters in order to communicate efficiently, while avoiding interference with other users [9]. Furthermore, multiple connected CRs may form a CRN. A CRN enables to establish communications among multiple cognitive radios through proper networking mechanisms and protocols [10].

CRNs have the capability of achieving high spectrum efficiency by enabling CR users to sense and learn the surrounding environment and correspondingly adapt their transceiver configurations. Thus cognitive radio can be considered as a *smart radio* which has an adaptive behavior than traditional communication devices do.

With respective to adaptive transmission, cellular phones are already equipped with dynamic power control capability which enables a device to dynamically adjust its output power level to avoid interference. Furthermore, mobile devices can also process incoming signals to overcome various distortion effects. However, the reconfigurable ability of a CR node in a CRN is far more than adjusting output power level or processing of incoming signals. CRs outperform those dynamic adaptation capabilities of current radio devices, thanks to several enhanced characteristics as mentioned below.

Cognitive radios are programmable wireless devices. They are designed with *self awareness* and the knowledge of transmission protocols, etiquette, and procedures. These features of CRs provide the ability to sense its RF environment and location in the CRN. With this sensing capability, CRs are able to dynamically adjust its power, frequency, modulation, and other operating parameters such as antenna beam pattern [11], [12]. An interesting feature of this dynamic adjustment procedure is that the adaptive process can be performed without any prior plans. When the CR has to change its parameters without a plan, CR can adopt past experiences and radio recognition patterns with a learning process. Accordingly, a CR can be regarded as a radio that is aware of the environment around it and can adapt its transmissions according to the interference it observes [13]. As already mentioned, SDR is a preclude to CR and CRNs [14], [15].

1.1.3 Basic Operation of CRNs

In a CRN, primary users (PUs) are the users who have license or legacy rights to occupy a specific band of the spectrum. On the other hand, secondary users (SUs) opportunistically exploit spectrum holes in such a way that they do not cause interference to PUs. A spectrum hole indicates a frequency band in which an SU is able to transmit without causing interference to other PUs. Therefore in a CRN, spectrum management becomes a fundamental task that needs to be performed effectively to facilitate efficient channel allocation for both primary and secondary networks. More specifically, four main functions for spectrum management in a CRN are identified, as spectrum sensing, spectrum decision, spectrum sharing and spectrum mobility [16]. Those functions are briefly summarized below.

In the literature, both cooperative and non-cooperative spectrum sensing methods are proposed for CRs [17], [18]. In cooperative sensing, the sensing results of each CR node are shared with other users to increase the accuracy of spectrum sensing although it incurs a communication overhead [19], [20]. Spectrum analysis in which the sensing results are analyzed to precisely locate the vacant spectrum bands is required for an effective spectrum decision [21]. Based on spectrum analysis, a decision is made for appropriate channel access. Spectrum sharing can be performed either in a centralized or a distributed manner. In centralized spectrum sharing, a central entity in the CRN controls the spectrum allocation and access processes. In distributed spectrum sharing, however, those functions are performed by each CR node in the CRN based on local policies. Spectrum mobility is another main function in CRNs that gives agility to CR nodes [22], [23]. Spectrum mobility deals with spectrum handover procedures due to PU's presence at the already occupied channel by a CR user. In the next two chapters we will discuss different forms of SUs' dynamic channel access in more details, regarding spectrum sharing and spectrum mobility.

1.1.4 CR Research Directions

With the rapid proliferation of CR research, a wide range of innovative research themes and standards have been pursued. Among them dynamic spectrum access (DSA) and advanced spectrum sensing techniques are two popular directions [24]. The other fields of interest in CRN include the design of MAC protocols and architectures [25–28], network security mechanisms [29–32], cooperative communications [33], [34] and efficient power allocation methods [35–38].

The research work performed so far on DSA in CRNs mainly involves spectrum management, resource allocation and spectrum sharing with license-exempt users to satisfy their transmission requirements. Additionally, QoS provisioning [39], transmission delay reduction [40] and capacity improvement are also important for DSA. In the next chapter, we will propose and discuss DSA strategies from both design and modeling perspectives.

To date, little work has been performed on the reliability and availability aspects of CRNs. Within the fifth generation (5G) mobile communication paradigm, the reliability of communication systems including connections and network components emerges as an important task. At the same time, CRNs are regarded as a promising technology which can support and maintain highly dependable future communication systems. Chap. 4 is dedicated to provide detailed information regarding this research area.

1.2 Research Objectives and Methodology

The concept of CR has been existing for more than one and a half decade, and the CR technology is heading from theory towards reality. However, commercial deployments of CRNs are still to emerge. An innovation gap exists between the research work and real implementation of CRNs due to various factors. Among them three factors have been identified as the most significant to bridge this innovation gap, i.e., 1) spectrum regulatory rules which limit the flexibility of channel access; 2) lack of technical feasibility in realistic networks; and 3) financial aspects. Other than the financial aspect, this dissertation makes efforts to address the other two factors via three steps, i.e., concept definitions, scheme design and performance analyses. In the following two subsections, we summarize them with respect to those two factors.

1.2.1 Improving Access Flexibility in CR Spectrum

As already mentioned, the CR concept is proposed to address the spectrum underutilization issues caused by fixed spectrum allocation policies. However, regulators tend to maintain a certain degree of inflexibility in order to impose communication standards. In general, it is challenging or even contradictory to enable spectrum sharing while satisfying the QoS requirement of SUs.

If licensed users have more flexibility to adopt most effective technology for their intended services and obtain benefits from spectrum sharing via for instance, spectrum leasing, the performance of secondary networks (SNs) would be improved while achieving better spectrum utilization. With this motivation, we develop channel occupancy-based spectrum leasing paradigms in Paper A in Part II of this dissertation. In addition, Paper C and Paper D propose static and dynamic spectrum reservation mechanisms respectively by partially relaxing the incumbent users' priority level over a limited licensed band in order to enhance the QoS of the secondary network. Furthermore, this dissertation work demonstrates that both spectrum leasing and spectrum reservation can be effectively integrated with DSA strategies for achieving better QoS performance for SUs while keeping PU's access privilege.

1.2.2 Integrating Realistic Features into CRN Modeling

The gap between research work and reality is caused by various assumptions made during system modeling for the purpose of analysis convenience. For instance, the homogeneity of channels and traffic that is assumed in many analytical models is not observed in practice since CR users are diverse in traffic types with distinct QoS requirements. Considering hybrid secondary traffic, a DSA scheme is proposed in Chap. 5 with integrated priority-based queuing schemes. The proposed queuing schemes differentiate SU traffic types when performing channel allocation to services buffered in a queue. A heterogeneous channel scenario is analyzed in Paper B for evaluating the secondary network's reliability performance from the perspectives of dependability theory. In addition, Paper B analyzes a CRN with hybrid SU traffic considering both elastic and real-time SU services.

In real-life networks, especially under heavy traffic load conditions, a wireless link encounters a certain probability of channel failure due to various reasons. The impact of these failures on the performance of CRNs is an overlooked area in existing research work. Indeed, the effect of link failures due to channel fading, shadowing or equipment and power failures in the system cannot be neglected in performance modeling. When analyzing performance measures like transmission delay or channel availability, link failures and repairs need to be considered in mathematical modeling. However, mathematical modeling of CRNs presented in many existing studies examines network performance merely under error-free channel conditions. Motivated by this observation, both Paper C and Paper D attempt to model a more realistic network scenario by considering channel failures and repairs of the system. Furthermore, these channel failures are diverse in nature. To capture this diversity, we consider another realistic feature of the networks in Paper D by investigating the heterogeneity of channel failures.

1.2.3 Performance Evaluation

Markovian analysis, probability theory and dependability theory form the basis for performance evaluation of the proposed strategies and algorithms in this dissertation. Performance evaluation is conducted with respect to two main categories of metrics, i.e., system-centric and dependability-oriented parameters. In one category, metrics that measure system-centric performance in terms of capacity, blocking probability, forced termination probability, transmission delay, utility and cost are considered. With another category, metrics that estimate reliability aspects of a network from dependability theory's perspective such as availability, retainability, unserviceable probability, time to failure and time to repair are evaluated.

In addition to the numerical results obtained based on the developed mathematical models, extensive simulations are carried out for performance evaluation in this dissertation work. The derived analytical models are generally restricted by the considered network topology and traffic characteristics such as interarrival and service time distributions. Moreover, assumption-based analytical models may not apply when hypotheses change. The simulation models approximate the analytical model to a more realistic scenario by releasing rigid assumptions. Therefore, we perform discrete-event simulations for evaluating the same performance measures independently of the developed analytical models. Specifically, simulations corresponding to non-exponential interarrival time distributions and log-normal service time distributions are performed in Paper B.

1.3 Research Questions and Contributions

Based on the above discussions, several research questions are raised and addressed with our proposed solutions throughout this dissertation. In this section, we summarize those research questions and the contributions made by the proposed solutions.

1.3.1 Research Questions

This dissertation makes efforts to answer the following research questions.

- **Question 1 (Q1):** In the presence of multiple channels in a CRN, how can we design dynamic channel access strategies and how to evaluate the performance of those strategies? Furthermore, how to enhance the QoS of SU services?

- **Question 2 (Q2):** How to model multi-channel CRNs under heterogeneous channel scenarios and hybrid traffic patterns? How about the performance of the proposed models under homogeneous and heterogeneous traffic conditions?

- **Question 3 (Q3):** What kinds of incentive approaches can be employed in order to encourage PU's interest for opportunistic spectrum access to mitigate spectrum under-utilization? Moreover, how to increase the licensed users revenue in such a model while satisfying unlicensed user's QoS requirements?

- **Question 4 (Q4):** How to model and analyze a CRN from a dependability perspective and define dependability performance metrics? Based on these definitions and models, how to improve the reliability of communications?

- **Question 5 (Q5):** Considering link failures in a CRN, how can we propose dynamic channel allocation schemes in order to minimize performance degradation due to channel impairment?

- **Question 6 (Q6):** When a tradeoff exists between two performance metrics for spectrum allocation in a CRN, how to develop a method for achieving optimal resource allocation?

Table 1.1: An overview of the identified research questions in different chapters and included papers of this dissertation. The check mark (✓) indicates the presence of each question and where the answers have been provided.

Question	Ch. 2	Ch. 3	Ch. 4	Ch. 5	Paper A	Paper B	Paper C	Paper D
Q1	✓	···	···	✓	✓	···	✓	✓
Q2	···	···	···	✓	✓	✓	···	···
Q3	···	✓	···	···	✓	···	···	···
Q4	···	···	✓	···	···	✓	✓	✓
Q5	···	···	···	···	···	···	✓	✓
Q6	···	···	···	✓	✓	···	✓	✓
Q7	···	···	···	✓	✓	✓	···	✓

- **Question 7 (Q7):** How to verify the correctness and the preciseness of the developed mathematical models?

Table 1.1 indicates the venues (chapters and papers) in this dissertation where solutions for the above mentioned research questions are proposed. To answer the first question, we develop dynamic channel access strategies with several enhanced features which lead to improved SU performance. Herein, channel aggregation and spectrum adaptation by dynamically adjusting the number of aggregated channels and spectrum handover are adopted. In addition, two queuing schemes are proposed to answer the second part of the question for providing QoS enhancement to the SU network. By taking into account both real-time and best-effort (BE) services as well as narrowband and wideband channels in a heterogeneous CRN, we provide an answer for Q2 in our developed models. The third question is answered at two levels: 1) first, we integrate three static leasing schemes into an existing dynamic spectrum access strategy and analyze their performance; 2) second, dynamic leasing schemes are proposed with different leasing algorithms based on QoS demand.

Both Q4 and Q5 are answered by employing dependability theory as the basis. New reliability metrics are defined in order to capture important system times for channel access in a CRN. Channel failure and repair characteristics are included in the mathematical modeling, and both static and dynamic channel reservations are adopted to provide answers to Q5. The answers to Q6 are provided via the developed optimal resource allocation algorithms under pre-defined QoS requirements and network constraints. To answer the last question, discrete-event simulations (DESs) are performed independently to verify the developed mathematical models.

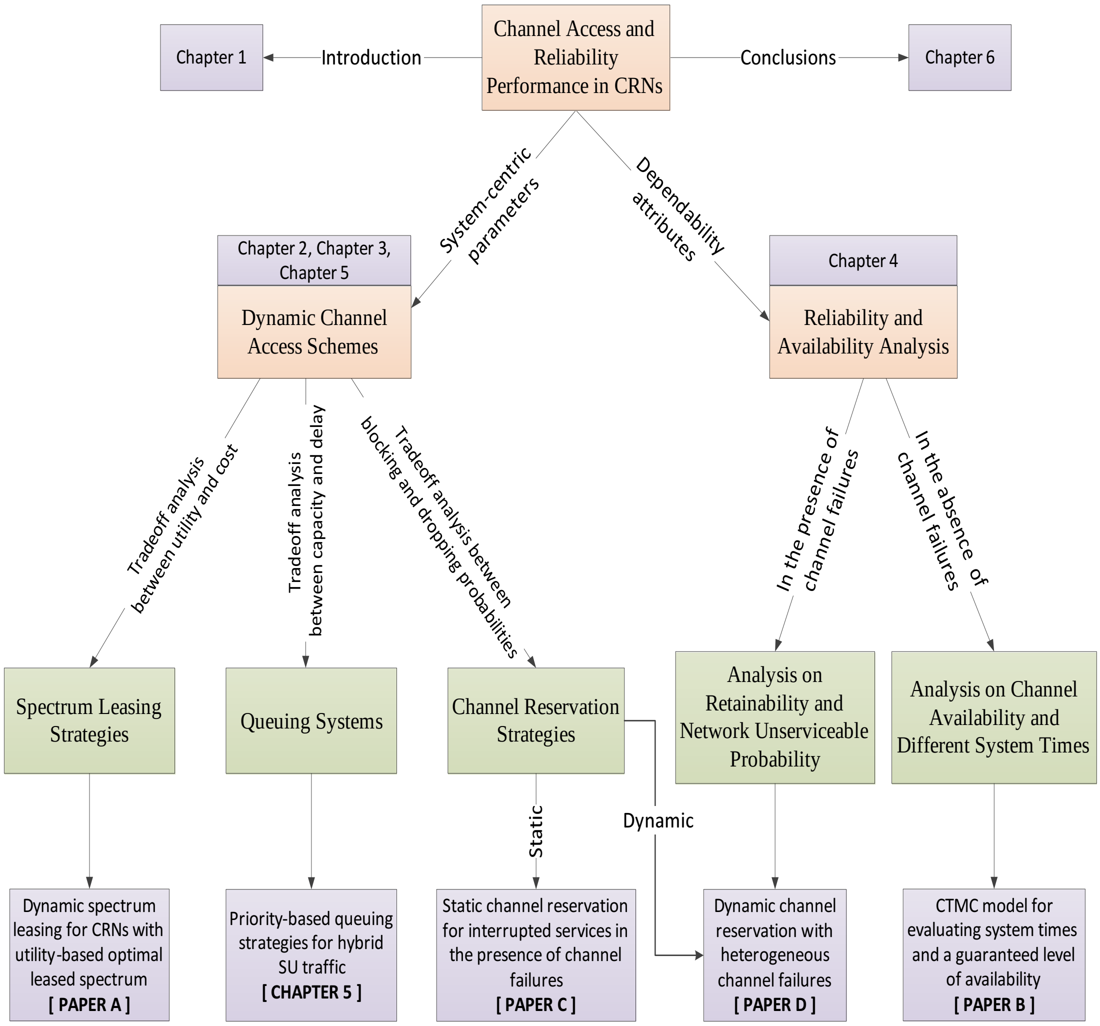


Figure 1.1: The structure of the dissertation and the connections of chapters and included papers to the main research directions.

The structure of this dissertation is also illustrated in Fig. 1.1 where the connections among the included papers and research topics are highlighted. As shown in this figure, the main focus of this dissertation ramifies over two scopes. On the one hand, efficient DSA schemes are explored by analyzing system-centric performance metrics such as capacity, blocking probability and forced termination probability. The diversity of the proposed DSA schemes spreads over approaches from spectrum leasing, channel reservation to queuing methods. On the other hand, a systematic framework is developed for reliability and availability analysis in CRNs based on dependability theory, including metric definition, mathematical analysis and simulation based validation.

1.3.2 Research Contributions

The contributions of this dissertation can be classified into two broad areas within the CRN research, i.e., dynamic spectrum access and reliability performance analysis. The main contributions are outlined as follows.

In this dissertation, three different spectrum allocation techniques, i.e., spectrum leasing, channel reservation and queuing mechanisms are proposed and their performance in multi-channel CRNs is investigated. The proposed spectrum leasing and channel reservation schemes support both static and dynamic leasing or reservation. These results have been published or under review in three IEEE Transactions and two IEEE Communication Society flagship conferences.

Another significant contribution of this dissertation is an in-depth study of the reliability aspect for channel access in CRNs. The developed systematic approach for reliability analysis includes metric definitions, mathematical analysis, simulation based validation, and it is applicable to existing channel access schemes. To the best of our knowledge, this is the first work which defines *system times* combined with channel access in CRNs and develops accordingly mathematical models to analyze them from the dependability theory's perspective.

1.4 Dissertation Outline

In general terms, this dissertation deals with the design and analysis of dynamic channel allocation schemes and reliability performance models in multi-channel CRNs. More specifically, the efforts concentrate on channel aggregation with priority-based queuing, spectrum leasing, channel reservation, reliability framework as well as performance evaluation in CRNs. The dissertation is organized in two parts where **Part I** provides an introduction, a summary of the research topics and the performed work, together with conclusions and future work. **Part II** presents the scientific contributions of this dissertation via four research papers. The rest of **Part I** is organized as follows:

- Chap. 2 overviews existing DSA strategies and points out associated research directions. Mathematical formulations for DSA strategies and performance metrics are summarized by using continuous time Markov chains. The chapter summarizes also available verification methods to substantiate the correctness and preciseness of the proposed mathematical models.

- Chap. 3 deals with spectrum leasing paradigms that can be employed in DSA strategies when licensed users are willing to lease their spectrum partially and temporarily. A short summary of existing spectrum leasing methods is also included, together with a brief introduction of the work performed in Paper A.

- Chap. 4 presents novel modeling and analysis on reliability performance of CRNs from a dependability perspective. Reliability metrics in accordance with classical dependability theory are discussed. The effect of channel failures and repairs on performance parameters is also discussed.

- Chap. 5 proposes two queuing schemes which can be applied in hybrid traffic scenarios when blocking probability and forced termination probability are considered as important parameters of QoS support.

- Chap. 6 concludes the dissertation and points out a few potential extensions of the proposed schemes and models.

Chapter 2

Dynamic Spectrum Access and CTMC Modeling

In order to allocate radio spectrum in an efficient and effective manner, dynamic spectrum access has been identified as a promising technique. DSA mitigates the effect of spectrum holes by providing means to allocate licensed spectrum in a flexible way and it increases spectrum availability for end users. Cognitive radio plays a major role to realize DSA in wireless communications. The capability of adaptively tuning communication parameters according to medium conditions leads to improved spectrum utilization in cognitive radio networks. To evaluate the performance of various DSA schemes, continuous time Markov chain analysis is one of the most popular tools.

According to the spectrum policy task force (SPTF) and the FCC, one of the core principles for effective spectrum management is to maximize the efficiency for radio spectrum utilization [41–43]. The spectrum efficiency working group of this task force has also found that at any given time and location, a considerable amount of allocated bandwidth lies idle without being utilized and a large portion of the assigned spectrum is used sporadically [44]. Therefore, the common impression on running out of usable spectrum does not hold. Since static frequency allocation methods raise the aforementioned spectrum under-utilization issues, more advanced approaches such as dynamic or flexible spectrum allocations are expected in order to increase spectrum efficiency in the next generation wireless communication systems. With the development of CR technologies, dynamic spectrum access is moving towards reality for this purpose [45].

2.1 Design of Dynamic Channel Access Strategies

Even though DSA leads to generally increased spectrum utilization, the effectiveness of DSA strategies heavily depends on both the availability of unused spectrum and the features of the adopted channel access scheme. A common conclusion which is drawn from recent studies is that the allocation of unused channels to unlicensed users needs to be facilitated with the following features in order to achieve effective dynamic channel access (DCA).

- The unused channels should be assigned to the requested secondary users so that the overall channel occupancy of a CRN can be maximized. Channel assembling[1] is one of the techniques for this purpose in order to increase the data rate or service rate of ongoing services. Using this technique, SUs assemble multiple channels together as one channel to transmit one SU service.
- An effective channel distribution procedure which assigns channel access opportunities fairly among a group of SUs is required. For instance, channel sharing with new SUs and interrupted SUs when the network lacks idle channels could be used to decrease blocking probability and forced termination probability.
- The DCA scheme must preserve PUs' access privilege to the respective licensed channels when required and as long as required. Precise spectrum sensing of SUs is an essential task regarding this feature.

Various techniques which can be integrated with the DSA concept have been proposed in CRNs for further enhancement of network performance. In the next subsections, four of them including channel assembling, priority-based queuing, cooperative communication and channel reservation are discussed with details.

2.1.1 Channel Aggregation and Fragmentation

Depending on the availability and the spectral occupancy of channels, a CR transmission may proceed over multiple contiguous, i.e., *bonded*, or non-contiguous, i.e., *aggregated* idle channels. More precisely, channel bonding (CB) refers to the combined occupancy of multiple *adjacent* channels to form a channel with wider

[1] The meaning of channel assembling is two-fold. When two or more channels in the frequency domain are available and contiguous with each other, they could be *bonded* as one channel. On the other hand, available channels located in different frequencies can be *aggregated* as a common channel [46].

bandwidth. Furthermore, channel aggregation (CA) extends this concept by allowing aggregation of non-contiguous multiple channels located in different spectrum bands. For instance, LTE-A networks already adopt both inter-band non-contiguous CA and intra-band contiguous CB [47], [48]. In this dissertation, we apply the same terminology *channel aggregation*, to both concepts since we are not studying the physical layer characteristics or guard-bands associated with those techniques. CA allows SUs to select a group of channels from multiple available channels in the CRN to provide services at higher data rates.

A new spectrum domain diversity scheme referred to as channel aggregation diversity was proposed in [49] to improve spectrum and energy efficiency in cognitive radio ad hoc networks (CRAHNs). An interesting feature of their scheme is that the imposed upper-bounded transmission power control for the selected channels depends on channel quality information. Unlike in conventional throughput maximization studies where throughput gain cannot match power consumption, the authors in [49] proposed their model under limited power resources. Furthermore, a distributed power-controlled and contention-based MAC protocol was proposed in [50] to enhance the throughput of a multi-channel CRN. Spectrum bonding and aggregation were distinguished in their developed model while investigating the impact of guard-band aware channel assignment. As shown in [50], by assigning adjacent channels to a single SU service rather than assigning non-contiguous channels, the number of required guard bands for a given transmission can be significantly reduced.

Contrary to CA, in channel fragmentation (CF), a channel is divided into multiple sub-channels and one or more sub-channels can be allocated for a single SU flow. A combined channel aggregation and fragmentation strategy was proposed in [51] to achieve flexible spectrum sharing in CRNs. The effect of CF on hybrid SU traffic was analyzed in one of our previous papers [52] when both CA and spectrum adaptation are enabled.

2.1.2 Integrating Queues in DSA

To reduce the blocking probability and the forced termination probability of PU and SU services in DCA schemes, various queuing-based system solutions have been proposed for CRNs [53–55]. Generally, in those queuing systems, the new service arrivals and interrupted ongoing services can wait in a queue until a vacant channel is found, given that the system has recorded the queuing delay and the remaining transmission volume of the interrupted SU services. An M/M/1 queuing model was proposed in [53] by introducing a probability-based channel selection scheme in order to reduce the average queuing delay. The queue was used to place the interrupted SUs into the head of the buffer. Multiple queuing systems were considered in [54] for video transmission sessions of SUs while considering PU's maximum tolerable delay on SU's performance. Moreover, [55] proposed a MAC protocol integrated with a buffer suitable for a CRAHN. According to that MAC protocol, a secondary node can stay on its reserved channel regardless of the PU signal's presence until its packet is transmitted, however, subject to a reservation period.

When queues are employed, the extra delay introduced to the CRN needs to be considered since long waiting time in queues is not acceptable, especially for real-time communications. Considering this fact and our previous research work on CA strategies, two priority-based queuing schemes are proposed in Chap. 5, targeting at hybrid SU traffic and their mathematical modeling.

2.1.3 Cooperative Communications for Cognitive Radio

The performance of a conventional CRN depends heavily on the state of the interfered channel and the activities of the primary network (PN). Therefore, a node or a cognitive network alone may not be capable of achieving higher throughput or further enhancing other performance. As a viable solution, a variety of collaborative techniques (spectrum sensing, transmission, interference coordination) recently gained attraction in the CR research community [56–60]. In cooperative CRNs, the licensed spectrum may be leased to SUs, and SUs may act as cooperative relays to enhance PUs' throughput [57], [58]. However, cooperative communication in CRNs also introduces several disadvantages, e.g., extra relay traffic and increased end-to-end delay.

2.1.4 Channel Reservation for Cognitive Radio

Channel reservation schemes have been studied in wireless networks, mainly for the purpose of reducing forced termination probability of handover calls. Consequently, blocking probability will increase in many cases. However, the percentage of such an increase depends on the employed channel access scheme. When it comes to CRNs, channel reservation can be efficiently adopted with DSA strategies in order to provide benefits for both PU and SU networks. When a set of channels are reserved for the exclusive usage of PUs, potential collisions with SU transmissions can be minimized. On the other hand, SUs can access vacant channels from a set of contiguous band (by channel bonding) [61]. Based on the spectrum occupancy information, the authors in [62] proposed a channel reservation policy which allows each SU to compete with other SUs in order to reduce collision probability. Different from those related work, we propose a static channel reservation (SCR) scheme in [63] for DCA in a multi-channel CRN by considering random channel failures. This work is presented as Paper C in Part II of this dissertation. In Paper D, the SCR scheme proposed in Paper C is extended to a dynamic channel reservation (DCR) scheme considering two working modes which target at channel availability and service retainability respectively.

2.2 An Overview of Markov Modeling

Markov analyses have been applied in modeling physical, biological, social as well as various engineering systems. Applying Markovian models to the fields of wireless communications and Internet traffic modeling has been very popular in the recent years [64]. This section gives a short overview on Markov processes involved in this dissertation. Prior to the discussion on CTMCs, we briefly explain discrete time Markov chains (DTMCs).

2.2.1 Discrete Time Markov Chains

Consider a stochastic process $X = \{X_n, n \in \mathbb{N}\}$ over a state space $\mathcal{S}$. X is a DTMC if the following two conditions are satisfied:

- for every $n \geq 0, X_n \in \mathcal{S}$,
- for every $n \in \mathbb{N}$ and for all $i_n, i_{n-1}, \cdots, i_0 \in \mathcal{S}$, we have

$$P\{X_n = i_n | X_{n-1} = i_{n-1}, \cdots, X_0 = i_0\} = P\{X_n = i_n | X_{n-1} = i_{n-1}\} \tag{2.1}$$

given that these two conditional probabilities can be defined.

The first condition states that all possible states in this system are always within the set of feasible state space $\mathcal{S}$. Moreover, the second condition indicates the well-known *memoryless property*, i.e., the future behavior of a system depends only on the present state, not on any past state. Moreover, in the case of a DTMC, state transitions are constrained to occur at time instants $0, 1, 2, \cdots$. A DTMC is characterized by $\mathcal{S}$ and the transition probabilities p_{ij} between the states where p_{ij} denotes the probability that the Markov chain is at the next time instant in state j, given that it is at the present time instant at state i. The transition probability matrix

$\boldsymbol{P}$ consists of all p_{ij}: $\boldsymbol{P} \equiv [p_{ij}]$ and the sum of elements in each row of $\boldsymbol{P}$ must be equal to 1.

2.2.2 Continuous Time Markov Chains

In a CTMC, the set of states $\mathcal{S}$ is still discrete as in a DTMC. However, the discrete time variable n is replaced by a continuous parameter, t. Therefore a stochastic process $X = \{X_t, t \in \mathbb{R}^+\}$, with values in a finite state space, $\mathcal{S}$, is a CTMC if for every $n \geq 0$, for all instants $0 \leq s_0 < \cdots < s_n < s < t$, and for all states $i_0, \cdots, i_n, i, j \in \mathcal{S}$, we have

$$P\{X_t = j | X_s = i, \cdots, X_{s_n} = i_n, \cdots X_{s_0} = i_0\} = P\{X_t = j | X_s = i\}. \quad (2.2)$$

Due to time continuation, a CTMC involves with transition rates instead transition probabilities between different states. If the system is in state i, then events that cause the system to make a transition to state j occur with a rate q_{ij} where $j \neq i$. The transition rate matrix $\boldsymbol{Q}$ with elements q_{ij} follows that $q_{ii} = -\sum_{j \neq i} q_{ij},\ i \in \mathcal{S}$ for diagonal elements. Consequently, the sum of elements in each row of $\boldsymbol{Q}$ must be equal to 0.

Let $\pi_j(t) = P(X_t = j)$ and $\pi_j = \lim_{t \to \infty} \pi_j(t)$ where π_j denotes the probability that the system resides in state j at the steady state. Accordingly, the steady state probability vector $\boldsymbol{\pi}$ is defined as $\boldsymbol{\pi} = [\pi_0, \pi_1, \cdots, \pi_N]$, given N as the size of state space. This vector is determined by solving the global balance equations, $\boldsymbol{\pi Q} = \boldsymbol{0}$, and the normalization equation, $\sum_{\text{all } j} \pi_j = 1$ where $\boldsymbol{0}$ is a row vector of all 0's of an appropriate size. More detailed theoretical formulations of Markov chains can be found in [65] and [66]. It is worth noting that the proposed channel access strategies in this dissertation including channel reservation and leasing schemes and reliability metrics are modeled by employing CTMCs. The following section is dedicated to explain a CTMC model of a simple DCA strategy considering a multi-channel CRN.

2.3 Dynamic Channel Access: Modeling and Performance Analysis

Analytical modeling of DCA schemes generally requires several assumptions since real-life systems are usually complicated and dynamic by nature. With simplified scenarios but reasonable assumptions, a CRN can be modeled as a system with a set of states which describe the system's behavior adequately. Moreover, the traffic

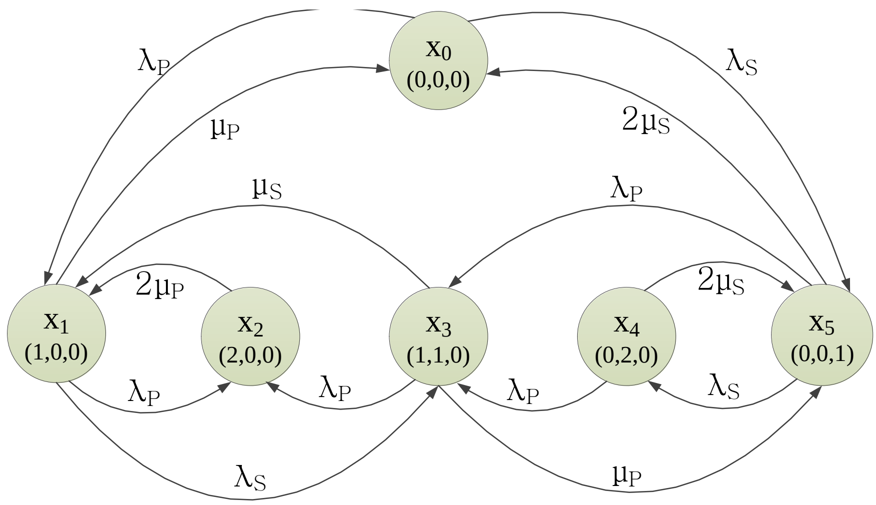


Figure 2.1: Illustration of the state transition diagram of channel access scheme. The CRN consists of two channels and one PU service occupies only a single channel while an SU service can occupy up to two channels.

in the CRN is modeled at the *session level* since modeling at this level captures the dynamic related to the arrival and departure of flows such as flow duration or number of active flows. Packet level analysis which considers performance measures such as packet drop ratio was investigated in [67]. Furthermore, a joint connection level and packet level analysis was studied in [68] for voice over IP (VoIP) traffic in CRNs.

2.3.1 A Representative CTMC Model for a Multi-Channel CRN

In order to illustrate the basic procedure adopted in our CTMC model, we consider a CRN with two homogeneous channels, i.e., $M = 2$ for illustration purpose. A simple DCA scheme for this CRN is explained as follows based on the *DFA* strategy proposed in [69]. In this dissertation, exponential interarrival times and service times are assumed for both PU and SU services to facilitate our CTMC analysis.

In this CRN, SU services opportunistically access the spectrum when no PUs operate on that frequency band. An SU can occupy a single channel or aggregate both channels depending on channel availability. However, if a new SU arrives while both channels are occupied by a single SU service, the ongoing SU will donate one of its occupied channels to the new SU. Moreover if a new PU arrives while there is a single SU service with two channels, the ongoing SU service will reduce its number of aggregated channels by one in order to release the other one to the new PU. If both channels are occupied, one of the ongoing SU services is forced to terminate upon a PU arrival.

In our CTMC model, the arrival rates of PU and SU services are denoted as λ_P and λ_S while the service rates per channel for those services are denoted as μ_P and μ_S respectively. The set of feasible states can be expressed as $\mathcal{S} = \{\boldsymbol{x}|i, j_1, j_2; (i + j_1 + 2j_2) \leq 2\}$ where $\boldsymbol{x} = \{i, j_1, j_2\}$ represents a general state of the system and i denotes the number of PU services. j_1 and j_2 represent the number of SU services with single channel and double channel occupancy respectively. Accordingly, the CTMC model consists of six states, i.e., $X_0, X_1, \cdots, X_5$, as shown in Fig. 2.1. In this model, the set of blocking states and the set of forced termination states of SUs can be expressed as $\mathcal{S}_{BL} = \{X_2, X_3, X_4\}$ and $\mathcal{S}_{FT} = \{X_3, X_4\}$. The global balance equations of the system are constructed as follows:

$$\begin{aligned}
\pi_0(\lambda_P + \lambda_S) &= \pi_1\mu_P + \pi_5(2\mu_S), \\
\pi_1(\lambda_P + \lambda_S + \mu_P) &= \pi_0\lambda_P + \pi_2(2\mu_P) + \pi_3\mu_S, \\
\pi_2(2\mu_P) &= \pi_1\lambda_P + \pi_3\lambda_P, \\
\pi_3(\lambda_P + \mu_P + \mu_S) &= \pi_4\lambda_P + \pi_5\lambda_P + \pi_1\lambda_S, \\
\pi_4(\lambda_P + 2\mu_S) &= \pi_5\lambda_S, \\
\pi_5(\lambda_P + 2\mu_S + \lambda_S) &= \pi_4(2\mu_S) + \pi_0\lambda_S + \pi_3\mu_P,
\end{aligned} \tag{2.3}$$

where $\pi_{\boldsymbol{x}}$ denotes the steady state probability of being in state $\boldsymbol{x}$. In order to solve the above linear equations we need the normalization equation, i.e.,

$$\pi_0 + \pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 = 1. \tag{2.4}$$

After obtaining the steady state probability distribution, i.e., $\pi = \{\pi_0, \pi_1, \cdots, \pi_5\}$, the next step is to derive the performance metrics of interest.

2.3.2 Derivation of Performance Metrics

From the steady state probability distribution of the system, different types of system-centric performance metrics can be derived, i.e.,

- state-based metrics (e.g., blocking probability and spectrum utilization),
- rate-based metrics (e.g., capacity and forced termination probability),
- other metrics (e.g., utility functions and reliability functions).

In state-based metrics, the probability that the system is in a set of states is evaluated. For instance, the blocking probability of a new SU service denoted as P_B in the above example is given by the sum of the steady state probabilities of all states in which the system blocks a new service. In the above example, the system blocks a newly arrived SU request only if both channels in the CRN are occupied and none of the ongoing SU services have two aggregated channels. Therefore,

$$P_B = \sum_{\boldsymbol{x} \in \mathcal{S},\ (i+j_1)=M} \pi_{\boldsymbol{x}} = \pi_2 + \pi_3 + \pi_4. \tag{2.5}$$

Define the capacity of the studied network as *the average number of service completions per time unit*. Therefore, the capacity is a rate-based measure which corresponds to a rate at which a service can be completed. Accordingly, the PU capacity, ρ_{PU}, and the SU capacity, ρ_{SU}, can be expressed as

$$\rho_{PU} = \sum_{k=0}^{5} i\mu_P \pi_k \text{ and } \rho_{SU} = \sum_{k=0}^{5}\sum_{r=1}^{2} r j_r \mu_S \pi_k. \tag{2.6}$$

In Papers A-D, both state-based and rate-based metrics are evaluated considering different DCA schemes. However, there may exist other metrics which do not belong to any of the above two categories. For instance, to calculate specific utility functions of CR users, the rates or state probabilities in the system are not needed.

Moreover in this thesis, we analyze reliability and availability of CRNs as we already pointed out in Subsec. 1.2.3 in order to calculate reliability metrics. For instance, to calculate the average value of time to first failure or call blocking, the stationary distribution is not required. Regarding reliability-oriented metrics, more details are provided in Chap. 4. The performance metrics studied in this thesis work include all these categories .

2.4 Model Verification

Within the field of wireless communications, a model can be considered as an analytical representation of a specific system or a process which can either be in a closed-form or an approximation based on a set of assumptions. As mentioned earlier, CTMCs are employed in this dissertation for modeling CRNs analytically. Modeling deals with the representation of a system and its functions, with the objective of obtaining an estimate of a system property, or a performance measure [70]. Regardless of the modeling tool or solution technique being used, the performance estimated from the analytical model must represent the behavior of the real systems to a sufficiently precise degree. Nevertheless, owing to application requirements of the performed modeling and various assumptions made for the system, a model can be more abstract than the real system it represents. Therefore, it is of significance to ascertain the correctness of the assumptions and validate them with respect to the real-life system. In the literature, three kinds of methods are identified to perform model validation, as listed below [71–73].

- Computer-based simulations;
- Trace-driven simulations [73], [74]; and
- Real-life test-bed experiments.

Analytical models, simulations or test-bed experiments alone have limited credibility. However, a result can be considered as reliable and trustworthy if the results obtained from two independent methods coincide with each other. In other words, we can state that our proposed analytical models are *precise enough* if the numerical results obtained based on the developed model match with the outcome of at least one of the aforementioned validation methods.

In this thesis, trace-driven simulations in which arrivals and departures collected from *actual traffic traces* and test-bed based experiments are not performed due to insufficient time and implementation difficulties. The main difficulties in possible trace-driven simulations are higher complexity due to required data gathered from real-life systems and its low level of representativeness. Since the traces from one real system cannot exactly reflect a general system, trace-driven simulations may not be representative. However, less randomness and the capability of evaluating minor changes in the system are motivating factors for performing trace-driven simulations. Instead, DES which is a form of computer-based modeling that provides a flexible approach to represent dynamical systems are employed in this dissertation to validate the correctness and the preciseness of the developed mathematical models. More details regarding how DES is performed in this dissertation are given in the following subsection.

2.4.1 Discrete-event Simulations

In our simulations, one can observe the following components.

- **Event queue:** The main events of the developed simulation models are arrivals and departures of PU and SU services. First we generate a finite number of those events with a time stamp based on a probability distribution and insert them in an event queue according to its time stamp.

- **Simulation clock:** In our simulations, the simulated time works as follows. Time is incremented to the instant of the next imminent event. After processing this event, the simulation clock is incremented again to the instant of the next imminent event.

- **State variables:** The variables which describe the state of the system at each time stamp are stored in an $m \times n$ matrix where m is the number of channels in the network and n is the number of required variables to represent the system.

- **Initialization variables:** At the beginning of each simulation run, we configure certain default values for state variables and statistical variables. For instance, the counter for recording PU arrivals is initially set to zero.

- **Event routines and input routines:** Upon the occurrence of events, an event routine sub-program is executed to handle the event and update the state variable matrix correspondingly. The event queue, the state variable matrix and other input variables supply the required inputs to the event routines.

Moreover, in the developed simulation codes/programs, the input variables follow a given probability distribution with configurable parameters. For instance, the arrival rate of PU services is fixed during one simulation run, but may vary for another simulation run. This fact and the above mentioned components collectively characterize a DES system [70].

Furthermore, DES has been conducted to validate the proposed analytical models developed in this dissertation as follows. By following the general rule of DES, inter-arrival times and service times of PU and SU services are generated as random variables (RVs). In our simulations, the state variables change instantaneously at separate points in time and these points are the ones at which an event occurs. These events indicate mainly PU and SU arrivals or departures of a multi-channel CRN. Moreover, the simulated time is recorded and updated (in a discrete manner) after each event and the current value of the simulated time can be tracked as the simulation proceeds. MATLAB programs are developed by following the above characteristics for obtaining required performance measures of the proposed models. The steps of those MATLAB simulation programs are summarized in the following subsection.

2.4.2 Simulation Steps

The following list indicates generally the simulation steps and some characteristics in the developed MATLAB simulation codes.

i A simulation round consists of L time units, as $L = 10000$. Each time unit is sub-divided into y sub-intervals with duration h. If a single time unit lasts for T duration, then $T = yh$.

ii Consequently, the probability of exactly one PU service arrival occurring in a sub-interval is $\lambda_P h$.

iii Since the minimum time unit in the system is h, the length of PU and SU services can be given by Zh where $Z \in \mathbb{Z}^+$. The remaining service time of an ongoing PU service or a real-time SU service is reduced by h after each h duration. However, for an ongoing elastic SU service with k aggregated channels, the service time is decreased by kh after each h duration.

iv The program checks the events of both PU and SU services during each sub-interval and updates the corresponding parameters accordingly. For instance, upon a new PU arrival, the value of a variable which represents the number of PU arrivals is raised by one.

v If the remaining service time of a PU or SU service reaches zero in a particular epoch, the service is considered as completed. Accordingly, the variable which records the number of service completions is raised by one.

vi The same procedure is repeated with multiple simulation rounds and the results from each round are recorded. The final simulation results which are reported are determined by obtaining the mean value of those recorded statistics from those simulation rounds. For instance, if the number of PU service completions over R simulation rounds is recorded as $N_1, N_2, \cdots, N_R$, then the simulation result for PU capacity is calculated as $\frac{1}{LR}\sum_{i=1}^{R} N_i$.

It is worth mentioning that the simulations are carried out independently, i.e., *not based on the derived mathematical expressions from the CTMC models at all.* For example, to obtain the blocking probability of SU services in our simulations, we count first the number of blocked SU services and record the number of total SU requests occurred during a simulation round. Then the ratio between them determines the blocking probability of SU services.

2.4.3 The Robustness and Applicability of the Developed Models

In order to have a robust mathematical model, the model should be able to perform effectively while its variables or assumptions are altered. Therefore the robustness of the derived mathematical models developed in this dissertation is studied considering different traffic pattern characteristics. In one aspect, log-normally distributed service times are considered for PU and SU services instead of exponential distribution. According to other related studies, it is widely shown that the application of the log-normal distribution for call duration provides a more realistic fit than the classical exponential distribution. Followed by this observation, simulations which are performed under log-noramally distributed service times are illustrated in [75]. However, the simulations performed in [75] remain with the assumption of exponential interarrival times for service arrivals.

Furthermore, recent work that investigates users' activity models has shown that the Poisson process fails to capture the temporal correlations and burstiness which have been observed in measurement-based studies, especially at the short time scales for packet level behavior [76–78]. These aspects are critical and cannot be disregarded when the focus of PU modeling is at short time scales as, for example, in spectrum sensing. However, the focus of this dissertation is significantly different from this type of studies in several respects. First, in this dissertation the time scales of interest are targeted at the session level, which are much longer than the time scales at the packet level. Second, although spectrum sensing is a fundamental function in CRNs, investigating the specifications and the detailed mechanisms of this function is out of the scope of this thesis work. We assume instead that there is an underlying block that takes care of spectrum sensing. However, in order to investigate the robustness of the models, another set of simulations in the presence of non-Poisson service arrivals is performed and they are illustrated in Paper B in this dissertation.

2.5 Chapter Summary

In this chapter, DSA schemes are discussed with different channel access features such as channel aggregation, channel reservation, queuing systems and cooperative communications. Except cooperative communication methods, other features are investigated in-depth in Papers A-D and in Chap. 5. In order to illustrate how CTMC modeling and performance evaluation are performed in the thesis work, an example CRN is illustrated and the stationary probabilities are obtained. Both state-based and rate-based performance metrics are introduced and analyzed based on this example. Furthermore, possible methods for model verification are discussed and the main steps of the adopted discrete-event simulations are summarized.

Chapter 3

Spectrum Leasing Schemes in CRNs

In this chapter, we study spectrum leasing and resource allocation in multi-channel CRNs based on already existing dynamic spectrum access strategies. The motivation to investigate spectrum leasing in CRNs is two-fold. First, dynamic spectrum leasing is still an emerging concept in the CRN research community. Second, effective spectrum leasing schemes integrated with DSA strategies are expected to provide significant performance improvement in both primary and secondary networks.

As already discussed in the previous chapter, DSA emerged as a promising approach for CRNs to meet its primary goals with the support of various constituent techniques. However, there are still challenges regarding the provision of required QoS for some applications due to the inflexibility of spectrum usage [79]. In order to cope with those challenges, licensed shared access (LSA) which relaxes the technical and commercial restrictions on existing licensees for spectrum access has been introduced by the electronic communications committee (ECC) [80]. Following the concept of LSA, SUs are allowed to occupy the licensed spectrum fully or partially as a temporary licensed user in accordance with the spectrum sharing rules. Those SUs are called as *LSA licensees*, i.e., as licensed users when they lease a part of the incumbent's spectrum band. LTE and LTE-Advanced have already been identified as supportive technologies for LSA realization [81]. Moreover, this spectrum management concept is identified as a pillar under dynamic exclusive use models in DSA as shown in Fig. 3.1 [82]. The next section is dedicated to describe the main concept of DSA in details.

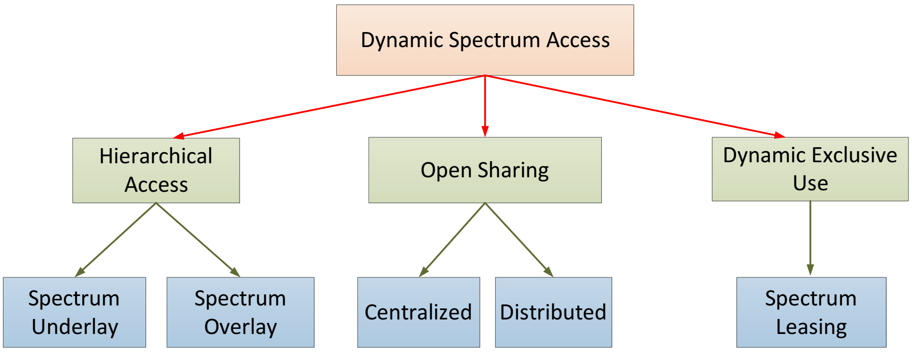


Figure 3.1: A taxonomy of dynamic spectrum access.

3.1 Dynamic Spectrum Access Categories

DSA strategies can be broadly classified into three categories as illustrated in Fig. 3.1. Within the hierarchical access category, there are two main CRN approaches: spectrum underlay and spectrum overlay [83]. Depending on the level of cognition about the operating CRN and the mode of operation, these two approaches are distinct from one another. In the *underlay* paradigm, SUs are allowed to access the licensed spectrum regardless of the PU's occupancy status as long as the interference level they generate to others' transmission is sufficiently low [84], [85]. Ultra-wideband (UWB) communication can be regarded as an already implemented solution for commencing *underlay* CRNs [86]. Under stringent transmission power constraints, SUs under UWB transceivers can potentially achieve a short-range high data rate by spreading transmission signals over a wide spectrum band. Differing from spectrum underlay, the *overlay* approach does not necessarily impose severe restrictions on transmission power of SUs since SUs can only access the spectrum when there is no presence of a PU signal [82], [87]. Joint overlay and underlay spectrum access mechanisms have also recently become a popular topic in hybrid transmission system design in CRNs in order to further improve spectrum utilization efficiency [88–90].

With *open sharing access*, a specific band of spectrum is opened to all users under certain spectrum etiquette [91]. Open sharing is motivated by the successful deployment of wireless networks in the ISM band such as wireless fidelity (WiFi). Both centralized and distributed spectrum sharing schemes are proposed to improve the performance of this category of spectrum access.

The third category in Fig. 3.1 is known as dynamic exclusive use. The main idea of proposing this category is to introduce a certain level of flexibility to improve spectrum utilization efficiency. Spectrum leasing has been considered as an approach under this category in which the spectrum bands can be sold and traded for the exclusive use of intended users (buyers). In the next sections, existing spectrum leasing schemes and leasing pricing mechanisms in CRNs are summarized.

3.2 Spectrum Leasing in CRNs

DSA network architectures under the hierarchical access category cannot provide any or sufficient revenues to PUs. Therefore, PUs are not willing, at least not eager, to cooperate with SUs and may even attempt to impose strict restrictions on SUs' spectrum access [92]. On the other hand, LSA approaches provide a certain level of QoS guarantee in terms of spectrum access and protect against harmful interference for both PUs and LSA licensees (i.e., SUs). Therefore, LSA appears as a promising technique in realizing DSA by performing spectrum trading between legacy spectrum owners and secondary users [93]

Indeed, spectrum leasing in communication networks has been proposed as an effective mechanism to improve spectrum utilization by sharing network resources [94–98]. When it comes to CRNs, spectrum leasing can be identified as an attractive mechanism in which the PUs lease a part of their licensed spectrum to SUs *in exchange for certain forms of rewards*. This reward could be in forms of, for instance, network performance enhancement through packet forwarding by SUs through cooperative transmission, or monetary rewards as leasing payments based on the leased bandwidth. In the following two subsections, these rewards are discussed in more details.

3.2.1 Cooperative Communication

In conventional CRN paradigms, potential collaboration between primary and secondary networks is not possible. SUs simply transmit their own data without causing interference to the licensed users. In recent studies, cooperative transmission schemes are jointly considered with leasing paradigms as a revenue for the PN. For example, if channel conditions for PUs are not good enough, SU's can establish another link towards the PU destination and help the PUs to relay data. Moreover, during the leasing period, SUs obtain more access to white spaces and therefore more transmission opportunities arise, leading to SU's throughput enhancement. Clearly, a certain level of coordination between PUs and SUs is required.

In [99], a portion of the spectrum is leased by PUs in both the time and the frequency domains as a benefit of cooperation to the SUs who act as relays for PUs' transmission. The adaptive spectrum leasing scheme proposed in [100] allows SUs to alter spectrum leasing scheme to cooperate with PUs for data transmission. As a reward, SUs gain access to the licensed channel for their own data transmission, achieving a higher data rate to the secondary network. A stochastic geometry based dynamic spectrum leasing scheme was proposed in [101] to improve primary spectrum users' bi-directional communication performance. A part of the spectrum is leased to geographically close SUs for a specific period of time during which the SUs have exclusive access to the leased spectrum. During the leasing period, SUs help PUs by relaying their data. Moreover, a cooperative diversity scheme was introduced in [102] as an enabling technique for spectrum leasing in CRNs.

3.2.2 Pricing Schemes

Other than the benefits through cooperative communication, PUs may also obtain monetary rewards as an incentive for their willingness to lease a portion of the licensed spectrum to SUs. For this purpose, various pricing schemes are proposed and analyzed by using game theory in recent studies. A linear pricing scheme based on leased spectrum was considered in [103] where PUs intend to offer a part of their spectrum to SUs to earn extra revenue. An interesting fact of their work is that PUs compete with each other to offer spectrum leasing while SUs compete for spectrum sharing. Knowing the characteristics of heterogeneous SUs, multiple pricing schemes were proposed in [104] instead of a fixed pricing scheme to improve PUs' profit and SU's satisfaction. An interesting feature in their work is that spectrum sensing, spectrum leasing, admission control and different pricing options are considered collectively for optimal decision making.

The tradeoff between the revenue from spectrum leasing and the cost due to possible performance loss from the leased region was investigated as an optimization problem in [105]. The proposed pricing scheme in that work suggests charging the lessee in proportional to the fraction of the admitted calls. A price-based spectrum access control analysis was performed in [106] between the network operator and delay-sensitive SUs in CRNs. In [107], different fine-grained pricing models were proposed to deal with the change of instantaneous number of active users. Accordingly, their pricing schemes adapt automatically to the traffic volume in the system.

In Paper A of this dissertation, we formulate an optimization problem to estimate the effective amount of leased spectrum considering a tradeoff between a utility function and a cost function. One of the key features of our model in Paper A is that the cost function is a non-linear function based on the received QoS. The pricing scheme proposed in Paper A is QoS dependent, in which the leasing price for L channels is determined by $C = \alpha L^{\beta}$ where α is the linear component and β is the exponent. The values of α and β depend on the blocking probability and the forced termination probability of the SN respectively. Moreover, this expression provides better flexibility to determine or bargain the leasing price according to the requirements of the spectrum seller as well as the buyer.

3.2.3 Existing Spectrum Leasing Schemes

Various spectrum leasing schemes exist, with different focus on PUs' and/or SUs' performance enhancement. In Table 3.1, we give a characteristic level comparison among those schemes in the literature by considering several performance aspects. The table entries denoted as "Y" and "N" represent the presence and the absence of the corresponding characteristic respectively. Please note that all the studies listed in this table, except the one proposed in our earlier work [108], are targeted at dynamic spectrum leasing (DSL) schemes.

Before outlining our work performed for dynamic spectrum leasing in Sec. 3.4 and Paper A, we present static spectrum leasing (SSL) and summarize our earlier work for SSL in Sec. 3.3.

3.3 Static Spectrum Leasing

Spectrum leasing can be performed either dynamically or statically. In SSL, PUs and SUs agree on an offline arrangement for SUs' access to a part of the licensed spectrum, whereas in DSL real-time negotiations between PUs and SUs are required [109]. In SSL schemes, the amount of leased spectrum is not dynamically adjustable during a specific time period with respect to the traffic conditions and channel status. Therefore, modeling of SSL schemes involves lower complexity compared to DSL schemes.

3.3.1 An SU Exclusive Access SSL Scheme

In our recent work [108], three SSL strategies were proposed considering hybrid SU traffic. Herein, we briefly revisit the secondary user exclusively access - leased CRN (SEL) strategy in [108], considering only the best-effort SU traffic (homogeneous traffic) in an M channel CRN. In the SEL strategy, L out of M channels are exclusively allocated for the BE secondary services through spectrum leasing. The set of leased channels of the CRN is denoted as L-CRN while the rest, i.e., the set of unleased channels of the CRN is denoted as N-CRN. The PUs can access only the N-CRN while the SUs can opportunistically access both. Each service occupies only a single channel. The channel access rules of the SEL strategy for a CRN for BE SUs are summarized as follows.

Table 3.1: Summary of the characteristics for spectrum leasing schemes

Ref.	Analysis of both PN and SN performance	System-centric parameter evaluation	Analysis on leasing price	Channel quality based pricing	Multiple PUs/SUs	Optimal leasing amount	Incentive for leasing[†]	Main focus
[110][‡]	Y	Y	Y	Y	Y	Y	Monetary rewards	PU and SU performance
[108]	Y	Y	N	N	Y	N	Monetary rewards	PU and SU performance
[111]	N	N	N	N	N	N	Service rewards	Network coding based DSL
[112]	N	N	Y	N	Two PUs and multiple SUs	N	Monetary rewards	Maximal profit based on operator's competition
[113]	Y	N	N	N	Y[§]	N	Interference-cap	Energy efficient transmissions
[114]	Y	N	N	N	Y[§]	N	Interference-cap	Utility of PUs and SUs
[115]	N	N	N	N	One PU and multiple SUs	Y	Monetary rewards and Interference-cap	Optimal power allocation
[116]	Y	Y	Y	Y	One PU and multiple SUs	N	Monetary rewards	PU's payoff
[117]	N	N	Y	N	One PU and multiple SUs	Y	Monetary rewards	Pricing schemes

[†]The incentives for spectrum leasing can be characterized into three main types: 1) monetary rewards; 2) service rewards; and 3) leasing is allowed as long as the interference from the CRs is below an interference threshold level.
[‡]The proposed DSL strategies in Paper A .
[§]In their analysis, only one PU and multiple SUs are considered. However, the model can be extended to more than one PU.

Table 3.2: Transitions from a generic state $\boldsymbol{x} = (i, j_n, j_r)$ of the channel access strategy upon PU/SU events.

Activity	Destination State	Tran. rate	Conditions
1. PU AR. A vacant channel exists in the N-CRN.	$(i+1, j_n, j_r)$	λ_P	$i + j_n < M - L.$
2. PU AR. An SU is forced to terminate.	$(i+1, j_n - 1, j_r)$	λ_P	$i+j_n = M-L, j_r = L.$
3. PU DP.	$(i-1, j_n, j_r)$	$i\mu_P$	$i > 0.$
4. SU AR. A vacant channel exists in the L-CRN.	$(i, j_n, j_r + 1)$	λ_S	$j_r < L.$
5. SU AR. A vacant channel exists in the N-CRN.	$(i, j_n + 1, j_r)$	λ_S	$j_r = L, i+j_n < M-L.$
6. SU DP from the L-CRN. An SU in the N-CRN performs spectrum handover to the L-CRN.	$(i, j_n - 1, j_r)$	$j_r\mu_S$	$j_n, j_r > 0.$
7. SU DP from the L-CRN. No spectrum handover.	$(i, j_n, j_r - 1)$	$j_r\mu_S$	$j_r > 0, j_n = 0.$
8. SU DP from the N-CRN.	$(i, j_n - 1, j_r)$	$j_n\mu_S$	$j_n > 0.$

The notations AR and DP indicate an arrival event and a departure event respectively.

Upon a PU arrival (or an SU arrival), one of the idle channels in the N-CRN (or L-CRN) is allocated. If all channels in both N-CRN and L-CRN are occupied while there are commenced SU services in the N-CRN, then one of those SU services is forced to terminate by re-allocating the channel to the newly arrived PU. Moreover, upon a new PU arrival when all the channels in the N-CRN are occupied by other PU services, the new PU is blocked even though there is one or more vacant channels in the L-CRN. If the whole L-CRN spectrum is occupied, a newly arrived SU can access a vacant channel in the N-CRN if available. Following the departure of an SU service in the L-CRN, one of the SU services in the N-CRN performs handover to the vacant channel.

Accordingly, a general state of this system is represented as $\boldsymbol{x} = \{i, j_n, j_r\}$ where j_n and j_r denote the number of SU services in the N-CRN and R-CRN respectively while i represents the number of PU services in the N-CRN. The set of feasible states becomes $\mathcal{S} = \{\boldsymbol{x}\,|(i+j_n) \leq M-L,\ j_r \leq L\}$. Table 3.2 summarizes the state transitions in this CRN where the notations, $\lambda_P, \lambda_S, \mu_P$ and μ_S, represent the PU/SU arrival and service rates as mentioned in Subsec. 2.3.1. The stationary distribution, $\boldsymbol{\pi}$, of this system is derived from the global balance equations as mentioned in Subsec. 2.3.1. The average number of SU service completions per time unit, i.e., the capacity of SU services, when L channels are leased is obtained as

$$\rho_S(L) = \sum_{\boldsymbol{x}\in\mathcal{S}} (j_n + j_r)\mu_S \pi_{\boldsymbol{x}}. \tag{3.1}$$

Furthermore, a new SU request will be blocked if both N-CRN and L-CRN are fully occupied. Correspondingly, the blocking probability of an SU service is given as

$$P_S^B(L) = \sum_{\substack{\boldsymbol{x}\in\mathcal{S},\\ i+j_n+j_r=M}} \pi_{\boldsymbol{x}}. \tag{3.2}$$

However, unlike in SU services, a new PU service is blocked if all channels in the N-CRN are occupied by PU services even though there are vacant channels in the L-CRN. Therefore, PU's blocking probability is given by

$$P_P^B(L) = \sum_{\substack{\boldsymbol{x}\in\mathcal{S},\\ i=M-L}} \pi_{\boldsymbol{x}}. \tag{3.3}$$

3.3.2 Utility Functions

The concept of utility functions has been widely used for CR research to analyze the QoS level received by PU and SU services which represents users' satisfaction [118–120]. Moreover, utility functions appear as popular tools when analyzing the optimal solutions for resource allocation problems [121–123] in communication networks. In general, no fixed utility function exists to represent a user's satisfaction. Based on the type of traffic, utility functions are different since each traffic type has its own QoS requirements. However, in many scenarios, especially for BE traffic, it is recommended that the utility of users should monotonically increase with the allocated resources [124]. Moreover, the marginal utility function, i.e., the first derivative of the utility function should be monotonically decreasing to avoid excessive resource assignment to users. At the same time, utility cannot hold negative values. To satisfy those conditions, we formulate the following sigmoid utility function to represent BE SU utility in a CRN [125].

$$U_S(R) = \frac{1}{1 + Be^{-CR}} \tag{3.4}$$

where B and C are the parameters of the utility function which affect the slope and the range of the utility curve respectively. For instance, one can maintain a normalized utility function, i.e., $0 \leq U_{BE}(R) \leq 1$ by selecting a proper value of B. R denotes the resource or a function corresponding to the resource assigned to the BE secondary services. A utility function can also incorporate different performance metrics depending on the traffic pattern [126]. In our work, the capacity achieved in the SN is incorporated with the assigned utility function since spectrum leasing has direct impact on the SU capacity. Therefore, following the sigmoid utility in (3.4), the utility function of the SN is defined as

$$U_S[\rho_S(L)] = \frac{1}{1 + Be^{-C[\rho_S(L)]}}. \tag{3.5}$$

This utility function reflects the benefit or profit obtained by the SN as a function of SU capacity through spectrum leasing. While achieving this benefit, the SN has to pay a certain price to the PN. In the following subsection, we propose a linear cost function to estimate the appropriate payment for the leased spectrum.

3.3.3 Leasing Cost Functions

In LSA based spectrum leasing schemes, PUs obtain a revenue due to the exercised price, or a rent over the leased spectrum [105]. On the other hand, this revenue is the associated leasing cost for the SN. Efficient pricing schemes need to be designed as they enhance the performance of users while improving network utilization. The cost functions can be defined as either linear or non-linear functions with respect to the leased bandwidth, capacity and leasing time. However, in this thesis, leasing price based on time is not studied. Consider a linear cost function which estimates the leasing price for L leased channels according to the amount of leased bandwidth and the provided QoS to the SUs as follows.

$$C_S(L) = \alpha[1 - P_S^B(L)]L \tag{3.6}$$

where $P_S^B(L)$ and L denote the blocking probability of SU services and the number of leased channels. α is a configurable parameter defined in order to normalize the cost function, such that $0 \leq C_{BE}(L) \leq 1$. In the next subsection, we discuss a utility-based resource allocation in CRNs considering the proposed SSL scheme, the utility function and the cost function.

3.3.4 Utility-based Leased Spectrum Allocation

To investigate the tradeoff between achieving a maximum utility and expending a minimum cost at the above mentioned CRN, the net utility function of the SN is defined as the difference between utility and cost given L leased channels

$$NU_S(L) \triangleq U_S[\rho_S(L)] - C_S(L). \tag{3.7}$$

Therefore,

$$NU_S(L) = \frac{1}{1 + Be^{-C[\rho_S(L)]}} - \alpha[1 - P_S^B(L)]L. \tag{3.8}$$

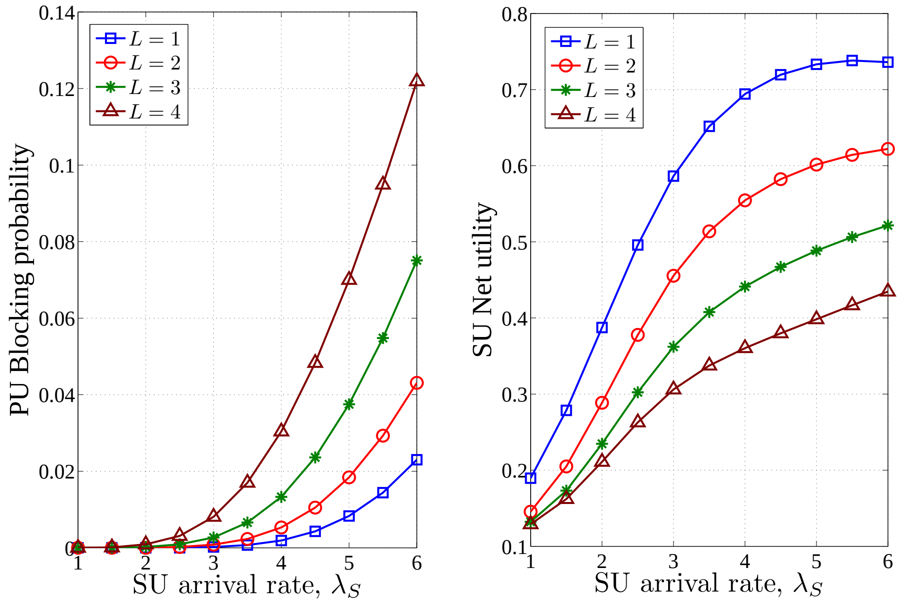


Figure 3.2: Blocking probability of the PN and the net utility of the SN as a function of λ_S. Note that $\lambda_P = \lambda_S$.

Herein, we study how many channels could be leased based on the derived net utility function. We first formulate an optimization problem and then analyze the solution for this problem. The objective is to maximize the net utility of the SN under the blocking probability constraint of the PN. Assume that the blocking probability of the PN must not exceed a certain threshold value given by $(P_P^B)_{max}$. Furthermore, assume that the network declares an upper bound for the number of leased channels as $L \leq \lfloor \frac{M}{A} \rfloor$ where $A > 1, A \in \mathbb{R}^+$. To that end, we formulate the following problem:

$$\begin{aligned} &\text{maximize}_{L=0,1,\cdots,\lfloor \frac{M}{A} \rfloor} && NU_S(L) \\ &\text{subject to:} && P_P^B(L) \leq (P_P^B)_{max}. \end{aligned} \tag{3.9}$$

Denote the optimal number of leased channels, i.e., the solution of this problem as L_{OPT}. It maximizes SU's net utility while ensuring the upper bound for PU's blocking probability. Consider a CRN with $M = 12$ channels as an example. The configurable parameters are set as follows: $\lambda_P = \lambda_S, \mu_P = \mu_S = 1.0$ service per unit time and $A = 3, B = 6.5, C = 1.0, \alpha = 1/6$. $A = 3$ implies that the system can lease maximum 4 channels since $L \leq \lfloor \frac{M}{A} \rfloor$. The numerical results corresponding to PU's blocking probability and the SU's net utility are shown in Fig. 3.2. As shown in

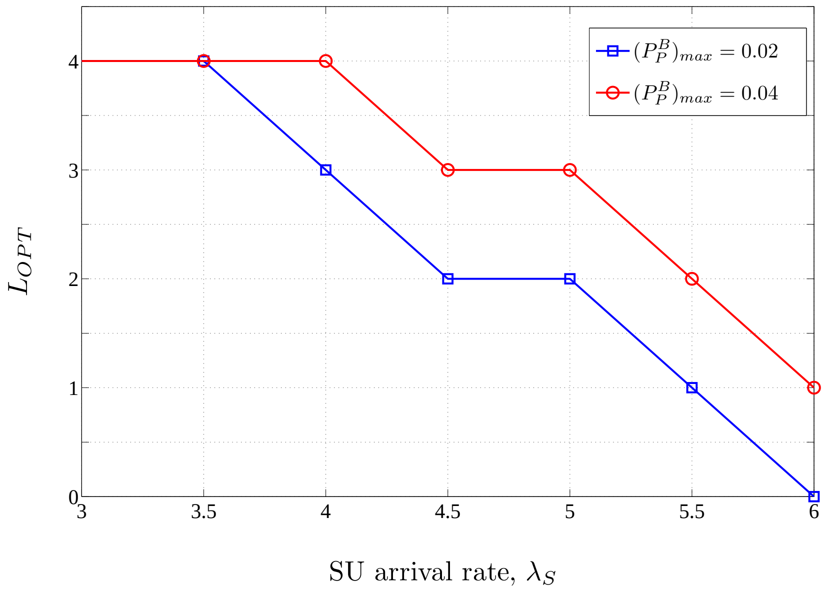


Figure 3.3: The optimal number of leased channels obtained in (3.9) based on the numerical results in Fig. 3.2.

this figure, within the range of $1 \leq \lambda_S \leq 6$, the net utility of the SN improves when more channels are leased. The reason is that the adopted utility function evaluates the SN's satisfaction in terms of SU capacity which is an increasing parameter with the resource, L.

However, the increase of L also leads to a monotonic rise of PU's blocking probability as illustrated in Fig. 3.2. This blocking probability can be bounded by configuring an appropriate value for $(P_P^B)_{max}$ in (3.9). For two given values of $(P_P^B)_{max}$, we obtain L_{OPT} and the results are illustrated in Fig. 3.3. Once the PU and SU arrival rates rise, the blocking probability of PUs becomes higher. Therefore, the system can lease only fewer number of channels at higher arrival rates. However, if the upper bound of the PU blocking probability is configured to a larger value, the system can allow more channels to be leased.

3.4 Dynamic Spectrum Leasing

The previous section is dedicated to an SSL scheme in our early work [108] which considers an SU exclusive access strategy together with a linear pricing scheme. The main advantage of SSL is represented by lower complexity of the dynamic access strategy. However, a CRN with SSL cannot utilize the leasing spectrum efficiently since the leased spectrum may not always be occupied by SUs during the entire leasing period. Therefore, the amount of leased spectrum needs to be adjusted according to the traffic conditions of the network in order to avoid spectrum under-utilization. As one option, if the SU arrival rate is significantly higher than that of PUs, the system may lease more channels to the SN and correspondingly provides a higher revenue for the PN. With this motivation, we propose two dynamic spectrum leasing schemes in Paper A of this dissertation by considering non-linear cost functions under hybrid SU traffic scenarios. The leasing schemes proposed in Paper A analyze both SU exclusive access and PU opportunistic access in the leased part of the spectrum. Those analyses are performed considering two leasing algorithms which assign the leased channels based on PU and SU channel occupancy status. The results based on dynamic adjustment of leased spectrum demonstrate improved performance, for both PU and SU services.

3.5 Chapter Summary

In this chapter, we present a concept in spectrum sharing known as licensed sharing access, which allows spectrum sharing between incumbent users and licensees in a flexible manner via spectrum leasing. The QoS assurance, more flexible spectrum utilization as well as the benefits brought to both PUs and SUs motivate the research on spectrum leasing as a promising mechanism towards CRN deployment. With spectrum leasing, both primary and secondary networks have exclusive privileges to access a portion of the CR spectrum for a given period of time. For spectrum leasing performance assessment, primary network's revenue and secondary network's utility are two major aspects of interest. This chapter analyzes an especial case based on an SSL strategy proposed in our early work [108], considering only the best-effort SU traffic. The chapter ends with a short introduction to the dynamic spectrum leasing schemes proposed in Paper A of this dissertation.

Chapter 4

Reliability and Availability Analysis in CRNs

Reliability, as part of dependability, has been widely deemed as an important aspect of 5G mobile and wireless networks despite the fact that it is extremely challenging to achieve ultra reliable communication in such networks. With its salient advantages for dynamic spectrum access in license and license-free bands, CRN is expected to play an important role for providing reliable communication services in future wireless networks. Therefore, it is imperative to study the dependability aspect of wireless cognitive radio networks. From our observation, this is indeed an overlooked topic even though a tremendous amount of research efforts have been made in the area of CRNs. In this chapter, we define a few reliability metrics combined with dynamic channel access in CRNs and develop accordingly mathematical models for performance evaluation from the dependability theory's perspective.

With a steadily growing of number of end users and the proliferation of content-rich applications, devices and the functions of communication systems are becoming more and more complex. Considering that today's wireless networks are interconnected and are dependent, a failure of one network component may affect other network components or even the entire network. On the other hand, reliable communication is an essential part of many industrial applications such as power system operations and smart grids [127], [128]. The QoS of real-time communications highly depends on link quality and network availability. Therefore, how to maintain a fault-tolerant network is an important task in real-world. Fault tolerance is defined as the ability of a system to continue performing its intended functions in

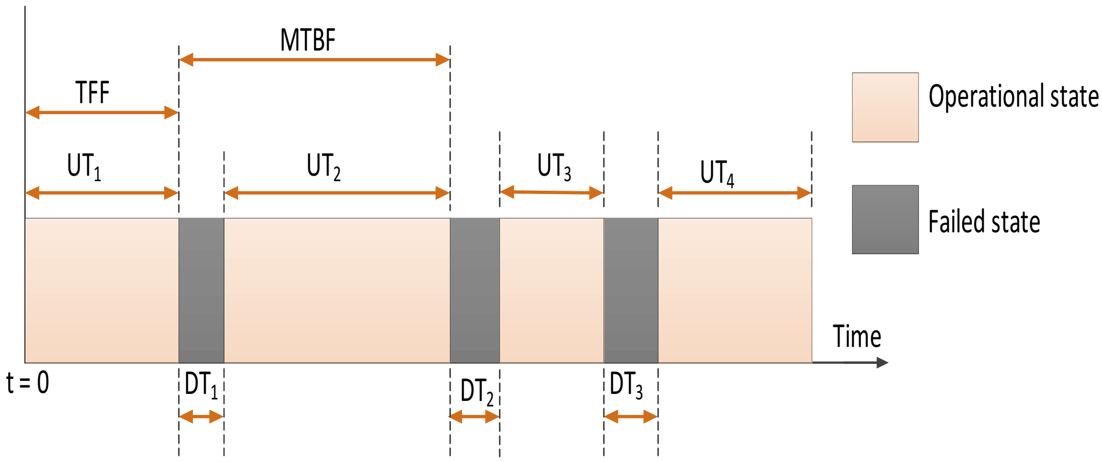


Figure 4.1: Illustration of system times in a repairable system assuming that the system is initially in the operational state.

the presence of faults [129]. One of the main targets for designing a *fault-tolerant* network is the development of a *dependable system*. The concept of dependability and the associated dependability attributes are described in the following section.

4.1 Dependability Basics

Dependability is the ability of a system to behave as it is specified when facing with failures of components to deliver its intended level of services to its users [130]. Therefore, failures and repairs are two keywords for dependability analysis of a system. *A failure* of a system can be considered as an event or a transition when the system's state changes from its specified operation to a state where it cannot perform as intended. *A repair* denotes the opposite transition.

4.1.1 MUT, MDT and MTFF

Generally, a system may undergo a series of failures and be repaired over time. Therefore operational time and failed time periods can be observed within the system, as illustrated in Fig. 4.1. The average duration of a system being in the operational state and the average duration of the system being in the failed state are known as mean up time (MUT) and mean down time (MDT) respectively. Moreover, the mean time until the first failure occurs is defined as the mean time to first failure (MTFF). By definition, the MUT and the MDT do not depend on the initial state distribution of the system. However, the MTFF may vary according to the system's initial state.

Fig. 4.1 illustrates the relations between different time averages considering failures and repairs of a system. In this figure, UT_i and DT_i denote the i^{th} up time and down time respectively. Time to first failure is denoted as TTF. Therefore, $MTFF = E\{TFF\}, MUT = E\{UT\}$ and $MDT = E\{DT\}$ where $E\{X\}$ denotes the expected value of a RV X. Mean time between failures (MTBF) is another measure of system's dependability which is generally used to estimate the lifetime of a system. MTBF is the mean time between two consecutive failures in a repairable system[2] and it follows $MTBF = MUT + MDT$.

There are three primary attributes of dependability which represent the properties that are expected from a system, i.e., reliability, availability and safety [129]. In the following subsection, the mathematical formulations of the first two attributes are presented. However, safety which is considered as another critical issue especially in engineering products and services is not a metric of interest in the context of this thesis work.

4.1.2 Reliability and Availability

When a system is available, users require reliable services with a sufficient level of quality. Therefore both reliability and availability are fundamental performance attibutes in dependability analysis. Reliability is increasingly becoming an important requirement for wireless communication networks, as human activities are more dependent than ever before on the successful operation of communication systems. The reliability of a system is defined as the *probability that a system will perform its intended functions without failure for a given interval of time under specified operating conditions* [131]. From this definition, four main elements can be identified, i.e., probability, failure, specific function and time. To obtain the mathematical definition of reliability, we need to consider these elements as follows. Assume that the time duration that a system failure is observed from its initial operation is a value of a RV, denoted as T. In other words, T represents the lifetime of a *non-repairable system*. Therefore, the probability of failure in the interval $[0, t]$ is given as

$$F_T(t) = P(T \leq t), \;\; t > 0 \tag{4.1}$$

where $F_T(t)$ is referred to as the unreliability at time t. Indeed, $F_T(t)$ is the cumulative distribution function of time to first failure. Consequently, the reliability function which represents the probability of no failure in the interval $[0, t]$ can be

[2]A repairable system is one which can be restored to intended operation by taking necessary repair actions such as replacement of failed components or modification of configurations. In non-repairable systems, a system cannot be restored after the first failure.

obtained as

$$R_T(t) = P(T > t) = 1 - F_T(t), \;\; t > 0. \tag{4.2}$$

When T is a continuous RV, $F_T(t)$ has a continuous distribution with a probability density function (PDF) $f_T(t)$. Then, the reliability function can be obtained as

$$R_T(t) = 1 - F_T(t) = 1 - \int_0^t f_T(x)dx = \int_t^\infty f_T(x)dx \tag{4.3}$$

where $\int_0^\infty f_T(x)dx = 1$ and $f_T(t) \geq 0$. The expected time to failure of a system which is denoted as MTTF is another important reliability parameter and it is derived as

$$MTTF = E\{T\} = \int_0^\infty tf(t)dt. \tag{4.4}$$

From (4.3), we can derive that $f(t) = -\frac{dR_T(t)}{dt}$. Therefore

$$MTTF = E\{T\} = \int_0^\infty t\Big(- \frac{dR_T(t)}{dt}\Big)dt. \tag{4.5}$$

Integrating the above equation by parts results in:

$$MTTF = -tR_T(t)|_0^\infty + \int_0^\infty R_T(t)dt = \int_0^\infty R_T(t)dt. \tag{4.6}$$

Moreover, the availability of a system can be analyzed mainly in two ways, i.e., instantaneous (point) availability and steady state availability. Instantaneous availability, denoted as $A(t)$, is the probability that the system is operating properly and is available to perform its functions at a specified time t, i.e.,

$$A(t) = P(Y(t) = 1) = E[Y(t)] \tag{4.7}$$

where $Y(t) = 1$ if the system is operating at time t, and 0 otherwise. The steady state availability, A_{SS}, can be obtained as the average availability over a period of time and it is expressed as

$$A_{SS} = \lim_{t\to\infty} A(t) = \frac{MUT}{MUT + MDT}. \tag{4.8}$$

4.2 Dependability Analysis in CRNs

To identify the weaknesses of a system and quantify the effect of failures and repairs of any system, dependability analysis is required. The *system* could be a wireless communication network such as a CRN. In CRNs, radio channels, software defined functions and hardware devices can be considered as components in a repairable system. Channel failures due to various reasons such as power failure and radio frequency (RF) failure and their recoveries are analogous to failure and repair events of a typical repairable system respectively. Upon channel failures due to one or more of those reasons, both PU and SU services can be subject to termination. In this case, how to improve channel availability in the presence of channel failures is a question to be answered from dependability perspectives. Furthermore, the SN of a CRN encounters forced terminations more frequently than the PN owing to PU's access privilege. Therefore, how to protect ongoing SUs from session failures upon new PU arrivals can also be considered as another dependability task.

Prior analyses on dependability can be used in order to efficiently design DSA strategies for CRNs. For instance, by identifying channels that cause longer delays on data transmission and more frequent errors for data transmission in a centralized CRN, a network administrator can regulate SU traffic over a specific channel set in such a way that the most important traffic does not go through those unreliable channels. Moreover, being aware of the mean operational time or MUT of a channel, routing and load balancing over multiple channels [132], [133] could provide more reliable solutions for QoS provisioning. For instance, the channels with longer MUTs can be allocated to services which have longer session duration to avoid forced terminations. On the other hand, when investigating specific QoS measures, dependability theory provides a more systematic methodology for performance analysis. For instance, while channel availability is a main QoS consideration for newly arriving users, the probability of successful completion of an already commenced service (referred to as *service retainability*) appears more important for ongoing users. From here onwards, we focus on dependability performance of multi-channel CRNs.

4.2.1 Related Work

Reliability and availability analysis of CRNs is still a quite new research area, although there exists certain amount of work related to this direction. While some of the existing work considered measurement-based studies to evaluate channel availability [134], the reliability aspect of CRNs in some other studies is evaluated via system-centric performance metrics such as packet delivery ratio, average congestion rate or reliable sensing [135–137]. Hereafter, we revisit some of them.

A steady state analysis of channel availability was performed in [138] by considering spectrum handoff functions in a multi-channel CRN. The PU signals were modeled therein as a two-state Markov chain to denote the presence and the absence of a PU signal. However, their analysis was not conducted from a dependability perspective. A novel method for channel availability analysis was proposed in [139] by considering time-slotted systems. A distributed online learning and adaptive channel access scheme was proposed in [140] by estimating channel availability with the help of SU's own sensing observation history on collision events. For accurate spectrum sensing, the authors of [140] proposed an ad hoc network based approach. Considering both reliability and delay, a reliable route selection scheme based on channel availability was proposed in [141] for a CRN to avoid unstable routes due to PU activities. For each optional route between the sender and the destination nodes, the reliability is evaluated and a route with higher availability is preferred for routing decision. An opportunistic scheduling policy was proposed in [136] to maximize SU's throughput by considering reliability guarantee for PU transmissions. Stationary PUs and the mobility of SUs were considered in their time-slotted model.

However, most of these studies were not pursued from a dependability perspective. The importance of employing *dependability theory* for reliability analysis is twofold. First and most importantly, the terminologies related to QoS of telecommunication services have been defined in the ITU-T E.800 Recommendation [142] based on the dependability theory. Correspondingly, our definitions of a few reliability metrics introduced in this dissertation conform with the ITU recommendations since this theory forms the basis for our study. Therefore we can guarantee the validity and the eligibility of these definitions as well as the models developed in this study. Second, the system times defined in this dissertation can be accurately and systematically adopted for reliability analysis in CRNs from the perspective of the dependability theory. In the following sections we define and analyze those system times in more details.

4.2.2 System Times

Let the dependable system of interest be a CRN with one or multiple channels. Without loss of generality, we study in this thesis work the reliability performance of the SN. However, a similar analysis can be performed for the PN as well. The above mentioned reliability and availability metrics are defined and adopted to DSA strategies in CRNs. More specifically, once the system is in a state in which the new SUs are blocked, the system is regarded as in the failed or channel unavailable mode. When the system is in a state in which a new SU can commence its transmission, the system is said to be in the operational or channel available mode. Correspondingly, the channel availability for the SN is defined as *the fraction of time that the CRN can allocate at least the minimum number of required channels for a new SU request.*

Moreover, the metrics which are related to time average quantities such as MUT, MDT, MTTF and MTFF are considered as *system times* in a CRN. MUT and MDT denote the mean channel available time and mean channel unavailable time of the CRN. Furthermore, MTFF represents the mean time to the occurrence of first channel unavailability with respect to the system's starting instant. Similarly, the MTTF of a CRN is defined as the mean time from a random instant during the system operation to the instant when channel unavailability occurs. In the next two sections, we revisit first the Laplace transformation based analysis of system times proposed in our recent work [143] and then introduce briefly the phase-type (PH) distribution based analysis which is adopted in Paper B.

4.3 CTMC Modeling for System Time Analysis

Consider a CRN with M non-overlapping channels and multiple PUs and SUs. The arrivals of PU and SU services follow a Poisson process and their service times are exponentially distributed. Without loss of generality, consider a DSA strategy such as the strategy given in Subsec. 2.3.1 or Subsec. 3.3.1. Furthermore, assume that the total number of states in this system is N. Let $\boldsymbol{x}$ represent a general state in the system and $\pi(\boldsymbol{x})$ be the steady state probability of being in state $\boldsymbol{x}$. The set of feasible states of the system is denoted as $\mathcal{S}$. Assume that out of N states, K states can provide the required number of channels for a new SU request. Rearrange the state space in a way such that these K states are $1, 2, \cdots, K$ and states $K+1,\ K+2,\ \cdots,\ N$ represent those states in which the system fails to admit new SU requests.

Let $\mathcal{S}_{\mathcal{A}}$ be the set of states that channels are available for new SU requests, i.e., the set of channel available states of the system and $\mathcal{S}_{\mathcal{B}}$ be the set of states that the system blocks new SU service requests, i.e., the set of channel unavailable states of the system. Thus, $\mathcal{S} = \mathcal{S}_{\mathcal{A}} \cup \mathcal{S}_{\mathcal{B}}$ and $\mathcal{S}_{\mathcal{A}} \cap \mathcal{S}_{\mathcal{B}} = \emptyset$. Since $n(\mathcal{S}_{\mathcal{A}}) = K$, then $n(\mathcal{S}_{\mathcal{B}}) = N - K$, given that $n(\mathcal{S}) = N$, where $n(\mathcal{A})$ denotes the cardinality of a set $\mathcal{A}$. Then we divide the system states into two parts and formulate the transition rate matrix $\boldsymbol{Q}$ in the following partitioned form.

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}, \tag{4.9}$$

where the $K \times K$ matrix $\boldsymbol{A}$ represents the transition rates from a state in $\mathcal{S}_{\mathcal{A}}$ to another state in $\mathcal{S}_{\mathcal{A}}$ and the $L \times (N-K)$ matrix $\boldsymbol{B}$ represents the transition rates from a state in $\mathcal{S}_{\mathcal{A}}$ to a state in $\mathcal{S}_{\mathcal{B}}$ and so on. The elements in the main diagonal of $\boldsymbol{A}$ (and $\boldsymbol{D}$) represent the negative sum of outgoing rates from the states in $\mathcal{S}_{\mathcal{A}}$ (respectively, $\mathcal{S}_{\mathcal{B}}$).

Let $P_{\boldsymbol{x}}(t)$ be the probability that the system is in state $\boldsymbol{x}$ at time t. Define $\boldsymbol{P}_A(t)$ as a K dimensional row vector with the transient probabilities of the channel available states and let $\boldsymbol{P}_B(t)$ denote the $N-K$ dimensional row vector with the transient probabilities of the channel unavailable states, as

$$\begin{aligned} \boldsymbol{P}_A(t) &\triangleq [P_1(t)\ P_2(t)\ \cdots\ P_K(t)], \\ \boldsymbol{P}_B(t) &\triangleq [P_{K+1}(t)\ P_{K+2}(t)\ \cdots\ P_N(t)]. \end{aligned} \tag{4.10}$$

Assume further that, at the system initial instant, i.e., $t=0$, the system can provide sufficient number of idle channels to new SU requests, i.e.,

$$\boldsymbol{P}_A(0)\boldsymbol{U}_K = 1 \text{ and } \boldsymbol{P}_B(0)\boldsymbol{U}_{N-K} = 0, \tag{4.11}$$

where $\boldsymbol{U}_K$ and $\boldsymbol{U}_{N-K}$ denote the column vector of K ones and $N-K$ ones respectively.

The mean time duration that the system can reside in the channel available state space until its first visit to the channel unavailable state space, i.e., MTFF, is an important parameter to evaluate different system times [144]. To calculate MTFF, we propose two approaches in this dissertation based on Laplace transforms (presented below) and the PH distribution (presented in Paper B) respectively.

4.3.1 Laplace Transform-based Approach to Calculate MTFF

Let the system stay initially in $\mathcal{S}_{\mathcal{A}}$. Recall that the system's visit to $\mathcal{S}_{\mathcal{B}}$ is considered as a failure of the network. By making all failure states in $\mathcal{S}_{\mathcal{B}}$ as absorbing states, we can compute the mean time to first visit to subset $\mathcal{S}_{\mathcal{B}}$. Using the forward Chapman-Kolmogorov equation [66], we have

$$\frac{d}{dt}\boldsymbol{P}_A(t) = \boldsymbol{P}_A(t)\boldsymbol{A}. \tag{4.12}$$

Similarly,

$$\frac{d}{dt}\boldsymbol{P}_B(t) = \boldsymbol{P}_A(t)\boldsymbol{B}. \tag{4.13}$$

The Laplace transform of a derivative, $f'(t)$, of an arbitrary differentiable function $f(t)$ is given by

$$\mathcal{L}[f'(t)] = s\widehat{f}(s) - f(0) \tag{4.14}$$

where $\mathcal{L}[f(t)]$ denotes the Laplace transform of function $f(t)$ and $\mathcal{L}[f(t)] = \widehat{f}(s)$ [145]. Therefore, from (4.12) we obtain

$$\begin{aligned} s\widehat{\boldsymbol{P}}_A(s) - \boldsymbol{P}_A(0) &= \widehat{\boldsymbol{P}}_A(s)\boldsymbol{A}, \\ (s\boldsymbol{I} - \boldsymbol{A})\widehat{\boldsymbol{P}}_A(s) &= \boldsymbol{P}_A(0), \\ \widehat{\boldsymbol{P}}_A(s) &= \boldsymbol{P}_A(0)(s\boldsymbol{I} - \boldsymbol{A})^{-1}, \end{aligned} \tag{4.15}$$

where $\widehat{\boldsymbol{P}}_A(s)$ indicates the Laplace transform of $\boldsymbol{P}_A(t)$ and $\boldsymbol{I}$ denotes the identity matrix of size K. Assume that the initial state of the system is a channel available state, we have $\boldsymbol{P}_B(0) = \boldsymbol{0}$. Then, the Laplace transform of (4.13) results in

$$\begin{aligned} s\widehat{\boldsymbol{P}}_B(s) - \boldsymbol{P}_B(0) &= \widehat{\boldsymbol{P}}_A(s)\boldsymbol{B}, \\ s\widehat{\boldsymbol{P}}_B(s) &= \boldsymbol{P}_A(0)(s\boldsymbol{I} - \boldsymbol{A})^{-1}B. \end{aligned} \tag{4.16}$$

The probability distribution of channel unavailable states at time t is given by $\boldsymbol{P}_B(t)$. Then, the probability that the CRN in a channel unavailable state at time t is given by $\boldsymbol{P}_B(t)\boldsymbol{U}_{N-K}$. Therefore, the Laplace transform of the probability density of failure or channel unavailability, $f_B(s)$, can be expressed as

$$f_B(s) = s\widehat{\boldsymbol{P}}_B(s)\boldsymbol{U}_{N-K} = \boldsymbol{P}_A(0)(s\boldsymbol{I} - \boldsymbol{A})^{-1}B\boldsymbol{U}_{N-K}. \tag{4.17}$$

However, the transition rate matrix follows $B\boldsymbol{U}_{N-K} = -\boldsymbol{A}\boldsymbol{U}_K$. Therefore,

$$f_B(s) = -\boldsymbol{P}_A(0)(s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{A}\boldsymbol{U}_K. \tag{4.18}$$

Now consider the following mathematical properties of Laplace transforms and moment generation function (MGF) of continuous RVs [146]. If $\mathcal{X}$ is a continuous RV with PDF $f_{\mathcal{X}}(x)$, then the MGF corresponding to this PDF is given by

$$M_{\mathcal{X}}(s) = \int_0^\infty e^{sx} f_{\mathcal{X}}(x)\, dx. \tag{4.19}$$

The Laplace transform of a real-valued function $f_{\mathcal{X}}(t)$, $t \geq 0$, is defined as

$$\widehat{f}_{\mathcal{X}}(s) = \int_0^\infty e^{-sx} f_{\mathcal{X}}(x)\, dx. \tag{4.20}$$

Therefore, the Laplace transform of $f_{\mathcal{X}}(t)$ is equal to the moment generating function evaluated at $-s$, i.e.,

$$\mathcal{L}[f_{\mathcal{X}}(s)] = M_{\mathcal{X}}(-s). \tag{4.21}$$

Furthermore, for a given RV $\mathcal{X}$, the MGF written as $M_{\mathcal{X}}(s)$, is defined for real values of s, as

$$M_{\mathcal{X}}(s) = E\{e^{s\mathcal{X}}\} \tag{4.22}$$

Hence,

$$\mathcal{L}[f_{\mathcal{X}}(s)] = E\{e^{-s\mathcal{X}}\}. \tag{4.23}$$

In a MGF, the r^{th} moment about the origin of $\mathcal{X}$ is obtained from the r^{th} derivative of the MGF evaluated at $s = 0$. It follows then that,

$$E\{\mathcal{X}^r\} = (-1)^r \frac{d^r}{ds^r} \mathcal{L}[f_{\mathcal{X}}(s)]\bigg|_{s=0} \quad \text{for } r = 1, 2, \cdots . \tag{4.24}$$

In our case, the Laplace transform of the PDF of failure is $f_B(s)$. Let T denote the RV which represents time to failure. Therefore, the r^{th} moment about the origin of the time to failure, i.e., T^r, can be obtained by

$$T^r = (-1)^r \frac{d^r}{ds^r} f_B(s)\bigg|_{s=0}. \tag{4.25}$$

By substituting $f_B(s)$ in (4.18), we obtain

$$T^r = r! \boldsymbol{P}_A(0)(-\boldsymbol{A})^{-r} \boldsymbol{U}_K. \tag{4.26}$$

The mean value of T, i.e., the first moment can be obtained at $r = 1$. Then

$$T^1 = \boldsymbol{P}_A(0)(-\boldsymbol{A})^{-1} \boldsymbol{U}_K. \tag{4.27}$$

It is obvious that the above derived mean value of time to failure equals to the MTFF of the SN. Therefore,

$$MTFF = \boldsymbol{P}_A(0)(-\boldsymbol{A})^{-1} \boldsymbol{U}_K. \tag{4.28}$$

Similar to (4.28), closed-form expressions for the other system times of the CRN, i.e., MUT, MDT and MTTF can also be derived. The corresponding procedure was discussed in our previous work in [143] with numerical results based on a simple DSA strategy.

4.4 MTFF Calculation based on the PH Distribution

Instead of using the Laplace transform method, we derive the same expression for calculating the mean time to first visit to a channel unavailable state in Paper B in Part II of this dissertation by using the PH distribution and its properties. Therein, the complementary cumulative distribution function (CCDF) of channel availability is also obtained for evaluating guaranteed levels of channel availability of the CRN. Moreover, the impact of DSA strategies on system times and channel availability is analyzed in Paper B under both homogeneous and heterogeneous channel scenarios.

4.5 Reliability Analysis under Channel Failures and Repairs

In the previous sections, our discussions on reliability analysis did not take link failures in a CRN into account. However in an error-prone channel environment, channel failures cannot be neglected, considering performance degradation of PU and SU services and insufficient QoS caused by such failures. Consequently, our analysis presented in this chapter has to be extended for modeling channel failures and repairs in order to capture the performance under more realistic channel conditions. Both Paper C and Paper D in Part II are dedicated for this purpose. In those two papers, we analyze two new reliability metrics referred to as service retainability and unserviceable probability in addition to channel availability. While Paper C analyzes those performance metrics in a *static* channel reservation scheme, Paper D proposes and analyzes a *dynamic* channel reservation scheme.

4.6 Chapter Summary

In this chapter, a mathematical framework to analyze CR systems from dependability theory's perspective is presented. The necessity of dependability theory for reliability analysis is emphasized with a summary of a few existing studies. The reliability and availability concepts in classical dependability theory are discussed and used to define reliability performance metrics for channel access in CRNs. In addition, important system times of a CRN are introduced considering CRN's channel availability variation in the time domain. To analyze those system times in a multi-channel CRN with Markovian PU/SU arrivals, a Laplace transform based approach is presented based on our recent work [143]. To reduce the complexity of the derivation process of system times, an alternative framework is systematically presented in Paper B by considering statistical properties of PH distributions. By considering channel failures and repairs in a CRN, Papers C and D evaluate also two other reliability metrics, i.e., retainability and network unserviceable probability, and propose channel reservation schemes. In those two papers, both static and dynamic channel reservation schemes are analyzed.

Chapter 5

Priority-based Queuing Schemes for Heterogeneous SU Traffic

In this chapter priority-based queuing schemes are studied for heterogeneous secondary traffic in CRNs. By introducing queues in a system, secondary services that would be blocked or forcibly terminated could be buffered and served later. Depending on the delay tolerance level of the interrupted secondary services, two queuing schemes are proposed. Numerical results demonstrate that the joint operation of queues and dynamic access schemes can further decrease blocking probability of secondary requests, however, at a cost of higher average delay. This chapter provides a summary and a further analysis of our early work published in [75].

Achieving efficient resource allocation in licensed frequency bands by dynamic spectrum access appears as one of the most promising approaches for CRN design [46, 69, 147–149]. How to resolve competition in a multi-user environment is one of the essential tasks related to DSA decision in a CRN. Especially when heterogeneous secondary traffic exists in a network, their preferences and QoS requirements will affect spectrum access decisions [150]. Among existing DSA approaches, channel aggregation which targets at increasing the capacity of a secondary network by allowing one SU service assemble multiple available channels for single transmission [52] is a popular technique[1]. Although CA leads to increased capacity of the secondary network, the blocking probability and the forced termina-

[1]CA has also been proposed as a key technique for LTE, LTE-A [151], [152] and WLANs [153], [154] whereby multiple carriers or channels can be flexibly aggregated in order to support higher data rates or higher throughput.

tion probability of SU services grow accordingly if the arrival rates of PU and SU services become higher [148]. Therefore, how to ensure low probabilities of blocking and forced termination for SU services despite higher user arrival rates appears as an interesting research question.

Consider conventional channel access in a CRN as a *loss system* where traffic flows would simply be rejected without being served if all channels are occupied. On the other hand, the newly arriving service requests or interrupted ongoing services can also be placed in a queue until the required number of channels is available in a *queuing system*. Motivated by this observation, two queue-based CA schemes are proposed in this chapter considering heterogeneous SU traffic in CRNs. Accordingly the performance of these schemes is evaluated.

5.1 Existing Queuing Schemes in CRNs

In several recent studies on CRN channel access, different queuing models have been introduced in order to enhance system performance. In [155], a queuing framework was developed to analyze the packet level performance of opportunistic access by CR users. Since their proposed spectrum scheduling is based on channel quality information (CQI), extra radio resources have to be allocated for CQI-based transmission. On the other hand, the spectrum scheduling scheme proposed in [155] provides less spectrum adaptation flexibility. In order to analyze the delay in CRNs, a stochastic fluid queue approach was adopted in [156]. In that paper, adaptive algorithms were developed for achieving minimum queuing delays in the SU network. However, the delay analysis presented in [156] did not differentiate between delay-sensitive and elastic SU traffic.

An essential feature of the model presented in [157] is that each channel in the network is considered as a single server queuing system. However, the network model considered in [157] is not suitable for delay-critical applications due to long waiting times. A two-class priority queuing network was modeled in [158] by characterizing average end-to-end delay of SU traffic in a CRN. Although the delay performance of multi-hop networks was analyzed for the proposed queuing networks in [158], the analysis was limited to a single channel CRN. Considering a multi-channel CRN, the probability mass function (PMF) of queue length was derived in [159]. Moreover, the work performed in [160–164] also analyzed queue performance in CRNs. Different from those existing studies, the queuing schemes proposed in this chapter are designed by jointly considering three main features, i.e., multi-channel CRNs, hybrid SU traffic and CA.

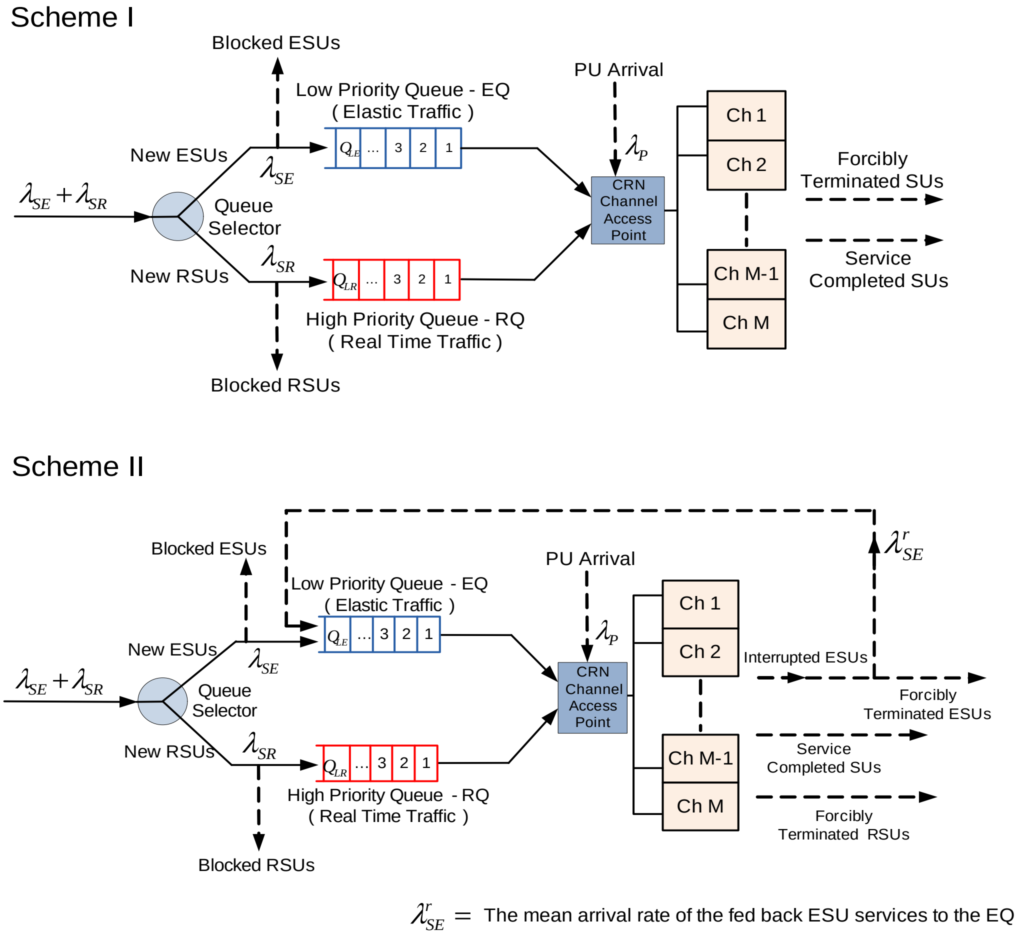


Figure 5.1: Illustration of the proposed queuing schemes where the forcibly terminated ESUs may re-join the EQ in SchemeII but not in Scheme I.

5.2 Two Proposed Priority-based Queuing Schemes

Consider a multi-channel CRN with two types of SU flows, i.e., elastic traffic and real-time traffic which can be characterized by distinct bandwidth requirement and service duration. While elastic flows such as file transfer and web browsing adjust their transmission rate according to network conditions [165], real-time applications, such as video conference and VoIP, have the same session duration even though the data rate may vary from time to time [166–168]. Herein, elastic and real-time SUs are denoted as *ESU* and RSU respectively while the term *SU* is used to represent both types of services. Both ESUs and RSUs are allowed to operate on the frequency band assigned to the PUs as long as the channel is not occupied by a licensed user. Moreover, we assume that one PU service occupies only a single channel while multiple channels can be aggregated by one SU service.

The proposed queuing model consists of two priority-based queues, a queue scheduling algorithm and a dynamic channel access strategy. The two queues are referred as EQ (for ESUs) and RQ (for RSUs) respectively. The queue referred to as *EQ* is exclusively allocated for elastic SU traffic while the queue referred to as *RQ* is dedicated for real-time SU services. The waiting services in the RQ will receive higher priority than the waiting services in the EQ during the process of idle channel allocation. In addition, two admission schemes are proposed for those two queues based on service arrivals to the queues. The main reason that we propose two distinct schemes is due to the consideration of hybrid SU traffic characteristics. If the network can serve preempted ESU services later without immediately terminating their connections, the forced termination probability of ESU services could be reduced. However, a cost of longer average waiting time for an *ESU service* is expected with this scheme. Therefore, for delay-sensitive applications like real-time traffic, it is better *not* to feed the preempted ESU services back to the EQ. The proposed queuing schemes, referred to as Scheme I and Scheme II respectively are illustrated in Fig. 5.1. Note that the only difference between these two schemes is that a queue access opportunity for interrupted ESU services is given in Scheme II, but not in Scheme I.

5.2.1 Scheme I: Queuing Scheme without Feedback Loop

In Scheme I, only the new services are allowed to enter the queuing system. Correspondingly, new ESU services will occupy waiting positions in the EQ while new RSU services will enter the RQ.

5.2.2 Scheme II: Queuing Scheme with Feedback Loop

In Scheme II the interrupted ESU services due to PU arrivals, which cannot find sufficient number of idle channels in the system by means of channel adaptation, are also inserted at the end of the EQ, together with the new ESU arrivals. However, if there are not enough waiting rooms in the EQ, the preempted ESUs are forced to terminate. Other than this feedback feature, the rest of the functions of Scheme II are the same as in Scheme I.

5.2.3 Dynamic Channel Access Strategy

The channel access strategy which is utilized in this chapter is enhanced from the dynamic CA strategy proposed in [148], i.e., $D(a, W, V)$. For the detailed description of the channel access strategy, please refer to Sec. IV-B in [75]. In $D(a, W, V)$,

Algorithm 5.1: Queue scheduling algorithm upon PU/SU departures and PU arrivals.

Input: n : Total number of newly idle channels
Input: β Ratio between priority levels assigned to the RQ and the EQ respectively
Input: j_{eq} : Total number of ESU services in the EQ
Input: j_{rq} : Total number of RSU services in the RQ
Input: W, V : Lower and upper bounds of the number of assembled channels for an ESU service
Input: a : Number of channels assembled by an RSU service
Output: n_E, n_R : Number of channels actually used by ESU and RSU services in the EQ and the RQ respectively
Output: n_O : Number of channels to be allocated to ongoing ESU services

if $n = 2k$ *where,* $k \in \mathbb{Z}^+$ **then**
    $n_r = \lfloor \frac{\beta}{\beta+1} n \rfloor$
    $n_e = n - n_r$
end
if $n = (2k+1)$ *where,* $k \in \mathbb{Z}^+$ **then**
    $n_r = \lceil \frac{\beta}{\beta+1} n \rceil$
    $n_e = n - n_r$
end
if $j_{eq} \geq 1$ **then**
    $n_{eq,\, max} = V \cdot j_{eq}$
    $n_{eq,\, min} = W$
else
    $n_{eq,\, max} = n_{eq,\, min} = 0$
end
$n_{rq,\, max} = a \cdot j_{rq}$
if $n_{rq,\, max} \geq n_r$ **then**
    $n_R = \lfloor \frac{n_r}{a} \rfloor \cdot a$
else
    $n_R = n_{rq,\, max}$
end
$n_e = n - n_R$
if $n_{eq,\, max} > n_e$ **then**
    if $n_e \geq n_{eq,\, min}$ **then**
        $n_E = n_e$
    else
        $n_E = 0$
    end
else
    $n_E = n_{eq,\, max}$
end
$n_Q = n_R + n_E$
$n_O = n - n_Q$

both spectrum handover and adjustable channel assembling are enabled and this strategy can be improved by integrating the priority-based queuing system presented above. The DSA strategy proposed in this chapter is denoted herein as *Dynamic*$(a, W, V, Q_{LE}, Q_{LR})$ where W and V denote the lower and upper bounds of the number of assembled channels for an ESU service and $a \in \mathbb{Z}^+$ represents the number of channels assembled by an RSU service. The queue lengths of EQ and RQ are denoted as Q_{LE} and Q_{LR} respectively. These parameters have the units as the number of services.

5.2.4 An Algorithm for Queue Scheduling

The above mentioned channel access strategy is implemented upon four different events, i.e., PU arrival, SU arrival, PU departure and SU departure. When channels become idle during one of those events, the vacant channels will be allocated for the waiting services in these two queues. A queue scheduling algorithm, Algorithm 5.1, is proposed for this channel allocation process on the waiting services.

One or several channels become idle when a PU or SU service departs or an ongoing SU service is forced to terminate upon PU arrivals. Those idle channels can first be allocated to the services waiting in the two queues. This can be done by dividing channel access opportunities between the EQ and the RQ according to the assigned priority levels [169]. The proposed allocation process of n newly available idle channels among the services in the EQ and the RQ is specified in Algorithm 5.1. In this algorithm, β: 1 represents the ratio between priority levels assigned to the RQ and the EQ respectively where, $\beta \in \mathbb{Q}^+$ and $\beta \geq 1$. Given the QoS requirements of elastic and real-time traffic, β needs to be properly assigned. Moreover, $\beta \geq 1$ is configured in order to give higher priority to the RSU services in the RQ. Implementing an adaptive scheme to adjust β online by increasing/decreasing β values based on the measurement of the target performance parameters can further add value to this study.

5.3 CTMC Modeling

For the sake of analysis convenience, we make a common assumption that Poisson arrivals and negative exponential distribution for service times apply to our studied CRNs. The arrival rates are denoted by λ_P, λ_{SE} and λ_{SR} for PU, ESU and RSU services respectively. The service times for PU and RSU services, and the volume of information been transferred in an ESU service are exponentially distributed, with corresponding service rates per channel μ_P, μ_{SR} and μ_{SE} respectively. The probability of occurring channel sensing errors in the system is assumed to be negligible [170], implying that the achieved secondary network capacity presented in this chapter is the ideal capacity. Let $\boldsymbol{x} = \{j_W, j_{W+1}, \cdots, j_k, \cdots, j_V, g_a, j_{pu}, j_{eq}, j_{rq}\}$ be the general state representation of the system and $\pi(\boldsymbol{x})$ be the steady state probability of being in state $\boldsymbol{x}$. Here, j_k denotes the number of ESU services with k aggregated channels where $W \leq k \leq V$, g_a is the number of RSUs with a aggregated channels, and j_{pu} denotes the number of PU services in the system. j_{eq} and j_{rq} denote the number of ESU services in the EQ and the number of RSU services in the RQ respectively.

The set of feasible states of the system is denoted as $\mathcal{S} = \{\boldsymbol{x} \mid j_W, j_{W+1}, \cdots j_V, g_a, j_{pu}, j_{eq}, j_{rq} \geq 0; b(\boldsymbol{x}) \leq M, j_{eq} \leq Q_{LE}, j_{rq} \leq Q_{LR}; \sum_{i=1}^{V-W} i j_{W+i} < W, \text{ if } j_{eq} > 0; \sum_{i=1}^{V-W} i j_{W+i} < a, \text{ if } j_{rq} > 0\}$, where $b(\boldsymbol{x})$ is the total number of utilized channels at state $\boldsymbol{x}$, i.e.,

$$b(\boldsymbol{x}) = j_{pu} + a g_a + \sum_{k=W}^{V} k j_k. \tag{5.1}$$

Correspondingly, $\boldsymbol{\pi}$, the steady state probability vector which can be calculated from the global balance equations, and the normalization equation, are given by

$$\boldsymbol{\pi Q} = \mathbf{0}, \ \sum_{\boldsymbol{x} \in \mathcal{S}} \pi(\boldsymbol{x}) = 1, \tag{5.2}$$

where $\boldsymbol{Q}$ denotes the transition rate matrix and $\mathbf{0}$ is a row vector of all 0's. The state transitions associated with different events with different conditions are summarized in Tables 1-4 in [75].

5.4 Performance Metrics - Scheme I

The following system-centric parameters are considered for evaluating the performance of the proposed queuing schemes.

5.4.1 Blocking Probability

A newly arrived ESU service is blocked if the system cannot find at least W channels from the set of idle channels and/or from the set of channels that could be donated by ongoing ESU services. Furthermore, the new ESU service can be blocked only if the EQ is fully occupied. Similarly, the blocking probability of RSU services can also be calculated given that they need exactly a channels to commence a service. Let P_{SE}^{B} and P_{SR}^{B} be the blocking probability of ESU and RSU services respectively. We have

$$P_{SE}^{B} = \sum_{\substack{\boldsymbol{x}\in\mathcal{S},\\ M-b(\boldsymbol{x})+\sum_{k=W+1}^{V}(k-W)j_k<W,\ j_{eq}=Q_{LE}}} \pi(\boldsymbol{x}), \text{ and} \tag{5.3}$$

$$P_{SR}^{B} = \sum_{\substack{\boldsymbol{x}\in\mathcal{S},\\ M-b(\boldsymbol{x})+\sum_{k=W+1}^{V}(k-W)j_k<a,\ j_{rq}=Q_{LR}}} \pi(\boldsymbol{x}). \tag{5.4}$$

5.4.2 Average Total Delay

Needless to say, delay is another important QoS metric in the proposed queuing model. The impact of the queue length is significant for delay performance and has to be investigated. In order to calculate the average delay of an SU service, both the waiting time in the queues and the transmission time of ongoing services are considered. Therefore, the total delay of an SU service equals to the average waiting time of SU services plus the average transmission time of an ongoing SU service. The total delay is averaged over the number of SU services. Furthermore, the following notations are used in this subsection.

N_{SE}^{s} and N_{SR}^{s}	:	Average number of ongoing ESU and RSU services,
T_{SE}^{s} and T_{SR}^{s}	:	Average transmission time of ongoing ESU and RSU services,
T_{SE}^{q} and T_{SR}^{q}	:	Average waiting time of ESU and RSU services in the queues,
T_{SE}^{t} and T_{SR}^{t}	:	Average total delay of ESUs and RSUs.

Therefore, we have the following expressions

$$N_{SE}^{s} = \sum_{\boldsymbol{x}\in\mathcal{S}} \sum_{k=W}^{V} \pi(\boldsymbol{x}) k j_k \quad \text{and} \quad N_{SR}^{s} = \sum_{\boldsymbol{x}\in\mathcal{S}} \pi(\boldsymbol{x}) a g_a.$$

Let L_{EQ} and L_{RQ} denote the mean queue length of ESU and RSU services respectively. They can be obtained as $L_{EQ} = \sum_{\boldsymbol{x}\in\mathcal{S}} \pi(\boldsymbol{x}) j_{eq}$ and $L_{RQ} = \sum_{\boldsymbol{x}\in\mathcal{S}} \pi(\boldsymbol{x}) j_{rq}$. Using Little's law [171], [172], the average total delay of ESU services, T_{SE}^{t}, and RSU services, T_{SR}^{t}, can be derived as

$$T_{SE}^{t} = T_{SE}^{q} + T_{SE}^{s} = \frac{L_{EQ}}{\lambda_{SE}^{*}} + \frac{N_{SE}^{s}}{\lambda_{SE}^{*}}, \tag{5.5}$$

$$T_{SR}^t = T_{SR}^q + T_{SR}^s = \frac{L_{RQ}}{\lambda_{SR}^*} + \frac{N_{SR}^s}{\lambda_{SR}^*} \tag{5.6}$$

where $\lambda_{SE}^* = \lambda_{SE}(1 - P_{SE}^B)$ and $\lambda_{SR}^* = \lambda_{SR}(1 - P_{SR}^B)$. Note that λ_{SE}^* and λ_{SR}^* represent respectively the mean admitted rate of new ESU and RSU services to the network.

5.5 Performance Metrics - Scheme II

As mentioned earlier, Scheme II extends Scheme I by allowing *interrupted ESUs' accessibility to the EQ*, given that the channel access rules and spectrum adaptation procedures for SUs are the same in both schemes. Thus the expressions for SU blocking probability, mean queue length and average delay of RSU services are the same in both cases. However, the average delay of an ESU service has to be analyzed separately for these two cases since the preempted ESU services in Scheme II is fed back into the EQ.

The service arrivals at the EQ in Scheme II consist of both new arrivals and interrupted ESU services. The mean admitted rate of new ESUs can be expressed as $\lambda_{SE}^* = \lambda_{SE}(1 - P_{SE}^B)$, while the mean admission rate of the preempted ESU services at the EQ is given as $\lambda_{SE}^q = R_{int} - R_f$ where R_{int} and R_f denote the mean PU interruption rate and the mean forced termination rate of the ESU services. They are obtained as

$$R_{int} = \sum_{\substack{\boldsymbol{x}\in\mathcal{S},\\ M=b(\boldsymbol{x}),\ j_{pu}<M,\ j_W>0,\ j_n=0;\ \forall n>W}} \frac{\lambda_P W j_W}{(M - j_{pu})}\pi(\boldsymbol{x}), \quad \text{and} \tag{5.7}$$

$$R_f = \sum_{\substack{\boldsymbol{x}\in\mathcal{S},\\ M=b(\boldsymbol{x}),\ j_{pu}<M,\\ j_W>0,\ j_n=0;\ \forall n>W,\ j_{eq}=Q_{LE}}} \frac{\lambda_P W j_W}{(M - j_{pu})}\pi(\boldsymbol{x}) \tag{5.8}$$

respectively. The forced termination probability of ESU services in Scheme II can therefore be expressed as the mean forced termination rate of ESU services, R_f, divided by the mean admitted ESU rate, i.e., λ_{SE}^*.

Correspondingly, the total arrival rate at the EQ, denoted as λ_{SE}^t, can be expressed as $\lambda_{SE}^t = \lambda_{SE}^* + \lambda_{SE}^q$. Based on Little's law, the average waiting time of the ESU services, T_{SE}^q, and the average transmission time of the ESU services, T_{SE}^s, can be obtained as

$$T_{SE}^q = \frac{L_{EQ}}{(\lambda_{SE}^* + \lambda_{SE}^q)} \quad \text{and} \quad T_{SE}^s = \frac{N_{SE}^s}{(\lambda_{SE}^* + \lambda_{SE}^q)}. \tag{5.9}$$

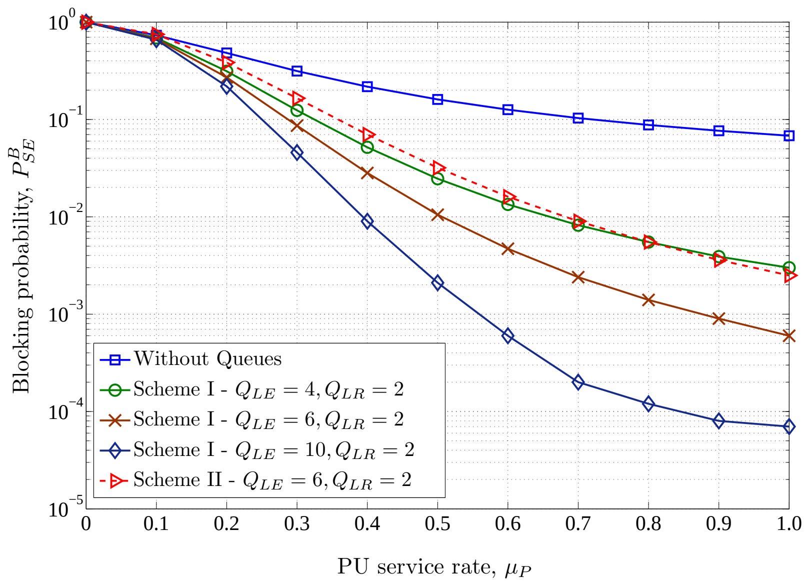


Figure 5.2: ESU blocking probability as a function of the PU service rate.

Define $U_r = \frac{\lambda_{SE}^q}{\lambda_{SE}^*}$ as the average number of times that an ongoing ESU service returns to the EQ due to a PU interruption. It gives

$$U_r = \frac{R_{int} - R_f}{\lambda_{SE}(1 - P_{SE}^B)}. \tag{5.10}$$

From (5.5), the average total delay of ESU services is given by $T_{SE}^t = (L_{EQ} + N_{SE}^s)/\lambda_{SE}^*$. Therefore we obtain

$$\begin{aligned} T_{SE}^t &= \frac{L_{EQ} + N_{SE}^s}{\lambda_{SE}^*}, \\ &= \frac{L_{EQ} + N_{SE}^s}{\lambda_{SE}^t} \cdot \frac{\lambda_{SE}^t}{\lambda_{SE}^*}, \\ &= \frac{L_{EQ} + N_{SE}^s}{\lambda_{SE}^t} \cdot \frac{\lambda_{SE}^* + \lambda_{SE}^q}{\lambda_{SE}^*}, \\ &= (T_{SE}^q + T_{SE}^s)(1 + U_r). \end{aligned} \tag{5.11}$$

5.6 Numerical and Simulation Results

In this section, the proposed queuing schemes are evaluated and the achieved system performance as a function of the users' arrival and service rates is illustrated. Furthermore, the optimal queue length for a given required QoS level is also studied.

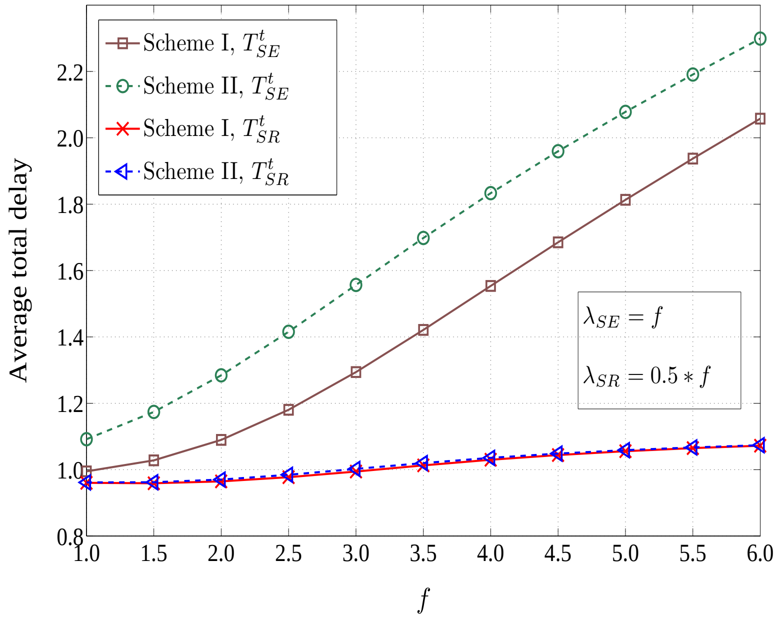


Figure 5.3: Average delay of SU services as a function of the service arrival rate.

5.6.1 Blocking Probability

Fig. 5.2 demonstrates the blocking probability of ESU services considering different queue length configurations. As a reference, the blocking probability in the case of without queues is also shown. It is observed that the blocking probability of SUs can be significantly reduced if the queues with a large number of waiting rooms are employed. Moreover, the blocking probability observed in Scheme II is higher than that of Scheme I for the same configuration. The reason is that the low priority queue in Scheme II is allocated for both new arrivals and interrupted services. Thus, the number of occupied waiting positions at the EQ is higher in Scheme II. Consequently, the probability of finding a vacant queue position becomes lower for new service arrivals. Therefore, more new ESU service requests are blocked and consequently the blocking probability in Scheme II is higher than in Scheme I.

5.6.2 Average Total Delay

In Fig. 5.3, a significant increase of delay in ESU services can be observed when the system employs Scheme II. The reason is that the EQ in Scheme II is more likely to be occupied than in Scheme I since the interrupted ESU services are also buffered. This will increase the waiting time of an ESU service in the EQ. However, there is no noticeable difference in RSU delay between these two schemes since Scheme II

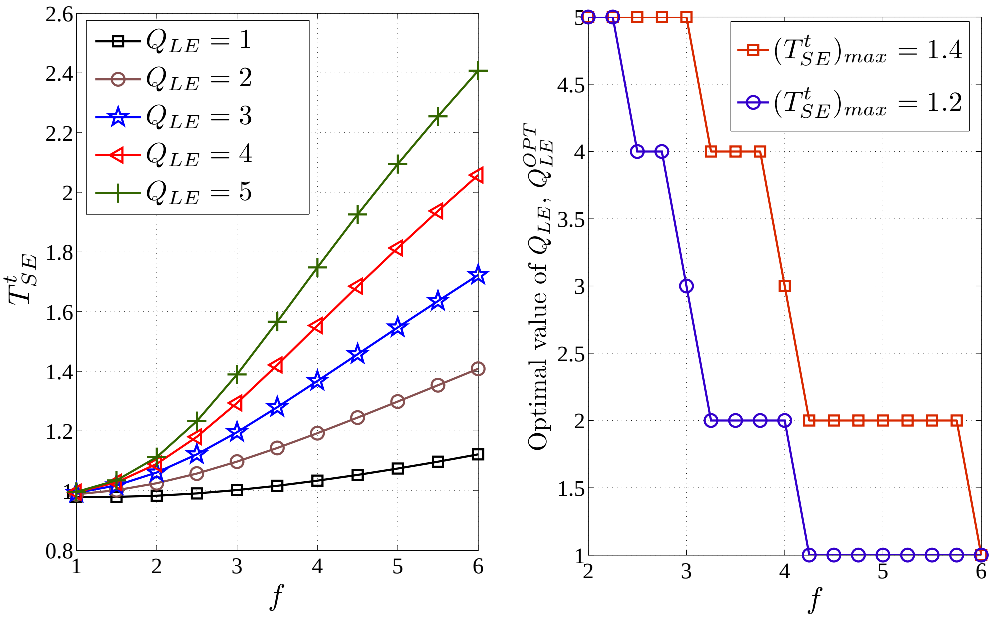


Figure 5.4: The optimal value of Q_{LE}, i.e., Q^{OPT}_{LE}, that can be allocated at the EQ as a function of f where $\lambda_{SE} = f$ and $\lambda_{SR} = f/2$. The value of Q^{OPT}_{LE} guarantees the average delay always being below a given threshold, $(T^t_{SE})_{max}$, while achieving the highest possible channel availability for SU services.

does not differ from Scheme I in terms of channel access for RSU services. Furthermore, the average waiting time of ESU services is much longer than that of RSU services owing to the assigned channel access priority as explained below. In the proposed schemes, more channel access opportunities are taken by RSU services in the RQ than ESU services in the EQ. According to the channel allocation procedure upon the occurrence of idle channels, a higher number of channels are allocated to the RSU services in the RQ as indicated in Lines 1-8 of Algorithm 5.1.

5.6.3 Optimal Queue Length

Although the blocking probability of new services can be reduced by extending the queue length, the experienced queuing delay becomes higher in the end. Therefore, it is worth noting that a long queue is not recommended for such a CRN due to the rapid increase of waiting time. Clearly, a further step needs to be performed to select the optimal queue length depending on the QoS requirements of users. For

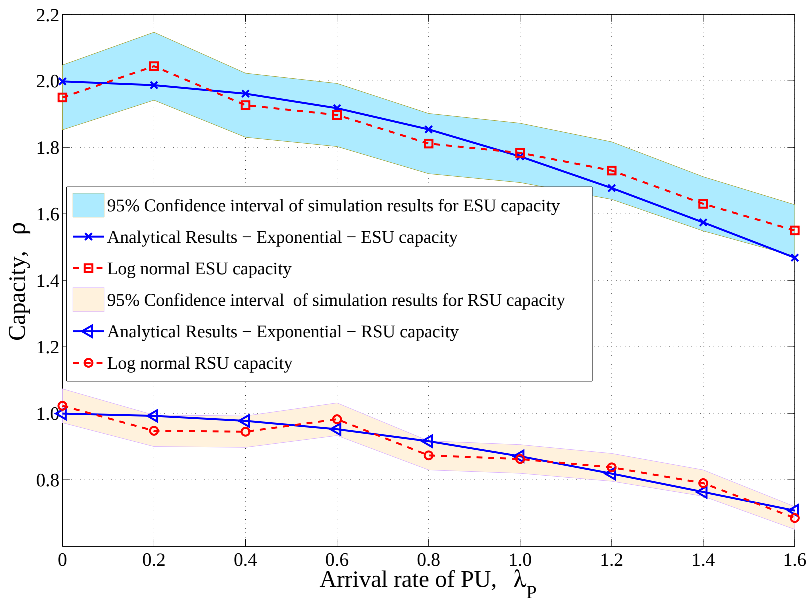


Figure 5.5: Capacity as a function of λ_P with the log-normal distribution with the same mean value and variance as used in the exponential distribution.

instance, consider the following optimization problem

$$\begin{aligned} &\text{maximize}_{\ Q_{LE}=0,1,\cdots,M} && \theta_{SE} \\ &\text{subject to:} && T^t_{SE} \leq (T^t_{SE})_{max} \end{aligned} \tag{5.12}$$

where $\theta_{SE} = 1 - P^B_{SE}$ and $(T^t_{SE})_{max}$ is the tolerable maximum average delay of an ESU service. Note that θ_{SE} is defined as the probability that the CRN can allocate at least the minimum number of required channels for a new SU request. In other words, θ_{SE} can be obtained by one minus the blocking probability of ESU services.

The value of the optimal queue length, denoted as Q^{OPT}_{LE}, given by the solution of (5.12), can be obtained under various given traffic conditions. As an example, Q^{OPT}_{LE} is determined while investigating the impact of the SU arrival rates on the optimal solution. Let the tolerable maximum average delay be $(T^t_{SE})_{max} = 1.4$ time unit when $0 \leq \lambda_{SE}, \lambda_{SR} \leq 6$ services per time unit. As observed in the left-side plot of Fig. 5.4, ESUs will experience $T^t_{SE} < 1.4$ of total average delay even with the configuration of $Q_{LE} = 5$ waiting positions when $\lambda_{SE} \leq 3$. Therefore, in order to maximize θ_{SE}, the system can allocate $Q_{LE} = 5$ unless the ESU arrival rate exceeds $\lambda_{SE} = 3$. Likewise, by analyzing the average delay of ESU services at different arrival rates, the value of Q^{OPT}_{LE} which satisfies the given QoS condition can be obtained for a given range of λ_{SE}. The obtained numerical results are illustrated

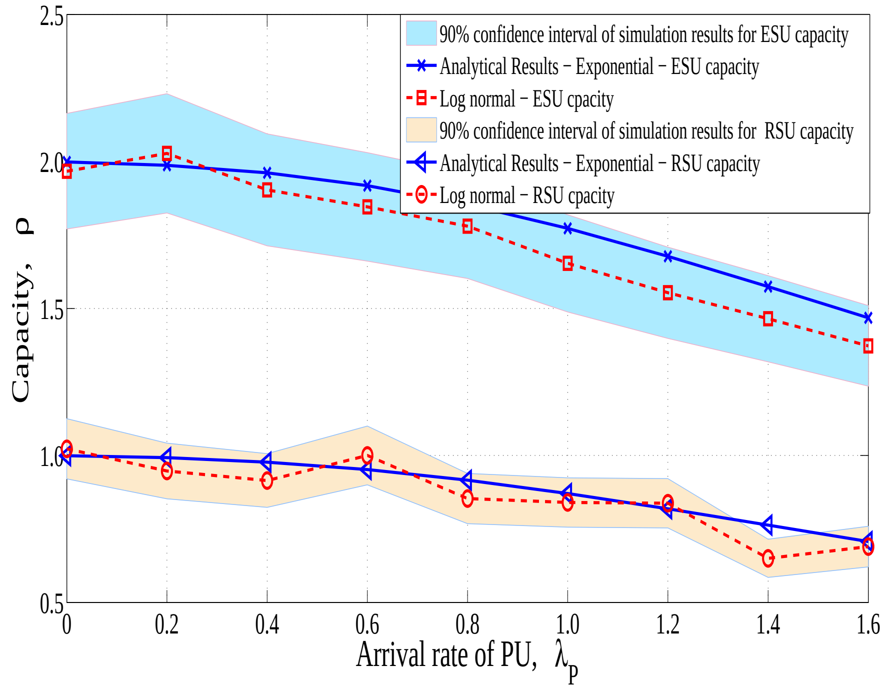


Figure 5.6: Capacity as a function of λ_P with the log-normal distribution with the same mean value and but a different variance from the exponential distribution.

in the right-side plot of Fig. 5.4 for $1 \leq \lambda_{SE} \leq 6$ for different values of the tolerable maximum average delay.

In addition, it is observed that at a given λ_{SE}, the optimal value for Q_{LE} needs to be decreased once the tolerable maximum average delay reduces. For instance, when $\lambda_{SE} = 5$, the length of the EQ should not exceed 2 if $(T_{SE}^t)_{max} = 1.4$. However, if the QoS constraint is made more stringent as $(T_{SE}^t)_{max} = 1.2$, the network cannot even accommodate $Q_{LE} = 2$ in the EQ.

5.6.4 Model Validation

In the developed CTMC model, one of the main assumptions we made is exponentially distributed service times for PU and SU services. In order to confirm the robustness and the applicability of the developed model, the sensitivity of the model to this assumption needs to be assessed. To this end, simulation results corresponding to log-normally distributed service times are illustrated in Fig. 5.5 with respect to the capacity of SU services. Many recent studies indicate that the most suitable models for call holding time distribution are log-normal and mixed log-normal distributions rather than an exponential distribution [173–175]. For comparison purposes, we include the corresponding analytical results of the original exponential distribution in the same figure. In this figure, *capacity* indicates the average number of service completions per time unit. Therefore the capacity of ESU and RSU services is obtained as

$$\rho_{SE} = \sum_{\boldsymbol{x} \in \mathcal{S}} \sum_{k=W}^{V} k j_k \mu_{SE} \pi(\boldsymbol{x}) \text{ and} \tag{5.13}$$

$$\rho_{SR} = \sum_{\boldsymbol{x} \in \mathcal{S}} a g_a \mu_{SR} \pi(\boldsymbol{x}) \tag{5.14}$$

respectively. The capacity curve is generated by simulations based on the log-normal distribution with the same mean value and the variance as used in the original exponential distribution for CTMC modeling, while the arrival process is still kept as a Poisson process. Evidently, the simulation results match very closely with the results obtained from the model, confirming that the proposed model is robust and applicable to other service time distributions as well.

Furthermore in Fig. 5.6, the same test is performed by re-configuring the variance of the log-normal distributions 20% higher than the variance of the original exponential distributions, while the mean value is kept the same. According to the obtained results, the theoretical exponential model still exhibits a good match with the simulation results even though the variance of the log-normal service time distribution is altered.

5.7 Chapter Summary

This chapter proposes two priority-based queuing schemes for heterogeneous traffic in a multi-channel CRN for improving QoS for secondary users in terms of blocking probability and forced termination probability. The proposed queuing schemes integrated together with a channel access strategy are modeled based on CTMC, and closed-form expressions for system performance are obtained. The simulation results based on both exponential and log-normal service time distributions are also presented. The numerical results obtained from analysis and simulations match closely with each other, confirming the correctness as well as the robustness of the proposed models. More details on the proposed queuing schemes, CTMC modeling and the numerical results can be found in [75].

Chapter 6

Conclusions and Future Work

This chapter is organized into three sections. In the first section, we summarize the dissertation work. The main contributions of this thesis work are listed in the second section. The third section points out a few potential research directions relevant to the topics addressed in this dissertation.

6.1 Conclusions

To accommodate the ever increasing demand on multimedia services in next generation wireless networks, more radio spectrum is required. Cognitive radio has been identified as a promising technology to meet this requirement since it can improve radio spectrum utilization significantly. In a cognitive radio network, unlicensed users or secondary users can temporarily utilize licensed users' or primary users' spectrum in an opportunistic manner. As discussed in Chap. 1 and Chap. 2, designing efficient channel access strategies becomes one of the fundamental tasks for CRNs.

Fixed channel allocation leads to spectrum under-utilization. Therefore developing smarter techniques for dynamic channel access is a basic requirement to improve spectrum utilization efficiency. On the other hand, the applicability of these channel access strategies depends heavily on their performance. The performance of channel access strategies needs to be analyzed from two aspects, i.e., system-centric performance evaluation and reliability-oriented performance evaluation. For performance assessment, heterogeneous traffic types and distinct QoS requirements of CRN users need to be considered. While a certain type of traffic may require shorter delay with higher throughput, another type of traffic may expect longer channel availability and higher reliability for its services. Although reliability and availability analysis of wireless networks is not a novel research area, it remains as a nearly un-chartered topic in the context of CRNs as stated in Chap. 4.

Based on the above observations, the research work performed in this dissertation covered two essential directions for designing cognitive radio networks. In one direction, dynamic spectrum access strategies are proposed and their performance is thoroughly investigated. On the other direction, the reliability and the availability of cognitive radio networks are analyzed from dependability theory's perspective. Both independent and joint considerations of those two research directions are presented in this thesis.

The designed dynamic channel access strategies in multi-channel CRNs include three different techniques, i.e., spectrum leasing, channel reservation and priority-based queuing. Another technique, channel aggregation, is also integrated in our priority-based queuing schemes. Spectrum leasing is appropriate when the licensed users are willing to lease part of their spectrum for obtaining an incentive from the secondary network. This dissertation work shows that dynamic spectrum leasing works better than static spectrum leasing for both primary and secondary networks in Chap. 3 and Paper A. Moreover, we observed that dynamic adjustment of spectrum leasing by considering existing channel occupancy status could minimize the increase of blocking probability for users. Meanwhile, an upper bound for the leasing bandwidth is also required in order to maintain primary network's access privilege.

The channel reservation schemes proposed in Paper C and Paper D mainly targeted at reducing forced termination probability for PU and SU services considering link failures. Ongoing SU services can be terminated before their service completion due to PU arrivals as well as channel failures in the network. However, ongoing PU services can only be terminated by channel failures. Although channel reservation schemes can reduce the forced termination probability for ongoing services, a rise of blocking probability is also observed. To deal with this tradeoff, an optimization problem could be formulated considering the required QoS levels of both PUs and SUs. It is shown that the allocation of an optimal number of reserved channels provides higher retainability to ongoing traffic without exceeding the upper limit of the blocking probability for newly incoming traffic.

The queuing schemes proposed in Chap. 5 employ several methods to enhance SU capacity. One of them allows channel aggregation for SU services and provides queue access opportunities for interrupted SU services in order to serve them later. These two features provide a significant performance enhancement for the secondary network without degrading primary network's performance. However, an increase of overall delay for SUs' transmission was observed consequently. The proposed priority-based queue with finite waiting rooms provides a solution to adequately control the increase of transmission delay, especially for delay-critical applications. In addition, we showed that the determination of an optimal queue size could guarantee an upper limit of overall transmission delay of services.

In this dissertation, especially in Paper B, the dependability theory constitutes the basis for our reliability and availability analysis in CRNs considering two facts. On the one hand, the recommendations provided by ITU E.800 related to QoS, reliability and network performance follow dependability theory. On the other hand, reliability and availability are two basic attributes of dependability. Furthermore, Paper C and Paper D also analyze reliability of CRNs by jointly considering channel access and dependability attributes. A set of reliability and availability performance metrics are defined, mathematically formulated and evaluated for a centralized multi-channel CRN. We show that the proposed analytical models can be applied to other channel access schemes under Markovian PU/SU arrivals, irrespective of the heterogeneity of traffic as well as channels. In addition, the CTMC models we developed take into account the random channel failures and channel repairs in CRNs. Therefore, we ascertain that the definitions and the models presented in this thesis collectively provide a systematic approach for reliability and availability analysis for channel access in multi-channel CRNs.

6.2 Contributions

As already mentioned, this dissertation concentrates mainly on two major directions within CRN research, i.e., dynamic channel access and reliability analysis. The work presented in this dissertation provides a comprehensive solution to the question on how to improve the performance of both primary and secondary networks by adopting various techniques that have been proposed for dynamic channel access. Moreover, the reliability metric definitions and analytical models proposed in this dissertation contribute to perform a systematic and accurate reliability analysis in CRNs. Accordingly, the main contributions of this thesis work are highlighted below.

- Two priority-based queuing schemes for secondary users are proposed so that the SU flows that would otherwise be blocked or forcibly terminated could be buffered and possibly served later. The queuing schemes are introduced based on the delay tolerance of the interrupted elastic services. Furthermore, CTMC models are developed to evaluate the performance of the proposed DSA strategy with queues, and the correctness as well as the preciseness of the derived theoretical models are verified through extensive simulations.

- Both static and dynamic spectrum leasing schemes are proposed to advance the state-of-the-art research on dynamic exclusive use in CRNs. Channel occupancy-based DSL is introduced and its performance is evaluated in terms of several system-centric parameters. A QoS-based non-linear cost function is also proposed to estimate the leasing cost. The tradeoff between secondary network's capacity and leasing cost is investigated via utility functions.

- Several reliability and availability performance measures for CRNs based on the dependability theory are defined and the mathematical expressions to evaluate them are derived. A few new terminologies for important system times in a CRN are introduced. In addition, extensive simulations are also conducted to ensure the robustness of these developed models in the presence of non-Markovian PU arrivals.

- A channel reservation scheme to improve the service retainability level of users in CRNs in the presence of random channel failures is designed. The proposed dynamic channel reservation and adjustment algorithms and the optimal channel reservation algorithm enhance the performance of both PNs and SNs while assuring the target QoS levels of PU and SU services. Furthermore, the performance differences between static and dynamic channel reservation schemes are also examined.

6.3 Future Work

Based on the research results achieved in this thesis work, a few potential topics can be identified. In brief, the following research directions are foreseen.

- In our developed framework for performance evaluation of spectrum leasing and channel reservation schemes, the analysis is performed in the frequency domain. So far, time based dynamic spectrum leasing schemes are not considered. Moreover, the impact of channel reservation is investigated considering both fixed and dynamically adjusted set of channels. We did not evaluate reserved time of a channel according to traffic intensity or/and channel status. Alternatively, time domain analysis for both spectrum leasing and channel reservation appears as an interesting topic.

- The queuing schemes proposed in this dissertation are restricted by introducing a fixed queue length. A variable length queue which can be adjusted according to the incoming traffic intensity could be a valuable extension to continue this work. To do so, the queue scheduling algorithm proposed in Chap. 5 needs to be improved in order to admit a variable queue length based on the current channel occupancy status and the arrival as well as service completion rates of services.
- In this dissertation, the network topology has been targeted on centralized CRNs. On the other hand, distributed network architectures of CRNs can be analyzed by modifying the proposed DSA schemes and techniques. When the proposed DSA schemes are applied, how the performance of the distributed CRNs differs from the centralized CRNs is another research question of interest. Moreover, one could adopt and possibly enhance the proposed reliability analysis to define and assess link reliability and routing path reliability in distributed CRNs.
- In 5G mobile and wireless networks, maintaining a high reliability level under tight latency constraints is considered as a main performance indicator. In order to achieve this, ultra reliable communication has recently been identified as an important emerging area for wireless communications. Therefore, how CRNs can be incorporated in other wireless networks in order to provide URC appears as another important research task.

REFERENCES

[1] ICT Data and Statistics Division, ITU-T Development Bureau, "ICT facts and figures," May 2015, [Online] Available: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf, Accessed on 30 Jun. 2016.

[2] J. Li, X. Wu, and R. Laroia, *OFDMA Mobile Broadband Communications: A Systems Approach*. New York, NY, USA: Cambridge University Press, 2013.

[3] R. Tandra, S. M. Mishra, and A. Sahai, "What is a spectrum hole and what does it take to recognize one?" *Proc. IEEE*, vol. 97, no. 5, pp. 824–848, May 2009.

[4] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal," in *Proc. Int. Workshop on Technology and Policy for Accessing Spectrum*, Boston, MA, USA, Sep. 2006, pp. 1–12.

[5] H. Karl and A. Willig, *Protocols and Architectures for Wireless Sensor Networks*. West Sussex, UK: John Wiley & Sons, 2005.

[6] E. Hossain, D. Niyato, and Z. Han, *Dynamic Spectrum Access and Management in Cognitive Radio Networks*. New York, NY, USA: Cambridge University Press, 2009.

[7] F. K. Jondral, "Software-defined radio: Basics and evolution to cognitive radio," *EURASIP J. Wirel. Commun. Netw.*, vol. 2005, no. 3, pp. 275–283, Aug. 2005.

[8] J. Mitola-III, "Cognitive radio: An integrated agent architecture for software defined radio," Ph.D. dissertation, Royal Institute of Technology (KTH), Sweden, May 2000, [Online] Available: http://www.diva-portal.org/smash/get/diva2:8730/FULLTEXT01.pdf, Accessed on 30 Jun. 2016.

[9] D. Ouattara, F. Krief, M. A. Chalouf, and O. Hamdi, *Spectrum Sensing Improvement in Cognitive Radio Networks for Real-Time Patients Monitoring*. Berlin, Heidelberg: Springer, 2012, pp. 179–188, [Online] Available: http://dx.doi.org/10.1007/978-3-642-37893-5_21, Accessed on 30 Jun. 2016.

[10] L. Gavrilovska, D. Denkovski, V. Rakovic, and M. Angjelichinoski, "Medium access control protocols in cognitive radio networks: Overview and general classification," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2092–2124, Fourth Quarter 2014.

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

[12] H. M. E. Misilmani, M. Y. Abou-Shahine, Y. Nasser, and K. Y. Kabalan, "Recent advances on radio-frequency design in cognitive radio," *Int. J. Antennas and Propagation*, vol. 2016, Article ID 9878475, pp. 1–16, Jan. 2016.

[13] E. Biglieri, A. J. Goldsmith, L. J. Greenstein, N. B. Mandayam, and H. V. Poor, *Principles of Cognitive Radio*. New York, NY, USA: Cambridge University Press, 2013, ISBN: 978-1-107-02875-3.

[14] V. Chakravarthy, X. Li, Z. Wu, M. A. Temple, F. Garber, R. Kannan, and A. Vasilakos, "Novel overlay/underlay cognitive radio waveforms using SD-SMSE framework to enhance spectrum efficiency-Part 1: Theoretical framework and analysis in AWGN channel," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3794–3804, Dec. 2009.

[15] Z. Chen, N. Guo, Z. Hu, and R. C. Qiu, "Experimental validation of channel state prediction considering delays in practical cognitive radio," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1314–1325, May 2011.

[16] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Survey on spectrum management in cognitive radio networks," *IEEE Commun. Mag.*, vol. 46, no. 4, pp. 40–48, Apr. 2008.

[17] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 1, pp. 116–130, First Quarter 2009.

[18] J. Lundén, M. Motani, and H. V. Poor, "Distributed algorithms for sharing spectrum sensing information in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4667–4678, Aug. 2015.

[19] X.-L. Huang, F. Hu, J. Wu, H.-H. Chen, G. Wang, and T. Jiang, "Intelligent cooperative spectrum sensing via hierarchical Dirichlet process in cognitive

radio networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 5, pp. 771–787, May 2015.

[20] D.-J. Lee, "Adaptive random access for cooperative spectrum sensing in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 831–840, Feb. 2015.

[21] F. Lin, R. C. Qiu, and J. P. Browning, "Spectrum sensing with small-sized data sets in cognitive radio: Algorithms and analysis," *IEEE Trans. Veh. Technol.*, vol. 64, no. 1, pp. 77–87, Jan. 2015.

[22] I. Christian, S. Moh, I. Chung, and J. Lee, "Spectrum mobility in cognitive radio networks," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 114–121, Jun. 2012.

[23] W.-Y. Lee and I. F. Akyildiz, "Spectrum-aware mobility managementin cognitive radio cellular networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 4, pp. 529–542, Apr. 2012.

[24] N. Janatian, S. Sun, and M. Modarres-Hashemi, "Joint optimal spectrum sensing and power allocation in CDMA-based cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 3990–3998, Sep. 2015.

[25] K. Hong, S. Sengupta, and R. Chandramouli, "SpiderRadio: A cognitive radio implementation using IEEE 802.11 components," *IEEE Trans. Mobile Comput.*, vol. 12, no. 11, pp. 2105–2118, Nov. 2013.

[26] A. A. Khan, M. H. Rehmani, and M. Reisslein, "Cognitive radio for smart grids: Survey of architectures, spectrum sensing mechanisms, and networking protocols," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 860–898, First Quarter 2016.

[27] M. Sooriyabandara, T. Farnham, P. Mähönen, M. Petrova, J. Riihijärvi, and Z. Wang, "Generic interface architecture supporting cognitive resource management in future wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 9, pp. 103–113, Sep. 2011.

[28] W. Cheng, X. Zhang, and H. Zhang, "Full-duplex spectrum-sensing and MAC-protocol for multichannel nontime-slotted cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 5, pp. 820–831, May 2015.

[29] N. Nguyen-Thanh, P. Ciblat, A. T. Pham, and V.-T. Nguyen, "Surveillance strategies against primary user emulation attack in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 4981–4993, Sep. 2015.

[30] M. Elkashlan, L. Wang, T. Q. Duong, G. K. Karagiannidis, and A. Nallanathan, "On the security of cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 8, pp. 3790–3795, Aug. 2015.

[31] G. Baldini, T. Sturman, A. R. Biswas, R. Leschhorn, G. Gódor, and M. Street, "Security aspects in software defined radio and cognitive radio networks: A survey and a way ahead," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 2, pp. 355–379, Second Quarter 2012.

[32] D. W. K. Ng, E. S. Lo, and R. Schober, "Multiobjective resource allocation for secure communication in cognitive radio networks with wireless information and power transfer," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3166–3184, May 2016.

[33] W. C. Ao, S.-M. Cheng, and K.-C. Chen, "Connectivity of multiple cooperative cognitive radio ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 2, pp. 263–270, Feb. 2012.

[34] M. G. Adian and H. Aghaeinia, "Resource allocation in MIMO-OFDM-based cooperative cognitive radio networks," *IEEE Trans. Commun.*, vol. 62, no. 7, pp. 2224–2235, Jul. 2014.

[35] R. Bordn, S. M. Snchez, S. B. Mafra, R. D. Souza, J. L. Rebelatto, and E. M. G. Fernandez, "Energy efficient power allocation schemes for a two-user network-coded cooperative cognitive radio network," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1654–1667, Apr. 2016.

[36] T. Zhang, W. Chen, Z. Han, and Z. Cao, "Hierarchic power allocation for spectrum sharing in OFDM-based cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 8, pp. 4077–4091, Oct. 2014.

[37] F. Zhou, N. C. Beaulieu, Z. Li, J. Si, and P. Qi, "Energy-efficient optimal power allocation for fading cognitive radio channels: Ergodic capacity, outage capacity, and minimum-rate capacity," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2741–2755, Apr. 2016.

[38] S. Dadallage, C. Yi, and J. Cai, “Joint beamforming, power, and channel allocation in multiuser and multichannel underlay MISO cognitive radio networks,” *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3349–3359, May 2016.

[39] Q. Jiang, V. C. M. Leung, M. T. Pourazad, H. Tang, and H.-S. Xi, “Energy-efficient adaptive transmission of scalable video streaming in cognitive radio communications,” *IEEE Syst. J.*, vol. 10, no. 2, pp. 761–772, Jun. 2016.

[40] Q. Liu, X. Wang, B. Han, X. Wang, and X. Zhou, “Access delay of cognitive radio networks based on asynchronous channel-hopping rendezvous and CSMA/CA MAC,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 3, pp. 1105–1119, Mar. 2015.

[41] R. Engelman, K. Abrokwah, G. Dillon, G. Foster, G. Godfrey, T. Hanbury, C. Lagerwerff, W. Leighton, M. Marcus, R. Noel, J. Payton, J. Tomchin, J. Williams, and A. Yang, “Report of the spectrum efficiency working group,” Spectrum Policy Task Force, Federal Communications Commission (FCC), Tech. Rep., Nov. 2002, [Online] Available: https://transition.fcc.gov/sptf/files/SEWGFinalReport_1.pdf, Accessed on 30 Jun. 2016.

[42] A. O. Bicen, E. B. Pehlivanoglu, S. Galmés, and O. B. Akan, “Dedicated radio utilization for spectrum handoff and efficiency in cognitive radio networks,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5251–5259, Sep. 2015.

[43] Y. Long, H. Li, H. Yue, M. Pan, and Y. Fang, “SUM: Spectrum utilization maximization in energy-constrained cooperative cognitive radio networks,” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 11, pp. 2105–2116, Nov. 2014.

[44] T. Jiang, Z. Wang, L. Zhang, D. Qu, and Y.-C. Liang, “Efficient spectrum utilization on TV band for cognitive radio based high speed vehicle network,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5319–5329, Oct. 2014.

[45] J. J. Lehtomäki, M. López-Benítez, K. Umebayashi, and M. Juntti, “Improved channel occupancy rate estimation,” *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 643–654, Mar. 2015.

[46] L. Jiao, V. Pla, and F. Y. Li, “Analysis on channel bonding/aggregation for multi-channel cognitive radio networks,” in *Proc. European Wireless (EW)*, Lucca, Italy, Apr. 2010, pp. 468–474.

[47] H. Lee, S. Vahid, and K. Moessner, "A survey of radio resource management for spectrum aggregation in LTE-Advanced," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 2, pp. 745–760, Second Quarter 2014.

[48] Z. Khan, H. Ahmadi, E. Hossain, M. Coupechoux, L. A. DaSilva, and J. J. Lehtomäki, "Carrier aggregation/channel bonding in next generation cellular networks: Methods and challenges," *IEEE Netw.*, vol. 28, no. 6, pp. 34–40, Nov. 2014.

[49] P. Ren, Y. Wang, and Q. Du, "CAD-MAC: A channel-aggregation diversity based MAC protocol for spectrum and energy efficient cognitive ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 237–250, Feb. 2014.

[50] H. A. B. Salameh, M. Krunz, and D. Manzi, "Spectrum bonding and aggregation with guard-band awareness in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 3, pp. 569–581, Mar. 2014.

[51] L. Li, K. Wang, and W. Zhou, "Queuing method in combined channel aggregation and fragmentation strategy for dynamic spectrum access," in *Proc. IEEE PIMRC*, Sydney, Australia, Sep. 2012, pp. 1214–1219.

[52] L. Jiao, I. A. M. Balapuwaduge, F. Y. Li, and V. Pla, "On the performance of channel assembling and fragmentation in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5661–5675, Oct. 2014.

[53] C. T. Do, N. H. Tran, C. S. Hong, S. Lee, J.-J. Lee, and W. Lee, "A lightweight algorithm for probability-based spectrum decision scheme in multiple channels cognitive radio networks," *IEEE Commun. Lett.*, vol. 17, no. 3, pp. 509–512, Mar. 2013.

[54] J. Hu, L.-L. Yang, and L. Hanzo, "Maximum average service rate and optimal queue scheduling of delay-constrained hybrid cognitive radio in Nakagami fading channels," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 2220–2229, Jun. 2013.

[55] A. Azarfar, J.-F. Frigon, and B. Sansò, "Delay analysis of multichannel opportunistic spectrum access MAC protocols," *IEEE Trans. Mobile Comput.*, vol. 15, no. 1, pp. 569–581, Jan. 2016.

[56] X. Chen, H.-H. Chen, and W. Meng, "Cooperative communications for cognitive radio networks from theory to applications," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1180–1192, Third Quarter 2014.

[57] M. B. Pandian, M. L. Sichitiu, and H. Dai, "Optimal resource allocation in random access cooperative cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 6, pp. 1245–1258, Jun. 2015.

[58] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Spectrum leasing as an incentive towards uplink macrocell and femtocell cooperation," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 478–488, Apr. 2012.

[59] M. Naeem, A. Anpalagan, M. Jaseemuddin, and D. C. Lee, "Resource allocation techniques in cooperative cognitive radio networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 2, pp. 729–744, Second Quarter 2014.

[60] P. Li, S. Guo, W. Zhuang, and B. Ye, "On efficient resource allocation for cognitive and cooperative communications," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 264–273, Feb. 2014.

[61] J. J. Alcaraz, M. López-Martínez, and J. Garcia-Haro, "Bandwidth reservation as a coexistence strategy in opportunistic spectrum access environments," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 3, pp. 478–488, Mar. 2014.

[62] S.-H. Wang, C.-Y. Hsu, and Y.-W. P. Hong, "Distributed exploitation of spectrum and channel state information for channel reservation and selection in interweave cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3458–3472, Jul. 2013.

[63] I. A. M. Balapuwaduge, F. Y. Li, and V. Pla, "Significance of channel failures on network performance in CRNs with reserved spectrum," in *Proc. IEEE ICC*, Kuala Lumpur, Malasia, May 2016.

[64] G. G. Yin and Q. Zhang, *Two-Time-Scale Methods and Applications*. New York, NY, USA: Springer New York, 2005, pp. 315–347, [Online] Available: http://link.springer.com/book/10.1007%2Fb138226, Accessed on 30 Jun. 2016.

[65] G. Rubino and B. Sericola, *Markov Chains and Dependability Theory*. New York, NY, USA: Cambridge University Press, 2014.

[66] C. G. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2008.

[67] S. Joshi, P. Pawełczak, D. Čabrić, and J. Villasenor, "When channel bonding is beneficial for opportunistic spectrum access networks?" *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 3942–3956, Nov. 2012.

[68] S. L. Castellanos-López, F. A. Cruz-Pérez, M. E. Rivero-Angeles, and G. Hernández-Valdez, "Joint connection level and packet level analysis of cognitive radio networks with VoIP traffic," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 3, pp. 601–614, Mar. 2014.

[69] L. Jiao, F. Y. Li, and V. Pla, "Dynamic channel aggregation strategies in cognitive radio networks with spectrum adaptation," in *Proc. IEEE GLOBECOM*, Houston, TX, USA, Dec. 2011, pp. 1–6.

[70] M. Guizani, A. Rayes, B. Khan, and A. Al-Fuqaha, *Network Modeling and Simulation: A Practical Perspective*. West Sussex, UK: Wiley, 2010.

[71] M. López-Benítez and F. Casadevall, "Time-dimension models of spectrum usage for the analysis, design, and simulation of cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 2091–2104, Jun. 2013.

[72] T. R. Newman, S. M. S. Hasan, D. DePoy, T. Bose, and J. H. Reed, "Designing and deploying a building-wide cognitive radio network testbed," *IEEE Commun. Mag.*, vol. 48, no. 9, pp. 106–112, Sep. 2010.

[73] I. C. Paschalidis and S. Vassilaras, "Importance sampling for the estimation of buffer overflow probabilities via trace-driven simulations," *IEEE/ACM Trans. Netw.*, vol. 12, no. 5, pp. 907–919, Oct. 2004.

[74] J. Banks, *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. New York, NY, USA: John Wiley & Sons, 1998.

[75] I. A. M. Balapuwaduge, L. Jiao, F. Y. Li, and V. Pla, "Channel assembling with priority-based queues in cognitive radio networks: Strategies and performance evaluation," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 630–645, Feb. 2014.

[76] B. Canberk, I. F. Akyildiz, and S. Oktug, "Primary user activity modeling using first-difference filter clustering and correlation in cognitive radio networks," *IEEE/ACM Trans. Netw.*, vol. 19, no. 1, pp. 170–183, Feb. 2011.

[77] N. Rahimian, C. N. Georghiades, M. Z. Shakir, and K. A. Qaraqe, "On the probabilistic model for primary and secondary user activity for OFDMA-based cognitive radio systems: Spectrum occupancy and system throughput perspectives," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 356–369, Jan. 2014.

[78] A. S. Alfa, V. Pla, J. Martinez-Bauset, and V. Casares-Giner, "Discrete time analysis of cognitive radio networks with imperfect sensing and saturated source of secondary users," *Computer Communications*, vol. 79, pp. 53–65, Jan. 2016.

[79] I. Sobron, W. A. Martins, M. L. R. de Campos, and M. Velez, "Incumbent and LSA licensee classification through distributed cognitive networks," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 94–103, Jan. 2016.

[80] Electronic Communications Committee, WG-Frequency Management, "ECC report 205, Licensed Shared Access," European Conference of Postal and Telecommunications Administrations (CEPT), Tech. Rep., Feb. 2014, [Online] Available: http://www.erodocdb.dk/Docs/doc98/official/pdf/ECCREP205.PDF, Accessed on 30 Jun. 2016.

[81] M. Mustonen, M. Matinmikko, M. Palola, S. Yrjölä, and K. Horneman, "An evolution toward cognitive cellular systems: Licensed shared access for network optimization," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 68–74, May 2015.

[82] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[83] H. Hakim, H. Boujemaa, and W. Ajib, "Performance comparison between adaptive and fixed transmit power in underlay cognitive radio networks," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4836–4846, Dec. 2013.

[84] G. Ozcan and M. C. Gursoy, "Optimal power control for underlay cognitive radio systems with arbitrary input distributions," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4219–4233, Aug. 2015.

[85] H. Chen, L. Liu, J. D. Matyjas, and M. J. Medley, "Cooperative routing for underlay cognitive radio networks using mutual-information accumulation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 7110–7122, Dec. 2015.

[86] H. Arslan and M. E. Şahin, "Cognitive UWB-OFDM: Pushing ultra-wideband beyond its limit via opportunistic spectrum usage," *J. Commun. Netw.*, vol. 8, no. 2, pp. 151–157, Jun. 2006.

[87] K.-C. Chen and R. Prasad, *Cognitive Radio Networks*. West Sussex, UK: John Wiley & Sons, 2009.

[88] F. Jasbi and D. K. C. So, "Hybrid overlay/underlay cognitive radio network with MC-CDMA," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2038–2047, Apr. 2016.

[89] G. Bansal, M. J. Hossain, V. K. Bhargava, and T. Le-Ngoc, "Subcarrier and power allocation for OFDMA-based cognitive radio systems with joint overlay/underlay spectrum access mechanisms," *IEEE Trans. Veh. Technol.*, vol. 62, no. 3, pp. 1111–1122, Mar. 2013.

[90] Y. Wang, P. Ren, F. Gao, and Z. Su, "A hybrid underlay/overlay transmission mode for cognitive radio networks with statistical quality-of-service provisioning," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1482–1498, Mar. 2014.

[91] J. Ai, "Dynamic resource allocation and stochastic scheduling strategies in wireless networks," Ph.D. dissertation, Rensselaer Polytechnic Institute, Troy, NY, Aug. 2008, [Online] Available: http://dl.acm.org/citation.cfm?id=1627261, Accessed on 30 Jun. 2016.

[92] C. Xin, M. Song, L. Ma, G. Hsieh, and C.-C. Shen, "An incentivized cooperative architecture for dynamic spectrum access networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5154–5161, Oct. 2013.

[93] A. W. Min, X. Zhang, J. Choi, and K. G. Shin, "Exploiting spectrum heterogeneity in dynamic spectrum market," *IEEE Trans. Mobile Comput.*, vol. 11, no. 12, pp. 2020–2032, Dec. 2012.

[94] J. Feng, R. Zhang, L. Hanzo, and S. X. Ng, "Cooperative medium access control based on spectrum leasing," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 297–307, Jan. 2014.

[95] A. Gavili and S. ShahbazPanahi, "Optimal spectrum leasing and resource sharing in two-way relay networks," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5030–5045, Oct. 2014.

[96] Y. Xu, L. Wang, C. Fischione, and V. Fodor, "Distributed spectrum leasing via vertical cooperation in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1588–1601, Feb. 2016.

[97] T. Elkourdi and O. Simeone, "Spectrum leasing via cooperative interference forwarding," *IEEE Trans. Veh. Technol.*, vol. 62, no. 3, pp. 1367–1372, Mar. 2013.

[98] M. Tao and Y. Liu, "Spectrum leasing and cooperative resource allocation in cognitive OFDMA networks," *J. Commun. Netw.*, vol. 15, no. 1, pp. 102–110, Feb. 2013.

[99] S. M. M. Toroujeni, S. M.-S. Sadough, and S. A. Ghorashi, "Spectrum leasing for OFDM-based cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 2131–2139, Jun. 2013.

[100] C. Zhai and W. Zhang, "Adaptive spectrum leasing with secondary user scheduling in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3388–3398, Jul. 2013.

[101] M. Hafeez and J. M. H. Elmirghani., "Dynamic spectrum leasing for bi-directional communication: Impact of selfishness," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2427–2437, Jun. 2016.

[102] F. Gómez-Cuba, R. Asorey-Cacheda, F. J. G.-C. no, and H. Huang, "Application of cooperative diversity to cognitive radio leasing: Model and analytical characterization of resource gains," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 40–49, Jan. 2013.

[103] P. Lin, J. Jia, Q. Zhang, and M. Hamdi, "Dynamic spectrum sharing with multiple primary and secondary users," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1756–1765, May 2011.

[104] X. Cao, Y. Chen, and K. J. R. Liu, "Cognitive radio networks with heterogeneous users: How to procure and price the spectrum?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1676–1688, Mar. 2015.

[105] A. A. Daoud, M. Alanyali, and D. Starobinski, "Pricing strategies for spectrum lease in secondary markets," *IEEE/ACM Trans. Netw.*, vol. 18, no. 2, pp. 462–475, Apr. 2010.

[106] N. H. Tran, C. S. Hong, Z. Han, and S. Lee, "Optimal pricing effect on equilibrium behaviors of delay-sensitive users in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 192–202, Nov. 2013.

[107] I. Sugathapala, I. Kovacevic, B. Lorenzo, S. Glisic, and Y. Fang, "Quantifying benefits in a business portfolio for multi-operator spectrum sharing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6635–6649, Dec. 2015.

[108] I. A. M. Balapuwaduge, A. Rajanna, M. Kaveh, and F. Y. Li, "Performance evaluation of three dynamic channel access strategies for spectrum leasing in CRNs," in *Proc. IEEE ICC*, London, UK, Jun. 2015, pp. 7570–7575.

[109] S. K. Jayaweera, *Signal Processing for Cognitive Radios*. Hoboken, NJ, USA: John Wiley & Sons, Nov. 2014.

[110] I. A. M. Balapuwaduge, F. Y. Li, A. Rajanna, and M. Kaveh, "Channel occupancy-based dynamic spectrum leasing in multichannel CRNs: Strategies and performance evaluation," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1313–1328, Mar. 2016.

[111] M. Hafeez and J. M. H. Elmirghani, "Analysis of dynamic spectrum leasing for coded bi-directional communication," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1500–1512, Sep. 2012.

[112] L. Deek, E. Garcia-Villegas, E. Belding, S.-J. Lee, and K. Almeroth, "Duopoly competition in dynamic spectrum leasing and pricing," *IEEE Trans. Mobile Comput.*, vol. 11, no. 11, pp. 1706–1719, Nov. 2012.

[113] S. K. Jayaweera and T. Li, "Dynamic spectrum leasing in cognitive radio networks via primary-secondary user power control games," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 3300–3310, Jun. 2009.

[114] S. K. Jayaweera, G. Vazquez-Vilar, and C. Mosquera, "Dynamic spectrum leasing: A new paradigm for spectrum sharing in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2328–2339, Jun. 2010.

[115] R. D. Taranto, P. Popovski, O. Simeone, and H. Yomo, "Efficient spectrum leasing via randomized silencing of secondary users," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3739–3749, Dec. 2010.

[116] I. Bajaj, Y. H. Lee, and Y. Gong, "A spectrum trading scheme for licensed user incentives," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4026–4036, Nov. 2015.

[117] R. Fan, Y. Zheng, J. An, H. Jiang, and X. Li, "Dynamic pricing over multiple rounds of spectrum leasing in cognitive radio," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1782–1789, Mar. 2016.

[118] D. Niyato and E. Hossain, "Competitive pricing for spectrum sharing in cognitive radio networks: Dynamic game, inefficiency of nash equilibrium, and

collusion," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 192–202, Jan. 2008.

[119] L. Zheng and C. W. Tan, "Cognitive radio network duality and algorithms for utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 500–513, Mar. 2013.

[120] H. Li, X. Cheng, K. Li, X. Xing, and T. Jing, "Utility-based cooperative spectrum sensing scheduling in cognitive radio networks," in *Proc. IEEE INFOCOM*, Turin, Italy, Apr. 2013, pp. 165–169.

[121] T.-H. Lee, Y.-W. Huang, and C.-N. Chen, "Optimal resource allocation for increasing strictly concave utility functions in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 61, no. 3, pp. 1306–1317, Mar. 2012.

[122] O. Semiari, W. Saad, S. Valentin, M. Bennis, and H. V. Poor, "Context-aware small cell networks: How social metrics improve wireless resource allocation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 5927–5939, Nov. 2015.

[123] M. Sheng, C. Xu, X. Wang, Y. Zhang, W. Han, and J. Li, "Utility-based resource allocation for multi-channel decentralized networks," *IEEE Trans. Commun.*, vol. 62, no. 10, pp. 3610–3620, Oct. 2014.

[124] L. Chen, B. Wang, X. Chen, X. Zhang, and D. Yang, "Utility-based resource allocation for mixed traffic in wireless networks," in *Proc. Computer Communications Workshops collocated with IEEE INFOCOM*, Shanghai, China, Apr. 2011, pp. 91–96.

[125] C. Yang and S. Jordan, "Resource allocation for semi-elastic applications with outage constraints in cellular networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1591–1606, Apr. 2015.

[126] Y. Zhao, S. Mao, J. O. Neel, and J. H. Reed, "Performance evaluation of cognitive radios: Metrics, utility functions, and methodology," *Proc. IEEE*, vol. 97, no. 4, pp. 642–659, Apr. 2009.

[127] Y. Cao, T. Jiang, M. He, and J. Zhang, "Device-to-device communications for energy management: A smart grid case," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 190–201, Jan. 2016.

[128] K. Mets, J. A. Ojea, and C. Develder, "Combining power and communication network simulation for cost-effective smart grid analysis," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1771–1796, Third Quarter 2014.

[129] E. Dubrova, *Fundamentals of Dependability*. New York, NY, USA: Springer New York, 2013, ch. Fault-Tolerant Design, pp. 5–20, [Online] Available: http://dx.doi.org/10.1007/978-1-4614-2113-9_2, Accessed on 30 Jun. 2016.

[130] G. Rubino and B. Sericola, *Markov Chains and Dependability Theory*. New York, NY, USA: Cambridge University Press, 2014, ISBN: 978-1-107-00757-4.

[131] H. Pham, *System Reliability Concepts*. London, UK: Springer London, 2006, ch. System Software Reliability, pp. 9–75, [Online] Available: http://link.springer.com/book/10.1007%2F978-0-387-68855-8, Accessed on 30 Jun. 2016.

[132] C. Qiu, H. Shen, S. Soltani, K. Sapra, H. Jiang, and J. O. Hallstrom, "CEDAR: A low-latency and distributed strategy for packet recovery in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 23, no. 5, pp. 1514–1527, Oct. 2015.

[133] Z. Ding and D. Wu, "Sliding-mode-based congestion control and scheduling for multiclass traffic over per-link queueing wireless networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 3, pp. 1276–1288, Mar. 2013.

[134] V. Kone, L. Yang, X. Yang, B. Y. Zhao, and H. Zheng, "The effectiveness of opportunistic spectrum access: A measurement study," *IEEE/ACM Trans. Netw.*, vol. 20, no. 6, pp. 2005–2016, Dec. 2012.

[135] N. Bouabdallah, B. Ishibashi, and R. Boutaba, "Performance of cognitive radio-based wireless mesh networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 1, pp. 122–135, Jan. 2011.

[136] R. Urgaonkar and M. J. Neely, "Opportunistic scheduling with reliability guarantees in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 8, no. 6, pp. 766–777, Jun. 2009.

[137] K. B. Letaief and W. Zhang, "Cooperative communications for cognitive radio networks," *Proc. IEEE*, vol. 97, no. 5, pp. 878–893, May 2009.

[138] D.-J. Lee and W.-Y. Yeo, "Channel availability analysis of spectrum handoff in cognitive radio networks," *IEEE Commun. Lett.*, vol. 19, no. 3, pp. 435–438, Mar. 2015.

[139] N. Jiang, J. Hua, Z. Shang, and K. Yang, "A new method for channel availability analysis and the associated policy design for selection of channel sensing order in CRNs," *IEEE Trans. Signal Process.*, vol. 64, no. 9, pp. 2443–2458, May 2016.

[140] M. Zandi, M. Dong, and A. Grami, "Distributed stochastic learning and adaptation to primary traffic for dynamic spectrum access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1675–1688, Mar. 2016.

[141] Y. Dai, J. Wu, and Y. Zhao, "Boundary helps: Reliable route selection with directional antennas in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4135–4143, Sep. 2015.

[142] "Definitions of terms related to quality of service: Recommendation ITU-T E.800," Telecommunication Standardization Sector of ITU, Sep. 2008, [Online] Available: http://www.itu.int/rec/T-REC-E.800-200809-I, Accessed on 30 Jun. 2016.

[143] I. A. M. Balapuwaduge and F. Y. Li, "System times and channel availability analyses in multi-channel cognitive radio networks," in *Proc. IEEE ICC*, Sydney, Australia, Jun. 2014, pp. 1320–1325.

[144] J. A. Buzacott, "Markov approach to finding failure times of repairable systems," *IEEE Trans. Rel.*, vol. R-19, no. 4, pp. 128–134, Dec. 1970.

[145] A. M. Cohen, *Numerical Methods and Algorithms*. Springer USA, 2007, ch. Numerical Methods for Laplace Transform Inversion, pp. 1–6, [Online] Available: http://link.springer.com/book/10.1007%2F978-0-387-68855-8, Accessed on 30 Jun. 2016.

[146] W. J. Stewart, *Probability, Markov Chains, Queues and Simulation: The Mathematical Basis of Performance Modeling*. Princeton, NJ, USA: Princeton University Press, 2009.

[147] E. W. M. Wong and C. H. Foh, "Analysis of cognitive radio spectrum access with finite user population," *IEEE Commun. Lett.*, vol. 13, no. 5, pp. 294–296, May 2009.

[148] L. Jiao, F. Y. Li, and V. Pla, "Modeling and performance analysis of channel assembling in multi-channel cognitive radio networks with spectrum adaptation," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2686–2697, Jul. 2012.

[149] B. Gao, Y. Yang, and J.-M. Park, "Channel aggregation in cognitive radio networks with practical considerations," in *Proc. IEEE ICC*, Kyoto, Japan, Jun. 2011, pp. 1–5.

[150] H.-P. Shiang, "Designing autonomic wireless multi-hop networks for delay-sensitive applications," Ph.D. dissertation, University of California, Oakland, CA, USA, 2009, [Online] Available: http://medianetlab.ee.ucla.edu/data/thesis/Dissertation_Shiang_final.pdf, Accessed on 30 Jun. 2016.

[151] Y. Rui, P. Cheng, M. Li, Q. Zhang, and M. Guizani, "Carrier aggregation for LTE-Advanced: Uplink multiple access and transmission enhancement features," *IEEE Wireless Commun.*, vol. 20, no. 4, pp. 101–108, Aug. 2013.

[152] R. Zhang, Z. Zheng, M. Wang, X. Shen, and L.-L. Xie, "Equivalent capacity in carrier aggregation-based LTE-A systems: A probabilistic analysis," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6444–6460, Nov. 2014.

[153] T. Li, Q. Ni, D. Malone, Y. Xiao, and T. Turletti, "Aggregation with fragment retransmission for very high-speed WLANs," *IEEE/ACM Trans. Netw.*, vol. 17, no. 2, pp. 591–604, Apr. 2009.

[154] L. Deek, E. Garcia-Villegas, E. Belding, S.-J. Lee, and K. Almeroth, "Intelligent channel bonding in 802.11n WLANs," *IEEE Trans. Mobile Comput.*, vol. 13, no. 6, pp. 1242–1255, Jun. 2014.

[155] M. M. Rashid, M. J. Hossain, E. Hossain, and V. K. Bhargava, "Opportunistic spectrum scheduling for multiuser cognitive radio: A queuing analysis," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5259–5269, Oct. 2009.

[156] S. Wang, J. Zhang, and L. Tong, "A characterization of delay performance of cognitive medium access," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 800–809, Feb. 2012.

[157] S. Tang and B. L. Mark, "Performance of a cognitive radio network with tolerable service degradation," in *Proc. Int. Workshop on Design of Reliable Communications Networks (DRCN)*, Washington DC, USA, Oct. 2009, pp. 11–20.

[158] D. Das and A. A. Abouzeid, "Spatial-temporal queuing theoretic modeling of opportunistic multihop wireless networks with and without cooperation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5209–5224, Sep. 2015.

[159] N. Tadayon and S. Aïssa, "Multi-channel cognitive radio networks: Modeling, analysis and synthesis," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 11, pp. 2065–2074, Nov. 2014.

[160] S. Li, E. Ekici, and N. Shroff, "Throughput-optimal queue length based CSMA/CA algorithm for cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 5, pp. 1098–1108, May 2015.

[161] T. M. C. Chu, H. Phan, and H.-J. Zepernick, "On the performance of underlay cognitive radio networks using M/G/1/K queueing model," *IEEE Commun. Lett.*, vol. 17, no. 5, pp. 876–879, May 2013.

[162] Y. Wang and K. Liu, "Statistical delay QoS protection for primary users in cooperative cognitive radio networks," *IEEE Commun. Lett.*, vol. 19, no. 5, pp. 835–838, May 2015.

[163] Z. Chang, T. Ristaniemi, and Z. Han, "Queueing game for spectrum access in cognitive radio networks," *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 2017–2020, Nov. 2015.

[164] F. Sheikholeslami, M. Nasiri-Kenari, and F. Ashtiani, "Optimal probabilistic initial and target channel selection for spectrum handoff in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 570–584, Jan. 2015.

[165] H. Yaïche, R. R. Mazumdar, and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 667–678, Oct. 2000.

[166] M. H. Cheung, A.-H. Mohsenian-Rad, V. W. S. Wong, and R. Schober, "Random access for elastic and inelastic traffic in WLANs," *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 1861–1866, Jun. 2010.

[167] M. Ge and S. Wang, "Fast optimal resource allocation is possible for multiuser OFDM-based cognitive radio networks with heterogeneous services," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1500–1509, Apr. 2012.

[168] T. Shu and H. Li, “QoS-compliant sequential channel sensing for cognitive radios,” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 11, pp. 2013–2025, Nov. 2014.

[169] K.-D. Wu and W. Liao, “On service differentiation for multimedia traffic in multi-hop wireless networks,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2464–2472, May 2009.

[170] A. A. El-Sherif and K. J. R. Liu, “Joint design of spectrum sensing and channel access in cognitive radio networks,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1743–1753, Jun. 2011.

[171] J. D. C. Little and S. C. Graves, *Little's Law*. Boston, MA, USA: Springer, 2008, pp. 81–100.

[172] A. S. Alfa, *Queuing Theory for Telecommunications: Discrete Time Modeling of a Single Node System*. New York, NY, USA: Springer, 2010.

[173] A. E. Xhafa and O. K. Tonguz, “Handover performance of priority schemes in cellular networks,” *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 565–577, Jan. 2008.

[174] S. Büyükçorak, G. K. Kurt, and O. Cengaver, “A probabilistic framework for estimating call holding time distributions,” *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 811–821, Feb. 2014.

[175] F. Barceló and J. Jordán, “Channel holding time distribution in public telephony systems,” *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 1615–1625, Jul. 2000.

PART II

A

Paper A

A

Title:	Channel Occupancy-based Dynamic Spectrum Leasing in Multi-channel CRNs: Strategies and Performance Evaluation
Authors:	Indika A. M. Balapuwaduge†, Frank Y. Li†, Amogh Rajanna‡, and Mostafa Kaveh‡
Affiliation:	† Dept. of Information and Communication Technology, University of Agder (UiA), N-4898 Grimstad, Norway ‡ Dept. of Electrical and Computer Engineering, University of Minnesota (UMN), Minneapolis, MN 55455, USA
Journal:	*IEEE Transactions on Communications*, vol. 64, no. 3, pp. 1313-1328, March 2016, DOI: 10.1109/TCOMM.2016.2521723.
Copyright ©:	IEEE

Channel Occupancy-based Dynamic Spectrum Leasing in Multi-channel CRNs: Strategies and Performance Evaluation

Indika A. M. Balapuwaduge, Frank Y. Li,
Amogh Rajanna, and Mostafa Kaveh

Abstract — **Spectrum leasing has been proposed as an effective approach for enabling more flexible spectrum utilization in CRNs. In CRNs, a primary network (PN) which consists of multiple primary users (PUs) can lease part of the licensed spectrum to secondary users (SUs) in exchange for operational benefits. The focus of this study is to investigate how and to what extent the PN allows spectrum leasing in CRNs, considering the QoS requirements of the PN and the secondary network (SN). Correspondingly, we propose two dynamic spectrum leasing strategies which can improve the QoS performance of SUs while ensuring sufficient remuneration for PUs. In order to dynamically adjust the portion of leased bandwidth, two forms of leasing algorithms which are based on channel occupancy of the PN and the SN respectively are proposed. Analytical models are developed to characterize system centric performance metrics including capacity, blocking probability and forced termination probability in both networks. Furthermore, a dynamic pricing scheme for spectrum leasing and utility-based resource allocation is introduced. Finally the performance of dynamic leasing is compared quantitatively with that of static leasing and approaches without leasing. Numerical results obtained from both analysis and simulations show that dynamically adjusted spectrum leasing improves network performance.**

Manuscript received April 28, 2015; revised August 28, 2015 and November 3, 2015; accepted January 16, 2016. The research leading to these results has received funding from the EU FP7-PEOPLE-IRSES program under grant agreement 247083, project acronym S2EuNet. This work was partially performed while Indika A. M. Balapuwaduge was visiting the Department of Electrical and Computer Engineering, University of Minnesota. This paper has been presented in part at the IEEE International Conference on Communications (ICC), London, June, 2015. The associate editor coordinating the review of this paper and approving it for publication was W. Saad.

I. A. M. Balapuwaduge and F. Y. Li are with the Department of Information and Communication Technology, University of Agder (UiA), N-4898 Grimstad, Norway (email: {indika.balapuwaduge; frank.li}@uia.no).

A. Rajanna and M. Kaveh are with the Department of Electrical and Computer Engineering, University of Minnesota (UMN), Minneapolis, MN 55455, USA (email: {raja0088; mos}@umn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier: 10.1109/TCOMM.2016.2521723

Keywords—Dynamic spectrum leasing, cognitive radio, spectrum adaptation, CTMC, utility function.

I. Introduction

With the emergence of cognitive radio (CR) technologies, the inefficiency of spectrum utilization due to outdated spectrum allocation policies is mitigated. Instead of static frequency allocation, CR enables much more flexible and effective spectrum utilization via dynamic spectrum access. Spectrum underlay and overlay techniques represent two basic forms for opportunistic spectrum access in cognitive radio networks (CRNs) [1]. In overlay CRNs, secondary (unlicensed) users (SUs) are allowed to access the licensed spectrum opportunistically when primary (licensed) users (PUs) are not occupying the licensed band [2]. In underlay CRNs, a PU and an SU can co-exist in a channel, as long as the induced interference at the PU receiver is below a given threshold [3].

In both overlay and underlay paradigms, assuring SUs' quality of service (QoS) is a challenging task due to PU activities and transmission constraints imposed on SUs. This is because in conventional CRNs, the access priority of PUs is always guaranteed while the QoS of SUs is mostly ignored. Exclusive channel allocation of the licensed spectrum for the secondary network (SN) is an approach to overcome the performance degradation of SUs caused by frequent spectrum handoffs and transmission limitations, and it requires higher *flexibility for CR channels* [4], [5]. For example, the U.S. Federal Communications Commission (FCC) accepts the development of *secondary market* allowing spectrum owners to sell a part of their spectrum.

Correspondingly, spectrum leasing has been proposed as a technique to improve SUs' QoS performance and the efficiency of wireless spectrum occupation while offering appropriate remuneration to PUs. This is an initiative for more efficient spectrum utilization when spectrum licensees are willing to sell or trade their spectra with third parties. Unlike legacy CR channel access schemes under the underlay and the overlay paradigms, spectrum leasing will relax the burden of interference management on the SN due to the fact that the channel access priority over the *leased spectrum* is given to the SUs. In the unleased spectrum band, however, the priority is still kept to PUs. Based on the concept of spectrum leasing, it is of interest to analyze how and to what extent the performance of the SN could be improved subject to a given set of PU's QoS objectives.

In this paper, we introduce a novel concept referred to as channel occupancy-based *dynamic* spectrum leasing considering the ongoing traffic from both PUs and

SUs in multi-channel CRNs. The proposed dynamic channel access (DCA) strategies assist the SN to achieve its QoS performance in a dynamically changing environment while the primary network (PN) gains a sufficient remuneration. Moreover, we study utility-based resource allocation by considering the QoS requirements of both networks.

A. Related Work

Spectrum leasing is an active topic in the CR research and many leasing paradigms have been proposed recently [6] - [26]. Several of these studies have focused on investigating cooperative transmission schemes [6] - [8] and pricing models [9] - [13] that can provide remunerations to the PN. How to perform dynamic spectrum leasing (DSL) in centralized networks where a central entity is responsible for communication between the PN and the SN was analyzed in [14], [15] whereas [16] proposed a game theory-based framework for modeling the competitive DSL in distributed networks. Two agent-based spectrum trading models were proposed in [17] and [18] where an agent plays a third-party role in the spectrum trading process. Therein, an agent coordinates the functions between the spectrum sellers and buyers instead of allowing direct spectrum trading between the PN and the SN. A comprehensive survey in [19] covers many existing studies on spectrum leasing in CRNs together with taxonomy definitions related to this subject.

Moreover, a novel network coding-based DSL technique was proposed in [6] in which the CRs assisted the PN in two-way communication. However, therein only a pair of primary nodes and a set of secondary nodes are considered, and no analysis was performed in [6] on possible negative effects to the PN. Another novel DSL scheme was proposed under a game theoretic framework in [20]. In that case, the amount of spectrum leasing is strongly controlled by a pre-defined interference level at the PU receiver. Accordingly, in their proposed model, SUs have to adjust their transmission power frequently depending on the interference level at PUs. Thereafter, a follow-up study [14] focused on dynamically adjusted spectrum leasing that PUs are willing to tolerate in response to the demand from secondary transmitters. However, the performance related to system centric parameters such as blocking probability, capacity of the PN and SN has not been investigated, neither in [14] nor in [20].

Recent research in [21] and [22] discussed a dynamic pricing strategy and a spectrum trading strategy respectively. However, none of them considered multiple PUs in their analyses. Although performance evaluation was done in both the PN and the SN, detailed discussions on the allocation of optimal amount of leased spectrum were not presented. The rationale behind spectrum leasing in CRNs was emphasized in [23] for the purpose of enhancing the QoS experienced by SU services. The proposed model suggests exclusively rent of sub-channels to the secondary transmission apart from the normal access channels. Consequently, the PN loses its accessibility over the rented sub-channels for its data traffic and spectrum utilization is also downgraded.

B. Contributions

The work presented in this paper is distinct from the above mentioned related work with respect to mainly three aspects. First, the proposed channel access schemes are based on the channel occupancy status of both the PN and the SN. A common assumption in many previous studies on DSL in CRNs is that the leased spectrum is occupied exclusively by the SN and the SN accesses only the leased spectrum. This approach considerably reduces the spectrum utilization of the network. In this paper, however, both opportunistic access and exclusive access of PUs and SUs over the leased and unleased bands are investigated when PUs lease a portion of their spectrum in *the frequency domain* to the SN. This approach will provide higher flexibility for channel access upon a sudden increase of spectrum demand of the PN or the SN. Second, heterogeneous traffic which has different QoS requirements is considered when proposing the access schemes. Third, the performance of those schemes is investigated in terms of several system centric parameters and the cost/benefit tradeoff between the PN performance loss and the SN performance gain is analyzed with respect to net utility. The main contributions of this paper are summarized as follows:

1. Our work addresses two main research questions in CRNs, i.e., how to design DCA strategies and how to develop mathematical models for evaluating the performance of these strategies. We propose two new DSL strategies which can be adopted for DCA in multi-channel CRNs including both homogeneous and heterogeneous channels. Each of these strategies can be applied to diverse scenarios considering the QoS requirements of PU and SU services. Unlike the existing literature on spectrum leasing, PU's opportunistic access to the leased spectrum is also introduced. The proposed strategies also facilitate resource allocation for heterogeneous traffic environments. Moreover, we derive the steady state probability vector of the continuous time Markov chain (CTMC) model and evaluate system performance via several system centric parameters.

2. To enable higher flexibility for DSL, we propose two forms of *leasing adjustment* algorithms based on the channel occupancy of PUs and SUs respec-

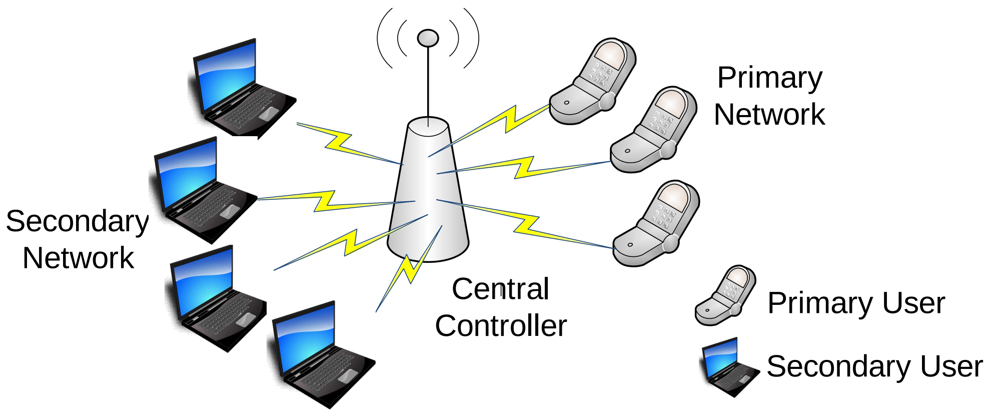


Figure A.1: Illustration of the centralized architecture for CRN.

tively. These two forms of leasing adjustments are adopted to the proposed DSL strategies, and are applicable to other DSL schemes as well.

3. To determine the optimal upper bound for the amount of leased spectrum given certain constraints, we evaluate the net utility of the SN by defining appropriate utility and cost functions. A dynamic pricing scheme is further proposed to develop the cost function for SUs. The price for the leased spectrum is determined based on the achieved QoS performance of SUs.

Furthermore, the correctness and the preciseness of the derived analytical models are validated through extensive discrete-event based simulations including simulations for different types of service time distributions. In brief, the novelty of this work lies on the developed systematic approach from channel access to performance analysis for dynamic spectrum leasing in CRNs. This systematic approach includes access strategy design, metric definition, CTMC modeling and mathematical analysis, as well as simulation based validation.

The rest of this paper is organized as follows. In Section II, the network scenario and assumptions are introduced. The proposed DSL strategies and two channel occupancy based dynamic leasing algorithms for the proposed DSL strategies are explained in Section III. The CTMC analysis is presented in Section IV. Furthermore, in Section V, we deduce an expression for the net utility of the SN in terms of defined utility and cost functions. In Section VI, we illustrate and discuss the obtained numerical results and finally the paper is concluded in Section VII.

II. System Overview

Consider a centralized CRN architecture in which the PN that owns the spectrum property rights is willing to lease a certain portion of its licensed spectrum to the

A

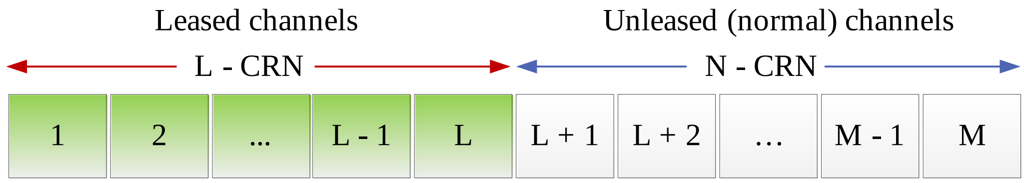


Figure A.2: Leased (L-CRN) and unleased (N-CRN) channel assignment in a CRN. The CR spectrum is divided into M channels and $L < M$ channels are leased to the SN.

SN. Fig. A.1 gives a pictorial representation of the PN and the SN arrangement. In this architecture, the decision for spectrum access is made by a central controller, e.g., a base station controller. For this purpose, the central controller needs to collect information about the spectrum usage of the licensed users as well as information about the transmission requirements of the unlicensed users. Based on this information and the results obtained from the DSL algorithms, a decision on dynamic channel access is made by the central controller. According to the given decision, unlicensed users may access the network or adjust their access status in the normal or leased portion of the CRN.

B. Network Scenario

A centralized CR architecture which allocates spectrum resources for both networks according to the assigned DSL strategy is studied. The licensed spectrum band in the CRN is divided into $M \in \mathbb{Z}^+$ frequency channels where $\mathbb{Z}^+$ is the set of positive integers. Out of these M channels, the PN leases $L \in \mathbb{Z}^+$ channels to the SN, as illustrated in Fig. A.2. Accordingly, the CRN which operates on the leased spectrum band is denoted as L-CRN and the rest, i.e., the unleased (normal) spectrum band is denoted as N-CRN. Here $L \leq L_{max} < M$ where L_{max} is the maximum number of channels that can be allocated to the L-CRN and $L_{max} \in \mathbb{Z}^+$. Since signaling will always consume certain amount of spectrum regardless of the user traffic load level in the CRN, keeping $L_{max} < M$, instead of allowing $L_{max} \leq M$, is a reasonable configuration. With a possibly very small amount of spectrum, the PN will maintain its access privilege as the owner of the spectrum. Furthermore, the configuration for L_{max} may be based on historical statistics regarding the traffic load injected into a network, the location of the network, as well as the time of the day. The number of leased channels, i.e., L, is adjusted dynamically depending on the instantaneous channel occupancy of the PN or the SN according to the algorithms proposed in this study.

C. Assumptions

In our analysis, two types of SU flows are considered, i.e., *low priority traffic* or elastic traffic such as file transfer and web browsing, and *high priority traffic* or real-time traffic such as video conference and voice over IP. However, for the sake of analysis simplicity, PU traffic is not further classified. Two abbreviations, *ESU* and *RSU*, are used to represent elastic and real-time traffic respectively whereas the term *SU* represents both SU services. The RSU services receive higher priority than the ESU services when there is an opportunity for upgrading spectrum occupancy. Moreover, if it is necessary to forcibly terminate an ongoing SU service upon a PU arrival, the ongoing ESU services are subject to termination prior to the RSU services. No matter it is a PU or SU service, one service occupies only one channel and this applies to both SU services. Moreover, the following assumptions are made as the basis for developing the analytical model presented in Section IV.

- The arrivals of both PU and SU services are Poisson processes with arrival rates λ_P, λ_{SE} and λ_{SR} for PU, ESU and RSU services respectively.
- The channels in the CRN could be either homogeneous or heterogeneous and the service times for PU, ESU and RSU services are exponentially distributed, with corresponding service rates per channel μ_P, μ_{SE} and μ_{SR} respectively. For homogeneous channels, the service rate per channel for each service type is the same across all channels. For heterogeneous channels, the service rate in the N-CRN is different from the one in the L-CRN.
- The sensing and spectrum adaptation latency is negligible in comparison with the duration between two consecutive service events. These service events indicate PU and SU arrivals or departures.

III. The Proposed DSL Strategies

In this section, we present the proposed spectrum leasing strategies[1] for multi-channel CRNs. These two strategies are different from each other when assigning the access priority and the opportunistic access for SU and PU services in the N-CRN and the L-CRN. In the *dynamic POL* (POL_D) strategy, PUs have the priority over the N-CRN channels and SUs have the priority over the L-CRN. However PUs can opportunistically access the L-CRN and SUs can also access the N-CRN in an opportunistic manner. In the *dynamic SEL* (SEL_D) strategy, the L-CRN is dedicated to SUs while the channel access in the N-CRN is similar to POL_D.

The selection of the most profitable strategy depends on the demand, QoS requirements of the PN and the SN as well as the leasing cost. For instance, when most of the SUs demand higher capacity and very low blocking probability, SEL_D

[1] The static versions of these strategies are defined in [27] as primary user opportunistically access L-CRN (POL) and secondary user exclusively access L-CRN (SEL) respectively.

Table A.1: Channel accessibility of PN and SN over N-CRN and L-CRN

Access Strategy	N-CRN		L-CRN	
	PU Access	SU Access	PU Access	SU Access
POL_D	Prioritized	Opportunistic	Opportunistic	Prioritized
SEL_D	Prioritized	Opportunistic	Denied	Dedicated

can be employed. On the other hand, if PUs demand more access in a busy network, POL_D is a better choice since in this case PUs can also access the leased spectrum opportunistically. Moreover, which strategy to select depends also on factors like geographical location, time, traffic intensity, etc. For instance, in a densely populated urban area where the PUs are more active, SEL_D is not an appropriate choice during the day time since it denies PUs access to the leased spectrum. Instead, POL_D provides better performance in terms of capacity and blocking probability for the PN by allowing PUs opportunistically access the channels in the leased spectrum. More details about these strategies are given below while their main features are summarized in Table A.1.

A. Basic Step: Dynamic Configuration for L

The proposed DSL strategies, i.e., both POL_D and SEL_D require a method to dynamically adjust the value for the number of channels in the L-CRN, i.e., L, during the system operation. In this paper, we propose two algorithms for this purpose. Once the channel access strategy is adopted, the system needs to select *one of those algorithms* to dynamically adjust the current value for L. This means that each proposed channel access strategy needs to run one of the proposed algorithms as an auxiliary step of the process by considering either PU or SU channel occupancy respectively. The main difference between those algorithms is that Algorithm A.1 calculates L based on the number of ongoing PUs in the whole CRN whereas Algorithm A.2 determines L based on the number of ongoing SUs in the whole CRN. Hereafter, the process presented in these algorithms is referred to as the *basic step*[2] for dynamic configuration of L.

The adjustment of L is not performed on a periodic basis or at fixed times of a day. Instead, the these algorithms are run only when there is a need for adjusting L, i.e., upon service arrivals or completions. If Algorithm A.1 is adopted as the basic step, it is run prior to channel allocation upon the arrival of a PU or an SU and after the departure of a PU. Since the number of ongoing PU services in the CRN will

[2]The *basic step* will add certain amount of computational cost when updating L and its effect on protocol efficiency can be studied through a complexity analysis. This detail is however beyond the scope of this paper.

not change upon an SU departure, it is not required to run Algorithm A.1 after an SU departure. Similarly, if Algorithm A.2 is adopted, it is run before any channel allocation upon the arrival of a PU or an SU and after the departure of an SU. More details about these algorithms are explained in the next subsection.

Algorithm A.1: SU occupancy independent dynamic spectrum leasing channel adjustment algorithm.

Input: Ψ : Number of channels occupied by PUs in the whole CRN including both L-CRN and N-CRN
Input: ζ : Number of channels occupied by both PUs and SUs in the N-CRN
Input: L_{max} : The upper bound of the number of leased channels
Input: $a, b : 0 < a < b < 1$
Input: $\eta_l, \eta_m : 0 < \eta_l < \eta_m < 1$
Output: The allowable number of leased channels that will be allocated to the L-CRN, i.e., L

if $\Psi \geq \eta_m$ **then**
  $L' = \lfloor aL_{max} \rfloor$: $L' \in \mathbb{Z}^+$ *is a local variable which is created to store the selected value of* L *where* $0 \leq L' \leq L_{max}$
  if $(L' + \zeta) > M$ **then**
    $L = M - \zeta$
  else
    $L = L'$
  end
else if $(\eta_l \leq \Psi < \eta_m)$ **then**
  $L' = \lfloor bL_{max} \rfloor$;
  if $(L' + \zeta) > M$ **then**
    $L = M - \zeta$
  else
    $L = L'$
  end
else
  $L' = L_{max}$;
  if $(L' + \zeta) > M$ **then**
    $L = M - \zeta$
  else
    $L = L'$
  end
end

B. Dynamics of Leased Spectrum Allocation

In both algorithms, the PU or SU ongoing traffic load is categorized by three levels as high, medium or low traffic load. To distinguish these levels, two configurable thresholds, η_m and η_l, are defined such that $0 < \eta_l < \eta_m < 1$. As illustrated in Algorithm A.1, the network is said to be in *high*, *medium* or *low* traffic mode respectively, representing the network load when PUs occupy more than η_m, between η_l and η_m, or lower than η_l of the whole CRN spectrum including both the L-CRN and the N-CRN. Similarly the same classification is done in Algorithm A.2, however, based on the SU's channel occupancy.

Algorithm A.2: SU occupancy aware dynamic spectrum leasing channel adjustment algorithm.

Input: Φ : Number of channels occupied by SUs in the whole CRN including both L-CRN and N-CRN

Input: ζ : Number of channels occupied by both PUs and SUs in the N-CRN

Input: L_{max} : The upper bound of the number of leased channels

Input: $a, b : 0 < a < b < 1$

Input: $\eta_l, \eta_m : 0 < \eta_l < \eta_m < 1$

Output: The allowable number of leased channels that will be allocated to the L-CRN, i.e., L

if $\Phi \geq \eta_m$ **then**
    $L' = L_{max}$: $L' \in \mathbb{Z}^+$ *is a local variable which is created to store the selected value of* L *where* $0 \leq L' \leq L_{max}$
    if $(L' + \zeta) > M$ **then**
        $L = M - \zeta$
    else
        $L = L'$
    end
else if $(\eta_l \leq \Phi < \eta_m)$ **then**
    $L' = \lfloor bL_{max} \rfloor$;
    if $(L' + \zeta) > M$ **then**
        $L = M - \zeta$
    else
        $L = L'$
    end
else
    $L' = \lfloor aL_{max} \rfloor$;
    if $(L' + \zeta) > M$ **then**
        $L = M - \zeta$
    else
        $L = L'$
    end
end

In these algorithms, a and b where $(0 < a < b < 1)$ are two configurable parameters, showing how much percent of L_{max} the PN is willing to lease to the SN for a given traffic load condition. When the network is in the high traffic mode for

PUs, Algorithm A.1 allocates fewer number of channels to the L-CRN in order to reserve more channels for the PN. If there are not many PUs, this algorithm allocates a larger number of channels to the L-CRN. In contrary, the number of channels allocated to the L-CRN is higher when there are more SU services if Algorithm A.2 is adopted. However, the maximum number of channels that can be allocated to the L-CRN in both approaches is bounded by L_{max}. Moreover, the selection of a proper value for L_{max} is dependent on various factors such as the QoS requirements of the PN and the SN. Note also that it is not necessary to configure the same value for L_{max} in both algorithms.

C. Dynamic POL (POL_D)

In the following, we describe the channel access procedure adopted in POL_D with respect to the SU and PU activities.

1) PU arrival: First, the *basic step* presented above is performed to determine L. If a new PU arrives at a moment when there are idle channels in the N-CRN, the new PU can start transmission in an idle channel. If all channels in the N-CRN are occupied while there are commenced ESU services in the N-CRN, then one of those ESUs is forced to terminate by re-allocating the channel to the newly arrived PU. If all channels in the N-CRN are occupied by PU or RSU services upon a new PU arrival, then one of the RSU services in the N-CRN is terminated. The purpose of the above mentioned procedure is to preserve ongoing RSU services prior to the ESU services. If the N-CRN is entirely occupied by PUs, then the new PU can access one of the leased channels in the L-CRN if it is available. Otherwise, the new PU request is blocked.

2) SU arrival: First, the *basic step* is performed to determine L. When a new SU service (either ESU or RSU) arrives while there are idle channels in the L-CRN, the new service is commenced in one of those idle channels. If all channels in the L-CRN are occupied, then the newly arrived SU will opportunistically access one of the idle channels in the N-CRN, if it is available. Otherwise the new request is blocked. Unlike in the static POL strategy in [27], PUs in the L-CRN are not forced to terminate upon an SU arrival in POL_D.

3) PU departure: If a PU service in the N-CRN departs when there are ongoing PU services in the L-CRN, one of them performs spectrum handover to the newly vacant channel in the N-CRN. A departure event of a PU service in the L-CRN merely creates a vacant channel. However, after a PU departure the *basic step* is run and L is dynamically configured to a new value if Algorithm A.1 is utilized.

4) SU departure: Following the departure of an SU service in the L-CRN, one of the SU services in the N-CRN performs handover to the vacant channel. This

step is performed for the purpose of reducing forced terminations of SU services. However, the first priority for handover is given to one of the ongoing RSU services. If there are no RSU services in the N-CRN, then one of the ESU services in the N-CRN performs handover to the L-CRN. If the SU service departure is from the N-CRN, the corresponding channel becomes vacant without any spectrum adaptation. Lastly, L is adjusted if Algorithm A.2 is used as the *basic step*.

D. Dynamic SEL (SEL_D)

Unlike in the POL_D strategy, PUs are not allowed to access channels in the leased spectrum according to SEL_D. Therefore, upon a new PU arrival when all channels in the N-CRN are occupied by other PU services, the new PU is blocked even though there is a channel available in the L-CRN. Other than that, the rest of the event handling process is the same as in the POL_D strategy.

IV. CTMC Modeling and Performance Analysis

In this section, we develop CTMC models to evaluate the performance of the proposed DSL strategies.

A. CTMC Models of the Proposed DSL Strategies

Let $\mathcal{S}$ be the set of feasible states and $\boldsymbol{x}$ represent a state of the system in the CTMC models. The total numbers of the occupied channels in the N-CRN and the L-CRN for a given state $\boldsymbol{x}$ are denoted by $b_n(\boldsymbol{x})$ and $b_l(\boldsymbol{x})$ respectively. Table A.2 lists the state space for the proposed strategies as well as for the system without leasing. The states of the CTMC model corresponding to POL_D can be represented by $\boldsymbol{x} = (i_n, j_n^E, j_n^R, i_l, j_l^E, j_l^R)$. A state in this model has six tuples (elements), representing the number of PUs, ESUs and RSUs in the N-CRN and the L-CRN respectively. More specifically, i_n, j_n^E and j_n^R denote the number of PU, ESU and RSU services in the N-CRN respectively. Similarly, i_l, j_l^E and j_l^R denote the number of these three types of services in the L-CRN respectively. In the model developed for the SEL_D strategy, there are only five tuples since the fourth term, i_l, is always null in this case. The other five tuples are the same as in POL_D. The state transition tables associated with different events are summarized with different conditions in Table A.3,Table A.4,Table A.5 and Table A.6 for these two strategies respectively. Note that the number of states in a CTMC may grow dramatically when for instance more complicated heterogeneous channels than the one reported in Subsection VI-F are considered. We focus however in this study on a manageable state space.

Denote $\pi(\boldsymbol{x})$ as the steady state probability of being in state $\boldsymbol{x}$. Given $\mathcal{S}$ and transition rate matrix $\boldsymbol{Q}$, the global balance equation and the normalization equation

Table A.2: CTMC analysis for the proposed access strategies

Parameter	POL_D	SEL_D	Without leasing
$\boldsymbol{x}$	$(i_n, j_n^E, j_n^R, i_l, j_l^E, j_l^R)$	$(i_n, j_n^E, j_n^R, j_l^E, j_l^R)$	(i_n, j_n^E, j_n^R)
$b_n(\boldsymbol{x})$	$i_n + j_n^E + j_n^R$	$i_n + j_n^E + j_n^R$	$i_n + j_n^E + j_n^R$
$b_l(\boldsymbol{x})$	$i_l + j_l^E + j_l^R$	$j_l^E + j_l^R$	–
$\mathcal{S}$	$\{\boldsymbol{x} \vert i_n, j_n^E, j_n^R, i_l, j_l^E, j_l^R \geq 0;$ $b_n(\boldsymbol{x}) \leq M - aL_{max}; b_l(\boldsymbol{x}) \leq L_{max};$ $b_n(\boldsymbol{x}) + b_l(\boldsymbol{x}) \leq M\}$	$\{\boldsymbol{x} \vert i_n, j_n^E, j_n^R, j_l^E, j_l^R \geq 0;$ $b_n(\boldsymbol{x}) \leq M - aL_{max}; b_l(\boldsymbol{x}) \leq L_{max};$ $b_n(\boldsymbol{x}) + b_l(\boldsymbol{x}) \leq M\}$	$\{\boldsymbol{x} \vert i_n, j_n^E, j_n^R \geq 0;$ $b_n(\boldsymbol{x}) \leq M\}$

are constructed as

$$\boldsymbol{\pi Q} = \boldsymbol{0}, \ \sum_{\boldsymbol{x} \in \mathcal{S}} \pi(\boldsymbol{x}) = 1, \tag{A.1}$$

where $\boldsymbol{\pi}$ is the steady state probability vector and $\boldsymbol{0}$ is a row vector of all 0's.

B. Performance Analysis of the Dynamic POL Strategy

In what follows, we derive the expressions for capacity and blocking probability of both PU and SU services. The forced termination probability of SUs is also analyzed. It is worth mentioning that the following expressions derived to calculate performance measures are valid for both algorithms presented in Subsection III-B.

1) Capacity: In this study, *capacity* is defined as the rate of service completions, i.e., the average number of service completions per time unit. Let ρ_{PU} be the capacity of PU services. Correspondingly, we obtain

$$\rho_{PU} = \sum_{\boldsymbol{x} \in \mathcal{S}} (i_n + i_l) \mu_P \pi(\boldsymbol{x}). \tag{A.2}$$

Similarly,

$$\rho_{SE} = \sum_{\boldsymbol{x} \in \mathcal{S}} (j_n^E + j_l^E) \mu_{SE} \pi(\boldsymbol{x}) \ \text{ and } \ \rho_{SR} = \sum_{\boldsymbol{x} \in \mathcal{S}} (j_n^R + j_l^R) \mu_{SR} \pi(\boldsymbol{x})$$

where ρ_{SE} and ρ_{SR} represent the capacity of ESU and RSU services respectively.

2) Blocking Probability: The blocking probability represents the probability with which an incoming SU or PU request is not allowed to enter the network due to lack of idle channels [28]. Blocking of a new PU service occurs when all the $M - L$ channels in the N-CRN are occupied by PUs and all the L channels in the L-CRN are also busy. Let P_B^{PU} denote the blocking probability of PU services. We obtain

$$P_B^{PU} = \sum_{\substack{\boldsymbol{x} \in \mathcal{S} \\ b_n(\boldsymbol{x}) + b_l(\boldsymbol{x}) = M,\ j_n^E = j_n^R = 0}} \pi(\boldsymbol{x}). \tag{A.3}$$

Moreover, a newly arrived ESU or RSU service will be blocked if all the M channels in the CRN are occupied. Therefore, the blocking probability of SU services is given by

$$P_B^{SE} = P_B^{SR} = \sum_{\substack{\boldsymbol{x} \in \mathcal{S} \\ [b_n(\boldsymbol{x}) + b_l(\boldsymbol{x})] = M}} \pi(\boldsymbol{x}). \tag{A.4}$$

2) Probability of forced termination: The probability of forced termination represents the probability that an ongoing SU service in the network is forced to terminate before its communication is finished [2]. Note that forced termination happens only to commenced SU services. Then, the forced termination probability of ESUs, P_F^{SE}, can be expressed as the ratio between the mean forced termination rate of ESU services, P_R^{SE}, and the effective rate in which a new ESU service is assigned a channel, Λ_{SE} [2]. P_R^{SE} is given by $\lambda_P \sum_{\substack{\boldsymbol{x} \in \mathcal{S} \\ [b_n(\boldsymbol{x}) + b_l(\boldsymbol{x})] = M,\ j_n^E > 0}} \pi(\boldsymbol{x})$ and $\Lambda_{SE} = (1 - P_B^{SE})\lambda_{SE}$. Therefore we obtain,

$$P_F^{SE} = \frac{\lambda_P}{(1 - P_B^{SE})\lambda_{SE}} \sum_{\substack{\boldsymbol{x} \in \mathcal{S} \\ [b_n(\boldsymbol{x}) + b_l(\boldsymbol{x})] = M,\ j_n^E > 0}} \pi(\boldsymbol{x}). \tag{A.5}$$

By following a similar procedure, the forced termination probability of RSU services is obtained by

$$P_F^{SR} = \frac{\lambda_P}{(1 - P_B^{SR})\lambda_{SR}} \sum_{\substack{\boldsymbol{x} \in \mathcal{S} \\ [b_n(\boldsymbol{x}) + b_l(\boldsymbol{x})] = M,\ j_n^E = 0,\ j_n^R > 0}} \pi(\boldsymbol{x}). \tag{A.6}$$

C. Performance Analysis of the Dynamic SEL Strategy

As mentioned earlier, SEL_D differs from POL_D by denying PUs' accessibility to the L-CRN but the channel access rules and spectrum adaptation procedures for SUs are the same in both strategies. Thus the expressions for PU/SU capacity, SU blocking probability and SU forced termination probability are the same in both cases provided that $i_l = 0$ in SEL_D.

Due to the absence of PUs in the L-CRN, the expression for the PU blocking probability in SEL_D is different from the corresponding expression derived in POL_D. PUs in SEL_D are not allowed to commence services in the CRN when 1) all $M - L$ channels allocated as the N-CRN are occupied by PUs, or 2) the number of allocated channels to the N-CRN is fewer than $M - L$ and all of them are occupied by PUs, and at the same time all the other channels in the CRN are occupied by SU services. The occurrence of the second condition is also possible in case that the algorithms in Subsection III-B assign L channels to the L-CRN while there are more than L channels already occupied by SU services in the L-CRN. Since forced termination of SUs in the L-CRN is not allowed, a new PU request is blocked when the second condition occurs. Correspondingly we obtain,

$$P_B^{PU} = \sum_{\substack{\boldsymbol{x} \in \mathcal{S} \\ [b_n(\boldsymbol{x})=M-L,\ j_n^E=j_n^R=0] \text{ OR} \\ [b_n(\boldsymbol{x})<M-L,\ b_n(\boldsymbol{x})+b_l(\boldsymbol{x})=M,\ j_n^E=j_n^R=0]}} \pi(\boldsymbol{x}). \tag{A.7}$$

V. Utility-based Resource Allocation

It is clear that the two algorithms proposed in Subsection III-B rely on L_{max} when dynamically adjusting the value of L. As the performance of such a network depends heavily on the value of L_{max}, it is of interest to further study how L_{max} is decided. In this section, we propose a utility-based approach to find the optimal configuration for the static parameter, L_{max}, under a given set of constraints.

The optimal value for L_{max} is investigated herein by means of a utility function and a cost function. Utility functions have been widely adopted in resource allocation and power control in wireless networks [29]. The QoS objectives of a group of users can be viewed as a utility function, which represents the degree of user satisfaction in terms of a given service quality level [30]. In other words, utility functions reflect the *benefit* gained by a user through the use of system resources. On the other hand, cost functions which represent a certain payment rule for channel occupancy are also introduced in spectrum trading. In what follows we estimate the SUs' benefit in terms of SU capacity and their cost in terms of the number of leased channels and the achieved QoS.

A. SU Network's Utility Function

In order to accurately reflect the user's benefit or satisfaction level, the utility function should satisfy certain properties. It must be a monotonically increasing and non-negative function with respect to the assigned resources [31]. According to the proposed leasing strategies, the *SU capacity achieved in the L-CRN* is a non-decreasing function with an increasing L_{max} and it always has a positive value. Therefore, we select the capacity of SU services in the L-CRN, $\rho_L(L_{max})$, as the preferred characteristic when calculating the utility function. ρ_{SE} and ρ_{SR} exhibit

A

similar variation with an increasing L_{max}. Thus when analyzing the utility of SUs, we consider only the ESU services for the sake of expression clarity. The capacity of ESU services achieved in the L-CRN is obtained as $\rho_L(L_{max}) = \sum_{\boldsymbol{x}\in\mathcal{S}} j_l^E \mu_{SE}\pi(\boldsymbol{x})$.

Let $U\{\rho_L(L_{max})\}$ be the utility function of ESU services with respect to the ESU capacity[3]. The exact expression of a utility function depends on traffic characteristics, and can be obtained by analyzing the behavior of users [29]. In many studies on resource allocation in wireless networks, the *sigmoid function* has been a popular expression to represent the utility for elastic and semi-elastic traffic [30]-[33]. Therefore, we adopt the following sigmoid function as the utility of ESU services.

$$U\{\rho_L(L_{max})\} = \frac{1}{1 + B\mathrm{e}^{-C[\rho_L(L_{max})-R]}} \tag{A.8}$$

where B, C and R are three parameters which determine the slope, the lower bound and the inflection of the utility function respectively and $\{B, C, R \in \mathbb{R} | B, C, R > 0\}$ where $\mathbb{R}$ denotes the set of real numbers.

According to (A.8), the demand degree of a user is reflected by C. When C is larger, the slope of the utility curve is higher, indicating that the user demands more resources. On the contrary, the smaller C, the weaker the user's demand. By varying parameter B, the utility of different traffic types can be represented. For instance, the value of B needs to be higher in a utility function for a best effort user than that of a QoS user [34]. Thus the utility functions of heterogeneous traffic flows like voice, video and web browsing which need different QoS treatments can be represented by selecting appropriate values for C, B and R respectively. The values of these parameters used in Section VI are not computed but configured using similar approaches as explained in [31] and [34].

From (A.8), it is clear that $U\{\rho_L(L_{max})\}$ is a monotonically increasing function with respect to ρ_{SE}, i.e., users feel more satisfied when they achieve higher capacity. Furthermore, since $\lim_{\rho_L(L_{max})\to 0} U\{\rho_L(L_{max})\} = \frac{1}{1+B\mathrm{e}^{CR}} > 0$ and $\lim_{\rho_L(L_{max})\to\infty} U\{\rho_L(L_{max})\} = 1$, the utility of the ESUs can be scaled as in the range of $U\{\rho_L(L_{max})\} \in (0, 1)$.

B. SU Network's Cost Function

In the previous subsection, we presented a method to estimate the benefit or profit in which the SN could gain through DSL. To encourage spectrum leasing from the PU's perspective, there must be a revenue obtained from the lessees and it

[3] In this section, the capacity achieved in the L-CRN, blocking probability and forced termination probability of ESU services are expressed as the functions of L_{max} as $\rho_L(L_{max})$, $P_B^{SE}(L_{max})$ and $P_F^{SE}(L_{max})$, since those values are dependent on the upper bound of the leased spectrum.

should be sufficiently compensated for the amount of the leased spectrum. On the other hand, the revenue obtained by the PN becomes the expense (the leasing cost) for the SN. In this study, we develop a flexible pricing scheme for SUs by taking into consideration two factors. First, the leasing price should be increasing if higher bandwidth is utilized. If the utilized bandwidth of the leased spectrum is L, the leasing price, P, should be in the form of αL where α is a variable which can be configured by the central controller. Additionally, we introduce an exponent, β, to represent the QoS satisfaction level of the SUs which rent the spectrum. Thus the cost function is defined as $P = \alpha L^{\beta}$.

In this expression, α and β are non-negative values which in this study represent the quality of the transmission in terms of the blocking probability and the forced termination probability respectively. Depending on the values taken by α and β, the price for the leased spectrum varies. By adjusting these two parameters, the leasing cost can be set as a fair and reasonable price for the leased channels as discussed below. In brief, when SUs achieve a high QoS level, i.e., once the blocking probability and the forced termination probability become low, the price for the leased spectrum shows a higher value, and vice versa.

Moreover, let us replace L by $\bar{L}$ which represents the *average number of leased channels* that an ESU service utilizes for the duration of network operation, i.e., $\bar{L} = \sum_{\boldsymbol{x}\in\mathcal{S}} j_l^E \pi(\boldsymbol{x})$. Thus, $P = \alpha \bar{L}^{\beta}$, and in the CTMC, $\{\bar{L} \in \mathbb{R} | 0 \leq \bar{L} \leq L_{max}\}$. Note that $\bar{L}$ can be a non-integer value since it is calculated as the *average value* of the number of leased channels in the steady state and can be lower than 1.

Furthermore, in order to define a pricing function that reflects the achieved QoS level of SUs, we configure the scaling parameters α and β as follows, where ω is a scaler introduced to normalize the cost function.

$$\alpha(L_{max}) = \omega\Bigg(1 - P_B^{SE}(L_{max})\Bigg) \quad \text{and} \tag{A.9}$$

$$\beta(L_{max}) = \begin{cases} 1 + P_F^{SE}(L_{max}), & \text{if } \bar{L} < 1 \\ 2 - P_F^{SE}(L_{max}), & \text{if } \bar{L} \geq 1. \end{cases} \tag{A.10}$$

In (A.9), $\alpha(L_{max})$ is defined in such a way that when the blocking probability increases the cost will decrease. Thus the form of $1 - P_B^{SE}(L_{max})$ is selected. When selecting a proper expression for the exponential scale, β, however, a careful consideration is required since the increase/decrease of $\bar{L}^{\beta}$ depends also on $\bar{L}$, considering that $\bar{L}$ may be greater or smaller than 1. When the forced termination probability becomes higher, the leasing charge must be reduced. To satisfy such a condition when $\bar{L} < 1$, the exponential coefficient should increase with an increasing rate of forced termination. Therefore, we define correspondingly, $\beta(L_{max}) = 1 + P_F^{SE}(L_{max})$ if $\bar{L} < 1$. When $\bar{L} \geq 1$, we need a decreasing function for $\beta(L_{max})$ with an increasing rate of forced terminations. Therefore, $\beta(L_{max}) = 2 - P_F^{SE}(L_{max})$ is defined when $\bar{L} \geq 1$. Since $0 \leq P_F^{SE}(L_{max}) \leq 1$, we can show that the exponential scale of the cost function satisfies $1 \leq \beta(L_{max}) \leq 2$. Moreover, $\beta(L_{max}) \geq 1$ indicates that the pricing function is convex [35].

Similarly, the linear component of the cost function is scaled as $0 \leq \alpha(L_{max}) \leq \omega$ since $0 \leq P_B^{SE}(L_{max}) \leq 1$. When deciding the final value of the cost function, its exponential component will affect more than the linear component. The reason to select the effect of the forced termination probability to the exponential component is due to the fact that the forced terminations of ongoing services is more annoying than blocking service requests from QoS provisioning's point of view [36]. Based on the above discussions, we define the cost function for ESU traffic as follows.

$$C(L_{max}) = \omega\left(1 - P_B^{SE}(L_{max})\right)\left(\sum_{\boldsymbol{x}\in\mathcal{S}} j_l^E \pi(\boldsymbol{x})\right)^{\beta(L_{max})}. \tag{A.11}$$

Here, the term $\sum_{\boldsymbol{x}\in\mathcal{S}} j_l^E \pi(\boldsymbol{x})$ represents the average number of channels utilized by the ESU services in the L-CRN[4]. Since the number of channels allocated to the L-CRN never exceeds L_{max}, $0 \leq \sum_{\boldsymbol{x}\in\mathcal{S}} j_l^E \pi(\boldsymbol{x}) \leq L_{max}$. Correspondingly, we can conclude that $C(L_{max}) \in (0, \omega(L_{max})^2)$.

C. The Net Utility Function

Lastly, we derive an expression for calculating the *net utility* [37] or the net benefit which is defined as the difference between the utility derived from the achieved capacity and the cost for spectrum leasing. Correspondingly, the net utility of ESU services is defined as $NU(L_{max}) = U\{\rho_L(L_{max})\} - C(L_{max})$ and it is given by (A.12). Following the feasible ranges obtained for the utility and cost functions, the net utility of ESUs services will be scaled as $-\omega(L_{max})^2 \leq NU(L_{max}) \leq 1$. When $NU(L_{max})$ is positive, it means that SN's benefit is larger than the cost. In (A.13) we deduce a condition in order to have a positive net utility by specifying $U\{\rho_L(L_{max})\} > C(L_{max})$. Furthermore, in Subsection VI-E, a utility-based approach is analyzed to obtain an optimum value for L_{max} subject to given conditions.

[4] It is worth mentioning that when computing the leasing cost, the term, $\sum_{\boldsymbol{x}\in\mathcal{S}} j_l^E \pi(\boldsymbol{x})$, considers only the L-CRN, since SUs do not need to pay for the unleased channels although they are allowed to opportunistically access the N-CRN.

$$NU(L_{max}) = \frac{1}{1 + B\mathrm{e}^{-C[\rho_L(L_{max})-R]}} - \omega\left(1 - P_B^{SE}(L_{max})\right)\left(\sum_{\boldsymbol{x}\in\mathcal{S}} j_l^E \pi(\boldsymbol{x})\right)^{\beta(L_{max})}. \tag{A.12}$$

$$\omega < \left[(1 + B\mathrm{e}^{-C[\rho_L(L_{max})-R]})\left(1 - P_B^{SE}(L_{max})\right)\left(\sum_{\boldsymbol{x}\in\mathcal{S}} j_l^E \pi(\boldsymbol{x})\right)^{\beta(L_{max})}\right]^{-1}. \tag{A.13}$$

VI. Numerical Results and Discussions

In this section, we provide numerical results to evaluate the performance of the proposed DSL strategies. The default configurations of the parameters are $M = 6$, $\lambda_P = \lambda_{SE} = \lambda_{SR} = 2$ and $\mu_P = \mu_{SE} = \mu_{SR} = 1$ respectively[5]. The units of these parameters could be services or flows per unit of time. When calculating L, the parameters are configured as $a = 0.4$ and $b = 0.7$ in Algorithm A.1 and Algorithm A.2. Since only three traffic load levels are considered, we configure $\eta_m = 70\%$ and $\eta_l = 35\%$ as an approximation for traffic load classification unless otherwise stated. The curves representing *Without Leasing* in this section are obtained by setting $L_{max} = 0$. Note that the number of channels allocated to the L-CRN in the static leasing strategies is denoted as L in the presented numerical results.

In the following subsections, we present the numerical results obtained from both the CTMC models and the discrete-event based simulations. For CTMC models, we construct the feasible states as well as the state transitions according to the procedures and tables presented in Section IV. For simulations, we made a custom-built MATLAB simulator which mimics the behavior of PU and ESU/RSU activities. The simulations were performed in a similar way as presented in [2]. It is worth mentioning that the calculation of the performance metrics in our simulations is not dependent on the derived mathematical expressions. In the simulator, the PUs/SUs arrive and depart according to a given distribution and the access principle adopted, and no transitions or state probabilities are needed when calculating the values of these performance metrics. To verify the preciseness of our model, we have also performed simulations based on the log-normal service time distribution.

[5]Note that the expressions for analyzing the performance measures given in Section IV and Section V are derived as general expressions and therefore they are applicable to any scenario within the proposed leasing paradigms given that $M > 1$ and $L < M$.

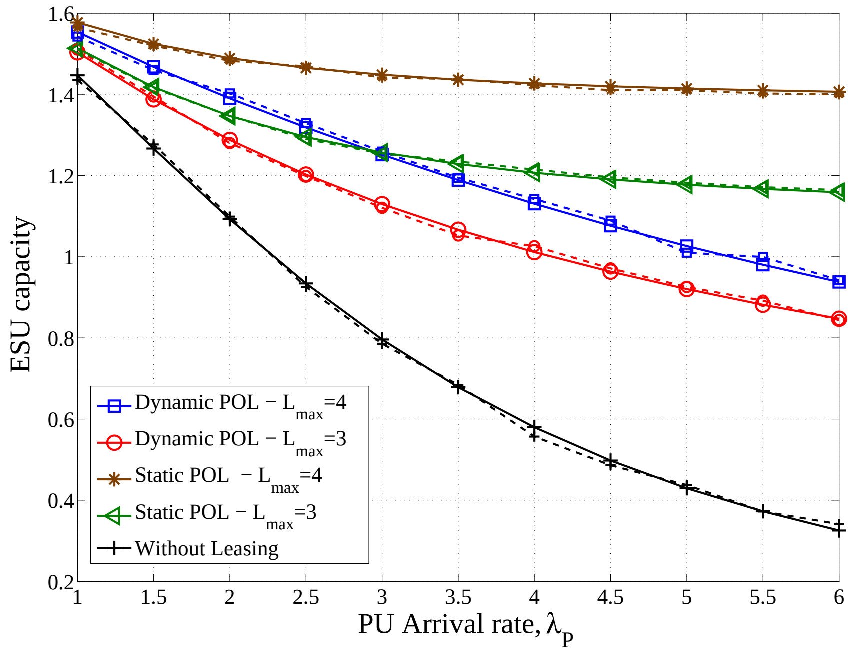


Figure A.3: ESU capacity when Algorithm A.1 is adopted. The dashed lines show the simulation results corresponding to the exponential service time distribution.

A. Network Capacity

Fig. A.3 and Fig. A.4 demonstrate the ESU capacity and the PU capacity respectively for different values of L_{max} corresponding to the POL_D strategy and its static version. These plots illustrate that both static and dynamic leasing strategies enhance the ESU capacity, however, with a loss of the PU capacity. In our previous work [27] where the static versions of the proposed DSL strategies were introduced, the PU capacity loss is a major negative effect for spectrum leasing. Fig. A.4 illustrates that PU capacity loss can be considerably reduced by employing a dynamic leasing strategy. This is because in DSL, the number of channels leased to SUs in the L-CRN is gradually reduced when there is an increasing demand for channel access from PUs. For instance, in Algorithm A.1, only 40% of L_{max}, is allocated to the L-CRN if $\eta_m \geq 70\%$ of the total bandwidth is occupied by PUs. The upper bound, i.e., L_{max}, is reached only when the PU occupancy level drops below $\eta_l = 35\%$.

The capacity improvement of the SN when employing a DSL scheme is more significant at a higher λ_P. When λ_P is low, a comparatively large number of SU services can be completed since more channels are available for SUs. However at a higher λ_P, the blocking probability of SUs becomes higher (see Fig. A.6) and consequently the SU capacity decreases dramatically (see Fig. A.3) unless a leasing

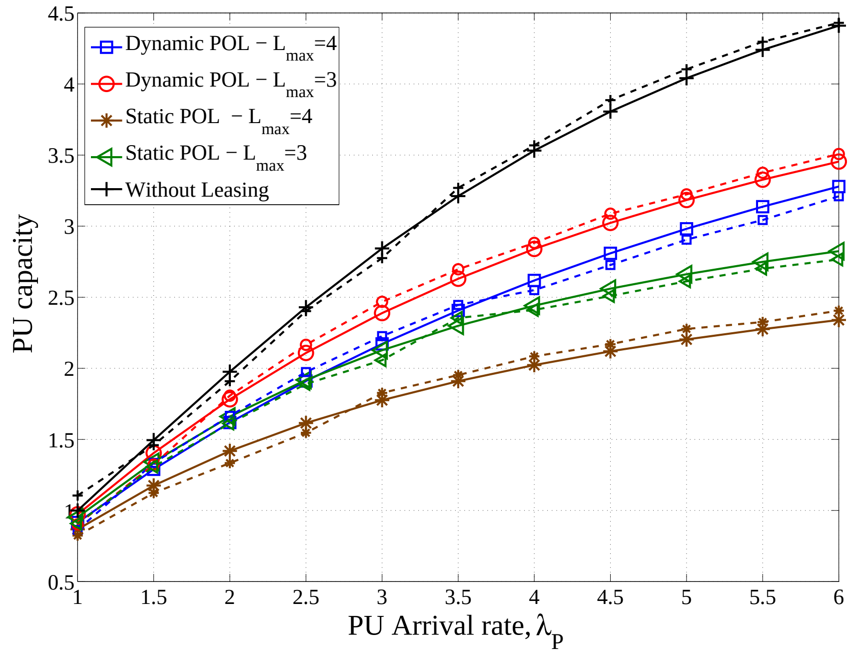


Figure A.4: PU capacity when Algorithm A.1 is adopted. The dashed lines show the simulation results obtained by adopting the log-normal service time distribution.

strategy is employed. Moreover, a larger L_{max} contributes to a further increase of the ESU capacity.

We also demonstrate the performance of the proposed two algorithms for DSL in Fig. A.5. In both POL_D and SEL_D, the PU capacity becomes higher when Algorithm A.2 is adopted as the *basic step* due to its SU occupancy awareness. In Algorithm A.1, a large number of channels can be allocated for the L-CRN even though there is light SU traffic in the system. In other words, to assign more leased channels, this algorithm does not care about the existence of SU traffic in the system. On the other hand, in Algorithm A.2, a higher number of channels will be allocated to the L-CRN only when the existing SU traffic is higher than a certain threshold. Thus, PUs sustain more channel access opportunities and therefore more PU capacity improvement is achieved in Algorithm A.2 than in Algorithm A.1. Following the same logic, SUs achieve higher capacity if Algorithm A.1 is adopted.

To validate the correctness of the mathematical analysis, we performed discrete-event based simulations. The obtained ESU capacity shown in Fig. A.3 confirms that the simulation results coincide with the analytical results. For preciseness verification, the PU capacity which is obtained based on the log-normal service time distribution (with the same mean value and the variance as used in the original exponential distribution) is illustrated in Fig. A.4. Again, the curves obtained from

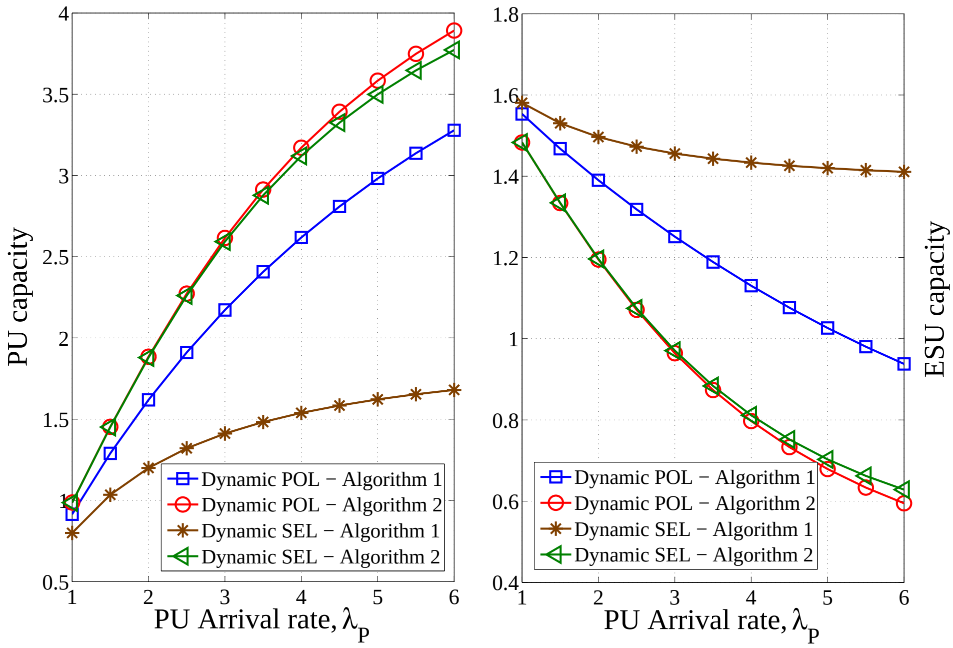


Figure A.5: PU and ESU capacity when $L_{max} = 4$.

analysis and simulations match very well with each other. This fact demonstrates that the developed CTMC model is robust as long as the Poisson arrival process assumption holds.

Comparing the PU and ESU capacity in Fig. A.5, SUs achieve increased benefits by SEL_D compared with those in the case of POL_D. Since SEL_D does not allow PU access in the L-CRN, the SU capacity in SEL_D is always higher than the one they could achieve in POL_D. On the other hand, POL_D shows the best performance in terms of the PU capacity. Once again, those observations are more evident when the PU arrival rate is high. From Fig. A.5 we can conclude that the selection of the proper leasing strategy needs to be based on the requirements for certain QoS measures of both networks.

B. Blocking Probability

Fig. A.6 and Fig. A.7 illustrate the blocking probability of PU and SU services under different configurations. For this CRN, the blocking probability of SU services can be reduced approximately by 15% and 6% when Algorithm A.1 and Algorithm A.2 are employed respectively, in contrast with the network *without leasing*. Therefore, Algorithm A.1 enabled DSL provides higher performance to the SN.

Furthermore, it is observed in Fig. A.6 that the static SEL strategy reduces the SU blocking probability by 30%, however, at a cost of much higher blocking probability for PUs. This is because that the reserved number of channels for PUs is

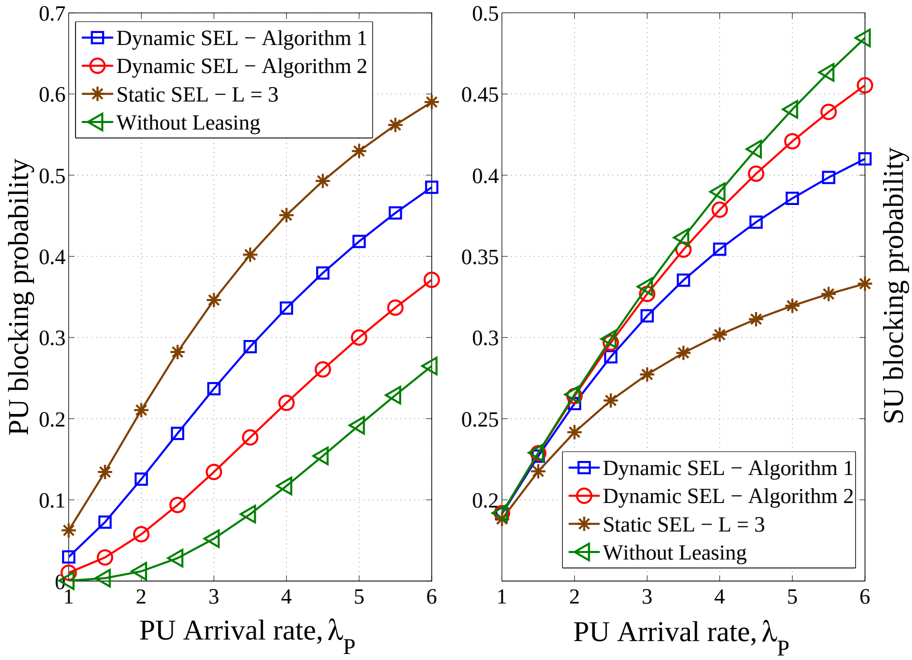
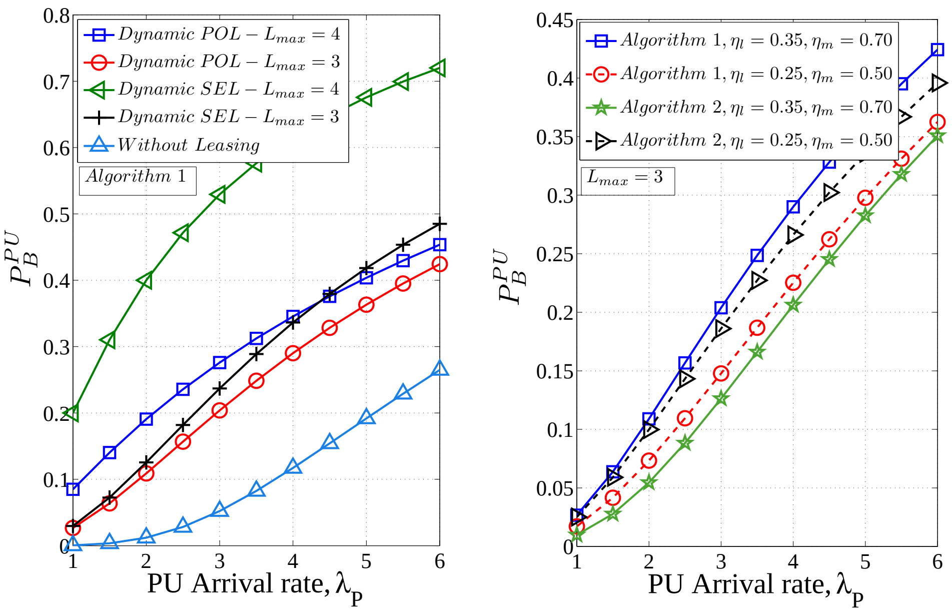


Figure A.6: Blocking probability in the SEL_D when $L_{max} = 3$.

too low in this case, as $M = 6$ and $L = 3$. On the other hand, when the proposed DSL is employed, the increase of P_B^{PU} due to spectrum leasing can be significantly reduced. For instance, when λ_P is very high, P_B^{PU} obtained by SEL_D is about 37% lower than that of static SEL when Algorithm A.2 is adopted. Furthermore, Fig. A.7(a) demonstrates that the increase of P_B^{PU} will be even slower once POL_D is employed instead of SEL_D. This is because that PUs can opportunistically access the L-CRN. Moreover, a low L_{max} is recommended, when there is a stringent QoS requirement for keeping P_B^{PU} low. From these results, it is clear that there is a tradeoff between the PU and the SU blocking probabilities. An appropriate combination of the three components, i.e., the DSL strategy, the DSL algorithm and L_{max} could satisfy both SU and PU QoS requirements on blocking probability.

Since the values of η_l and η_m are configurable, setting these two thresholds to different values may have impact on the performance of PU and SU services. In Fig. A.7(b), the blocking probability of PU services under POL_D is observed under two different sets of threshold values, as $\eta_l = 0.35$, $\eta_m = 0.70$ and $\eta_l = 0.25$, $\eta_m = 0.50$ respectively. With a lower η_m, the CRN reaches the high traffic mode faster, then fewer channels are allocated to the L-CRN for SUs when Algorithm 1 is employed. Consequently, the blocking probability of PUs becomes lower with a lower η_m. On the other hand, an opposite result is observed when Algorithm 2 is applied.



(a) (b)

Figure A.7: PU blocking probability as a function of λ_P.

C. Forced Termination Probability

From Fig. A.8 it can be observed that the proposed DSL strategies exhibit a significant reduction of the forced termination probability of ESU services in comparison with the legacy scheme which does not support spectrum leasing. As shown in Fig. A.8, P_F^{SE} has been reduced from 53% to about 17% and 12% at $\lambda_P = 4$ respectively when the dynamic and static leasing strategies are employed. Fig. A.9 plots the forced termination probability of both ESU and RSU services when POL_D is employed with Algorithm A.1. In both POL_D and SEL_D, the forced termination probability of RSU services can be kept at a lower level compared with ESU services since ESU services have lower priority than RSU services.

Moreover, DSL outperforms static leasing when a higher L_{max} is configured. For instance, we observe a lower forced termination probability of SUs under POL_D when the upper bound for DSL is set as $L_{max} = 4$ in comparison with the static POL with $L = 3$. This observation concludes that depending on the available resources and the required QoS level for a given network scenario, a suitable leasing scheme with appropriate parameter configurations should be selected. In the rest of this section, we consider a CRN with $M = 8$ channels and the utility and the cost function parameters are configured as $B = 3.5, C = 1.0$, $R = 0$ and $\omega = {}^5\!/_{18}$ respectively. Note that the x-axis values corresponding to the static leasing illustrated in Fig. A.10 and Fig. A.11 indicate L.

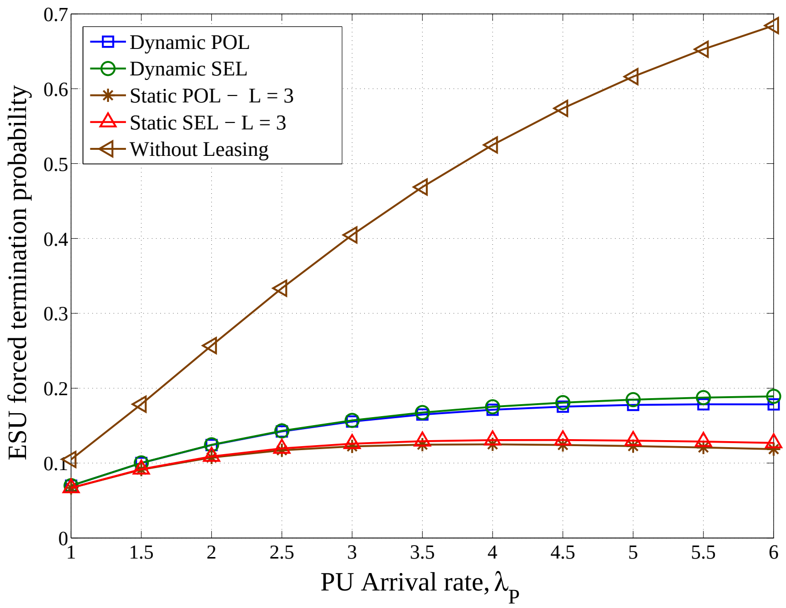


Figure A.8: Forced termination probability of ESU services when Algorithm A.1 is adopted and $L_{max} = 3$.

D. Utility and Cost Analysis

Plot first the utility function of ESUs in Fig. A.10 for two values of λ_P. By observing the left-side plot, we note that the utility of ESUs is higher when a static leasing scheme is adopted. However, according to the results shown in Fig. A.11, the leasing cost associated with the static leasing is also significantly higher than that of DSL. Since the number of leased channels is fixed and the channels are in a way dedicated to SUs in static leasing strategies, SUs will get channel access in the L-CRN even though there are more PU arrivals to the network. As a result, SUs utilize more leased channels and correspondingly the capacity of the SN and the leasing cost also increase. The tradeoff between the utility and the cost will be analyzed in terms of net utility in Subsection VI-E.

From the right-side plot in Fig. A.10, it is clear that Algorithm A.1, which does not take into account SU's channel occupancy, achieves better performance than Algorithm A.2 does, in terms of the achieved utility since Algorithm A.1 is more likely to allocate a higher number of channels to the SN. However, as depicted in both graphs in Fig. A.10, the increase of the PU arrival rate diminishes the ESU utility significantly. This is due to the fact that when there are more PU arrivals, the network with DSL strategies dynamically adjusts the number of leased channels downwards, in both algorithms.

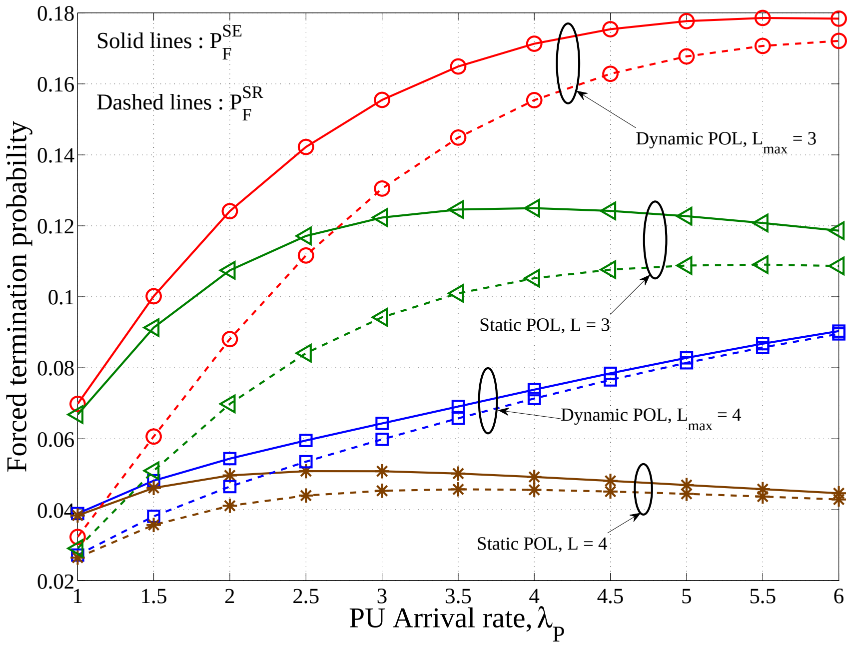


Figure A.9: Forced termination probability when Algorithm A.1 is adopted. The solid lines and the dashed lines represent the ESU and the RSU forced termination probability respectively.

Fig. A.12 illustrates both utility and cost functions of ESU services when the POL_D strategy is adopted. When L_{max} is configured to a lower value, ESUs experience higher *net utility* since the leasing cost takes a considerably low value. However, if Algorithm A.1 is employed, the net utility of the system at $\lambda_P = 2$ becomes negative when $L_{max} > 4$. Under the same conditions, ESUs can still maintain a positive net utility at $\lambda_P = 2$, if Algorithm A.2 is adopted as the *basic step*. The reason for this result is as follows. Since Algorithm A.1 allocates more channels to the L-CRN than Algorithm A.2 does, spectrum leasing with SU occupancy independent DSL leads to higher cost for SUs, resulting in a negative net utility level. Therefore it is of essential importance to analyze how to achieve a positive net utility in the network by leasing an optimal number of channels to the SUs. This point will be discussed in the next subsections.

In order to investigate the behavior of ESU's utility with respect to parameters C and R given in (A.8), we illustrate in Fig. A.13 and Fig. A.14 the utility achieved in the SEL_D strategy. As already mentioned, parameters C and R reflect the demand degree and the resource requirement of the users respectively and both have impact on the utility level. When PUs lease more channels, the utility of SUs becomes higher. Furthermore, it can be observed that at a given L_{max}, utility becomes lower

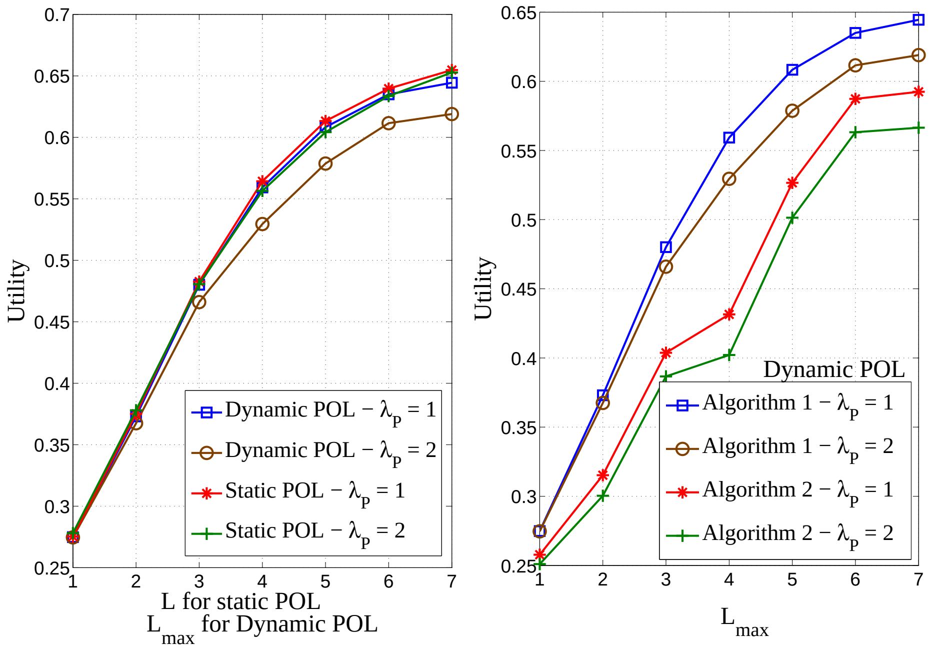


Figure A.10: Utility of ESU services at different PU arrival rates. The results shown in the left-side graph are obtained by adopting Algorithm A.1 and the results shown in the right-side graph illustrate ESUs' utility using POL_D.

as λ_{SE} is smaller, i.e., when fewer SUs arrive. When C becomes larger, the ESU utility also increases, showing that utility is an increasing function with respect to the demand degree. Moreover, once R becomes smaller, the utility will improve towards a higher value. This is due to the fact that if users' resource requirement becomes lower, the network will be able to allocate sufficiently the requested resources without blocking a request. Consequently, the satisfaction level of the users becomes higher.

E. Utility-based Spectrum Leasing

Although the allocation of the leased spectrum in the proposed strategies is regulated via the algorithms presented in Subsection III-B, the number of channels which could be allocated to the L-CRN always depends on a static parameter, L_{max}, which plays a major role for DSL configuration. In this subsection, we identify an optimal value of L_{max} for resource allocation for the proposed strategies. The objective is to determine the highest possible configuration of L_{max} for a given set of conditions. Consider the following two conditions:

$$\text{C1: } U\{\rho_L(L_{max})\} > C(L_{max}),$$

$$\text{C2: } L_{max} < \lfloor k \cdot M \rfloor, \;\; a \leq 0.5, \;\; b \leq 0.8,$$

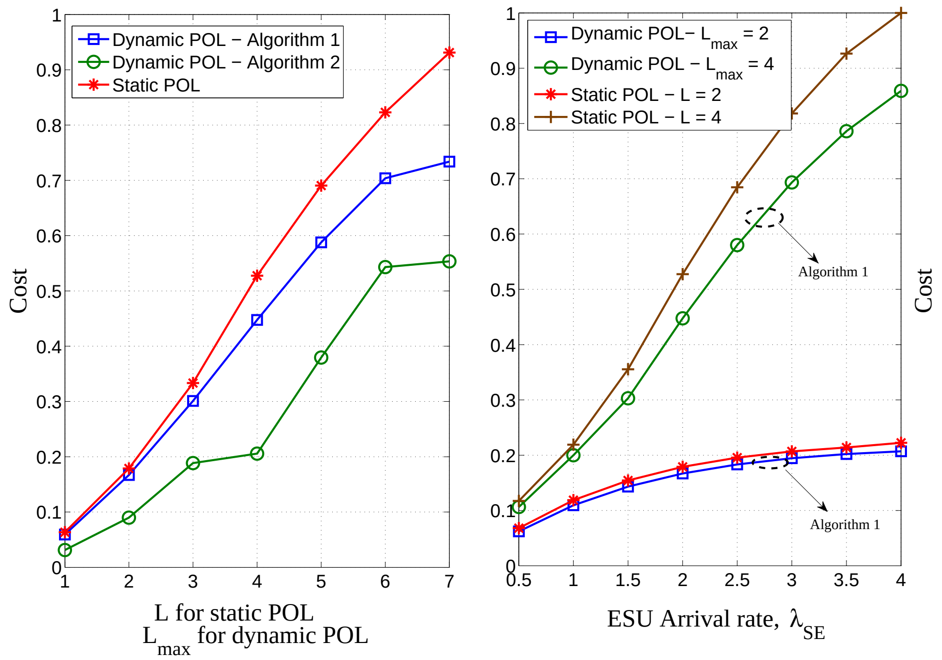


Figure A.11: The leasing cost in the primary user opportunistic leasing strategies.

where k is a scalar which constraints the amount of spectrum to be leased and $0 < k \leq 1$ [27].

Condition *C1* indicates the expectation by the SN in order to achieve a positive utility. The PN instead prefers to lease their spectrum with Condition *C2* in order to maintain a certain level of channel access priority, with two main considerations. First, the PN never leases the whole spectrum. Second, the PN does not lease more than 50% or 80% of L_{max} to the L-CRN when the sytem is in the *high* or *medium* traffic load levels respectively. Note however that the values given in Condition *C2* are configurable and that $a \leq 0.5$ and $b \leq 0.8$ are set only for illustration purpose. Moreover, when assigning values to a and b, the central controller can check the channel requirements of the PUs and SUs. For instance, when Algorithm A.1 is employed in a CRN and there are high PU activities, a and b need to be configured to lower values since PUs have higher priority in this case.

To identify the optimal L_{max} value, we evaluate the utility and cost functions of ESUs with respect to L_{max} with a variable λ_P. The following observations for the plots in Fig. A.15 are made. An increase of both utility and leasing cost is observed when the allocated resource is increased. However, beyond a certain L_{max} value, the leasing cost exceeds the achieved utility level, resulting in a negative net utility. Therefore, the optimal value for L_{max} is obtained by observing the intersection point after which the net utility becomes negative. The obtained optimal value of

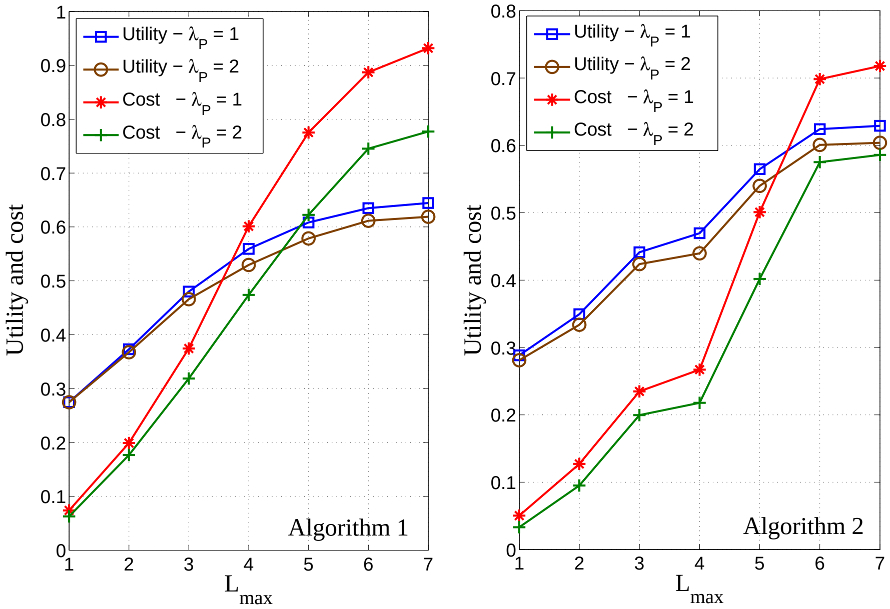


Figure A.12: Utility and cost functions of ESU services in the POL_D strategy.

L_{max} is illustrated with respect to the PU arrival rate in Fig. A.16.

Interestingly, in Fig. A.16, we observe that when the PU arrival rate increases, the system intends to configure a larger value to L_{max}. The reason can be explained considering the behavior of Algorithm A.1 and Condition *C1*. At a higher λ_P, the fraction of channel occupancy by PUs becomes larger and therefore the CRN allocates fewer number of channels to the L-CRN. For instance, once the system is at the high traffic load level, Algorithm A.1 assigns only 40% of L_{max} to the L-CRN given that $a = 0.4$. Moreover, the QoS parameters in the cost function, i.e., the blocking probability and the forced termination probability of SUs, become larger at a higher λ_P. Correspondingly, the leasing cost will be reduced as discussed in Subsection V-B. Therefore, in order to improve the utility level and to obtain a positive net utility in such cases, the system must configure a higher value for L_{max}. That is, the larger the value of λ_P, the larger the required L_{max} in order to satisfy Condition *C1*.

F. Performance of the SN as a function of QoS for the PN

As already observed, the performance of the SN can be significantly improved by employing the proposed DSL strategies. Meanwhile, the QoS of the PN needs to be maintained to a satisfactory level. To investigate this, a CRN consisting of $M = 6$ channels is considered where the PN can lease L ($L \leq L_{max} < M$) channels to the SN. Let $L_{max} \leq 4$. In this example, the optimal value for L_{max},

A

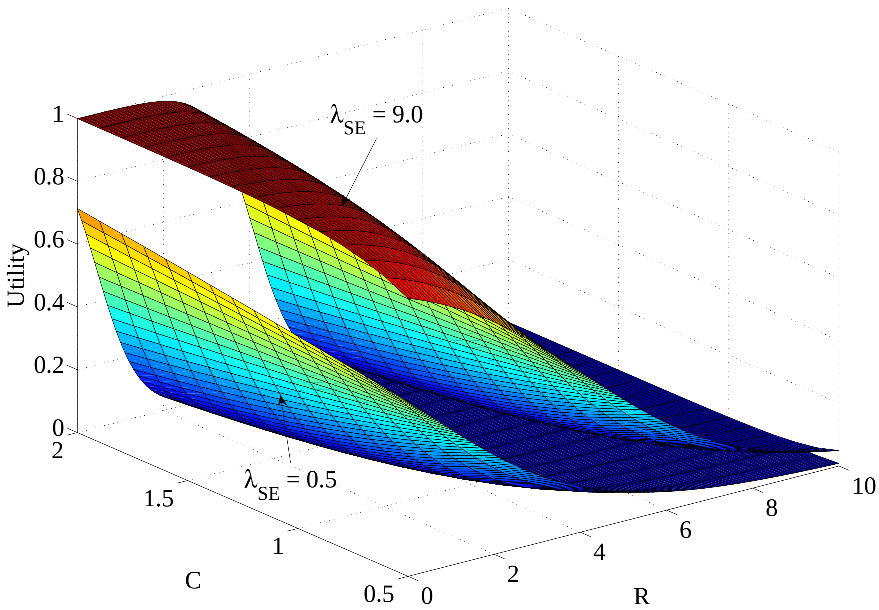


Figure A.13: Utility variation in the SEL_D strategy for different configurations of C and R when $B = 1.0$ where Algorithm A.1 is adopted.

denoted as L_{max}^{OPT} needs to be found in order to maximize ρ_{SE}, provided that the maximum tolerable PU blocking probability is bounded by $(P_B^{PU})_{max}$. Assuming $(P_B^{PU})_{max} = 0.1$, and 0.2 respectively, and observing the variation of P_B^{PU} in Fig. A.17, L_{max}^{OPT} can be determined. The results for L_{max}^{OPT} and the corresponding ρ_{SE} are illustrated in Fig. A.18(a) and Fig. A.18(b) respectively considering these two values of $(P_B^{PU})_{max}$, together with the ρ_{SE} without leasing. The curves in Fig. A.18(b) show that higher ρ_{SE} is achieved with spectrum leasing and ρ_{SE} can be further increased if the PN can tolerate a higher blocking probability.

G. DSL in CRNs with Heterogeneous Channels

Finally, we consider a heterogeneous channel scenario where two types of channels with distinct service rates are allocated to the N-CRN and the L-CRN respectively. For performance evaluation, we employ the same CTMC model with the same state definition and transitions but different service rates, as $\mu_N = 1.0$ and $\mu_L = 0.5$ for the N-CRN and the L-CRN respectively. In Fig. A.19, we illustrate the achieved capacity and blocking probability for ESU services in this heterogeneous channel environment in comparison with the results under the homogeneous channel condition where $\mu_L = \mu_N = 1.0$. Comparing these obtained analytical and simulation results, we observe the same trend for capacity and blocking probability variations of ESU services in both environments, however with different values for

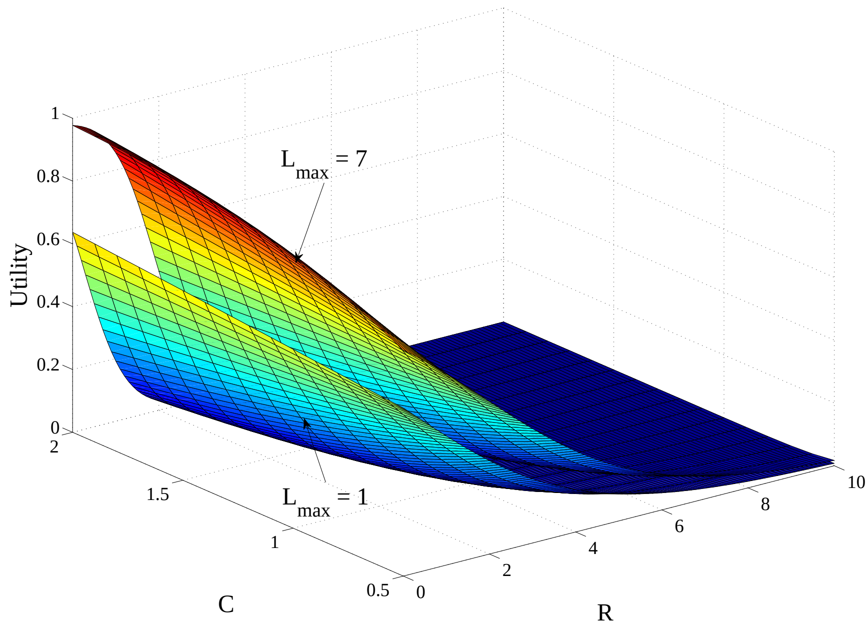


Figure A.14: Utility variation in the SEL_D strategy for different configurations of C and R when $B = 1.0$ where Algorithm A.1 is adopted.

ρ_{SE} and P_B^{SE}. The achieved capacity in the SN becomes lower in the heterogeneous case since the service rate of the leased channels is only one half of the service rate as in the N-CRN. Since ESU services in the heterogeneous case with lower service rate need longer time to complete their services, higher blocking probability is observed. Therefore, the results shown in Fig. A.19 confirm the robustness and the applicability of the developed models in other kinds of network scenarios.

VI. Conclusions

This paper proposes two dynamic spectrum leasing strategies, POL_D and SEL_D, for multi-channel CRNs. DSL provides insight on the design of efficient spectrum access with both performance enhancement and cost effective solutions. The proposed two strategies differ from each other with distinct considerations on channel accessibility of primary users into the leased spectrum. We also develop two algorithms which are associated with a DSL strategy in order to dynamically adjust the number of leased channels in the CRN. The numerical results reveal that the main disadvantages encountered in static leasing schemes such as drop of the PU capacity and higher leasing cost can be mitigated significantly through the employment of DSL. Furthermore, an appropriate utility function and a cost function are defined to analyze the tradeoff between the benefit and the cost associated with DSL. By

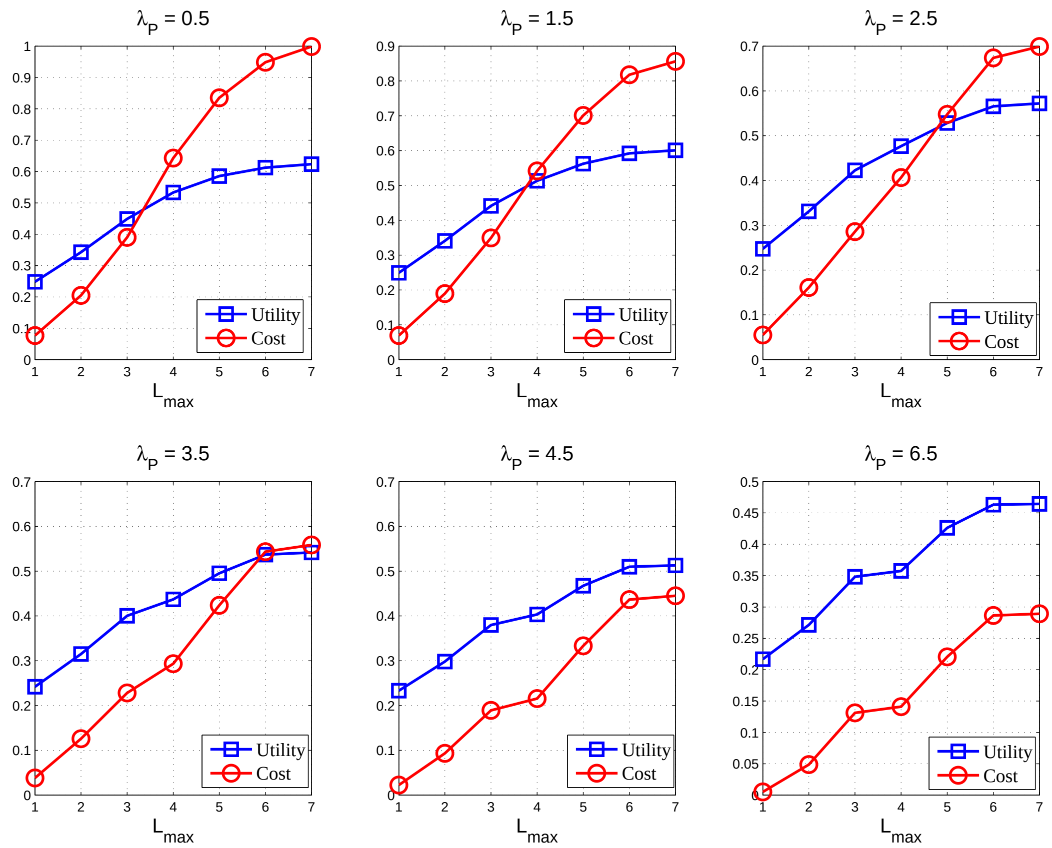


Figure A.15: Utility and cost functions of ESU services based on the POL_D strategy when Algorithm A.1 is employed with $M = 8, a = 0.4$ and $b = 0.7$.

deriving an expression for the net utility of secondary users, we obtain the optimal value for the number of leased channels based on a given set of conditions in order to achieve a positive net utility. When deciding which strategy to employ, a network operator may select SEL_D when SUs demand high capacity with low blocking probability, or POL_D when giving priority to performance stability of the primary network.

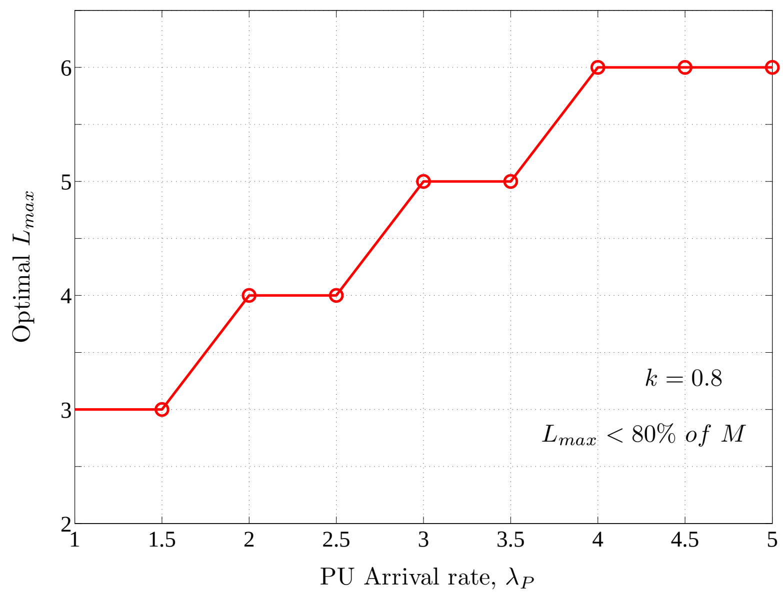


Figure A.16: The maximum L_{max} that can be set in POL_D strategy to have a positive net utility. The curve is obtained based on the results from Fig. A.15.

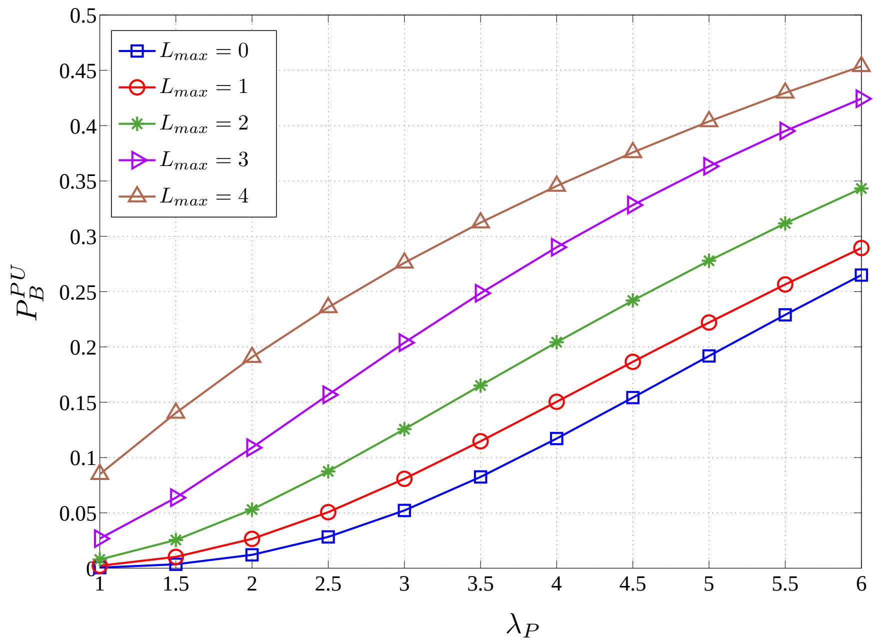


Figure A.17: Blocking probability of PU services in the POL_D strategy when Algorithm A.1 is employed.

A

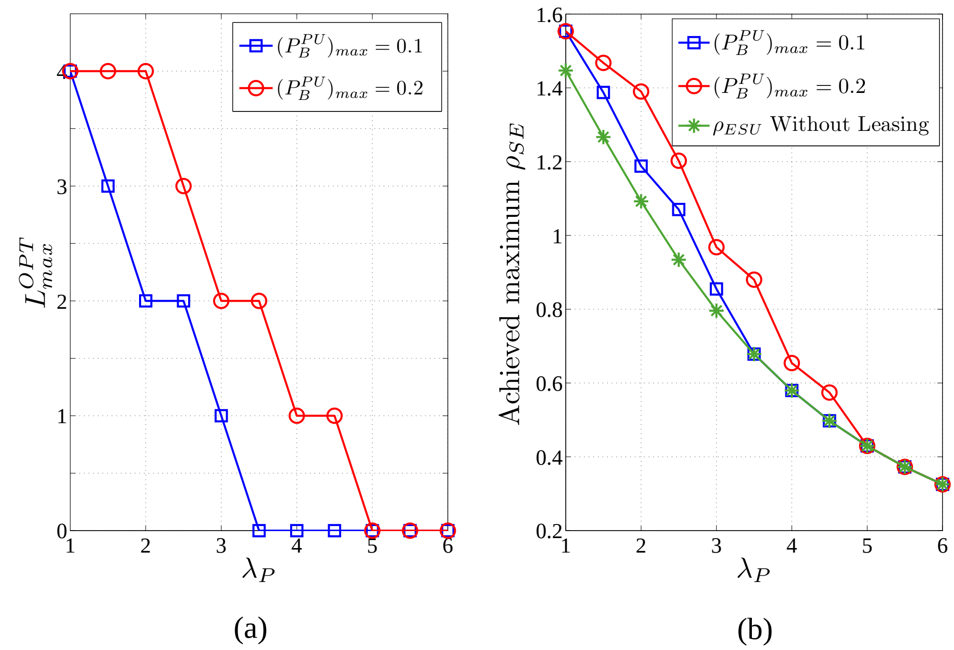


Figure A.18: Configurable maximum L_{max} and achievable maximum ESU capacity given that the maximum tolerable P_B^{PU} in the POL_D strategy when Algorithm A.1 is employed.

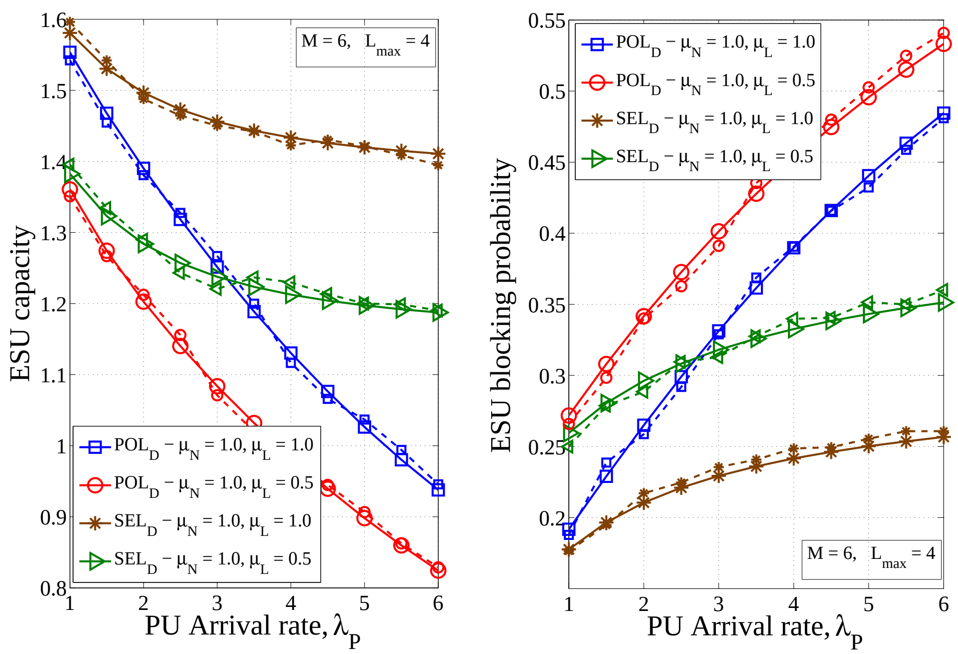


Figure A.19: Capacity and blocking probability analysis of ESU services in the heterogeneous channel scenario when Algorithm 1 is adopted. The dashed lines show the simulation results corresponding to the exponential service time distribution.

Table A.3: States and transitions for the POL_D strategy from a generic state $\boldsymbol{x} = \{i_n, j_n^E, j_n^R, i_l, j_l^E, j_l^R\}$ upon PU events.

Activity	Dest. State	Tran. rate	Conditions
PU AR. A vacant channel exists in the N-CRN.	$\boldsymbol{x}+\boldsymbol{e}_1$	λ_P	$b(\boldsymbol{x}) < M, b_n(\boldsymbol{x}) < M-L$
PU AR. No vacant channels exist in the N-CRN. An ESU_N performs handover to the L-CRN.	$\boldsymbol{x}+\boldsymbol{e}_1-\boldsymbol{e}_2+\boldsymbol{e}_5$	λ_P	$b(\boldsymbol{x}) < M, b_n(\boldsymbol{x}) \geq M-L; j_n^E > 0; b_l(\boldsymbol{x}) < L$
PU AR. No vacant channels exist in the N-CRN. An RSU_N performs handover to the L-CRN .	$\boldsymbol{x}+\boldsymbol{e}_1-\boldsymbol{e}_3+\boldsymbol{e}_6$	λ_P	$b(\boldsymbol{x}) < M, b_n(\boldsymbol{x}) \geq M-L; j_n^E = 0; j_n^R > 0; b_l(\boldsymbol{x}) < L$
PU AR. No vacant channels exist in the N-CRN. The new PU accesses a channel in the L-CRN .	$\boldsymbol{x}+\boldsymbol{e}_4$	λ_P	$b(\boldsymbol{x}) < M, b_n(\boldsymbol{x}) \geq M-L; j_n^E = 0; j_n^R = 0; b_l(\boldsymbol{x}) < L$
PU AR. No vacant channels exist in the N-CRN. An ESU_N is forced to terminate.	$\boldsymbol{x}+\boldsymbol{e}_1-\boldsymbol{e}_2$	λ_P	$b(\boldsymbol{x}) = M; j_n^E > 0$
PU AR. No vacant channels exist in the N-CRN. An RSU_N is forced to terminate.	$\boldsymbol{x}+\boldsymbol{e}_1-\boldsymbol{e}_3$	λ_P	$b(\boldsymbol{x}) = M; j_n^E = 0; j_n^R > 0$
PU DP from the N-CRN.	$\boldsymbol{x}-\boldsymbol{e}_1$	$i_n\mu_P$	$i_n > 0; i_l = 0$
PU DP from the N-CRN. A PU in the L-CRN performs handover to the N-CRN.	$\boldsymbol{x}-\boldsymbol{e}_1-\boldsymbol{e}_4+\boldsymbol{e}_1 = \boldsymbol{x}-\boldsymbol{e}_4$	$i_n\mu_P$	$i_n > 0; i_l > 0$
PU DP from the N-CRN. An RSU_N performs handover to the L-CRN.	$\boldsymbol{x}-\boldsymbol{e}_1-\boldsymbol{e}_3+\boldsymbol{e}_6$	$i_n\mu_P$	$i_n > 0; b_l(\boldsymbol{x}) < L; j_n^R > 0$
PU DP from the N-CRN. An ESU_N performs handover to the L-CRN.	$\boldsymbol{x}-\boldsymbol{e}_1-\boldsymbol{e}_2+\boldsymbol{e}_5$	$i_n\mu_P$	$i_n > 0; b_l(\boldsymbol{x}) < L; j_n^R = 0; j_n^E > 0$
PU DP from the L-CRN.	$\boldsymbol{x}-\boldsymbol{e}_4$	$i_l\mu_P$	$i_l > 0$

In this table and Table A.4, $\boldsymbol{e}_k$ denotes a six element vector whose k^{th} component is equal to 1 while the other components are 0. Here, $b(\boldsymbol{x}) = b_n(\boldsymbol{x}) + b_l(\boldsymbol{x})$. Note that in departure events L is determined when the activity is finished whereas in arrival events it is determined prior to handling the activity. The notations, AR and DP indicate an arrival event and a departure event respectively. An ESU and an RSU service in the N-CRN are denoted as ESU_N and RSU_N respectively.

A

A

Table A.4: States and transitions for the POL_D strategy from a generic state $\boldsymbol{x} = \{i_n, j_n^E, j_n^R, i_l, j_l^E, j_l^R\}$ upon SU events.

Activity	Dest. State	Tran. rate	Conditions
ESU AR. A vacant channel exists in the L-CRN.	$\boldsymbol{x} + \boldsymbol{e}_5$	λ_{SE}	$b(\boldsymbol{x}) < M; b_l(\boldsymbol{x}) < L$
ESU AR. A PU in the L-CRN performs handover to the N-CRN.	$\boldsymbol{x} + \boldsymbol{e}_5 - \boldsymbol{e}_4 + \boldsymbol{e}_1$	λ_{SE}	$b(\boldsymbol{x}) < M; b_l(\boldsymbol{x}) = L; i_l > 0$
ESU AR. The new ESU accesses a channel in the N-CRN.	$\boldsymbol{x} + \boldsymbol{e}_2$	λ_{SE}	$b(\boldsymbol{x}) < M; b_l(\boldsymbol{x}) = L; i_l = 0$
RSU AR. A vacant channel exists in the L-CRN.	$\boldsymbol{x} + \boldsymbol{e}_6$	λ_{SR}	$b(\boldsymbol{x}) < M; b_l(\boldsymbol{x}) < L$
RSU AR. A PU in the L-CRN performs handover to the N-CRN.	$\boldsymbol{x} + \boldsymbol{e}_6 - \boldsymbol{e}_4 + \boldsymbol{e}_1$	λ_{SR}	$b(\boldsymbol{x}) < M; b_l(\boldsymbol{x}) = L; i_l > 0$
RSU AR. The new RSU accesses a channel in the N-CRN.	$\boldsymbol{x} + \boldsymbol{e}_3$	λ_{SE}	$b(\boldsymbol{x}) < M; b_l(\boldsymbol{x}) = L; i_l = 0$
ESU DP from the N-CRN.	$\boldsymbol{x} - \boldsymbol{e}_2$	$j_n^E \mu_{SE}$	$j_n^E > 0$
ESU DP from the L-CRN. An RSU_N performs a handover to the L-CRN.	$\boldsymbol{x} - \boldsymbol{e}_5 - \boldsymbol{e}_3 + \boldsymbol{e}_6$	$j_l^E \mu_{SE}$	$j_l^E > 0, j_n^R > 0$
ESU DP from the L-CRN. An ESU_N performs a handover to the L-CRN.	$\boldsymbol{x} - \boldsymbol{e}_5 - \boldsymbol{e}_2 + \boldsymbol{e}_5 = \boldsymbol{x} - \boldsymbol{e}_2$	$j_l^E \mu_{SE}$	$j_n^E > 0$
RSU DP from the N-CRN.	$\boldsymbol{x} - \boldsymbol{e}_3$	$j_n^R \mu_{SR}$	$j_n^R > 0$
RSU DP from the L-CRN. An RSU_N performs a handover to the L-CRN.	$\boldsymbol{x} - \boldsymbol{e}_6 - \boldsymbol{e}_3 + \boldsymbol{e}_6 = \boldsymbol{x} - \boldsymbol{e}_3$	$j_l^R \mu_{SR}$	$j_l^E > 0, j_n^R > 0$
RSU DP from the L-CRN. An ESU_N performs a handover to the L-CRN.	$\boldsymbol{x} - \boldsymbol{e}_6 - \boldsymbol{e}_2 + \boldsymbol{e}_5$	$j_l^R \mu_{SR}$	$j_n^E > 0; j_l^R > 0$

Table A.5: States and transitions for the SEL_D strategy from a generic state $\boldsymbol{x} = \{i_n, j_n^E, j_n^R, j_l^E, j_l^R\}$ upon PU events.

Activity	Dest. State	Tran. rate	Conditions
PU AR. A vacant channel exists in the N-CRN.	$\boldsymbol{x}+\boldsymbol{e}_1$	λ_P	$b(\boldsymbol{x}) < M, b_n(\boldsymbol{x}) < M-L$
PU AR. No vacant channels exist in the N-CRN. An ESU_N performs handover to the L-CRN.	$\boldsymbol{x}+\boldsymbol{e}_1-\boldsymbol{e}_2+\boldsymbol{e}_4$	λ_P	$b(\boldsymbol{x}) < M, b_n(\boldsymbol{x}) \geq M-L; j_n^E > 0; b_l(\boldsymbol{x}) < L$
PU AR. No vacant channels exist in the N-CRN. An RSU_N performs handover to the L-CRN.	$\boldsymbol{x}+\boldsymbol{e}_1-\boldsymbol{e}_3+\boldsymbol{e}_5$	λ_P	$b(\boldsymbol{x}) < M, b_n(\boldsymbol{x}) \geq M-L; j_n^E = 0; j_n^R > 0; b_l(\boldsymbol{x}) < L$
PU AR. No vacant channels exist in the N-CRN. An ESU_N is forced to terminate.	$\boldsymbol{x}+\boldsymbol{e}_1-\boldsymbol{e}_2$	λ_P	$b(\boldsymbol{x}) = M; j_n^E > 0$
PU AR. No vacant channels exist in the N-CRN. An RSU_N is forced to terminate.	$\boldsymbol{x}+\boldsymbol{e}_1-\boldsymbol{e}_3$	λ_P	$b(\boldsymbol{x}) = M; j_n^E = 0; j_n^R > 0$
PU DP from the N-CRN.	$\boldsymbol{x}-\boldsymbol{e}_1$	$i_n\mu_P$	$i_n > 0; b_l(\boldsymbol{x}) = 0$
PU DP from the N-CRN. An RSU_N performs handover to the L-CRN.	$\boldsymbol{x}-\boldsymbol{e}_1-\boldsymbol{e}_3+\boldsymbol{e}_5$	$i_n\mu_P$	$i_n > 0; b_l(\boldsymbol{x}) < L; j_n^R > 0$
PU DP from the N-CRN. An ESU_N performs handover to the L-CRN.	$\boldsymbol{x}-\boldsymbol{e}_1-\boldsymbol{e}_2+\boldsymbol{e}_4$	$i_n\mu_P$	$i_n > 0; b_l(\boldsymbol{x}) < L; j_n^R = 0; j_n^E > 0$

In this table and Table A.6, $\boldsymbol{e}_k$ denotes a five element vector whose k^{th} component is equal to 1 while the other components are 0. Here, $b(\boldsymbol{x}) = b_n(\boldsymbol{x}) + b_l(\boldsymbol{x})$. Note that in departure events L is determined when the activity is finished whereas in arrival events it is determined prior to handling the activity. The notations, AR and DP indicate an arrival event and a departure event respectively. An ESU and an RSU service in the N-CRN are denoted as ESU_N and RSU_N respectively.

Table A.6: States and transitions for the SEL_D strategy from a generic state $\boldsymbol{x} = \{i_n, j_n^E, j_n^R, j_l^E, j_l^R\}$ upon SU events.

Activity	Dest. State	Tran. rate	Conditions
ESU AR. A vacant channel exists in the L-CRN.	$\boldsymbol{x} + \boldsymbol{e}_4$	λ_{SE}	$b(\boldsymbol{x}) < M; b_l(\boldsymbol{x}) < L$
ESU AR. A vacant channel exists in the N-CRN.	$\boldsymbol{x} + \boldsymbol{e}_2$	λ_{SE}	$b(\boldsymbol{x}) < M; b_l(\boldsymbol{x}) \geq L$
RSU AR. A vacant channel exists in the L-CRN.	$\boldsymbol{x} + \boldsymbol{e}_5$	λ_{SR}	$b(\boldsymbol{x}) < M; b_l(\boldsymbol{x}) < L$
RSU AR. A vacant channel exists in the N-CRN.	$\boldsymbol{x} + \boldsymbol{e}_3$	λ_{SR}	$b(\boldsymbol{x}) < M; b_l(\boldsymbol{x}) \geq L$
ESU DP from the N-CRN.	$\boldsymbol{x} - \boldsymbol{e}_2$	$j_n^E \mu_{SE}$	$j_n^E > 0$
ESU DP from the L-CRN. An RSU_N performs a handover to the L-CRN.	$\boldsymbol{x} - \boldsymbol{e}_4 - \boldsymbol{e}_3 + \boldsymbol{e}_5$	$j_l^E \mu_{SE}$	$j_l^E > 0, j_n^R > 0$
ESU DP from the L-CRN. An ESU_N performs a handover to the L-CRN.	$\boldsymbol{x} - \boldsymbol{e}_4 - \boldsymbol{e}_2 + \boldsymbol{e}_4 = \boldsymbol{x} - \boldsymbol{e}_2$	$j_l^E \mu_{SE}$	$j_l^E > 0, j_n^R = 0, j_n^E > 0$
RSU DP from the N-CRN.	$\boldsymbol{x} - \boldsymbol{e}_3$	$j_n^R \mu_{SR}$	$j_n^R > 0$
RSU DP from the L-CRN. An RSU_N performs a handover to the L-CRN.	$\boldsymbol{x} - \boldsymbol{e}_5 - \boldsymbol{e}_3 + \boldsymbol{e}_5 = \boldsymbol{x} - \boldsymbol{e}_3$	$j_l^R \mu_{SR}$	$j_l^R > 0, j_n^R > 0$
RSU DP from the L-CRN. An ESU_N performs a handover to the L-CRN.	$\boldsymbol{x} - \boldsymbol{e}_5 - \boldsymbol{e}_2 + \boldsymbol{e}_4$	$j_l^R \mu_{SR}$	$j_l^R > 0, j_n^R = 0, j_n^E > 0$

PAPER A: REFERENCES

[1] L. B. Le and E. Hossain, "Resource allocation for spectrum underlay in cognitive radio networks," *IEEE Trans. Wireless Commun.*, pp. 5306-5315, vol. 7, no. 12, Dec. 2008.

[2] L. Jiao, F. Y. Li, and V. Pla, "Modeling and performance analysis of channel assembling in multi-channel cognitive radio networks with spectrum adaptation," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2686-2697, Jul. 2012.

[3] H. Hakim, H. Boujemaa, and W. Ajib, "Performance comparison between adaptive and fixed transmit power in underlay cognitive radio networks," *IEEE Trans. Commun.*, pp. 4836-4846, vol. 61, no. 12, Dec. 2013.

[4] J. M. Peha, "Approaches to spectrum sharing," *IEEE Commun. Mag.*, vol. 43, no. 2, pp. 10-12, Feb. 2005.

[5] Federal Communications Commission, "Second report and order: Promoting efficient use of spectrum through elimination of barriers to the development of secondary markets (FCC 04-167)," Washington, DC, USA, Sep. 2004. [Online.] Available: https://apps.fcc.gov/edocs_public/attachmatch/FCC-04-167A1.pdf.

[6] M. Hafeez and J. M. H. Elmirghani, "Analysis of dynamic spectrum leasing for coded bi-directional communication," *IEEE J. Sel. Areas Commun.*, pp. 1500-1512, vol. 30, no. 8, Sep. 2012.

[7] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Spectrum leasing as an incentive towards uplink macrocell and femtocell cooperation", *IEEE J. Sel. Areas Commun.*, pp. 617-630, vol. 30, no. 3, Apr. 2012.

[8] O. Simeone, I. Stanojev, S. Savazzi, Y. B.-Ness, U. Spagnolini, and R. Pickholtz, "Spectrum leasing to cooperating secondary ad hoc networks," *IEEE J. Sel. Areas Commun.*, pp. 203-213, vol. 26, no. 1, Jan. 2008.

[9] D. Niyato and E. Hossain, "Competitive pricing for spectrum sharing in cognitive radio networks: Dynamic game, inefficiency of Nash equilibrium, and collusion," *IEEE J. Sel. Areas Commun.*, pp. 192-202, vol. 26, no. 1, Jan. 2008.

[10] H.-J. Lim, M.-G. Song, and G.-H. Im, "Cooperation-based dynamic spectrum leasing via multi-winner auction of multiple bands," *IEEE Trans. Commun.*, pp. 1254-1263, vol. 61, no. 4, Apr. 2013.

[11] C. Yi and J. Cai, "Two-stage spectrum sharing with combinatorial auction and Stackelberg game in recall-based cognitive radio networks," *IEEE Trans. Commun.*, pp. 3740-3752, vol. 62, no. 11, Nov. 2014.

[12] L. Duan, J. Huang, and B. Shou, "Duopoly competition in dynamic spectrum leasing and pricing," *IEEE Trans. Mobile Comput.*, pp. 1706-1719, vol. 11, no. 11, Nov. 2012.

[13] A. A. Daoud, M. Alanyali, and D. Starobinski, "Pricing strategies for spectrum lease in secondary markets," *IEEE Trans. Networking*, pp. 462-475, vol. 18, no. 2, Apr. 2010.

[14] S. K. Jayaweera, G. V.-Vilar, and C. Mosquera, "Dynamic spectrum leasing: A new paradigm for spectrum sharing in cognitive radio networks," *IEEE Trans. Veh. Technol.*, pp. 2328-2339, vol. 59, no. 5, Jun. 2010.

[15] R. D. Taranto, P. Popovski, O. Simeone, and H. Yomo, "Efficient spectrum leasing via randomized silencing of secondary users," *IEEE Trans. Wireless Commun.*, pp. 3739-3749, vol. 9, no. 12, Dec. 2010.

[16] K. Zhu, D. Niyato, P. Wang, and Z. Han, "Dynamic spectrum leasing and service selection in spectrum secondary market of cognitive radio networks," *IEEE Trans. Wireless Commun.*, pp. 1136-1145, vol. 11, no. 3, Mar. 2012.

[17] L. Qian, F. Ye, L. Gao, X. Gan, T. Chu, X. Tian, X. Wang, and M. Guizani, "Spectrum trading in cognitive radio networks: An agent-based model under demand uncertainty," *IEEE Trans. Commun.*, pp. 3192-3203, vol. 59, no. 11, Nov. 2011.

[18] Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy, "A scalable collusion-resistant multi-winner cognitive spectrum auction game," *IEEE Trans. Commun.*, pp. 3805-3816, vol. 57, no. 12, Dec. 2009.

[19] A. R. Syed and K.-L. A. Yau, "Spectrum leasing in cognitive radio networks: A survey," *Int. J. Distributed Sensor Networks*, vol. 2014, article ID 329235, 22 pages, 2014, Hindawi, doi:10.1155/2014/329235.

[20] S. K. Jayaweera and T. Li, "Dynamic spectrum leasing in cognitive radio networks via primary-secondary user power control games," *IEEE Trans. Wireless Commun.*, pp. 3300-3310, vol. 8, no. 6, Jun. 2009.

[21] I. Bajaj, Y. H. Lee, and Y. Gong, "A spectrum trading scheme for licensed user incentives," *IEEE Trans. Commun.*, pp. 4026-4036, vol. 63, no. 11, Nov. 2015.

[22] R. Fan, J. An, H. Jiang, and X. Li, "Dynamic pricing over multiple rounds of spectrum leasing in cognitive radio," *IEEE Trans. Veh. Technol.*, pp. 1782-1789, vol. 65, no. 3, Mar. 2016.

[23] S. L. C.-Lopez, F. A. C.-Pérez, M. E. R.-Angeles, and G. H.-Valdez, "Erlang capacity in coordinated cognitive radio networks with stringent-delay applications," in *Proc. IEEE PIMRC*, London, UK, Sep. 2013, pp. 3166-3170.

[24] R. Murawski and E. Ekici, "Utilizing dynamic spectrum leasing for cognitive radios in 802.11-based wireless networks," *Computer Networks*, vol. 55, no. 11, pp. 2646-2657, Elsevier, Aug. 2011.

[25] I. Stanojev, O. Simeone, U. Spagnolini, Y. Bar-Ness, and R. L. Pickholtz, "Cooperative ARQ via auction-based spectrum leasing," *IEEE Trans. Commun.*, pp. 1843-1856, vol. 58, no. 6, Jun. 2010.

[26] J. J. Alcaraz and M. V. D. Schaar, "Coalitional games with intervention: Application to spectrum leasing in cognitive radio," *IEEE Trans. Wireless Commun.*, pp. 6166-6179, vol. 13, no. 11, Nov. 2014.

[27] I. A. M. Balapuwaduge, A. Rajanna, M. Kaveh, and F. Y. Li, "Performance evaluation of three dynamic channel access strategies for spectrum leasing in CRNs," in *Proc. IEEE ICC*, London, UK, 8-12 Jun. 2015, pp. 7570-7575.

[28] V. K. Tumuluru, P. Wang, D. Niyato, and W. Song, "Performance analysis of cognitive radio spectrum access with prioritized traffic," *IEEE Trans. Veh. Tech.*, pp. 1895-1906, vol. 61, no. 4, May 2012.

[29] W.-H. Kuo and W. Liao, "Utility-based resource allocation in wireless networks," *IEEE Trans. Wireless Commun.*, pp. 3600-3606, vol. 6, no. 10, Oct. 2007.

[30] M. Xiao, N. B. Shroff, and E. K. P. Chong, "A utility-based power-control scheme in wireless cellular systems," *IEEE Trans. Networking*, pp. 210-221, vol. 11, no. 2, Apr. 2003.

[31] M. Sheng, C. Xu, X. Wang, Y. Zhang, W. Han, and J. Li, "Utility-based resource allocation for multi-channel decentralized networks," *IEEE Trans. Commun.*, pp. 3610-3620, vol. 62, no. 10, Oct. 2014.

[32] S. Shenker, “Fundamental design issues for the future Internet,” *IEEE J. Sel. Areas Commun.*, pp. 617-630, vol. 13, no. 7, Sep. 1995.

[33] C. Yang and S. Jordan, “Resource allocation for semi-elastic applications with outage constraints in cellular networks,” *IEEE Trans. Veh. Tech.*, pp. 1591-1606, vol. 64, no. 4, Apr. 2015.

[34] L. Chen, B. Wang, X. Zhang, and D. Yang, “Utility-based resource allocation for mixed traffic in wireless networks,” in *Proc. IEEE Computer Communications Workshops Collocated with IEEE INFOCOM*, Shanghai, China, Apr. 2011, pp. 91-96.

[35] D. Niyato and E. Hossain, “A Game-theoretic approach to competitive spectrum sharing in cognitive radio networks,” in *Proc. IEEE WCNC*, pp. 16-20, Hong Kong, Mar. 2007.

[36] J. Wang, Q.-A. Zeng, and D. P. Agrawal, “Performance analysis of a preemptive and priority reservation handoff scheme for integrated service-based wireless mobile networks,” *IEEE Trans. Mobile Comput.*, pp. 65-75, vol. 2, no.1, Jan.-Mar. 2003.

[37] J. Zheng and M. Ma, “A utility-based joint power and rate adaptive algorithm in wireless ad Hoc networks,” *IEEE Trans. Commun.*, pp. 134-140, vol. 57, no. 1, Jan. 2009.

Paper B

Title:	System Times and Channel Availability for Secondary Transmissions in CRNs: A Dependability Theory based Analysis
Authors:	Indika A. M. Balapuwaduge[†], Frank Y. Li[†], and Vicent Pla[‡]
Affiliation:	[†] Dept. of Information and Communication Technology, University of Agder (UiA), N-4898 Grimstad, Norway [‡] Dept. of Communications, Universitat Politècnica de València (UPV), 46022 València, Spain
Journal:	*IEEE Transactions on Vehicular Technology*, Early access available in IEEE Xplore, June 2016, DOI:10.1109/TVT.2016.2585200.
Copyright ©:	IEEE

B

System Times and Channel Availability for Secondary Transmissions in CRNs: A Dependability Theory based Analysis

Indika A. M. Balapuwaduge, Frank Y. Li, and Vicent Pla

Abstract — **Reliability is of fundamental importance for the performance of secondary networks in cognitive radio networks (CRNs). To date, most studies have focused on predicting reliability parameters based on prior statistics of traffic patterns from user behavior. In this paper, we define a few reliability metrics for channel access in multi-channel CRNs which are analogous to the concepts of reliability and availability in the classical dependability theory. Continuous time Markov chains (CTMC) are employed to model channel available and unavailable time intervals based on channel occupancy status. The impact on user access opportunities based on channel availability is investigated by analyzing the steady state channel availability and several system times such as mean channel available time and mean time to first channel unavailability. Moreover, the complementary cumulative distribution function for channel availability is derived by applying the uniformization method and it is evaluated as a measure of guaranteed availability for channel access by secondary users. The preciseness and the correctness of the derived analytical models are validated through discrete-event based simulations. We believe that the reliability metric definitions and the analytical models proposed in this paper have their significance for reliability and availability analysis in CRNs.**

B

Keywords—Cognitive radio networks, spectrum access, channel availability, system times, guaranteed availability, CTMC.

 Manuscript recieved November 30, 2015; revised April 22, 2016; accepted June 14, 2016. The research of V. Pla was partially supported by the Ministry of Economy and Competitiveness of Spain under Grant TIN2013-47272-C2-1-R. The review of this paper was coordinated by Prof. B. Canberk.

Indika A. M. Balapuwaduge, and Frank Y. Li are with the Department of Information and Communication Technology, University of Agder (UiA), N-4898 Grimstad, Norway (email: {indika.balapuwaduge; frank.li}@uia.no).

Vicent Pla is with ITACA, Universitat Politècnica de València (UPV), 46022 València, Spain (email: vpla@upv.es).

I. Introduction

During the past decade, the research interests on cognitive radio networks (CRNs) have been growing tremendously due to CRN's capability of exploiting the unused spectrum in dynamically changing environments [1], [2]. The cognition and reconfigurability features of cognitive radios (CRs) tackle the problem of spectrum scarcity, improving the quality of service (QoS) as well as communication reliability. Dynamic spectrum access is one of the key features for CRNs which allows secondary users (SUs) occupy unused channels in an opportunistic manner. In this way, primary users (PUs) who own the license have priority for spectrum access and SUs can also access the channel when the spectrum is not occupied by PUs. Consequently, the system performance of the secondary network (SN) depends on the dynamics of the primary network.

In order to improve the performance of SNs, various techniques have been proposed in the literature. The focus of these approaches has been on techniques including spectrum sensing [3], channel access schemes and resource allocation [2], [5], spectrum management [6], etc. However, little attention has been paid so far to the reliability and availability performance of SU networks from the perspective of dependability theory. Indeed, *dependability* is one of the four major performance attributes[1] for any computing and communication systems since it has direct impact on the performance of the system [7]. Therefore, studying the reliability and availability aspects of CRNs is of essential importance for real-life deployment of CRNs in the near future. This paper makes an effort towards this direction by defining and investigating several reliability and availability metrics for channel access in multi-channel CRNs based on dependability theory.

The dependability of a computing or communication system can be described as its ability to deliver services that can justifiably be trusted [8]. Therein, reliability, availability, safety and confidentiality are the main attributes which describe the properties of a system with respect to dependability. In a CRN, PUs always have priority for channel access while SUs can only opportunistically access the idle channels. Once all the channels are occupied at a given instant, a new secondary request will simply be blocked and the CRN is said to be unavailable for new SUs. Thus, channel availability is an important metric for SNs to be considered when accessing channels in a CRN. Furthermore, the analysis of channel idleness from the *dependability perspective* provides a systematic approach to evaluate the time intervals related to channel availability and unavailability periods.

[1]In general, computing and communication systems are characterized by four main properties: functionality, performance, cost and dependability.

The importance of employing *dependability theory* for reliability analysis is twofold. First and most importantly, the terminologies related to QoS of telecommunication services have been defined in the *ITU-T E.800 Recommendation* [9] based on the dependability theory. Correspondingly, our definitions of reliability metrics introduced in this study conform with the ITU recommendations since dependability theory forms the basis for our study. Second, the four types of system times (mean time to first channel unavailability, mean time to channel unavailability, mean channel available time and mean channel unavailable time) defined in this study can be adopted for accurate and systematic reliability analysis in CRNs from the perspective of the dependability theory rather than using other empirical rules or conventional performance metrics like packet error probability or packet delivery ratio. Therefore the validity and the eligibility of the definitions for reliability measures are confirmed and the model developed in this study provides a systematic approach for reliability analysis in CRNs and it is applicable to realistic scenarios.

In this paper, we focus particularly on two categories of dependability measures, i.e., availability and system times, for spectrum access in CRNs. The goal of this study is to define those dependability measures which constitute the basis for determining channel availability in CRNs and to develop analytical models to calculate system availability and its distribution. In brief, the main contributions of this paper are summarized as follows:

- A number of reliability and availability metrics for channel access in CRNs are defined from *the perspective of dependability theory*, including mean time to first channel unavailability, mean time to channel unavailability, mean channel available and unavailable times, and steady state channel availability.
- A Markov process based approach is proposed to calculate channel availability as well as related system times in a multi-channel CRN. In addition to channel availability, we introduce a scale of guaranteed availability and deduce the complementary cumulative distribution function (CCDF) of channel availability of CRNs.
- The proposed model is evaluated by adopting an existing channel access scheme from our early work [4] to investigate how the reliability of a CRN is influenced by traffic load under both homogeneous and heterogeneous channel conditions. Furthermore, the dependability measures of the SN are evaluated using a queuing scheme as an extension to the existing scheme. The proposed model is however independent of the channel access schemes employed.

- The correctness and the preciseness of the derived analytical models are validated through discrete-event based simulations. Additionally, channel availability under log-normal PU interarrival time and SU service time distributions is also validated by means of simulations in order to assess the applicability of our model under non-exponentially distributed interarrival and service times.

The remainder of this paper is organized as follows. In Sec. II, we provide background information on the concept of reliability in communication systems and translate it into CRNs. The related work is also summarized therein. Sec. III outlines the network scenario and the dynamic channel access scheme employed in this study. Then the procedure of the proposed analytical model to calculate channel availability and system times in a CRN is presented in Sec. IV, followed by the detailed description of the numerical approach to determine the CCDF of channel availability in Sec. V. In Sec. VI, the numerical results are presented and discussed. Finally, the conclusions are drawn in Sec. VII.

II. Reliability and Availability in CRNs: Background and Related Work

In this section, we first revisit the classical reliability concepts in communication networks and present briefly the concepts of reliability in CRNs introduced in this study. Afterward, the related work is summarized.

A. Background Information on Reliability Metrics

In communications networks and electronic systems, the time to first failure is the time from the system initiation instant until it fails for the first time. For *non-repairable* systems or components like integrated circuits, the mean time to first failure (MTFF) is a critical dependability parameter which indicates the system's expected lifetime. Denote a random variable which represents the time to first failure as TFF and the cumulative distribution function of time to first failure as $F_{TFF}(t)$. We have

$$F_{\text{TFF}}(t) = P(\text{TFF} \leq t) = \int_0^t f_{\text{TFF}}(x)\,dx \quad \text{for} \quad t \geq 0, \tag{B.1}$$

where $f_{\text{TFF}}(t)$ is the probability density function of the random variable TFF. The reliability function $R(t)$ is defined as the probability that a system can provide its required services under stated conditions for a given time interval. Therefore,

$$R(t) = P(\text{TFF} > t) = 1 - F_{\text{TFF}}(t) = \int_t^\infty f_{\text{TFF}}(x)\,dx. \tag{B.2}$$

In *repairable* systems, the mean time to failure (MTTF) is another parameter indicating the time from a *random instant* when the system is working and is in its steady state until it fails. The available time or uptime is the time during which the system is operational. The unavailable time or downtime is the time during which the system is not operational. Observe the system for a cycle which includes a period of uptime and a period of downtime respectively. Then the mean cycle time,

$$\text{MCT} = \text{MUT} + \text{MDT}$$

where MUT and MDT represent mean uptime and mean downtime respectively. Correspondingly, the steady state availability, A_{ss}, is obtained as the ratio between the *long run average* of the uptime and the cycle time, expressed as,

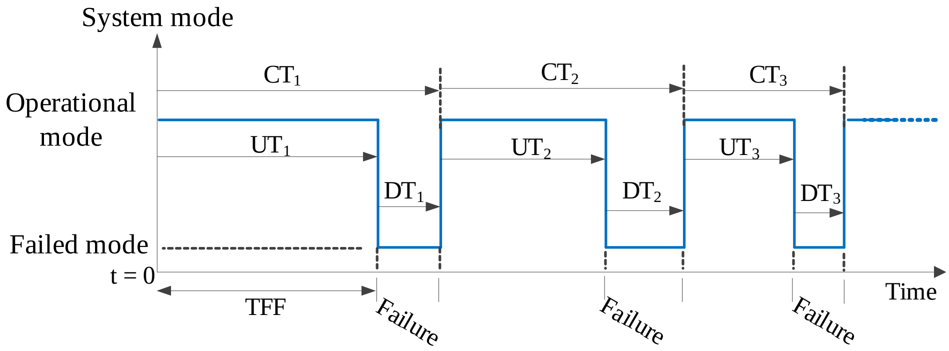
$$A_{ss} = \frac{\text{MUT}}{\text{MUT} + \text{MDT}}. \tag{B.3}$$

B. The Concepts of Reliability in CRNs

In this subsection, we initiate the concepts of reliability in CRNs. Later on in Subsec. IV-C, we will give the definitions of channel availability and the related system times mathematically. To do so, the above mentioned reliability metrics are revised such that they can be applied in multi-channel CRNs without deviating from the original concept.

Let the *system* referred to in the aforementioned classical reliability model be a CRN with one or multiple channels. The network is considered to be in the *operational mode* if there are enough channels available to commence a new SU request. A *system failure* represents the instant when there is no channel access opportunity for a newly arrived SU request to start its service. Correspondingly, the network is considered to be in the *failed mode* if the system has to block the new SU requests.

Similar to the concept of MUT, we define the mean channel available time, $\bar{T}_{UT}$, in a CRN as the average time duration during which the network resides in the operational mode. Accordingly, the mean channel unavailable time, $\bar{T}_{DT}$, is defined as the average time duration during which the network resides in the failed mode. Moreover, the MTFF in the considered CRN is represented by the mean time to first channel unavailability, $\bar{T}_{FF}$, indicating the expected time interval from network initiation to the instant when a new SU request is blocked for the first time. The concepts of UT, DT and TFF in CRNs as well as their relationships with channel availability are illustrated in Fig. B.1. Note that the average values of UT_i and DT_i terms in Fig. B.1, where $\text{i} = 1, 2, 3, \cdots$, determine the $\bar{T}_{UT}$ and the $\bar{T}_{DT}$ values respectively.



TFF = Time to first channel unavailability
UT_i = Channel available time
DT_i = Channel unavailable time
CT_i = $UT_i + DT_i$ = Cycle period

Figure B.1: Illustration of channel availability time intervals and dependability metrics in CRNs.

C. Related Work

Although the notion of dependability has been an important aspect in communication systems since long time ago, very little work has been done on the dependability aspects of CRNs. In [10], a resource management controller was designed for primary and secondary traffic in vehicular networks. The proposed controllers in that paper provide reliable guarantees to PU traffic in terms of aggregate goodput and collision rate. Moreover, the same authors proposed a reliable adaptive resource management controller in [11] for cognitive cloud vehicular networks in order to provide hard reliability guarantees to PU traffic in the presence of mobility and fading induced changes in vehicular networks. However, the considered reliability metrics in those two papers are distinct from the metrics defined in this paper. To maximize the network performance of a *single channel CRN*, a combined scheme considering PU activity and spectrum hole detection was proposed in [76] to address the potential drawbacks of the Poisson model.

The availability and reliability of wireless multi-hop networks, not of CRNs, were evaluated in [13] by considering stochastic link failures. By placing redundant nodes at appropriate locations in the existing network, the availability of wireless links is improved. The availability of a multi-cell CRN under time varying channels was studied in [14] and accordingly an efficient spectrum allocation mechanism was proposed. In the proposed network model therein, the average number of available channels is determined based on busy probabilities of channels. The authors of [15] demonstrated that deterministic channel failure models might result in a significant over-estimation or under-estimation of network reliability. Therefore, a reliability assessment was performed for wireless mesh networks in [15] by considering a probabilistic regional failure model which analyzed the geographical location of the network. Another recent paper [16] focused on investigating the security-reliability tradeoff of cognitive relay transmission in the presence of realistic spectrum sensing. Therein reliability was characterized in terms of outage probability.

In [17], the probability of the availability for a spectrum band in CRNs was evaluated by incorporating the results of an algorithm which can predict the arrival patterns of PUs. Thereafter, a similar but improved technique was proposed by the same authors in [18] to evaluate channel availability in CRNs. In our earlier work [19], three availability metrics were defined from the perspective of dependability theory. In [20], the authors analyzed the feasibility of opportunistic spectrum access in CRNs to provide a robust infrastructure for wireless networks. Moreover, the work done in [21] and [22] also analyzed the reliability aspects in CRNs.

However, the research work presented in this paper is distinct from the above related work with respect to mainly two aspects. First, by defining important system times from *the perspective of dependability theory*, we propose a systematic approach for analyzing channel availability in multi-channel CRNs. Second, the presented work thoroughly analyzes channel availability of CRNs via multiple metrics including system times, asymptotic channel availability and the distribution of interval availability instead of studying merely the steady state availability as in [19] and [20]. To the best of our knowledge, this is the first work which defines the reliability metrics combined with channel access in multi-channel CRNs and develops accordingly mathematical models to analyze them from the dependability theory's perspective.

III. Network Scenario and Access Schemes

In this section, we present the network scenario considered in this study and the representative channel access schemes used to obtain numerical results to be presented in Sec. VI.

A. Network Scenario and Assumptions

In this study we consider both homogeneous and heterogeneous channel environments for a CRN. In the homogeneous channel scenario, all channels have equal bandwidth and equal transmission data rate per channel. In contrast, in the heterogeneous channel scenario, we consider two types of channels which have different bandwidth and transmission rates.

In the following sections, CRNs with homogeneous channels and heterogeneous channels are referred to as *Scenario I* and *Scenario II* respectively. In *Scenario I*, we

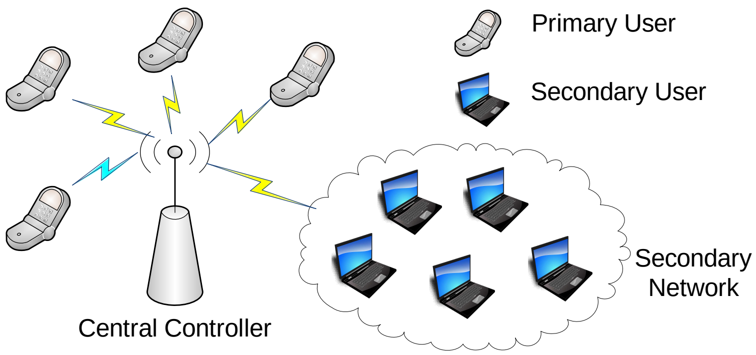


Figure B.2: A centralized architecture for CRN.

consider a CRN with $M \in \mathbb{Z}^+$ channels with equal bandwidth where, $\mathbb{Z}^+$ is the set of positive integers. In *Scenario II*, a CRN with two types of channels is considered, namely, $M_1 \in \mathbb{Z}^+$ wider bandwidth channels hereafter referred to as wideband channels, and correspondingly $M_2 \in \mathbb{Z}^+$ narrowband channels. For presentation clarity, we denote the set of wideband channels as *WBC* and the set of narrowband channels as *NBC*.

In both scenarios, a centralized CRN architecture is considered. Fig. B.2 illustrates the infrastructure based network architecture envisaged in this study. Resource allocation for the primary and secondary networks is coordinated by the base station which functions as the central controller in the network. Only one user can be allocated to a specific channel at a time. When a channel is occupied by an SU or a PU, the channel is said to be in the *busy mode* whereas it is in the *idle mode* if it is not occupied by any user. PUs have the priority of accessing all channels and SUs may occupy the channels which are in the idle mode in an opportunistic way. For user traffic considered in this study, we model traffic at the *flow level* since modeling at this level captures the dynamics related to the arrival and departure of flows such as flow duration or number of active flows. While a PU service occupies only one channel for each service, an SU may assemble several channels in order to complete its service at a higher data rate. Moreover, the following assumptions are made in order to develop our analytical model to be presented later.

- The arrivals of both PU and SU services are Poisson processes with arrival rates λ_P, λ_S for PU and SU services respectively.
- The service times for PU and SU services are exponentially distributed. In *Scenario I*, the service rates per channel for PU and SU services are μ_P and μ_S respectively. In *Scenario II*, the corresponding service rates per channel in the WBC are μ_{P1} and μ_{S1} respectively while those values in the NBC take μ_{P2} and μ_{S2} respectively.

- The sensing and spectrum adaptation latency is negligible in comparison with the time between two consecutive service events. These service events indicate PU and SU arrivals and departures.

It is worth mentioning that the proposed continuous time Markov chain (CTMC) models are applicable to other CRN channel access schemes as well, given that the PU and SU activity models are Markovian.

B. Dynamic Channel Access Scheme for Scenario I

1) Access scheme: The channel access scheme employed in the considered CRN with homogeneous channels is one of the dynamic channel access strategies proposed in [23], namely, dynamic fully adjustable (*DFA*). In the *DFA* scheme, channel aggregation is adopted, meaning that one SU service can utilize multiple channels if available. Two parameters, W and V, are introduced in *DFA* to represent the lower bound and the upper bound of the number of aggregated channels for an SU service respectively. Moreover, spectrum adaptation is allowed in order to protect ongoing SU services. Therefore, if an ongoing SU is interrupted *upon a PU arrival* and there are idle channels available, the interrupted SU will immediately release the occupied channel to the licensed PU and continue its service on one of the idle channels. In the following, we revisit the channel access procedure adopted in *DFA* with respect to the SU and PU activities.

- SU Arrival: When a new SU service arrives, the system should offer at least W channels to accommodate the new arrival. Otherwise the request is rejected. Moreover, the new SU service may aggregate up to V channels if there are enough channels available. If the number of idle channels upon an SU arrival is fewer than W, the ongoing SU service with the maximum number of channels will donate one or several channels as long as it can still keep W channels after donation. If the donated number is not enough, other ongoing SU services will collectively donate channels. If the number of idle channels plus the number of channels that can be donated by ongoing SU services is still fewer than W, the new SU request is blocked.

- PU Arrival: If a new PU arrives at a moment when there are idle channels in the CRN, the new PU can start transmission in an idle channel. If there is no idle channel, the SU service which has the maximum number of aggregated channels will donate one channel to the interrupted SU service, given that it has more aggregated channels than the interrupted SU service and its remaining number is still not fewer than W. In the worst case, if all ongoing

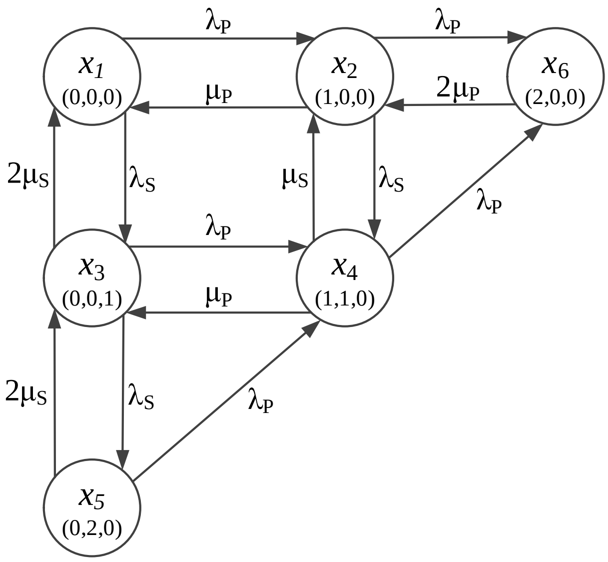


Figure B.3: State transition diagram of a CRN employing the *DFA* scheme when $M = 2, W = 1$ and $V = 2$ as well as the state transition rates.

SU services have exactly W channels upon a PU arrival, the interrupted SU service is forced to terminate.

- PU and SU Departures: As a result of *service departures of both PUs and SUs, or forced termination of SU services*, channels become idle. The ongoing SU service which has the minimum number of aggregated channels will then utilize those idle channels up to V. If the SU service with the minimum number reaches the upper bound V after spectrum adaptation and there are still idle channels, other SU services will occupy the remaining ones according to the same principle.

2) CTMC model for *DFA*: The states of the CTMC model corresponding to *DFA* can be represented by $\boldsymbol{x} = (i_{pu}, j_W, j_{W+1}, \cdots, j_V)$ where i_{pu} denotes the number of PU services in the CRN. The number of SU services with k aggregated channels in the CRN is denoted as j_k where $k = W, W+1, \cdots, V$. Let $\mathcal{S}_1$ be the set of feasible states of the system such that $\mathcal{S}_1 = \{\boldsymbol{x}|i_{pu}, j_W, j_{W+1}, \cdots, j_V \geq 0; b(\boldsymbol{x}) \leq M\}$ where $b(\boldsymbol{x}) = i_{pu} + \sum_{k=W}^{V} kj_k$. The state transitions associated with different service events are summarized by considering various traffic conditions in Table B.3, Table B.4, Table B.5 and Table B.6.

C. An Example to Illustrate Channel Availability in DFA

To illustrate the system times and channel availability described in Subsec. II-A, we present an example topology of Scenario I. Consider a *DFA* employed CRN with $M = 2$ channels, where $W = 1$ and $V = 2$. Accordingly, to commence

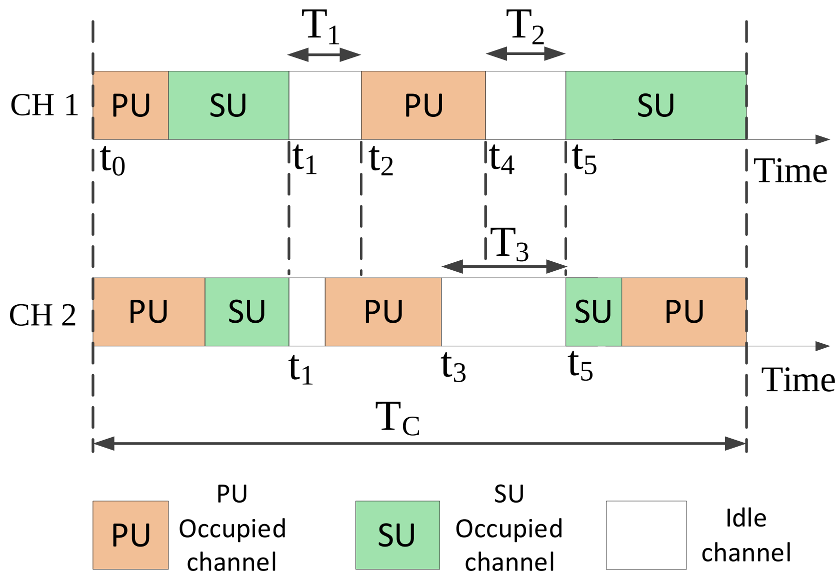


Figure B.4: Channel occupancy representation of a two channel CRN (*CH 1* and *CH 2*) where one SU service may occupy one or two channels if available.

a new SU request, the CRN should be able to provide at least one channel. A general state, $\boldsymbol{x}$, of this system can be expressed as (i_{pu}, j_1, j_2), where i_{pu} denotes the number of PUs in the system, j_1 and j_2 denote the number of SUs which have single channel occupancy and two channel occupancy respectively. The states of this system are represented in Fig. B.3 as $\boldsymbol{x_r}$ where $r = 1, 2, \cdots, 6$, and the state transitions corresponding to this example scenario are also shown.

Consider state $\boldsymbol{x_2}(1, 0, 0)$ where only one channel is occupied by a PU service. Indeed, it is a channel available state for new SU requests, since the other channel can be allocated upon a new SU request. Similarly, states $\boldsymbol{x_1}$ and $\boldsymbol{x_3}$ are also channel available states. However, in state $\boldsymbol{x_6}(2, 0, 0)$, both channels are occupied by two PUs and therefore $\boldsymbol{x_6}$ is *not* a channel available state for new SU requests. Likewise, states $\boldsymbol{x_4}$, and $\boldsymbol{x_5}$ are channel unavailable states since new SU services cannot be admitted to the network in those two states. An example of the channel occupancy status for this CRN during a continuous time interval T_C is shown in Fig. B.4. As shown in this figure, a new SU service can be commenced with at least one channel from time instant t_1 to t_2 and time instant t_3 to t_5. Thus, during those time intervals the CRN is available for a new SU service. Let T_1 and T_3 represent the duration between time instants t_1 and t_2 and time instants t_3 and t_5 respectively, i.e., $T_1 = t_2 - t_1$ and $T_3 = t_5 - t_3$. In this paper, channel availability is regarded as *the fraction of time that the CRN can allocate at least the minimum number of required channels for a new SU request.* As defined in (B.3), the steady state

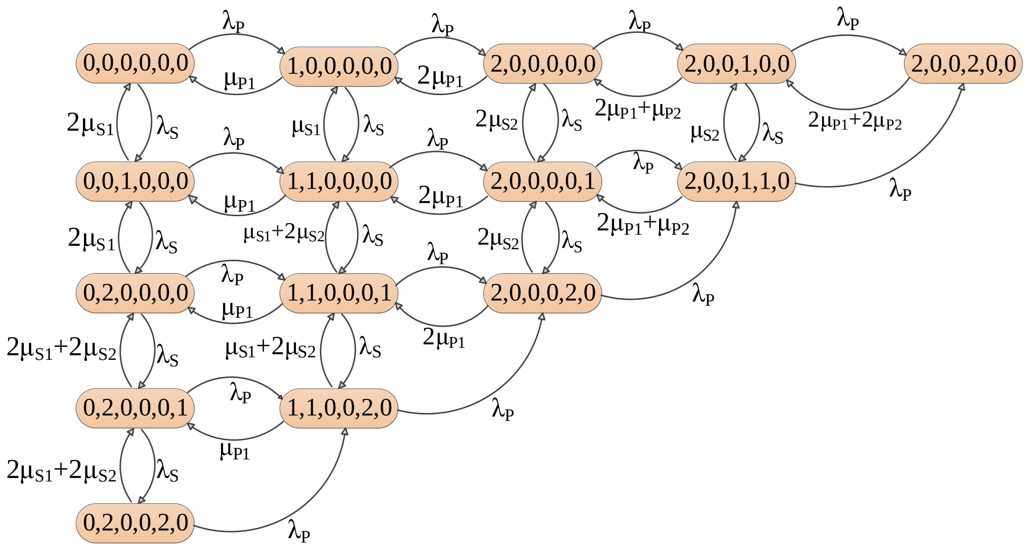


Figure B.5: State transition diagram for the *DFA*$_H$ scheme when $M_1 = 2, M_2 = 2, W = 1, V = 2$ given that $\mu_{S1} > \mu_{S2}$.

channel availability of this CRN, A_{ss}, can be estimated as $^{(T_1+T_3)}/_{T_C}$, given that the observation period T_C is sufficiently long.

D. Dynamic Channel Access Scheme for Scenario II

1) Access scheme: The basic principles of the channel access policy for *Scenario II* are the same as those highlighted in Sec. III-B for Scenario I, except the modifications explained herein. The proposed scheme for *Scenario II* is referred to as heterogeneous channel access for *DFA* (*DFA*$_H$).

Upon an arrival of a PU or an SU, the system first searches for an opportunity in the WBC. If there is no channel access opportunity in the WBC, then access in the NBC will be attempted. Although channel aggregation in *DFA*$_H$ is also enabled, it is performed either among the WBC or the NBC channels. More complicated channel access which allows the aggregation of heterogeneous channels is not considered to avoid increased complexity in our model. Anyhow, we have also studied a more flexible channel access in CRNs by allowing both channel aggregation and channel fragmentation in another paper [24].

Upon a departure of a service or a forced termination of a service in the WBC, the resultant vacant channels are allocated according to the following rule. The first priority for accessing those vacant channels is given to the ongoing PUs and SUs in the NBC successively. Thus, the PUs and SUs in the NBC perform spectrum handover to the newly vacant channels in the WBC. However, the ongoing services in the NBC handover to the WBC only if the current aggregated service rate is lower than that of the new opportunity. If more vacant channels exist even after this handoff process, the SUs with minimum number of aggregated channels in the WBC can aggregate them. Nevertheless, the vacant channels in the NBC will *not* be allocated to SUs in the WBC considering that users with high bandwidth resources prefer to continue their services in the WBC rather than performing spectrum handover to a NBC and channel aggregation.

2) CTMC model for DFA_H: The states of the CTMC model can be represented by $\boldsymbol{x} = (i^w, j^w_W, j^w_{W+1}, \cdots, j^w_V, i^n, j^n_W, j^n_{W+1}, \cdots, j^n_V)$. Here i^w and j^w_k denote, respectively, the number of PU services and the number of SU services with $k(k \in \mathbb{Z}^+, W \leq k \leq V)$ aggregated channels in the WBC. Similarly, i^n and j^n_k denote the number of those services in the NBC. Let $\mathcal{S}_2$ be the set of feasible states of the system, $\mathcal{S}_2 = \{\boldsymbol{x} | i^w, j^w_W, j^w_{W+1}, \cdots, j^w_V, i^n, j^n_W, j^n_{W+1}, \cdots, j^n_V \geq 0; b_w(\boldsymbol{x}) \leq M_1; b_n(\boldsymbol{x}) \leq M_2\}$ where $b_w(\boldsymbol{x}) = i^w + \sum_{k=W}^{V} k j^w_k$ and $b_n(\boldsymbol{x}) = i^n + \sum_{k=W}^{V} k j^n_k$. The state transitions associated with different events in DFA_H are summarized with different conditions in Table B.4. As an example, the complete state transition diagram for a CRN with $M_1 = 2, M_2 = 2, W = 1$ and $V = 2$, can be found in Fig. B.5, which illustrates the state transitions associated with different PU and SU activities. Moreover, in Scenario II, channel availability is regarded as *the fraction of time that the CRN can allocate at least the minimum number of required channels for a new SU request either in the WBC or in the NBC.*

IV. CTMC Model for Dependability Analysis

CTMC models are quite often used as an analytical tool in reliability engineering [25]. Since it can capture the important interdependence and dynamic relationships between system components, it is more efficient than the other analytical techniques such as reliability block diagrams and fault tree analysis [26]. In this section, we present a stepwise procedure of the proposed analytical model to evaluate system times and channel availability based on CTMC analysis. Note that the analytical model developed in this section applies to both Scenario I and Scenario II.

A. State Space and Partitioned State Transition Matrix

The set of feasible states of the system is denoted as $\mathcal{S}$. Without loss of generality, assume that the total number of states in the system is N. Let i represent a general state in the system where $1 \leq i \leq N, \ \ i \in \mathbb{Z}^+$. Here, $N = n(\mathcal{S})$, where $n(\mathcal{A})$ denotes the cardinality of the set $\mathcal{A}$. Furthermore, the steady state probability of being in state i is denoted as π_i. The stationary probabilities, π_i, can be calculated from the global balance equations and the normalization equation [27], which are given as

$$\boldsymbol{\pi Q} = \boldsymbol{0}, \ \sum_{i \in \mathcal{S}} \pi_i = 1, \tag{B.4}$$

where $\boldsymbol{\pi}$ is the steady state probability vector, $\boldsymbol{Q}$ denotes the transition rate matrix, and $\mathbf{0}$ denotes a row vector of 0's of an appropriate size. Furthermore, the entire state space, $\mathcal{S}$, can be divided into two subsets, i.e., channel available states *(operational mode)* and channel unavailable states *(failed mode)*. For convenience, we consider that the states in $\mathcal{S}$ are ordered (lexicographically) and refer to them by their order number. In other words, the feasible states are re-arranged in a way that in subset $\mathcal{S}_{\mathcal{A}} = \{1, 2, ..., L\}$, the system can provide at least the minimum number of required channels for a new SU request, while in subset $\mathcal{S}_{\mathcal{B}} = \{L+1, ..., N\}$, the system failed to do so. Thus, $\mathcal{S} = \mathcal{S}_{\mathcal{A}} \cup \mathcal{S}_{\mathcal{B}}$ and $\mathcal{S}_{\mathcal{A}} \cap \mathcal{S}_{\mathcal{B}} = \varnothing$. Since $n(\mathcal{S}_{\mathcal{A}}) = L$, then $n(\mathcal{S}_{\mathcal{B}}) = N - L$.

By re-arranging the rows and columns, the transition rate matrix $\boldsymbol{Q}$ can be written in the following partitioned form

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}, \tag{B.5}$$

B

where the $L \times L$ matrix $\boldsymbol{A}$ represents the transition rates from a state in $\mathcal{S}_{\mathcal{A}}$ to another state in $\mathcal{S}_{\mathcal{A}}$ and the $L \times (N-L)$ matrix $\boldsymbol{B}$ represents the transition rates from a state in $\mathcal{S}_{\mathcal{A}}$ to a state in $\mathcal{S}_{\mathcal{B}}$. Similarly, the $(N-L) \times L$ matrix $\boldsymbol{C}$ represents the transition rates from a state in $\mathcal{S}_{\mathcal{B}}$ to another state in $\mathcal{S}_{\mathcal{A}}$ and the $(N-L) \times (N-L)$ matrix $\boldsymbol{D}$ represents the transition rates from a state in $\mathcal{S}_{\mathcal{B}}$ to another state in $\mathcal{S}_{\mathcal{B}}$. Note that the entries in the main diagonal of $\boldsymbol{A}$ (and $\boldsymbol{D}$) represent the total outgoing rates from the states in $\mathcal{S}_{\mathcal{A}}$ (respectively, $\mathcal{S}_{\mathcal{B}}$) and, as such, they also include the transition rates to the states in $\mathcal{S}_{\mathcal{B}}$ (respectively, $\mathcal{S}_{\mathcal{A}}$).

Let $P_i(t)$ be the probability that the system is in state i at time t. Define

$$\boldsymbol{P}_A(t) \equiv [P_1(t)\ P_2(t)\ \cdots\ P_L(t)]$$

as an L-dimensional row vector with the transient probabilities of the channel available states with i^{th} component. Similarly, $\boldsymbol{P}_B(t)$ denotes the $N - L$ dimensional row vector representing the channel unavailable states, and

$$\boldsymbol{P}_B(t) \equiv [P_{L+1}(t)\ P_{L+2}(t)\ \cdots\ P_N(t)].$$

Assume further that, at the system initial instant, i.e., $t = 0$, the system can provide sufficient number of idle channels to new SU requests, i.e.,

$$\boldsymbol{P}_A(0)\boldsymbol{U}_L = 1 \text{ and } \boldsymbol{P}_B(0)\boldsymbol{U}_{N-L} = 0, \tag{B.6}$$

where $\boldsymbol{U}_L$ denotes the column vector of L ones.

B. Distribution of the Time until the First Visit to Subset $\mathcal{S}_{\mathcal{B}}$

Let the system be initially at the state space, $\mathcal{S}_{\mathcal{A}}$. Once the system transits to one of the states in $\mathcal{S}_{\mathcal{B}}$ for the first time, the elapsed time from the system initiation until such a transition is the time to first channel unavailability. Therefore, in order to determine the distribution of the time until the first channel unavailability, we consider that all the states in $\mathcal{S}_{\mathcal{B}}$ as a *single absorbing state*. Now, consider a new CTMC in which all the states in $\mathcal{S}_{\mathcal{B}}$ are lumped into a single absorbing state $L+1$. Let the stochastic process $\{X(t) \in \mathcal{S} | t \in \mathbb{R}\}$ be a homogeneous Markov process defined on a discrete and finite state space $\mathcal{S} = \{1, 2, \cdots, L, L+1\}$. Therefore, in the new CTMC, $\mathcal{S}_{\mathcal{A}} = \{1, 2, \cdots, L\}$ and $\mathcal{S}_{\mathcal{B}} = \{L+1\}$. Since state $L+1$ is absorbing and all other states are transient, the generator matrix of the Markov chain can be written as,

$$\boldsymbol{Q}' = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{A}^0 \\ \boldsymbol{0} & 0 \end{bmatrix}. \tag{B.7}$$

In (B.7), $\boldsymbol{A}^0$ is a column vector such that $\boldsymbol{A}\boldsymbol{U}_L + \boldsymbol{A}^0 = \boldsymbol{0}_L$ where $\boldsymbol{0}_L$ denotes the column vector of L zeros. Let T be a random variable representing the time until the CTMC reaches the absorbing state. It is said that T follows a continuous phase-type (PH) distribution with the representation $(\boldsymbol{P}_A(0), \boldsymbol{A})$ [28]. For PH random variables, the probability distribution of the time until absorption in the Markov chain is given by $F_T(t) = 1 - \boldsymbol{P}_A(0)e^{\boldsymbol{A}t}\boldsymbol{U}_L$, and correspondingly the probability density function is expressed as,

$$f_T(t) = \frac{d}{dt}F_T(t) = \boldsymbol{P}_A(0)e^{\boldsymbol{A}t}\boldsymbol{A}^0. \tag{B.8}$$

The r^{th} moment about the origin of the time until absorption (time to system failure), m'_r, is given by

$$m'_r = E[T^r] = r!\boldsymbol{P}_A(0)(-\boldsymbol{A})^{-r}\boldsymbol{U}_L,$$

where $E[\mathcal{X}]$ denotes the *expected value* or *mean* of the continuous-type random variable $\mathcal{X}$. When $r = 1$, the mean value, i.e., m'_1, can be obtained as follows.

$$m'_1 = E[T] = \boldsymbol{P}_A(0)(-\boldsymbol{A})^{-1}\boldsymbol{U}_L. \tag{B.9}$$

To determine the mean time until the CTMC reaches the absorbing state, (B.9) can be used and it gives the basis for determining the system times introduced in the next subsection. Alternatively, Laplace transforms can also be used to derive (B.9), as presented in our earlier work [19]. In this paper, however, we adopt a simpler method to obtain (9) by means of lumping all absorbing states into a single state.

C. System Times and Steady State Channel Availability

In this subsection we derive analytical expressions for the dependability metrics introduced in Sec. II-B. In what follows, we give the mathematical definitions for the channel available and unavailable time in a CRN. Based on these time duration definitions, an expression for steady state channel availability is derived.

1) Mean time to first channel unavailability, $\bar{T}_{FF}$: It is defined as *the expected time from system initiation to the instant when a new user request is blocked by the system for the first time*. Denote now by $\boldsymbol{t} \equiv [t_1 \; t_2 \; \cdots \; t_L]$ the row vector representing the mean time until the CTMC reaches the absorbing state. Here t_i represents the mean total time spent in state i until absorption. The i-th row and j-th column element of $-\boldsymbol{A}^{-1}$ is the expected total time spent in state j during the time until absorption, given that the initial state is i. Then, by unconditioning the initial probabilities, we obtain

B

$$\boldsymbol{t} = \boldsymbol{P}_A(0)(-\boldsymbol{A})^{-1}, \tag{B.10}$$

given that the initial probability row vector of $\boldsymbol{P}_A(0) \equiv [P_1(0) \; P_2(0) \; \cdots \; P_L(0)]$. By solving (B.10), all the elements in $\boldsymbol{t}$ can be determined. Now, by summing up the elements in $\boldsymbol{t}$ (i.e., right multiplying by $\boldsymbol{U}_L$) we obtain

$$\bar{T}_{FF} = \boldsymbol{P}_A(0)(-\boldsymbol{A})^{-1}\boldsymbol{U}_L. \tag{B.11}$$

2) Mean time to channel unavailability, $\bar{T}_{TF}$: The mean time to channel unavailability is defined as *the average time interval from an instant when the channels are available for new SUs, to the next channel unavailable instant*. The channel available instant is assumed to be selected when the system is in the steady state. To calculate $\bar{T}_{TF}$, the steady-state probabilities of the state space are required. The steady state probability vector of $\mathcal{S}$, can be partitioned as $\boldsymbol{\pi} = [\boldsymbol{\pi}_A \; \boldsymbol{\pi}_B]$ where $\boldsymbol{\pi}_A \equiv [\pi_1 \; \pi_2 \; \cdots \pi_L]$ and $\boldsymbol{\pi}_B \equiv [\pi_{L+1} \; \pi_{L+2} \; \cdots \pi_N]$. In this subsection, our focus is to calculate the mean time to channel unavailability elapsed from an instant which is randomly chosen when the system is in a channel available state. Given that the system is in $\mathcal{S}_A$, the probability that it is in state i is $(\boldsymbol{\pi}_A\boldsymbol{U}_L)^{-1}\pi_i$ and correspondingly, we can show that, $\boldsymbol{P}_A(0) = (\boldsymbol{\pi}_A\boldsymbol{U}_L)^{-1}\boldsymbol{\pi}_A$. Then following (B.9), we obtain

$$\bar{T}_{TF} = \frac{\boldsymbol{\pi}_A(-\boldsymbol{A})^{-1}\boldsymbol{U}_L}{\boldsymbol{\pi}_A\boldsymbol{U}_L}. \tag{B.12}$$

3) Mean channel available time, $\bar{T}_{UT}$: Once the system transits to a state in $\mathcal{S}_{\mathcal{A}}$ from a state in $\mathcal{S}_{\mathcal{B}}$, the system becomes available. *The average duration that the system resides in $\mathcal{S}_{\mathcal{A}}$ before making a transition back to a state in $\mathcal{S}_{\mathcal{B}}$* is defined as the mean channel available time. A channel available time duration will start in state $j \in \mathcal{S}_{\mathcal{A}}$, if the system was in any state $i \in \mathcal{S}_{\mathcal{B}}$ and then made a transition from state i to state j. Correspondingly, the probability of initiating a channel available time in state j can be expressed as,

$$P_j = \frac{\sum_{i=L+1}^{N} \pi_i q_{ij}}{\sum_{h=1}^{L}\{\sum_{i=L+1}^{N} \pi_i q_{ih}\}}, \tag{B.13}$$

where q_{ij} is the instantaneous transition rate from state i to state j. This result implies that

$$\boldsymbol{P}_A(0) = \frac{\boldsymbol{\pi}_B \boldsymbol{C}}{\boldsymbol{\pi}_B \boldsymbol{C} \boldsymbol{U}_L}. \tag{B.14}$$

Since $\boldsymbol{\pi}\boldsymbol{Q} = \boldsymbol{0}$, we have $\boldsymbol{\pi}_A\boldsymbol{A} + \boldsymbol{\pi}_B\boldsymbol{C} = \boldsymbol{0}$. Furthermore, as the sum of the elements in every row in $\boldsymbol{Q}$ is zero, we have $\boldsymbol{A}\boldsymbol{U}_L + \boldsymbol{B}\boldsymbol{U}_{N-L} = \boldsymbol{0}_L$. Therefore, we obtain

$$\boldsymbol{P}_A(0) = \frac{\boldsymbol{\pi}_A \boldsymbol{A}}{\boldsymbol{\pi}_A \boldsymbol{A} \boldsymbol{U}_L} = \frac{-\boldsymbol{\pi}_A \boldsymbol{A}}{\boldsymbol{\pi}_A \boldsymbol{B} \boldsymbol{U}_{N-L}}.$$

By substituting the above expression of $\boldsymbol{P}_A(0)$ into (B.9), we obtain

$$\bar{T}_{UT} = \frac{-\boldsymbol{\pi}_A \boldsymbol{A}(-\boldsymbol{A})^{-1} \boldsymbol{U}_L}{\boldsymbol{\pi}_A \boldsymbol{B} \boldsymbol{U}_{N-L}} = \frac{\boldsymbol{\pi}_A \boldsymbol{U}_L}{\boldsymbol{\pi}_A \boldsymbol{B} \boldsymbol{U}_{N-L}}. \tag{B.15}$$

4) Mean channel unavailable time, $\bar{T}_{DT}$: Once the system transits to a state in state space $\mathcal{S}_{\mathcal{B}}$ from a state in state space $\mathcal{S}_{\mathcal{A}}$, the system becomes unavailable. *The average time duration during which the system resides in $\mathcal{S}_{\mathcal{B}}$ before making a transition back to a state in $\mathcal{S}_{\mathcal{A}}$* is defined as the mean channel unavailable time. By following a similar procedure as in the previous $\bar{T}_{UT}$ calculation, we can derive that, $\bar{T}_{DT} = \frac{\boldsymbol{\pi}_B \boldsymbol{U}_{N-L}}{\boldsymbol{\pi}_B \boldsymbol{C} \boldsymbol{U}_L}$. Since we can show that, $\boldsymbol{\pi}_B \boldsymbol{C} \boldsymbol{U}_L = \boldsymbol{\pi}_A \boldsymbol{B} \boldsymbol{U}_{N-L}$, $\bar{T}_{DT}$ can be expressed as

$$\bar{T}_{DT} = \frac{\boldsymbol{\pi}_B \boldsymbol{U}_{N-L}}{\boldsymbol{\pi}_A \boldsymbol{B} \boldsymbol{U}_{N-L}}. \tag{B.16}$$

5) Steady state channel availability, A_{ss}: As already stated in (B.3), *the steady state availability is equal to the mean uptime divided by the summation of the mean uptime and the mean downtime*. Since $\bar{T}_{UT}$ and $\bar{T}_{DT}$ are analogous to the mean uptime and the mean downtime respectively, we show straightforwardly that

$$A_{ss} = \frac{\bar{T}_{UT}}{\bar{T}_{UT} + \bar{T}_{DT}} = \frac{\boldsymbol{\pi}_A \boldsymbol{U}_L}{\boldsymbol{\pi}_A \boldsymbol{U}_L + \boldsymbol{\pi}_B \boldsymbol{U}_{N-L}}. \tag{B.17}$$

Following the normalization equation in (B.4),

$$\boldsymbol{\pi}_A \boldsymbol{U}_L + \boldsymbol{\pi}_B \boldsymbol{U}_{N-L} = 1.$$

Therefore, (B.17) is simplified as

$$A_{ss} = \boldsymbol{\pi}_A \boldsymbol{U}_L. \tag{B.18}$$

It is obvious that in order to calculate the above performance measures, we need to obtain the sets of channel available and unavailable states, the transition rate matrix and and their stationary probabilities, $\boldsymbol{\pi}$, of the system. It is worth mentioning that the procedure to obtain channel available and unavailable states and $\boldsymbol{\pi}$, does not rely on the adopted channel access scheme, as long as the PU/SU activity model is Markovian. An example which shows how to determine $\bar{T}_{DT}$ in a single channel CRN is illustrated in the Appendix.

V. CHANNEL AVAILABLE TIME DISTRIBUTIONS

In the previous section we analyzed the steady state channel availability and several system times corresponding to the commonly used metrics in dependability studies. Moreover, channel availability can also be considered as an appropriate measure for QoS analysis in CR systems since the profit which the SN can gain depends on the channel availability level. However, the steady state availability alone is not sufficient when there is an associated penalty if *a pre-specified level of availability over a finite time interval is not met* [29]. In such a situation, the probability that the CRN cannot meet a specified channel availability level needs to be calculated. This observation triggered our motivation for further determining the distribution function of channel availability.

Note that the *interval availability* over a period $(0, t)$ is defined as *the fraction of time in which the system is in operation during* $(0, t)$, and it is another useful metric related to availability [26]. Hereafter, we present the details of a numerical approach which can determine the complementary cumulative distribution function of the channel availability in a CRN by computing *the probability of gaining an interval availability greater than a certain value,* p. Before presenting that approach, we briefly introduce a solution technique known as *uniformization* which can be used to numerically calculate the distribution of the channel availability during a finite observation period.

A. Uniformization Technique

Studying the transient regime of a CTMC involves solving a system of linear first-order differential equation, or equivalently evaluating a matrix exponential. Direct evaluation of a matrix exponential is not considered as a practical method to solve dependability models at the transient phase due to its high computation cost [30]. Alternatively, the *uniformization* or *randomization* method has been introduced to perform transient analysis of finite state CTMCs in many studies [31], [32]. The uniformization process can be applied to any discrete state Markov process with a bounded transition rate matrix $\boldsymbol{Q}$ and it reduces a CTMC to an equivalent discrete-time Markov chain (DTMC) subordinate to a Poisson process. In this method, the transition probability matrix $\boldsymbol{P}$ of the uniformized Markov chain is given by $\boldsymbol{P} = I + {}^{\boldsymbol{Q}}/_{\Delta}$, where Δ is referred to as the *uniformization rate* which is equal to or greater than the largest magnitude of the diagonal elements of $\boldsymbol{Q}$, i.e., $\Delta \geq \max\{|q_{ii}|, i \in \mathcal{S}\}$ and I denotes the identity matrix of the same size as $\boldsymbol{Q}$. For more details about uniformization, read [27] and [33]. In the next subsection we apply the uniformized Markov chain to calculate the interval availability distribution.

B

B. Interval Availability Distribution

The states of a system at time t can be described by a CTMC, $X = \{X_t, t \geq 0\}$, over a discrete state space $\mathcal{S}$ [30]. The state space is assumed to be finite and of size N. As already mentioned in Sec. IV-A, the total state space can be divided into two disjoint subsets, $\mathcal{S}_{\mathcal{A}}$ and $\mathcal{S}_{\mathcal{B}}$. Let $O(t)$ be the cumulative amount of channel available time during $(0, t)$. We can express $O(t)$ as

$$O(t) = \int_0^t r(s)\, ds, \text{ where } r(s) \triangleq \begin{cases} 1, & \text{if } X_s \in \mathcal{S}_{\mathcal{A}} \\ 0, & \text{otherwise.} \end{cases} \tag{B.19}$$

Then the interval availability of channel available time over $(0, t)$, $A_{int}(t)$, can be expressed as $A_{int}(t) = \frac{O(t)}{t}$. Consider that n transitions occurred during $(0, t)$ and therein the uniformized Markov chain visited $\mathcal{S}_{\mathcal{A}}$, k times, where $0 \leq k \leq n + 1$. Note that k can be equal to $n + 1$ since initially the system may reside in one of the channel available states before any transition occurs. Let us now briefly recall how the cumulative distribution function of the interval availability is derived [33], [35].

Consider a Poisson counting process, $\{N(t), t \geq 0\}$, split into two processes $\{N_1(t); t \geq 0\}$ and $\{N_2(t), t \geq 0\}$: each arrival in $\{N(t), t \geq 0\}$ is sent to the first process with probability p and to the second process with probability $1 - p$ (see Fig. B.6). Out of the first n arrivals to the combined process, the probability that

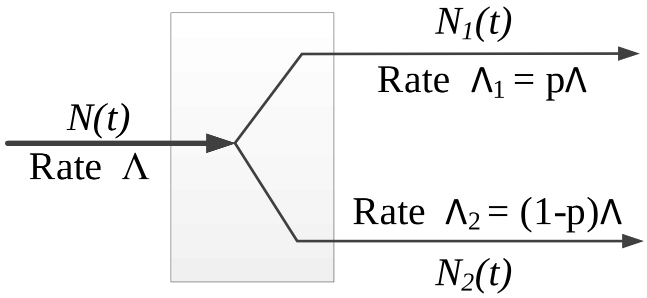


Figure B.6: Subdividing a Poisson process into two processes where each arrival is independently sent to process $N_1(t)$ with probability p and to process $N_2(t)$ with probability $1-p$.

B

k or more arrivals occur to the first process is given by $\sum_{i=k}^{n}\binom{n}{i}p^i(1-p)^{n-i}$ [34]. Consider that the above mentioned first and second processes are the uniformized Markov chain visits to $\mathcal{S}_\mathcal{A}$ and $\mathcal{S}_\mathcal{B}$ respectively. By following these properties of the Poisson process, it can be shown that the following conditional probability holds

$$P\left(A_{int}(t)\leq p\middle|\begin{array}{c}n \text{ state changes in } (0,t)\\ k \text{ visiting occurrences to } \mathcal{S}_\mathcal{A}\end{array}\right)=\sum_{i=k}^{n}\binom{n}{i}p^i(1-p)^{n-i}, \quad \text{(B.20)}$$

where $\binom{n}{k}=\frac{n!}{k!(n-k)!}$, and $p<1$ is a pre-specified value of the required channel availability for the system. Hereafter, we refer to the parameter, p, as the *guaranteed level* of channel availability. To remove the condition on the number of visiting occurrences to the states of $\mathcal{S}_\mathcal{A}$, define $Y_{n,k}$, $0\leq k\leq n+1$ to be the probability that the uniformized Markov chain visits $\mathcal{S}_\mathcal{A}$, k times before the $(n+1)^{th}$ transition occurs. Unconditioning on the number of visits to $\mathcal{S}_\mathcal{A}$, we have

$$P(A_{int}(t)\leq p|\ n \text{ state changes in } (0,t)\ )=\sum_{k=0}^{n}Y_{n,k}\sum_{i=k}^{n}\binom{n}{i}p^i(1-p)^{n-i}. \quad \text{(B.21)}$$

When uniformization is applied to the CTMC, we have a Poisson process with rate equal to the uniformization rate Δ. Considering this Poisson process, we can further perform unconditioning on the number of transitions in $(0,t)$ since parameter n is a counting process in $(0,t)$. Then we obtain,

$$P(A_{int}(t)\leq p)=\sum_{n=0}^{+\infty}e^{-\Delta t}\frac{(\Delta t)^n}{n!}\sum_{k=0}^{n}Y_{n,k}\sum_{i=k}^{n}\binom{n}{i}p^i(1-p)^{n-i}. \quad \text{(B.22)}$$

It is obvious that the infinite sum in (B.22) cannot be implemented for the numerical evaluation. However an approximate solution with adequate accuracy is sufficient for CCDF evaluation. Therefore, we truncate the above series by limiting the outer summation to $N_c + 1$ terms such that,

$$P(A_{int}(t) \leq p) = err(N_c) + \sum_{n=0}^{N_c} e^{-\Delta t} \frac{(\Delta t)^n}{n!} \sum_{k=0}^{n} Y_{n,k} \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}, \tag{B.23}$$

where, $err(N_c)$ is estimated as the first term left out of the summation, i.e., the term $n = N_c + 1$, and that N_c has been chosen so that $err(N_c)$ is sufficiently small (with $err(N_c) \leq 10^{-4}$) and it is given as

$$err(N_c) = \sum_{n=N_c+1}^{+\infty} e^{-\Delta t} \frac{(\Delta t)^n}{n!} \sum_{k=0}^{n} Y_{n,k} \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}. \tag{B.24}$$

Consequently, we can deduce that

$$P(A_{int}(t) \leq p) \approx \sum_{n=0}^{N_c} e^{-\Delta t} \frac{(\Delta t)^n}{n!} \sum_{k=0}^{n} Y_{n,k} \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}. \tag{B.25}$$

In this study, N_c is selected in an iterative way by calculating $P(A_{int}(t) \leq p)$ corresponding to certain values of N_c and $N_c + 1$. Once we obtain $P_{N_c+1} - P_{N_c} \leq 10^{-4}$, the calculation process is terminated and the corresponding N_c is selected. Here P_{N_c} denotes the value of $P(A_{int}(t) \leq p)$ after the series is truncated into N_c terms.

C. Calculation of $Y_{n,k}$ and CCDF of Channel Availability

Based on (B.25), the only parameter left to calculate $P(A_{int}(t) \leq p)$ is $Y_{n,k}$. In this study, we revise the approach proposed in [35] to recursively determine $Y_{n,k}$. Firstly, we define $\boldsymbol{Y}^{\mathcal{S}}_{n,k}$ as a row vector of N entries where N is the size of the total state space $\mathcal{S}$. The i^{th} entry of $\boldsymbol{Y}^{\mathcal{S}}_{n,k}$ is the probability that the uniformized Markov chain visits the states of $\mathcal{S}_{\mathcal{A}}$, k times during its first n transitions and the n^{th} transition occurs into state i. We partition $\boldsymbol{Y}^{\mathcal{S}}_{n,k}$ according to the subsets $\mathcal{S}_{\mathcal{A}}$ and $\mathcal{S}_{\mathcal{B}}$ as $\boldsymbol{Y}^{\mathcal{S}}_{n,k} = [\boldsymbol{Y}^{\mathcal{S}_{\mathcal{A}}}_{n,k}, \ \boldsymbol{Y}^{\mathcal{S}_{\mathcal{B}}}_{n,k}]$ where $\boldsymbol{Y}^{\mathcal{S}_{\mathcal{A}}}_{n,k}$ is a row vector with L entries and $\boldsymbol{Y}^{\mathcal{S}_{\mathcal{B}}}_{n,k}$ is a row vector with $N - L$ entries. Let $\boldsymbol{P}(0)$ represent the initial state probability vector such that $\boldsymbol{P}(0) = [\boldsymbol{P}_A(0), \ \boldsymbol{P}_B(0)]$ where $\boldsymbol{P}_A(0)$ and $\boldsymbol{P}_B(0)$ have the same meaning as mentioned in Sec. IV-A. Then the following initial conditions can be formulated,

$$\boldsymbol{Y}^{\mathcal{S}_{\mathcal{A}}}_{0,1} = \boldsymbol{P}_A(0), \ \ \boldsymbol{Y}^{\mathcal{S}_{\mathcal{B}}}_{0,0} = \boldsymbol{P}_B(0), \ \ \boldsymbol{Y}^{\mathcal{S}_{\mathcal{A}}}_{0,0} = \boldsymbol{0}, \ \ \boldsymbol{Y}^{\mathcal{S}_{\mathcal{B}}}_{0,1} = \boldsymbol{0}. \tag{B.26}$$

For instance, $\boldsymbol{Y}_{0,1}^{\mathcal{S}_{\mathcal{A}}}$ in (B.26) is the probability that the uniformized Markov chain visits one channel available state without any state transition occurred (i.e., $k = 1$ and $n = 0$). This could happen only if the system resides in one of the channel available states initially. Thus, $\boldsymbol{Y}_{0,1}^{\mathcal{S}_{\mathcal{A}}}$ equals to the initial probability vector, $\boldsymbol{P}_A(0)$. The following recursion formulas are used to calculate the whole set of $\boldsymbol{Y}_{n,k}^{S}$.

$$\boldsymbol{Y}_{n,k}^{\mathcal{S}_{\mathcal{A}}} = \boldsymbol{Y}_{n-1,k-1}^{\mathcal{S}_{\mathcal{A}}}\boldsymbol{A'} + \boldsymbol{Y}_{n-1,k-1}^{\mathcal{S}_{\mathcal{B}}}\boldsymbol{C'}, \tag{B.27}$$

$$\boldsymbol{Y}_{n,k}^{\mathcal{S}_{\mathcal{B}}} = \boldsymbol{Y}_{n-1,k}^{\mathcal{S}_{\mathcal{A}}}\boldsymbol{B'} + \boldsymbol{Y}_{n-1,k}^{\mathcal{S}_{\mathcal{B}}}\boldsymbol{D'}, \tag{B.28}$$

where $\boldsymbol{A'}$, $\boldsymbol{B'}$, $\boldsymbol{C'}$ and $\boldsymbol{D'}$ are the sub-matrices of the uniformized Markov chain corresponding to (B.5) in Sec. IV-A.

In order to understand the reasoning behind the above mentioned recursive formulas, we analyze (B.27) as an example. The i-th element of $\boldsymbol{Y}_{n,k}^{\mathcal{S}_{\mathcal{A}}}$ in (B.27) is the probability of the uniformized Markov chain being k times in channel available (operational) states out of n transitions in total, and the state visited after the last transition is the i-th state of $\mathcal{S}_{\mathcal{A}}$. The last transition can occur from a state either within $\mathcal{S}_{\mathcal{A}}$ or from $\mathcal{S}_{\mathcal{B}}$. If it occurs from $\mathcal{S}_{\mathcal{A}}$, the corresponding transition probabilities are given by $\boldsymbol{A'}$ and if it occurs from a state in $\mathcal{S}_{\mathcal{B}}$, the corresponding transition rate is $\boldsymbol{C'}$. In both cases, the number of transitions increases by one due to the state transition. Correspondingly, the number of visited operational states increases by one since the transition is into $\mathcal{S}_{\mathcal{A}}$. In contrast, in (B.28), the sub-index n is increased by one, and k is not increased. The reason is that the n^{th} transition occurred into a channel unavailable state in $\mathcal{S}_{\mathcal{B}}$. Once $\boldsymbol{Y}_{n,k}^{\mathcal{S}}$ is determined, we obtain

$$Y_{n,k} = \boldsymbol{Y}_{n,k}^{\mathcal{S}}\boldsymbol{U}_N, \tag{B.29}$$

where $\boldsymbol{U}_N$ denotes the column vector of N ones. After the above calculation, $P(A_{int}(t) \leq p)$ in (B.25) can be deduced. Finally, the CCDF of the channel availability given by $A(t,p)$ is obtained by

$$A(t,p) = P(A_{int}(t) > p) = 1 - P(A_{int}(t) \leq p). \tag{B.30}$$

VI. Numerical Results and Discussions

In this section we evaluate numerically each dependability measure presented in Sec. IV and Sec. V based on the proposed analytical models. The numerical results obtained from the analytical models are validated through extensive simulations which are performed by using a custom-built discrete-event simulator based on MATLAB. The centralized CRN topology described in Sec. III is considered.

Table B.1: Four initial conditions used for performance evaluation of *DFA* (the set of feasible states of the system is denoted as $\mathcal{S} = (y, x_1, x_2, x_3)^\dagger$ where $M = 6$, $W = 1$ and $V = 3$.)

Initial condition	*DFA* states, (y, x_1, x_2, x_3)
C1: All channels are idle	$(0, 0, 0, 0)$
C2: One PU service and one SU service are in the system	$(1, 0, 0, 1)$
C3: System can allocate at most W channels upon a new SU arrival	$(y, x_1, 1, 0)^\ddagger$ and $(5, 0, 0, 0)$
C4: Only one PU service is in the system	$(1, 0, 0, 0)$

† Here y denotes the number of PU services each with single channel occupancy and x_k denotes the number of SU services with $k(1 \leq k \leq 3)$ aggregated channels.
‡ There are five states which satisfy $y + x_1 = 4$: $(0, 4, 1, 0)$, $(1, 3, 1, 0)$, $(2, 2, 1, 0)$, $(3, 1, 1, 0)$ and $(4, 0, 1, 0)$.

The network has $M = 6$ channels and the default values of the parameters mentioned in Sec. III-A are configured as $\lambda_P = 1.0,\ \lambda_S = 2.0,\ \mu_P = 0.5$ and $\mu_S = 1.0$ respectively. The units of these parameters could be services or flows per unit of time. *DFA* or *DFA*$_H$ explained in Sec. III-B is employed as the channel access scheme, with $W = 1$ and $V = 3$. In order to analyze the system from different angles, various initial conditions of the system are configured. The initial system conditions adopted in this study are listed in Table B.1. Unless otherwise stated, *initial condition C1* is adopted in numerical evaluations. In Subsec. VI-K, *DFA*$_H$ is employed in *Scenario II* whereas all other subsections in Sec. VI are dedicated to analyze *Scenario I* based on *DFA*.

B

A. Steady State Channel Availability

Fig. B.7 shows the steady state channel availability variation as a function of SUs' arrival rate when the offered SU traffic load, i.e., the ratio $\frac{\lambda_S}{\mu_S}$, is constant. This figure illustrates A_{SS} for two different offered traffic values. The results in Fig. B.7 show that higher channel availability is achieved in the CRN when the offered traffic of SUs is low. A low offered traffic load implies that the channels are less likely to be occupied by existing users. As a result, the new SU arrivals have more channel access opportunities. Therefore it is evident that A_{SS} holds a comparatively high value when $\frac{\lambda_S}{\mu_S}$ is low. However, the channel availability decreases with a higher SU arrival rate although the offered traffic load is kept constant. Given that the offered load, i.e., λ_S/μ_S, is constant, SUs would utilize the amount of resources (measured as mean number of occupied channels) if they had unconstrained access to the channels.

However, the number of channels is limited and SUs can only access the system resources when those are not occupied by PUs. As a consequence of keeping the offered load constant but with more frequent requests (a higher λ_S) and shorter service

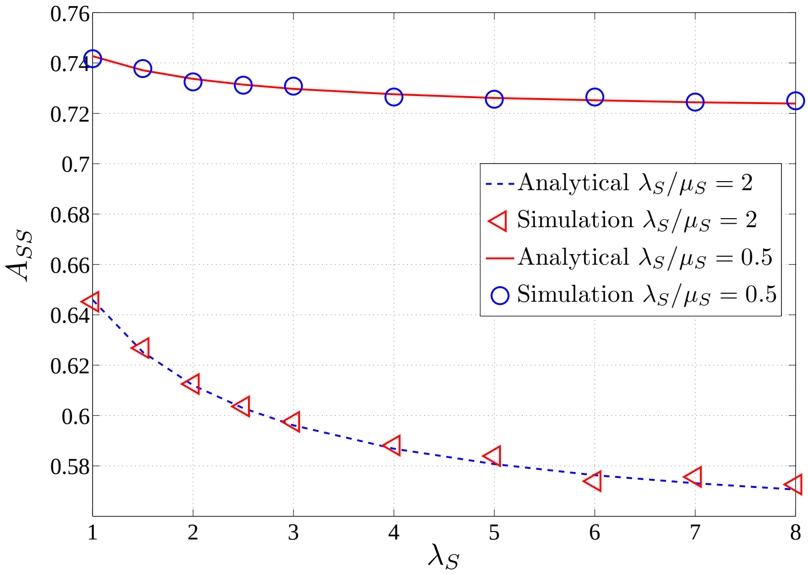


Figure B.7: A_{SS} as a function of λ_S while keeping the offered traffic, ${}^{\lambda_S}/_{\mu_S}$, constant.

B

duration (a higher μ_s), the likelihood that SUs complete their services before they are forcibly terminated is increased. Thus a lower forced termination probability is achieved. This reduction of the forced termination probability clearly increases the carried load by the system, leading to lower A_{SS}. At higher offered traffic loads, the above mentioned observation is more evident as can be observed from Fig. B.7.

To validate the correctness of the mathematical analysis, we present the simulation results obtained for channel availability measures in Figs. B.7-B.9 and Fig. B.12 together with the analytical results. It is worth mentioning that the simulations are performed independently of the developed analytical models. In other words, the calculation of the system times and availability values in our simulations is not dependent on the derived mathematical expressions at all, nor are the state transition tables used in these calculations. From these curves, we observe that the results from the analytical model coincide precisely with the obtained simulation results. Later on in Sec. VI-G, we further check the preciseness of the model by employing different PU interarrival time and SU service time distributions.

B. Mean Channel Available Time

Next we investigate the impact of the PU arrival rate on mean channel available time, $\bar{T}_{UT}$, which corresponds to MUT in classical reliability analysis. Fig. B.8 shows the mean channel available time variation as a function of PUs' arrival rate when the PU offered traffic load, i.e., the ratio $\frac{\lambda_P}{\mu_P}$, is constant. As illustrated in this figure, the mean channel available time decreases as the offered traffic load becomes

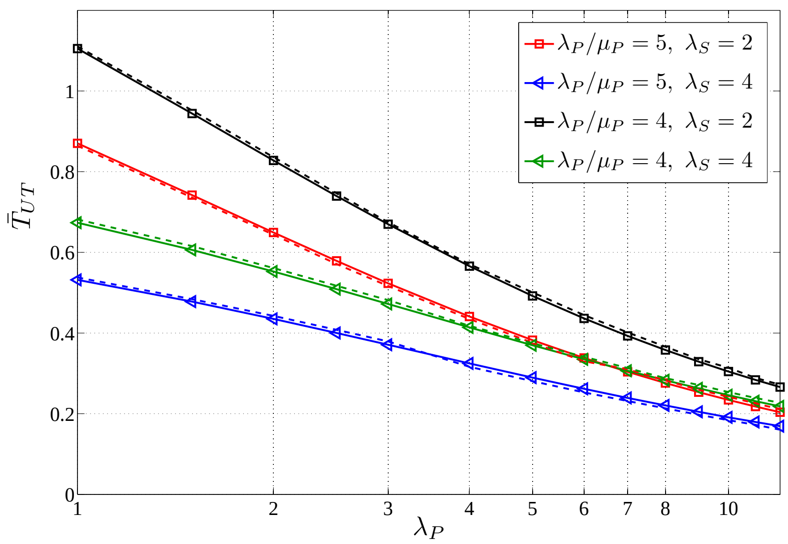


Figure B.8: $\bar{T}_{UT}$ as a function of λ_P while keeping the offered traffic, λ_P/μ_P, constant. The dashed lines indicate the simulation results.

higher. The reasons for this behavior are explained below. When the offered traffic load is high, the system has a large number of active PU services. The time duration in which a particular channel being in the idle state, i.e., the available time of the channel, becomes shorter when PUs become more active. The results also clearly demonstrate that the channel available time duration becomes longer when λ_S is at a lower value due to the same reason. The reason for the descending trend of $\bar{T}_{UT}$ with a larger λ_P value can be explained in a similar way as described in the previous subsection.

C. Mean Channel Unavailable Time

The mean downtime of a system is a critical parameter since long downtime duration causes dissatisfaction of customers. The mean channel unavailable time as a function of the PU service rate is plotted in Fig. B.9, given that the steady state channel availability of the system is set as 0.8 by adjusting the SU arrival rate, λ_S. From this figure, we observe a continuous descent in the channel unavailable time as the service rate of PUs increases while keeping the arrival rate of PUs constant. When the PU service rate is low, fewer PUs finish their services per unit of time. Then, most of the channels are occupied by respective PUs for a longer period of time. That means, the time interval that the system will reside in the channel unavailable state space is comparatively long. It is clear that when the μ_P is high, the probability of being in the space $\mathcal{S}_B$ is low, thus the value of $\boldsymbol{\pi}_B \boldsymbol{U}_{N-L}$ in (B.16) becomes low, resulting in a shorter $\bar{T}_{DT}$.

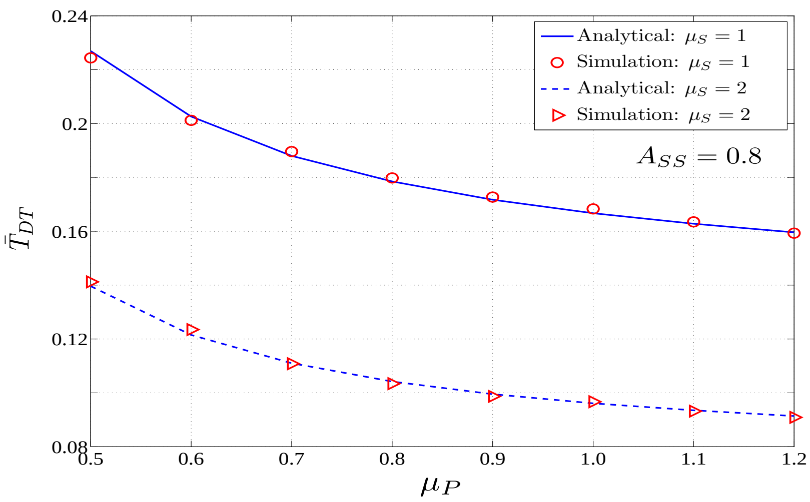


Figure B.9: $\bar{T}_{DT}$ as a function of μ_P when $A_{ss} = 0.8$.

B

On the other hand, at a higher PU service rate, a comparatively large number of PU services can be finished during a unit of time. Thus, the time period that the system will reside in the channel unavailable state space becomes shorter with a higher μ_P. For this reason, the mean channel unavailable time becomes shorter when PUs finish their services faster, as shown in Fig. B.9. Additionally, for a given μ_P value, the channel unavailable time is longer at a lower SU service rate compared with a larger SU service rate. Again, this is due to the fact that the state holding time of the system being in a busy state decreases with a higher μ_S.

D. Mean Time to First Channel Unavailability

In Fig. B.10, the plots indicate the mean time to first channel unavailability corresponding to the three initial conditions indicated in Table B.1. At *C1*, all channels are in the idle state while at *C2*, fewer channels are available in the beginning. At *C3*, the system can allocate merely $W = 1$ channel upon a new SU arrival. Thus, this initial condition, *C3*, represents a sort of *worst-case* scenario. As expected, when PUs arrive more often, all $\bar{T}_{FF}$ curves decrease monotonically. This is because that at lower PU arrival rates, more channels are likely to be idle and the newly arrived user requests can be accommodated with the required number of channels. Therefore, the network can operate without blocking any new users over a longer period. This result is in sharp contrast with the result under the high PU arrival rate circumstances.

As mentioned earlier, the time taken for the first channel unavailability is depending on the initial state of the system. If the system does not have any existing services in the beginning, $\bar{T}_{FF}$ will last longer compared with the system with al-

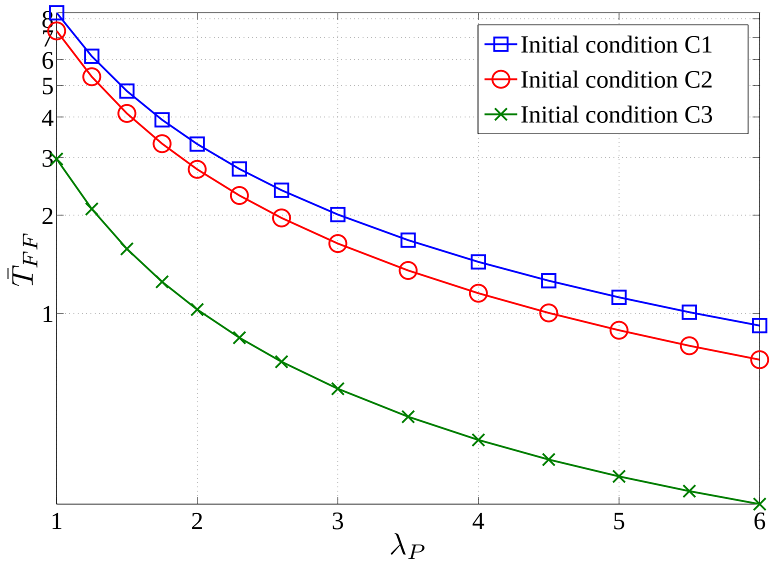


Figure B.10: $\bar{T}_{FF}$ as a function of λ_P. For the definitions of initial conditions C1, C2 and C3, refer to Table B.1.

B

ready commenced PU and/or SU services, as illustrated in Fig. B.10. When a single PU service and a single SU service already exist in a CRN with 6 channels, 4 channels are already occupied at the system initialization in *DFA* since $V = 3$. In other words, only 2 idle channels are available in the network for new users. As a consequence, there is a higher probability that this system will move to a channel unavailable state within a shorter period of time in comparison with the system with 6 idle channels in the initial phase.

E. Mean Time to Channel Unavailability

Fig. B.11 presents the results for the mean time to channel unavailability, $\bar{T}_{TF}$, with respect to the arrival rate of SUs when $A_{SS} = 0.8$. As expected, the mean time for the next system failure (channel unavailability) also shows a descending trend with a higher SU arrival rate. For a given steady state channel availability, the higher the number of SU arrivals to the system, the higher the channel occupancy. As a consequence of high channel occupancy, the remaining time in the channel available state space will reduce. In other words, at higher channel occupancy, the mean total time spent in $\mathcal{S}_A$ and the steady state probabilities in $\boldsymbol{\pi}_A$ show lower values. Then the numerical value of the term $\boldsymbol{\pi}_A(-\boldsymbol{A})^{-1}\boldsymbol{U}_L$ in (B.12) is lower, leading to the decreasing trend of $\bar{T}_{TF}$ with a larger λ_S. In order to investigate the impact of PU arrival rate on the behavior of $\bar{T}_{TF}$, Fig. B.11 depicts $\bar{T}_{TF}$ for two different PU arrival rates. As can be observed from this figure, $\bar{T}_{TF}$ further reduces when more PUs arrive.

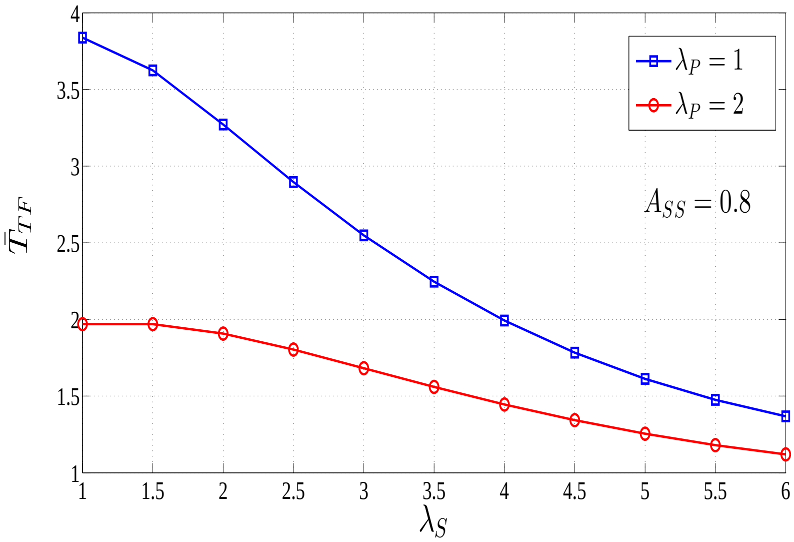


Figure B.11: $\bar{T}_{TF}$ as a function of λ_S when $A_{ss} = 0.8$.

F. Channel Availability Improvement via a Queuing Scheme

So far we investigate the channel availability as a reliability metric by considering the CRN as a *loss system*. In this subsection we adopt a queuing scheme which is proposed in our earlier work in [36] as a technique to improve channel availability of CRNs. We deploy a first in first out queue for the newly arrived secondary traffic flows mentioned in the *DFA* scheme. Note that this modified scheme is denoted as *DFAQ* in this paper for illustration convenience and the queue has a finite buffer size, Q_L. Using *DFA* discussed in Sec. III-B, a new SU request is blocked once the network is fully occupied.

However, with the proposed queuing scheme *DFAQ*, the new SU request which would be blocked by *DFA* is fed back into the buffer until the buffer is full. If the queue is fully occupied, then the new request is blocked. As we already mentioned, the derived analytical models in Sec. IV-C is independent of the channel access scheme since the derived mathematical expressions in that section do not depend on the parameters of the channel access scheme. Correspondingly, in order to analyze the steady state channel availability in *DFAQ*, we need only to determine the correspondent subsets, $\mathcal{S}_{\mathcal{A}}$ and $\mathcal{S}_{\mathcal{B}}$ of *DFAQ*, and evaluate the steady state probability vector $\boldsymbol{\pi}$ by using (B.4). Then, the summation of the steady state probabilities of $\mathcal{S}_{\mathcal{A}}$, i.e., $\sum_{i \in \mathcal{S}_{\mathcal{A}}} \pi_i$, equals to the steady state channel availability of *DFAQ* as given in (B.18).

Fig. B.12 depicts the comparison between the *DFA* scheme and the *DFAQ* scheme on the performance of steady state channel availability as a function of the SU arrival rate. From this figure, we observe a monotonic descent for channel availability

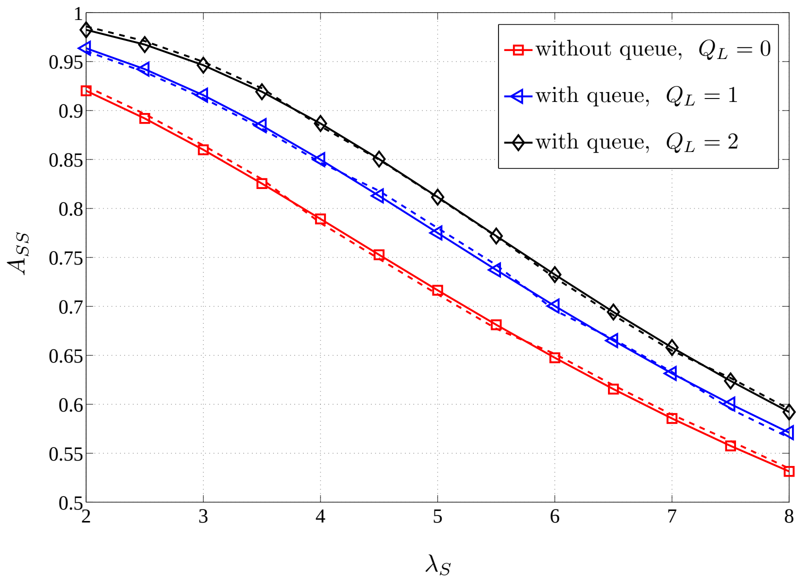


Figure B.12: A_{SS} as a function of λ_S for *DFA* and *DFAQ* schemes. The dashed lines indicate the simulation results.

as the arrival rate of SUs increases. Furthermore, the channel availability using *DFA* is lower than that of *DFAQ* for the same configuration. Clearly, the channel availability becomes higher using the queuing scheme since the blocked services due to insufficient channels are buffered into the queue until they are possibly offered with the required number of channels. With a larger queue size, more SUs can be queued instead of being blocked, resulting in higher channel availability. However, this higher channel availability is achieved at a cost of queuing delay [36]. Therefore, the buffer size should not be too large in order to minimize the associated queuing delay.

G. Steady State Channel Availability for Traffic Pattern with Log-normal Distribution

To further assess the preciseness as well as the applicability of the developed model which is based on the assumption of exponential service time and interarrival time (Poisson arrival process), we adopt another distribution and obtain the steady state availability by simulations. Log-normal distribution is a more realistic model for service times of real-life traffic patterns [37] as well for interarrival times. In Fig. B.13, the steady state availability is depicted as a function of the *PU arrival rate based on the log-normal distribution* with the same mean value and the variance as used in the original exponential distribution for service times.

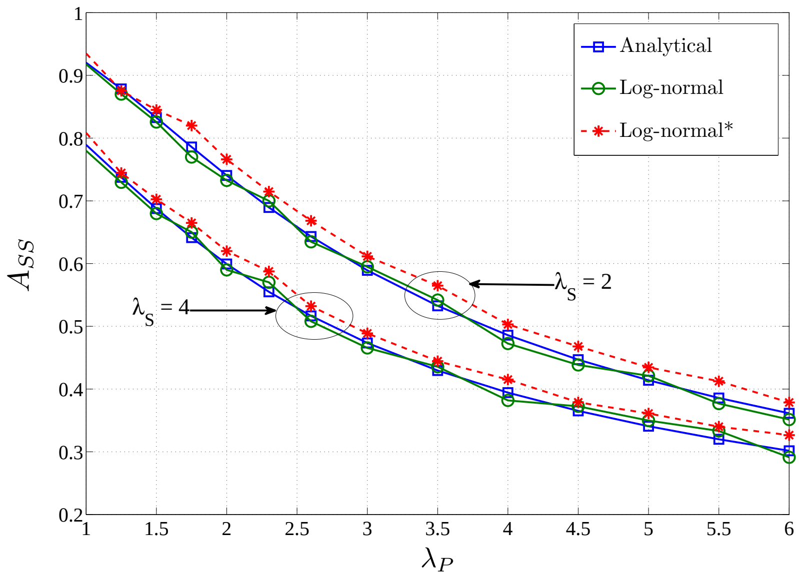


Figure B.13: Steady state availability, A_{SS}, as a function of the primary user arrival rate when the log-normal distribution is assumed for SU service time and PU interarrival time. The distributions of time sequences in the different experiments are listed in Table B.2.

Table B.2: The distribution of time sequences in the different experiments (cases) illustrated in Fig. B.13.

Case	PU ST	SU ST	PU IAT	SU IAT
Analytical	EXP	EXP	EXP	EXP
Log-normal	LN	LN	EXP	EXP
Log-normal*	EXP	EXP	LN	EXP

ST and *IAT* in this table denote service time and interarrival time respectively. *EXP* and *LN* denote the exponential distribution and log-normal distribution respectively.

For PU interarrival times, although the mean value of the log-normal distribution is kept the same as in the original exponential distribution, the variance is larger as its squared coefficient of variation ($SCV = {}^{\text{variance}}/_{\text{mean}^2}$) varies such that $1 \leq$ SCV ≤ 4.6. For comparison, we depict also the corresponding analytical results which are obtained based on the original exponential distribution in the same figure. The curves obtained from analyses and simulations match with each other precisely, as illustrated in Fig. B.13, confirming that the proposed model is robust and applicable to other service time distributions and non-Poisson PU arrival processes as well.

In the above subsections from VI-A to VI-G, we have investigated the dependability metrics related to system times and the *steady state availability* of a CRN.

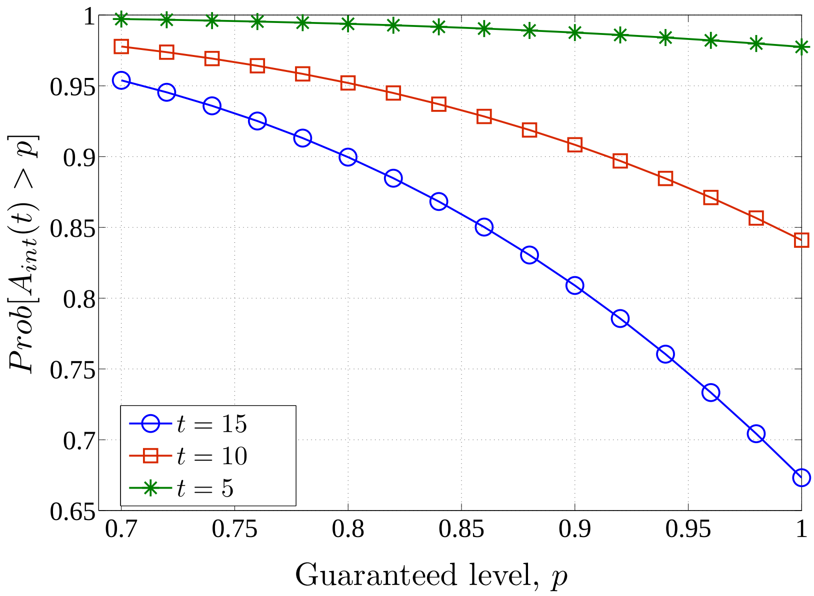


Figure B.14: Probability of achieving at least p channel availability as a function of the guaranteed level.

In the next two subsections, the CCDF of the *interval availability* is studied as a function of the guaranteed level, p, and the duration of the observation period, t, respectively. *The purpose of those two subsections is to analyze the CRN's ability to satisfy the requirements of the SN on channel availability over a given period.* The numerical results presented in the following subsections are obtained by selecting the truncation parameter introduced in (B.23) such that $P_{N_c+1} - P_{N_c} \leq 0.0001$.

H. Channel Availability Distribution as a Function of a Guaranteed Level

In Fig. B.14, the behavior of the CCDF for channel availability obtained by (B.30) is illustrated with different observation time periods where t represents the observation time in time units. As shown in the figure, the probability of spectrum availability for the SN for a guaranteed level monotonically decreases with a growing guaranteed level p. For instance, the probability that the spectrum will be available for more than 75% of the 15 time units during the observation period is around 0.93. However, this probability decreases to 0.81 if we need more than 90% of channel availability. In other words, the possibility of obtaining a higher channel availability level becomes lower when the system requires a higher availability level. Note that the probability of spectrum availability for higher than a specific guaranteed level also decreases when the duration of the observation period, i.e., t, becomes longer. The reason for this behavior can be explained as follows. If we look at an extreme case, i.e., an observation period of 0 time units, the entire spectrum is available for

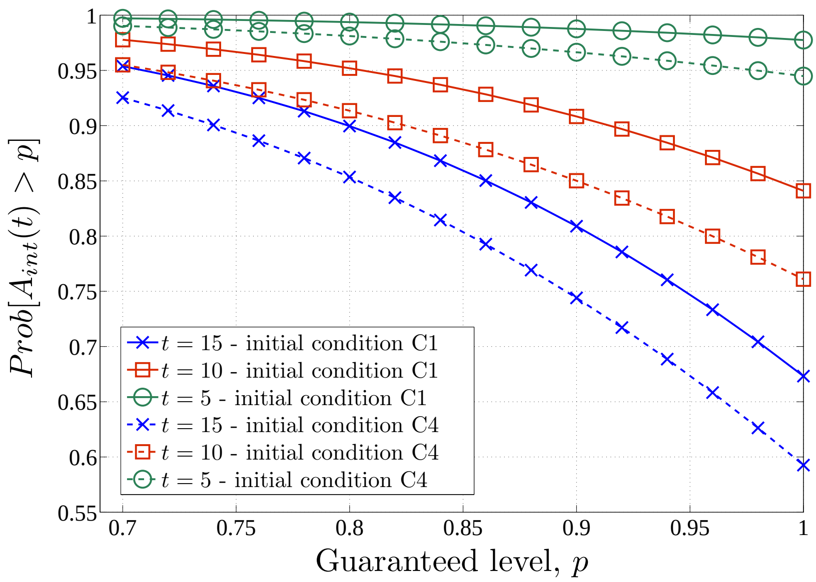


Figure B.15: Probability of achieving at least p channel availability as a function of the guaranteed level under different initial conditions. For the definitions of initial conditions C1, C2 and C3, refer to Table B.1.

SUs due to the initial condition of idle channels. As t lasts longer, however, part of the spectrum will likely be occupied by PUs and/or SUs, leading to lower channel availability.

In order to further investigate the above observation, each curve presented in Fig. B.14 is analyzed under two different initial conditions as shown in Fig. B.15. The solid lines in Fig. B.15 represent the CCDF at *C1* while the dashed lines illustrate the CCDF at *C4*. In contrast to *C1*, the channel availability for the SUs in *C4* is lower since one channel is already occupied in the beginning according to *C4*. Thus the probability of achieving channel availability higher than a specified guaranteed level under *C4* is always lower than that under *C1*.

I. Channel Availability Distribution as a Function of Observation Time

Furthermore, we show the probability of achieving a channel availability value which is greater than p, i.e., $P(A_{int}(t) > p)$, as a function of the duration of the observation period for different values of p. Basically two different types of behavior can be observed depending on whether the guaranteed level p can be reached in the asymptotic case ($p < A_{ss}$) or not ($p \geq A_{ss}$). Fig. B.16 and Fig. B.17 illustrate channel availability distribution as a function of the observation time under *C1*. In both illustrations *the steady state channel availability, A_{ss}, of the particular system is* fixed as 0.82.

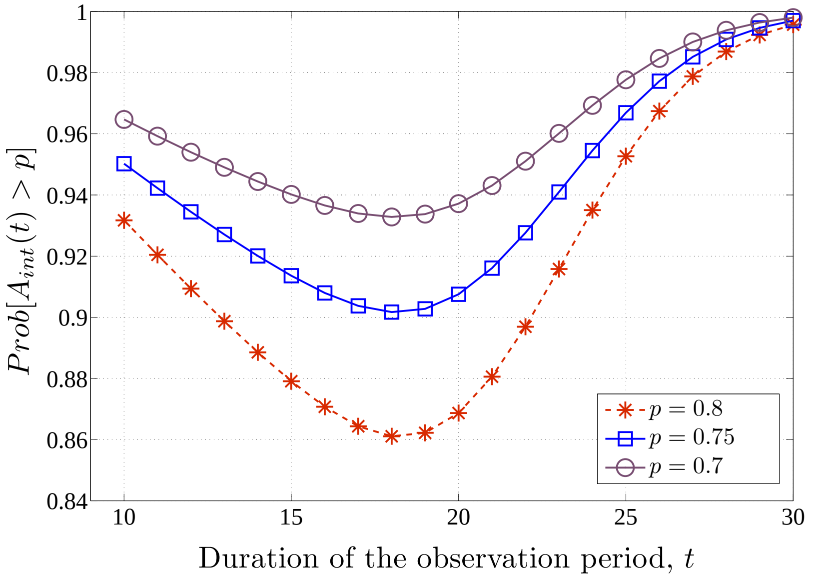


Figure B.16: Probability of achieving at least p channel availability when $p < A_{ss}$ as a function of observation period.

B

1) When $p < A_{ss}$: Fig. B.16 depicts $P(A_{int}(t) > p)$, as a function of the observation period when $p < A_{ss}$. As shown in those curves, the probability first decreases as the duration of the observation period becomes longer and then it increases up to one. Since the whole spectrum is available at time 0, $P(A_{int}(t) > p)$ is close to one for small values of t. As the duration of the observation period lasts longer, i.e., with a larger value of t, the possibility of channel access decreases due to the commencement of PU and SU services. That is, both channel availability and $P(A_{int}(t) > p)$ decrease with a larger t. However, further extending the observation time will increase the value of $P(A_{int}(t) > p)$ asymptotically to one. The reason for such behavior is due to the steady state availability of the system. From the analysis of the asymptotic behavior of the interval availability, it will converge to the value of steady state availability as $t \to \infty$. In other words, *with a sufficiently long observation period,* $A_{int}(t) \approx A_{ss}$. Therefore, for a guaranteed level of p given that $p < A_{ss}$, the probability of achieving at least p interval availability tends to one as $t \to \infty$. Nevertheless, when the specified guaranteed level p increases, the probability of achieving an interval availability of at least p becomes lower. For instance, when the length of the observation period is 22 time units, the probability of achieving at least $p = 0.7$ interval availability is 95%. This probability will be only 90% if $p = 0.8$.

1) When $p \geq A_{ss}$: If the guaranteed level $p \geq A_{ss}$, the behavior of the CCDF is disparate from the previously obtained results as illustrated in Fig. B.17. In this

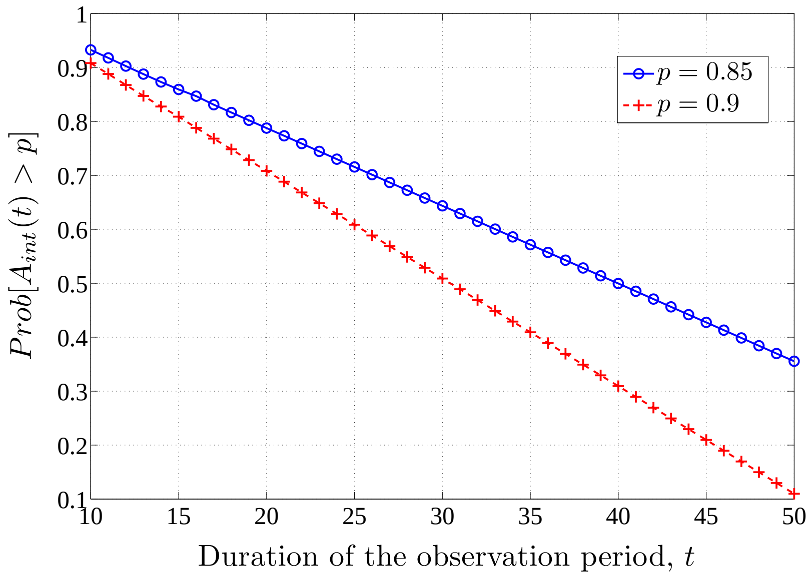


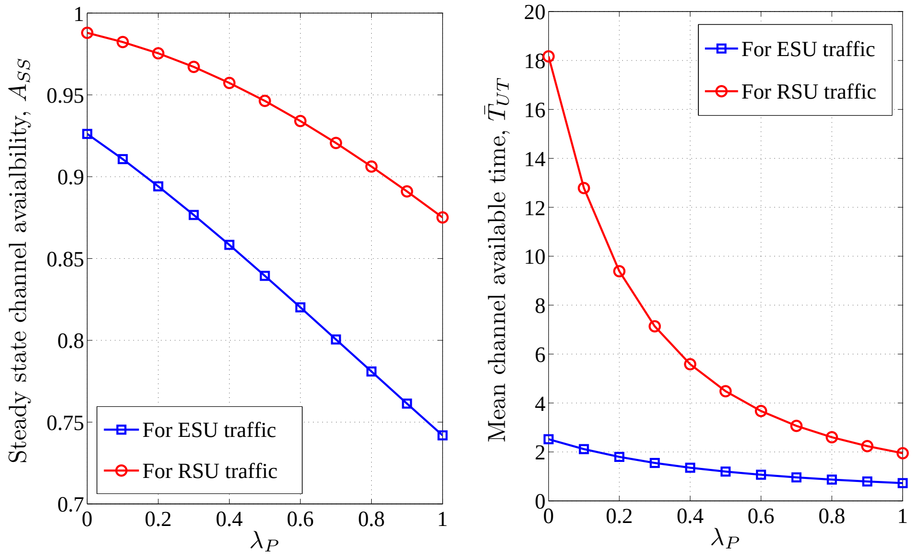
Figure B.17: Probability of achieving at least p channel availability when $p > A_{ss}$ as a function of observation period.

B

case, $P(A_{int}(t) > p)$ is monotonically decreasing when the observation period increases, as observed in Fig. B.17. The reason for this behavior is explained as follows. Since $A_{int}(t)$ converges to the steady state channel availability, A_{ss}, when $t \to \infty$, the probability of achieving an interval availability level which is higher than A_{ss} is lower. For small values of t, this probability is higher in contrast with the large values of t. This is because that the fraction of the accumulated time which the system is residing in channel available states is higher in the beginning. However this fraction goes down monotonically as we observe the system over a longer period. Therefore, $\lim_{t\to\infty} P(A_{int}(t) > p) = 0$ if $p \geq A_{ss}$.

J. Reliability Analysis in CRNs with Hybrid SU Traffic

In this subsection, we consider a heterogeneous SU traffic scenario which consists of two types of SU flows, i.e., elastic traffic and real-time traffic. While elastic flows adjust their transmission rate according to network conditions, real-time applications have the same session duration even though the data rate may vary from time to time. In this scenario, the number of channels allocated to a real-time SU (RSU) service is considered to be fixed to a single channel while an elastic SU (ESU) service can occupy from $W = 1$ up to $V = 3$ channels, the same as in the single-type SU scenario. An RSU service has higher priority than an ESU service if there is a need to release a channel to a PU arrival by reducing channel occupancy of an SU service or even forcibly terminating it. In addition, an ESU service can be preempted by an incoming RSU service if no idle channel ex-



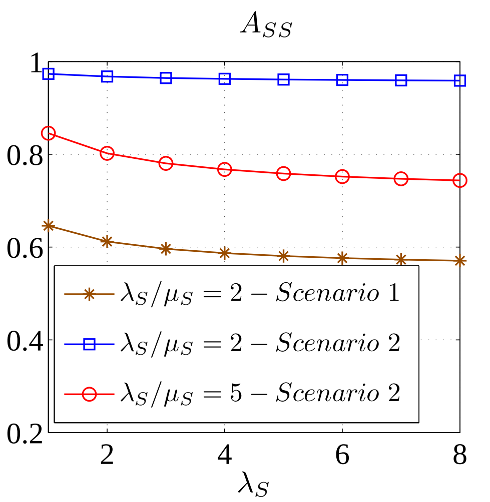
(a) (b)

Figure B.18: Steady state availability and mean channel available time for hybrid SU traffic when $M = 6$.

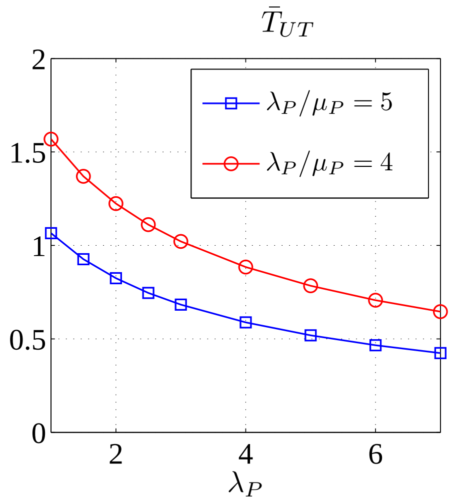
ists upon an RSU arrival. Fig. B.18 shows the steady state channel availability and mean channel available time for ESU and RSU services when the arrival rates (of ESU and RSU) and the service rates (of ESU and RSU) are set as follows: $\lambda_{ESU} = 2.0, \lambda_{RSU} = 1.5, \mu_{ESU} = 1.0$ and $\mu_{RSU} = 0.75$. Evidently, both the steady state channel availability and the mean channel available time decrease as the PU arrival rate becomes higher. However, the RSU services outperform the ESU services with respect to those reliability metrics since they have higher channel access privilege.

K. Reliability Analysis in CRNs with Heterogeneous Channels

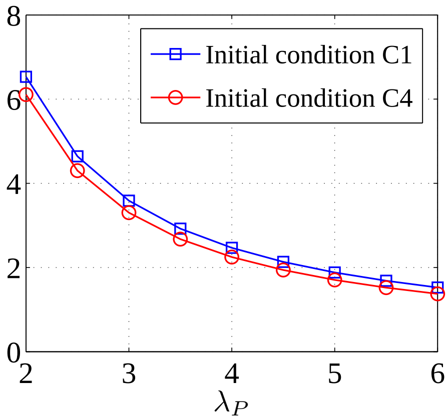
Finally, we evaluate the performance of Scenario II introduced in Sec. III-D with $W = 1$ and $V = 3$. Consider now a CRN with $M_1 = 3$ and $M_2 = 6$. The channel allocation in our evaluation herein is selected in such a way that the total bandwidth for Scenario II is the same as in Scenario I with $M = 6$. More specifically, the bandwidth for a wideband channel is equal to the bandwidth of a channel in Scenario I while one channel in the NBC occupies half bandwidth of a channel in Scenario I. Correspondingly, the service rate per channel is configured as $\mu_{S1} = 1.0, \mu_{S2} = 0.5, \mu_{P1} = 0.5$ and $\mu_{P2} = 0.25$ services per unit of time respectively. In Fig. B.19, we demonstrate three channel availability measures corresponding to the aforementioned heterogeneous channel environment. Comparing



(a) A_{SS} as a function of λ_S while keeping the offered traffic, $\frac{\lambda_S}{\mu_S}$, constant.



(b) $\bar{T}_{UT}$ as a function of λ_P while keeping the offered traffic, $\frac{\lambda_P}{\mu_P}$, constant.



$\bar{T}_{FF}$

(c) $\bar{T}_{FF}$ as a function of λ_P. The initial conditions C1 and C4 are explained in Table I.

Figure B.19: Channel availability analysis in the heterogeneous channel scenario. Please note that Scenario I in Fig. B.19(a) represents the homogeneous channel scenario employed in Fig. B.7.

Fig. B.19(a) with Fig. B.7, we observe similar variations of the steady state channel availability in the heterogeneous channel scenario. Moreover, as illustrated in Fig. B.19(b) and Fig. B.19(c), $\bar{T}_{UT}$ and $\bar{T}_{FF}$ show also the same trend as depicted in Fig. B.8 and Fig. B.10 respectively. Therefore, the results shown in Fig. B.19 clearly confirm the robustness and the applicability of the developed models in other kinds of network scenarios.

In order to compare the performance for both scenarios, we depict the steady state channel availability in Fig. B.19(a) under both homogeneous and heterogeneous conditions. As we observe from this figure, subject to the same arrival and departure rate configurations, channel availability in Scenario II shows a higher value than that of Scenario I. The reason for this result is as follows. Consider a CRN with $M = 6$ homogeneous channels in Scenario I. If three out of the six channels are occupied, then only three channels are available for newly arrived services. Re-configure this CRN to the heterogeneous scenario with the same wideband and narrowband relationship mentioned above, we end up with a CRN with $M_1 = 3$ and $M_2 = 6$. This means that there are still six narrowband channels available in the CRN even if the three wideband channels are occupied. In other words, channel fragmentation leads to an increased number of channels. Consequently, the steady state channel availability is increased.

VII. CONCLUSIONS

The reliability and availability aspects of CRNs remain largely un-chartered despite tremendous research efforts within the area of CR during the past decade. In this paper, we define five reliability and availability metrics for channel access in both homogeneous and heterogeneous CRNs from the perspective of dependability theory and develop Markov chain based models to analyze these metrics. The mathematical expressions for those metrics as well as the distribution of channel availability are derived using the developed models. The numerical results show that the steady state channel availability, the mean time to channel unavailability and the mean channel available time decrease with a higher PU arrival rate and increase with a higher PU service rate. Furthermore the distribution of channel availability is derived based on a uniformization tool. We propose thereafter a measure to calculate the probability that the channel availability during the observation period t is not lower than a pre-defined guaranteed level p and evaluate it with respect to t and p. We believe that the definitions and the models presented in this paper collectively provide a systematic approach for availability analysis of channel access in multi-channel CRNs.

APPENDIX

Calculation of $\bar{T}_{DT}$ in a single channel CRN

Consider a CRN with a single channel, i.e., $M = 1$ and a simple channel access scheme where SUs opportunistically access the channel while PUs have the priority for channel access. Assuming Markovian PU/SU arrivals and departures, the states of the corresponding CTMC model are represented by $\boldsymbol{x} = (i_{pu}, j_{su})$ where i_{pu} and j_{su} respectively denote the number of PU and SU services in the CRN. The feasible state space of this system is represented by $\mathcal{S} = \{(0,0),(0,1),(1,0)\}$. A new

SU arrival will access the CRN only when the system is in state $(0,0)$. Therefore, $\mathcal{S}_{\mathcal{A}} = \{(0,0)\}$ and $\mathcal{S}_{\mathcal{B}} = \{(0,1),(1,0)\}$ where $\mathcal{S}_{\mathcal{A}}$ and $\mathcal{S}_{\mathcal{B}}$ denote the set of channel available and unavailable states respectively. Accordingly, the transition rate matrix of this network is constructed as

$$\boldsymbol{Q} = \begin{matrix} \text{States} \\ (0,0) \\ (0,1) \\ (1,0) \end{matrix} \begin{matrix} \begin{matrix} (0,0) & & (0,1) & & (1,0) \end{matrix} \\ \begin{pmatrix} -(\lambda_S+\lambda_P) & \lambda_S & \lambda_P \\ \mu_S & -(\mu_S+\lambda_P) & \lambda_P \\ \mu_P & 0 & -\mu_P \end{pmatrix} \end{matrix}.$$

Assume that $\lambda_P = \lambda_S = 2$ and $\mu_P = 0.5, \mu_s = 1$. We have

$$\boldsymbol{Q} = \begin{bmatrix} -4 & 2 & 2 \\ 1 & -3 & 2 \\ 0.5 & 0 & -0.5 \end{bmatrix}.$$

Since state $(0,0)$ is the only channel available state in this example, the sub-matrices in the partitioned form in (B.5) are obtained respectively as follows

$$\boldsymbol{A} = \begin{bmatrix} -4 \end{bmatrix}, \;\; \boldsymbol{B} = \begin{bmatrix} 2 & 2 \end{bmatrix},$$

$$\boldsymbol{C} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}, \;\; \boldsymbol{D} = \begin{bmatrix} -3 & 2 \\ 0 & -0.5 \end{bmatrix}.$$

The steady state probability vectors are obtained by solving (B.4) and they are given as $\boldsymbol{\pi}_A = \begin{bmatrix} 0.12 \end{bmatrix}$ and $\boldsymbol{\pi}_B = \begin{bmatrix} 0.08 & 0.8 \end{bmatrix}$ for channel available states and channel unavailable states respectively. Once we reach this step, all necessary parameters that are needed to obtain the values of the performance metrics given in Sec. IV-C have been obtained. For instance, substituting these values into (B.16) yields the *mean channel unavailable time* as follows:

$$\begin{aligned} \bar{T}_{DT} &= \frac{\boldsymbol{\pi}_B \boldsymbol{U}_{N-L}}{\boldsymbol{\pi}_A \boldsymbol{B} \boldsymbol{U}_{N-L}} \\ &= \frac{\begin{bmatrix} 0.08 & 0.8 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}}{\begin{bmatrix} 0.12 \end{bmatrix} \begin{bmatrix} 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}} = \frac{0.88}{0.12 \cdot 4} = 1.8333 \text{ time units.} \end{aligned}$$

Table B.3: Transitions from a generic state $\boldsymbol{x} = (i_{pu}, j_W, j_{W+1}, \cdots j_V)$ of DFA upon PU events.

Activity	Destination State	Tran. rate	Conditions
PU AR. A vacant channel exists.	$(i_{pu}+1, j_W, j_{W+1}, \cdots, j_k, \cdots, j_V)$	λ_P	$b(\boldsymbol{x}) < M$.
PU AR. No vacant channel exists. An SU with k channels reduces its aggregated channels.	$(i_{pu}+1, j_W, j_{W+1}, \cdots j_{k-1}+1, j_k-1, 0, \cdots, 0)$	λ_P	$b(\boldsymbol{x}) = M; j_k > 0, k = \max\{r \mid j_r > 0, W+1 \leq r \leq V\}$.
PU AR. An SU service is terminated. No spectrum adaptation is executed.	$(i_{pu}+1, j_W-1, j_{W+1}, \cdots, j_V)$	λ_P	$b(\boldsymbol{x}) = M; ((j_W > 0, W = V.)$ Or $(j_W \geq 1; W = 1, j_k = 0$ for $k > W+1.)$ Or $(j_W = 1; j_k = 0, \forall k > W.))$
PU AR. An SU service is terminated and provides idle channels. Another SU service uses the idle channels.	$(i_{pu}+1, j_W-2, 0, \cdots j_l, 0, \cdots, 0)$	λ_P	$b(\boldsymbol{x}) = M; j_W > 1; j_l = 1, l = 2W-1 \leq V; W > 1; j_k = 0, \forall k > W$.
$\cdots$	$\cdots$	$\cdots$	$\cdots$
PU AR. An SU service is terminated. The remaining ongoing SU services use the idle channels and achieve the upper bound V.	$(i_{pu}+1, 0, \cdots, 0, j_V+q)$	λ_P	$b(\boldsymbol{x}) = M; q = j_W-1; j_W > 1; W-1 \geq (V-W)(j_W-1); V > W > 1$.
PU DP. An SU with k channels uses the idle channel.	$(i_{pu}-1, 0, \cdots, 0, j_k-1, j_{k+1}+1, \cdots j_V)$	$i\mu_P$	$j_k > 0, k = \min\{r \mid j_r > 0, W \leq r \leq V-1\}; i_{pu} > 0; V > 1$.
PU DP. SUs cannot use the idle channel.	$(i_{pu}-1, j_W, j_{W+1}, \cdots, j_k, \cdots j_V)$	$i\mu_P$	$V = W; i_{pu} > 0$. Or $j_k = 0, \forall k < V; i_{pu} > 0$.

In this table, Table B.4, Table B.5 and Table B.6, the notations, AR and DP, indicate an arrival event and a departure event respectively.

B

Table B.4: Transitions from a generic state $\boldsymbol{x} = (i_{pu}, j_W, j_{W+1}, \cdots j_V)$ of DFA upon SU events.

Activity	Destination State	Tran. rate	Conditions
SU AR. Enough idle channels exist.	$(i_{pu}, j_W, j_{W+1}, \cdots, j_k + 1, \cdots, j_V)$	λ_S	$k = \min\{M - b(\boldsymbol{x}), V\} \geq W$.
SU AR. The ongoing SU service with the maximum number of occupied channels, m, donates channels to the newly arrived service.	$(i_{pu}, j_W + 1, \cdots, j_n + 1, \cdots, j_m - 1, 0, \cdots, 0)$	λ_S	$M - b(\boldsymbol{x}) < M; m = \max\{r \| j_r > 0, W+1 \leq r \leq V\}; n = m - [W - (M - b(\boldsymbol{x}))] \geq W; V > 1$.
SU AR. Two ongoing SU services using m and h channels respectively, donate channels to the newly arrived service.	$(i_{pu}, j_W + 2, \cdots, j_n + 1, \cdots, j_h - 1, \cdots, j_m - 1, 0, \cdots, 0)$	λ_S	$M - m = \max\{r \| j_r > 0, W+1 \leq r \leq V\}; h = \max\{r \| j_r > 0, W+1 \leq r \leq m-1\}$ if $j_m = 1$, Or $h = m$ if $j_m > 1; n = h + m - 2W + M - b(\boldsymbol{x}) \geq W; V > 1; W > 1$.
$\cdots$	$\cdots$	$\cdots$	$\cdots$
SU AR. All ongoing SU services that aggregate more than W channels donate channel(s) to the newly arrived SU service.	$(i_{pu}, j_W + q, \cdots,, j_n + 1, 0, \cdots, 0)$	λ_S	$q = \sum_{m=W+1}^{V} j_m; n = \sum_{m=W+1}^{V}(m - W)j_m + M - b(\boldsymbol{x}), W \leq n < \min\{r \| j_r > 0, W+1 \leq r \leq V\}; V > 1$.
SU with k channels DP, $j_k = 1$. Other SU services, if exist, cannot use the idle channel(s).	$(i_{pu}, 0, \cdots, 0, j_k - 1, 0, \cdots, 0, j_V)$	$kj_k\mu_S$	$j_k = 1, k < V, j_m = 0\ \forall m \neq k, V$.
SU with k channels DP, $j_k > 0$. Other SU services, if exist, cannot use the idle channel(s).	$(i_{pu}, 0, \cdots, 0, j_V - 1)$	$kj_k\mu_S$	$j_k > 0, k = V; j_m = 0,\ \forall m < V$.
SU with k channels DP. An SU service with minimum number of aggregated channels, h, uses all the idle channel(s).	$(i_{pu}, j_W, \cdots, j_h - 1, \cdots, j_k - 1, \cdots, j_l + 1, \cdots, j_V)$	$kj_k\mu_S$	$j_k > 1, h = \min\{r \| j_r > 0, W \leq r \leq V-1\}; l = k + h \leq V; V > 1$. Or $j_k = 1; h = \min\{r \| j_r > 0, r \neq k, W \leq r \leq V-1\};\ l = k + h \leq V; V > 1$.
$\cdots$	$\cdots$	$\cdots$	$\cdots$
SU with k channels DP. The remaining SU services with fewer than V channels use the idle channel(s) and achieve the upper bound V.	$(i_{pu}, 0, \cdots, 0, j_V + q)$	$kj_k\mu_S$	$q = \sum_{m=W}^{V-1} j_m - 1; k \geq \sum_{m=W}^{V-1} j_m(V - m); V > 1$.

Table B.5: Transitions from a generic state $\boldsymbol{x} = (i^w, j^w_W, j^w_{W+1}, \cdots, j^w_V, i^n, j^n_W, j^n_{W+1}, \cdots j^n_V)$ of DFA_H upon PU events.

Activity	Destination State	Tran. rate	Conditions
PU AR. A vacant channel exists in the WBC.	$(i^w+1, j^w_W, \cdots, j^w_V, i^n, j^n_W, j^n_{W+1}, \cdots j^n_V)$	λ_P	$b_w(\boldsymbol{x}) < M_1$.
PU AR. No vacant channel exists in the WBC. An SU_W with k channels reduces its aggregated channels.	$(i^w + 1, j^w_W, j^w_{W+1}, \cdots j^w_{k-1} + 1, j^w_k - 1, \cdots j^w_V, i^n, j^n_W, j^n_{W+1}, \cdots j^n_V)$	λ_P	$b_w(\boldsymbol{x}) = M_1; j^w_k > 0, k = \max\{r \| j^w_r > 0, W+1 \le r \le V\}$.
PU AR. No vacant channel exists in the WBC. An SU_W with W channels performs spectrum handover to NBC.	$(i^w+1, j^w_W - 1, 0, \cdots, 0, i^n, 0, \cdots, 0, j^n_h + 1, \cdots j^n_V)$	λ_P	$b_w(\boldsymbol{x}) = M_1; j^w_W > 0, j^w_k = 0\ \forall k > W; h = \min\{M_2 - b_n(\boldsymbol{x}), V\} \ge W$.
PU AR. No vacant channel exists in the WBC. An SU_W with W channels performs spectrum handover to NBC and SU_N with k channels reduces its aggregated channels.	$(i^w+1, j^w_W-1, 0, \cdots, 0, i^n, j^n_W+1, \cdots, j^n_h+1, \cdots, j^n_m - 1, 0, \cdots, 0)$	λ_P	$b_w(\boldsymbol{x}) = M_1, b_n(\boldsymbol{x}) = M_2; j^w_W > 0, j^w_l = 0\ \forall l > W; m = \max\{r \| j^n_r > 0, W+1 \le r \le V\}; h = m - W, W \le h < m; V > 1$.
$\cdots$	$\cdots$	$\cdots$	$\cdots$
PU AR. No vacant channels exist in the CRN. An SU_W is forced to terminate.	$(i^w+1, j^w_W - 1, 0, \cdots, 0, i^n, j^n_W, 0, \cdots, 0)$	λ_P	$b_w(\boldsymbol{x}) = M_1, b_n(\boldsymbol{x}) = M_2; j^w_W > 0; j^w_k = 0\ \forall k > W; j^n_h = 0\ \forall h > W$.
PU AR. A vacant channel exists in the NBC.	$(i^w, 0, \cdots, 0, i^n+1, j^n_W, \cdots, j^n_V)$	λ_P	$b_w(\boldsymbol{x}) = M_1; j^w_k = 0\ \forall k \ge W; b_n(\boldsymbol{x}) < M_2$.
PU AR. No vacant channel exists in the CRN. An SU_N with k channels reduces its aggregated channels.	$(i^w, 0, \cdots, 0, i^n+1, j^n_W, \cdots j^n_{k-1} + 1, j^n_k - 1, 0, \cdots, 0)$	λ_P	$b_w(\boldsymbol{x}) = M_1, b_n(\boldsymbol{x}) = M_2; j^w_k = 0\ \forall k \ge W; j^n_k > 0; k = \max\{r \| j^n_r > 0, W+1 \le r \le V\}$.
PU AR. No vacant channels exist in the CRN. An SU_N is forced to terminate.	$(i^w, 0, \cdots, 0, i^n+1, j^n_W - 1, 0, \cdots, 0)$	λ_P	$b_w(\boldsymbol{x}) = M_1, b_n(\boldsymbol{x}) = M_2; j^w_k = 0\ \forall k \ge W; j^n_W > 0, j^n_h = 0\ \forall h > W$.
PU_W DP. A PU_N performs spectrum handover to the WBC. An SU_N with k channels aggregates the idle channel.	$(i^w, j^w_W, j^w_{W+1}, \cdots j^w_V, i^n - 1, 0, \cdots, 0, j^n_k - 1, j^n_{k+1} + 1, \cdots j^n_V)$	$i^w \mu_{P1}$	$i^w > 0, i^n > 0; j^n_k > 0, k = \min\{r \| j^n_r > 0, W \le r \le V-1\}$.
PU_W DP. An SU_N performs spectrum handover to the WBC. An SU_N with k channels aggregates the idle channel.	$(i^w - 1, j^w_W + 1, j^w_{W+1}, \cdots j^w_V, i^n, j^n_W - 1, 0, \cdots, 0, j^n_k - 1, j^n_{k+1} + 1, \cdots j^n_V)$	$i^w \mu_{P1}$	$i^w > 0, i^n = 0; j^n_W > 0, W = 1; j^n_k > 0, k = \min\{r \| j^n_r > 0, W \le r \le V-1\}$.
PU_W DP. An SU_W with k channels aggregates the idle channel.	$(i^w - 1, 0, \cdots, 0, j^w_k - 1, j^w_{k+1} + 1, \cdots j^w_V, i^n, j^n_W, j^n_{W+1}, \cdots j^n_V)$	$i^w \mu_{P1}$	$i^w > 0, i^n = 0; j^n_W = j^n_{W+1} = 0, W = 1; j^w_k > 0, k = \min\{r \| j^w_r > 0, W \le r \le V-1\}$.
PU_N DP. An SU_N with k channels aggregates the idle channel.	$(i^w, j^w_W, j^w_{W+1}, \cdots j^w_V, i^n - 1, 0, \cdots, 0, j^n_k - 1, j^n_{k+1} + 1, \cdots j^n_V)$	$i^n \mu_{P2}$	$i^n > 0; j^n_k > 0, k = \min\{r \| j^n_r > 0, W \le r \le V-1\}$.

The notations, AR and DP, indicate an arrival event and a departure event respectively. An SU service and a PU service in the NBC are denoted as SU_N and PU_N respectively. An SU service and a PU service in the WBC are denoted as SU_W and PU_W respectively.

B

B

Table B.6: Transitions from a generic state $\boldsymbol{x} = (i^w, j^w_W, j^w_{W+1}, \cdots, j^w_V, i^n, j^n_W, j^n_{W+1}, \cdots j^n_V)$ of DFA_H upon SU events.

Activity	Destination State	Tran. rate	Conditions
SU AR. Enough idle channels exist in the WBC.	$(i^w, j^w_W, j^w_{W+1}, \cdots j^w_k + 1, j^w_{k+1}, \cdots j^w_V, i^n, j^n_W, j^n_{W+1}, \cdots j^n_V)$	λ_S	$k = \min\{M_1 - b_w(\boldsymbol{x}), V\} \geq W$.
SU AR. SU_W with the maximum number of occupied channels, m, donates channel(s) to the newly arrived service.	$(i^w, j^w_W + 1, j^w_{W+1}, \cdots, j^w_n + 1, \cdots, j^w_m - 1, 0, \cdots, 0, i^n, j^n_W, j^n_{W+1}, \cdots, j^n_V)$	λ_S	$b_w(\boldsymbol{x}) = M_1; m = \max\{r \mid j^w_r > 0, W+1 \leq r \leq V\}; n = m - [W - (M_1 - b_w(\boldsymbol{x}))], W \leq n < m; V > 1$.
SU AR. Enough idle channels exist in the NBC.	$(i^w, j^w_W, 0, \cdots, 0, i^n, j^n_W, j^n_{W+1}, \cdots, j^n_k + 1, \cdots j^n_V)$	λ_S	$b_w(\boldsymbol{x}) = M_1; j^w_l = 0\ \forall l > W; k = \min\{M_2 - b_n(\boldsymbol{x}), V\} \geq W$.
SU AR. SU_N with the maximum number of occupied channels, m, donates channel(s) to the newly arrived service.	$(i^w, j^w_W, 0, \cdots, 0, i^n, j^n_W + 1, j^n_{W+1}, \cdots j^n_n + 1, \cdots, j^n_m - 1, 0, \cdots, 0)$	λ_S	$b_w(\boldsymbol{x}) = M_1, b_n(\boldsymbol{x}) = M_2; j^w_l = 0\ \forall l > W; m = \max\{r \mid j^n_r > 0, W+1 \leq r \leq V\}; n = m - W, W \leq n < m; V > 1$.
$\cdots$	$\cdots$	$\cdots$	$\cdots$
SU AR. All ongoing SU_N services that aggregate more than W channels donate channel(s) to the newly arrived SU service.	$(i^w, j^w_W, 0, \cdots, 0, i^n, j^n_W + q, 0, \cdots j^n_n + 1, 0, \cdots, 0)$	λ_S	$b_w(\boldsymbol{x}) = M_1, b_n(\boldsymbol{x}) = M_2; j^w_l = 0\ \forall l > W; q = \sum_{m=W+1}^{V} j^n_m; n = \sum_{m=W+1}^{V}(m - W)j^n_m + M_2 - b_n(\boldsymbol{x}), W \leq n < \min\{r \mid j^n_r > 0, W+1 \leq r \leq V\}; V > 1$.
SU_W with k channels DP. Other SU services, if exist, cannot use the idle channel(s).	$(i^w, j^w_W, j^w_{W+1}, \cdots j^w_k - 1, j^w_{k+1}, \cdots j^w_V, i^n, j^n_W, j^n_{W+1}, \cdots j^n_V)$	$kj^w_k\mu_{S1}$	$(j^w_k = 1, k < V; j^w_h = 0\ \forall h < V$ and $h \neq k)$ Or $(j^w_k > 0, k = V; j^w_h = 0\ \forall h < V); j^n_m = 0\ \forall m < 2k$.
SU_W with k channels DP. An SU_N with l channels performs spectrum handover to the WBC. An SU_N with h channels aggregates the idle channel.	$(i^w, j^w_W, \cdots j^w_k, \cdots j^w_V, i^n, j^n_W, \cdots, j^n_l - 1, \cdots, j^n_h - 1, \cdots, j^n_q + 1, \cdots j^n_V)$	$kj^w_k\mu_{S1}$	$(j^n_l > 1, l = \min\{r \mid j^n_r > 0, W \leq r < V\}; h = l)$ Or $(j^n_l = 1, l = \min\{r \mid j^n_r > 0, W \leq r < V\}; h = \min\{r \mid j^n_r > 0, r > l\}); q = h + l \leq V; V > 1$.
SU_W with k channels DP. An SU_W with minimum number of aggregated channels, h, uses all the idle channel(s).	$(i^w, j^w_W, \cdots j^w_h - 1, \cdots j^w_k - 1, \cdots j^w_l + 1, \cdots j^w_V, i^n, j^n_W, j^n_{W+1}, \cdots j^n_V)$	$kj^w_k\mu_{S1}$	$j^w_k > 0, j^n_m = 0\ \forall m < 2k; h = \min\{r \mid j^w_r > 0, W \leq r \leq V-1\}; l = k + h \leq V; V < 1$.
$\cdots$	$\cdots$	$\cdots$	$\cdots$
SU_W with k channels DP. All other SU_W services with fewer than V channels use the idle channel(s) and achieve the upper bound V.	$(i^w, 0, \cdots, 0, j^w_V + q, i^n, j^n_W, j^n_{W+1}, \cdots j^n_V)$	$kj^w_k\mu_{S1}$	$j^n_l = 0\ \forall l < 2k; q = \sum_{m=W}^{V-1} j^w_m - 1; k \geq \sum_{m=W}^{V-1}(V-m)j^w_m - (V-k); V > 1$
SU_N with k channels DP. Other SU services, if exist, cannot use the idle channel(s).	$(i^w, j^w_W, j^w_{W+1}, \cdots j^w_V, i^n, j^n_W, \cdots, j^n_k - 1, j^n_{k+1}, \cdots j^n_V)$	$kj^n_k\mu_{S2}$	$j^n_k = 1, j^n_m = 0\ \forall m < V$and $m \neq k$. Or $j^n_k > 0, k = V; j^n_m = 0\ \forall m < V$.
SU_N with k channels DP. An SU_N with minimum number of aggregated channels, h, uses all the idle channel(s).	$(i^w, j^w_W, j^w_{W+1}, \cdots j^w_V, i^n, j^n_W, \cdots j^n_h - 1, \cdots j^n_k - 1, j^n_{k+1}, \cdots, j^n_l + 1, \cdots, j^n_V)$	$kj^n_k\mu_{S2}$	$j^n_k > 0; h = \min\{r \mid j^n_r > 0, W \leq r \leq V-1\}; l = k + h \leq V; V > 1$.
$\cdots$	$\cdots$	$\cdots$	$\cdots$
SU_N with k channels DP. The remaining SU_N services with fewer than V channels use the idle channel(s) and achieve the upper bound V.	$(i^w, j^w_W, j^w_{W+1}, \cdots j^w_V, i^n, 0, \cdots, 0, j^n_V + q)$	$kj^n_k\mu_{S2}$	$q = \sum_{m=W}^{V-1} j^n_m - 1; k \geq \sum_{m=W}^{V-1}(V-m)j^n_m - (V-k); V > 1$

PAPER B: REFERENCES

[1] R. V. Prasad, P. Pawelczak, J. A. Hoffmeyer, and H. S. Berger, "Cognitive functionality in next generation wireless networks: Standardization efforts," *IEEE Commun. Mag.*, vol. 46, no. 4, pp. 72-78, Apr. 2008.

[2] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201-220, Feb. 2005.

[3] H. Kim and K. G. Shin, "Efficient discovery of spectrum opportunities with MAC-layer sensing in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 5, pp. 533-545, May 2008.

[4] L. Jiao, F. Y. Li, and V. Pla, "Modeling and performance analysis of channel assembling in multi-channel cognitive radio networks with spectrum adaptation," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2686-2697, Jul. 2012.

[5] 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 J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 406-416, Mar. 2013.

[6] P. Kaur, A. Khosla, and M. Uddin, "Spectrum management models for cognitive radios," *J. of Commun. and Netw.*, vol. 15, no. 2, pp. 222-227, Apr. 2013.

[7] U. Varshney and A. D. Malloy, "Improving the dependability of wireless networks using design techniques," in *Proc. IEEE LCN*, Tampa, FL, USA, Nov. 2001, pp. 122-131.

[8] A. Avižienis, J.-C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," *IEEE Trans. Dependable and Secure Comput.*, vol. 1, no. 1. pp. 11-33, Mar. 2004.

[9] Telecommunication Standardization Sector of ITU, "Definitions of terms related to quality of service," *Recommendation ITU-T E.800*, 2009. Available at: http://www.itu.int/rec/T-REC-E.800-200809-I/en.

[10] N. Cordeshi, D. Amendola, and E. Baccarelli, "Hard and soft optimal resource allocation for primary and secondary users in infrastructure vehicular networks," in *Proc. IEEE CCNC*, Las Vegas, NV, USA, Jan. 2015, pp. 708-713.

[11] N. Cordeshi, D. Amendola, and E. Baccarelli, "Reliable adaptive resource management for cognitive cloud vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2528-2537, Jun. 2015.

[12] B. Canberk, I. F. Akyildiz, and S. Oktug, "Primary user activity modeling using first-difference filter clustering and correlation in cognitive radio networks," *IEEE/ACM Trans. Netw.*, vol. 19, no. 1, pp. 170-183, Feb. 2011.

[13] G. Egeland and P. E. Engelstad, "The availability and reliability of wireless multi-hop networks with stochastic link failures," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1132-1146, Sep. 2009.

[14] H. B. Salameh, "Efficient resource allocation for multi-cell heterogeneous cognitive networks with varying spectrum availability," *IEEE Trans. Veh. Technol.*, Early access available in IEEE Xplore, DOI: 10.1109/TVT.2015.2477432, Jun. 2016.

[15] J. Liu, X. Jiang, H. Nishiyama, and N. Kato, "Reliability assessment for wireless mesh networks under probabilistic region failure model," *IEEE Trans. Veh. Tech.*, pp. 2253-2264, vol. 60, no. 5, Jun. 2011.

[16] Y. Zou, B. Champagne, W.-P. Zhu, and L. Hanzo, "Relay-selection improves the security-reliability trade-off in cognitive radio systems," *IEEE Trans. Commun.*, vol. 63, no. 1, pp. 215-228, Jan. 2015.

[17] X. Li, and S. A. Zekavat, "Traffic pattern prediction and performance investigation for cognitive radio systems," in *Proc. IEEE WCNC*, Las Vegas, NV, USA, Apr. 2008, pp. 894-899.

[18] ———, "Cognitive radio based spectrum sharing: Evaluating channel availability via traffic pattern prediction," *J. of Commun. and Netw.*, vol. 11, no. 2, pp. 104-114, Apr. 2009.

[19] I. A. M. Balapuwaduge and F. Y. Li, "System times and channel availability analyses in multi-channel cognitive radio networks," in *Proc. IEEE ICC*, Sydney, Australia, 10-14 Jun. 2014, pp. 1320-1325.

[20] A. Azarfar, J.-F. Frigon, and B. Sansò, "Reliability analysis of a channel restoration mechanism for opportunistic spectrum access," *Computer Communications*, vol. 36, no. 8, pp. 947-961, Elsevier, May 2013.

[21] ———, "Improving the reliability of wireless networks using cognitive radios," *IEEE Commun. Surveys & Tuts.*, vol. 14, no. 2, pp. 338-354, Second Quarter, May 2012.

[22] C. Ghosh, S. Pagadarai, D. P. Agrawal, and A. M. Wyglinski, “Queuing theory representation and modeling of spectrum occupancy employing radio frequency measurements,” in *Proc. IEEE VTC Fall*, Anchorage, AK, USA, Sep. 2009, pp. 1-5.

[23] L. Jiao, F. Y. Li, and V. Pla, “Dynamic channel aggregation strategies in cognitive radio networks with spectrum adaptation,” in *Proc. IEEE GLOBECOM*, Houston, TX, USA, Dec. 2011, pp. 1-6.

[24] L. Jiao, I. A. M. Balapuwaduge, F. Y. Li, and V. Pla, “On the performance of channel assembling and fragmentation in cognitive radio networks,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5661-5675, Oct. 2014.

[25] J. A. Buzacott, “Markov approach to finding failure times of repairable systems,” *IEEE Trans. Rel.*, vol. 19, no. 4, pp. 128-134, Nov. 1970.

[26] J. Muppala, R. Fricks, and K. S. Trivedi, *“Techniques for System Dependability Evaluation,”* in Computational Probability, pp. 445-480, Springer, 2000.

[27] C. G. Cassandras and S. Lafortune, *“Markov Chains”* in Introduction to Discrete Event Systems, 2nd Ed., pp. 405-411, Springer, New York, USA, 2008.

[28] G. Latouche and V. Ramaswami, *“PH Distributions”*, in Introduction to Matrix Analytic Methods in Stochastic Modeling, pp. 33-60, SIAM, Philadelphia, PA, USA, 1999.

[29] A. Goyal and A. N. Tantawi, “A measure of guaranteed availability and its numerical evaluation,” *IEEE Trans. Comput.*, vol. 37, no. 1, pp. 25-32, Jan. 1988.

[30] A. Reibman and K. Trivedi, “Numerical transient analysis of Markov models,” *Computers and Operations Research*, vol. 15, no. 1, pp. 19-36, Elsevier, Jan. 1988.

[31] W. Grassman, “Means and variances of time averages in Markovian environments,” *European J. of Operational Research*, vol. 31, no. 1, pp. 132-139, Elsevier, Jul. 1987.

[32] B. Melamed and M. Yadin, “Randomization procedures in the computation of cumulative-time distributions over discrete state Markov processes,” *Operations Research*, vol. 32, no. 4, pp. 926-944, Aug. 1984.

[33] G. Rubino and B. Sericola, *Markov Chains and Dependability Theory*, 1st Ed., pp. 148-159, Cambridge University Press, New York, USA, 2014.

[34] R. G. Gallager, *Stochastic Processes: Theory for Applications*, 1st Ed., pp. 86-89, Cambridge University Press, New York, USA, 2013.

[35] E. De Souza E Silva and H. R. Gail, "Calculating cumulative operational time distributions of repairable computer systems," *IEEE Trans. Comput.*, vol. 35, no. 4, pp. 322-332, Apr. 1986.

[36] I. A. M. Balapuwaduge, L. Jiao, V. Pla and F. Y. Li, "Channel assembling with priority-based queues in cognitive radio networks: Strategies and performance evaluation," *IEEE Trans. Wireless Commun.*, vol. 13, no.2, pp. 630-645, Feb. 2014.

[37] F. Barcelo and J. Jordan, "Channel holding time distribution in public telephony systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 1615-1625, Sep. 2000.

Paper C

Title:	Significance of Channel Failures on Network Performance in CRNs with Reserved Spectrum
Authors:	Indika A. M. Balapuwaduge[†], Frank Y. Li[†], and Vicent Pla[‡]
Affiliation:	[†] Dept. of Information and Communication Technology, University of Agder (UiA), N-4898 Grimstad, Norway [‡] Dept. of Communications, Universitat Politècnica de València (UPV), 46022 València, Spain
Conference:	*IEEE International Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, May 2016.
Copyright ©:	IEEE

C

Significance of Channel Failures on Network Performance in CRNs with Reserved Spectrum

Indika A. M. Balapuwaduge, Frank Y. Li, and Vicent Pla

Abstract — **It is well understood that in wireless networks, *channel failures*, which are typically caused by equipment or power failures as well as intrinsic features in radio transmissions, such as fading and shadowing, can easily result in network performance degradation. Therefore, fast recovery from channel failures is an important measure that should be incorporated with those networks. Consequently, in a cognitive radio network (CRN), channel failures can cause significant performance degradation in both primary and secondary networks. Instead, retainability, i.e., the capability of providing continuous connection for users must be guaranteed even if a significant network element is disrupted. In this paper, we develop an analytical model to analyze the system performance of CRNs with a focus on random channel failures. In addition to performance analysis under channel failures, a channel reservation scheme is proposed. The scheme can be adopted in dynamic channel access strategies in order to prioritize ongoing sessions over new user requests. Furthermore, optimal channel reservations subject to different performance requirements are also studied.**

I. Introduction

The fifth generation (5G) cellular network is now on its way. With a focus on fundamental challenges in the design of 5G networks, new research directions are already identified. One of the main challenges of future 5G wireless systems is to provide an ultra reliable communication (URC) whenever and wherever needed and to efficiently utilize radio spectrum. URC refers to provision of certain level of communication service almost 100% of the time and it is seen as a key enabler for 5G mobile communication systems [1]. Cognitive radio (CR) is a key component which is envisaged as a part of 5G and it is regarded as a promising technology which can tackle the challenges in 5G such as URC. Therefore reliability and availability of CR channels are of major importance to its successful operation in future 5G wireless networks.

Generally speaking, channels in a cognitive radio network (CRN) could be unavailable for primary users (PUs) and secondary users (SUs) mainly due to two fundamental reasons. First the licensed channels can be already occupied by other PUs and SUs in the system. Second the channels themselves could fail due to a variety of factors. Those factors include equipment or power failures, software malfunctioning, channel fading and shadowing. In recent work which studied network performance of CRNs, the aforementioned channel failures are almost neglected and merely the first reason, i.e., channel unavailability due to traffic conditions is considered [2], [3]. Unlike those studies in the literature, the goal of this paper is to study the performance of dynamic channel access in CRNs focusing on the impact of channel failures and their recovery.

Typically, forced termination of PU sessions has not been considered in legacy CRNs since ongoing PU sessions cannot be preempted by new user arrivals. Once channel failures occur, however, the ongoing PU sessions can also be terminated before service completion. Therefore in realistic scenarios, forced termination of services is not negligible when channels are prone to failures. Accordingly, *retainability*, which is obtained by one minus the forced termination probability is considered as one of the key performance indicators (KPIs) for E-UTRAN [4] and in an EU H2020 research project METIS [5]. Motivated by this fact, retainability-based resource allocation in CRNs is investigated in this paper.

By considering a topology in a CRN, the spreading of random failures was investigated in a recent paper [6]. Even though correlations among failures and cascading effects are studied, the results presented in [6] are limited to the analysis on failure statistics and network percolation. A cognitive relaying scheme was proposed in [7] to enhance the PU and SU performance in the event of transmission failures. Although the secondary throughput maximization is targeted in [7], forced termination of connections has not been investigated. Instead, forced termination and blocking probability analyses were conducted in [8] considering both resource insufficiency and link unreliability. However the traffic model considered in [8] is a multi-cellular system and the analysis does not directly apply to CRNs with reserved channels. In addition to the analysis on significance of channel failures, this paper proposes a channel reservation scheme as a network survivability mechanism to reduce forced terminations due to insufficient resources and channel failures. A continuous time Markov chain (CTMC) model is developed in this study to evaluate the system performance for *both primary and secondary networks* with respect to their capacity, blocking probability and retainability. Furthermore, optimal channel reservation is also a main consideration of this paper.

The rest of this paper is as follows. Sec. II describes our system model with assumptions and the employed dynamic channel access strategy. By providing the-

C

oretical analysis in Sec. III, CTMC models are developed to analyze system performance. Sec. IV and Sec. V present selected numerical results and an optimal channel reservation algorithm. Finally, Sec. V concludes the paper.

II. SYSTEM MODEL

In this paper we consider a centralized CRN architecture which contains a base station and a number of SUs and PUs. Resource allocation for the primary and secondary networks is coordinated by the base station which functions as the central controller in the network.

A. Network Scenario and Assumptions

The total licensed spectrum of the CRN is divided into $M \in \mathbb{Z}^+$ equal bandwidth channels where $\mathbb{Z}^+$ is the set of positive integers. As already discussed, all of these channels are prone to different kinds of failures which will interrupt ongoing communication sessions. As an essential measure to protect ongoing services, a certain, typically small, number of channels are reserved for the exclusive usage of the services which are interrupted due to channel failures or priority users' access in order to avoid forced terminations. In brief, *those reserved channels can only be accessed by the SU and PU services that are interrupted due to channel failures or the preempted SU services upon new PU arrivals.* As depicted in Fig. C.1, $R \in \mathbb{Z}^+$ channels out of M channels are selected to be reserved.

However, when channels are reserved for interrupted ongoing services, the blocking probability will increase as a consequence. Moreover, the overall capacity will degrade due to channel under-utilization. In order to reduce this capacity loss and channel unavailability for new users, an upper bound for R is set as $R \leq \lfloor \frac{M}{A} \rfloor$ where $A > 1$. The scalar A is a parameter which constraints the amount of spectrum to be reserved. For instance, if $A = 4$, the system cannot reserve more than 25% of its total spectrum for interrupted users.

In this study, we model traffic at the *service level* to capture the dynamic related to the arrival and departure of flows. Each PU or SU service occupies only one channel in the CRN. Moreover, the following assumptions are made as the basis to develop the analytical model presented later in Sec. III.

- The arrivals of both PU and SU services are Poisson processes with rates λ_P and λ_S for PU and SU services respectively. Moreover, the service times for PU and SU services are exponentially distributed, with corresponding service rates per channel μ_P and μ_S respectively.
- The on-time duration (during which the channel is working properly until a failure occurs) of a channel is exponentially distributed with failure rate

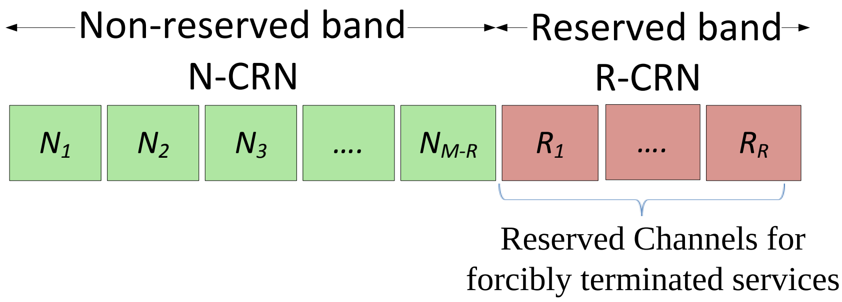


Figure C.1: Reserved (R-CRN) and non-reserved (N-CRN) channel assignment. N_i where $i \in \{1, 2, \cdots, M - R\}$ represents the channels in the N-CRN and R_j where $j \in \{1, 2, \cdots, R\}$ represents the channels in the R-CRN.

per channel λ_F in both N-CRN and R-CRN. Moreover, we assume that both occupied and idle channels are prone to failures.

- The repair time of a failed channel in both N-CRN and R-CRN is exponentially distributed, with a repair rate per channel denoted as μ_R. Moreover, all failed channels can be repaired simultaneously.

B. Dynamic Channel Access Strategy

C

This subsection provides a detailed overview of the proposed dynamic channel access strategy and the channel reservation scheme. It is worth mentioning that in this strategy, channels in the R-CRN are allocated *only for the interrupted services* to avoid forced terminations and newly arrived users will not be assigned to the R-CRN no matter whether it is an SU or a PU service. The reason to apply such a restriction is due to the fact that maintaining a higher retainability level for established connections is one of the main KPIs in future wireless systems such as 5G. Moreover, it is considered that forced termination of ongoing services is more annoying than blocking new service requests from quality of service (QoS) provisioning's point of view [9], [10]. The proposed strategy is explained under six events as follows.

1) PU arrival: If there is an idle channel in the N-CRN, a newly arrived PU will commence transmission on that idle channel. In the case where the whole N-CRN is occupied, one of the SU services in the N-CRN is interrupted and perform spectrum handover to a vacant channel in the R-CRN by giving the channel access to the new PU. If the interrupted SU service cannot find an idle channel in the R-CRN, it is forced to terminate. When all channels in the N-CRN are occupied by PUs, the new PU request is blocked.

2) SU arrival: Upon an SU arrival, the system allocates a channel to the new SU in the N-CRN. If all channels in the N-CRN are occupied, the new SU request is blocked without considering the channel availability in the R-CRN.

3) PU/SU departure: Once a PU or SU service in the N-CRN departs, one of the ongoing PU or SU services in the R-CRN will perform handover to the newly vacant channel in the N-CRN. However, the priority of accessing the channel is given to the PUs. No actions are performed after a departure of PU or SU services from the R-CRN.

4) Channel failure in the N-CRN: If an idle channel in the N-CRN or R-CRN failed, there is no need of any spectrum adaptation. However when a channel occupied by a PU in the N-CRN fails, the system will assign an idle channel if available from the N-CRN to continue its service. Otherwise, an ongoing SU service in the N-CRN will perform spectrum handover to a vacant channel in the R-CRN by giving its channel to the interrupted PU service. However, if no vacant channel is found in the R-CRN, the SU service is forced to terminate. Nevertheless, if there are no SU services in the N-CRN, the interrupted PU checks for available channels in the R-CRN. If there are no vacant channels, one of the SU services is preempted. In this way, the interrupted PU service is forced to terminate only if all the non-failed channels in both N-CRN and R-CRN are occupied by other PUs.

On the other hand, if the failed channel is occupied by an SU service, it will get channel access in the R-CRN for service continuation if all channels in the N-CRN are busy or failed. If the R-CRN is also fully occupied or failed, there is no option other than forcibly terminating the SU service.

5) Channel failure in the R-CRN: Once a PU occupied channel in the R-CRN fails, another idle channel in the R-CRN is allocated for the interrupted PU service. If there are no idle channels in the R-CRN, one of the ongoing SU services in the R-CRN is preempted by the interrupted PU. If there are neither idle channels nor ongoing SU services in the R-CRN, the interrupted PU service is forced to terminate. On the other hand, upon a failure of an SU occupied channel, the system will terminate the interrupted SU service if it could not find a vacant channel in the R-CRN.

6) Channel failure in the R-CRN: Once a failed channel is re-established to its normal conditions and ready for carrying services, it is said that the channel is repaired. In this strategy, the spectrum adaptation procedure applied to a newly repaired channel in the N-CRN is similar to the adaptation process applied to a newly vacant channel after a PU departure from the N-CRN. However, no spectrum adaptation is required when a reserved channel is recovered.

III. CTMC Modeling and Performance Analysis

We develop an analytical model for the CRN under the proposed channel reservation strategy using a CTMC.

A. CTMC Modeling

The CTMC model can be described by its state transition characteristics. A general state of the CTMC for the proposed strategy with R reserved channels is denoted as $\boldsymbol{x} = (i_n, j_n, f_n, i_r, j_r, f_r)$ where $i_n,\ j_n,\ f_n = 0, 1, \cdots, M - R$ and $i_r,\ j_r,\ f_r = 0, 1, \cdots, R$. $i_n,\ j_n$ and f_n represent, respectively, the number of ongoing PU services, ongoing SU services, and failed channels in the N-CRN. Similarly, $i_r,\ j_r$ and f_r represent the same parameters in the R-CRN. The set of feasible states of the system is denoted as $\mathcal{S} = \{\boldsymbol{x} \mid i_n, j_n, f_n, i_r, j_r, f_r \geq 0; B_n(\boldsymbol{x}) \leq M - R,\ B_r(\boldsymbol{x}) \leq R,\ B_r(\boldsymbol{x}) - f_r = 0$ if $B_n(\boldsymbol{x}) < M - R\}$ where $B_n(\boldsymbol{x}) = i_n + j_n + f_n$ and $B_r(\boldsymbol{x}) = i_r + j_r + f_r$. The state transitions associated with different events under different conditions are summarized in Table C.2 and Table C.3.

Denote $\pi(\boldsymbol{x})$ as the steady state probability of being in state $\boldsymbol{x}$. Given $\mathcal{S}$ and transition rate matrix $\boldsymbol{Q}$, the global balance equation and the normalization equation are constructed as

$$\boldsymbol{\pi Q} = \boldsymbol{0},\ \sum_{\boldsymbol{x} \in \mathcal{S}} \pi(\boldsymbol{x}) = 1, \tag{C.1}$$

where $\boldsymbol{\pi}$ is the steady state probability vector and $\boldsymbol{0}$ is a row vector of all 0's. The vector $\boldsymbol{\pi}$ is determined by solving (C.1).

C

B. Retainability Definition

The telecommunication standardization sector of ITU presents in its recommendation E.800 [11] a comprehensive set of terms and definitions relating to QoS and dependability standards for telecommunication networks. According to [11], retainability is defined as the probability that a connection, once established, will operate within specified transmission tolerance without interruption for a given time interval. Mathematically, the retainability of a service, θ, can be expressed as $\theta = 1 - P_F$ where P_F stands for the forced termination probability of that service.

C. Retainability Analysis for PU and SU Services

In a CRN where both PUs and SUs are active in operation, an ongoing SU service can be forced to terminate due to new PU arrivals. Whenever a PU preempts an SU and the preempted SU cannot find an idle channel in the N-CRN or R-CRN, the SU service is forced to terminate. Denote the rate of forced terminations of SUs due to PU arrivals as R_{SU}. Then

$$R_{SU} = \lambda_P \sum_{\substack{\boldsymbol{x}\in\mathcal{S} \\ B_n(\boldsymbol{x})+B_r(\boldsymbol{x})=M,\ j_n>0}} \pi(\boldsymbol{x}). \tag{C.2}$$

On the other hand, ongoing SU services can also be terminated upon a channel failure when all other channels in the CRN are busy. Denote the rate of forced terminations of SUs due to channel failures as R'_{SU}. It is obtained by

$$R'_{SU} = \sum_{\substack{\boldsymbol{x}\in\mathcal{S} \\ B_n(\boldsymbol{x})+B_r(\boldsymbol{x})=M,\ j_n+j_r>0}} (M - f_n - f_r)\lambda_F \pi(\boldsymbol{x}). \tag{C.3}$$

The forced termination probability of SUs, P_{SU}^F, can be expressed as the ratio between the mean forced termination rate of SU services, and the effective rate in which a new SU service is assigned a channel, $\Lambda_S = (1 - P_{SU}^B)\lambda_S$ [12]. P_{SU}^B and P_{PU}^B denote the blocking probability of SU and PU services respectively. Therefore, $P_{SU}^F = \frac{R_{SU}+R'_{SU}}{\Lambda_S}$. As stated earlier, retainability reflects the probability that a service will not be terminated before service completion. Correspondingly, the retainability of SU services, θ_{SU}, can be expressed as

$$\theta_{SU} = 1 - \left(\frac{R_{SU} + R'_{SU}}{\Lambda_S}\right). \tag{C.4}$$

Similarly, the retainability of PU services, θ_{SU}, is given by

$$\theta_{PU} = 1 - \frac{R'_{PU}}{\Lambda_P}, \tag{C.5}$$

where R'_{PU} and Λ_P are given by

$$R'_{PU} = \sum_{\substack{\boldsymbol{x}\in\mathcal{S} \\ B_n(\boldsymbol{x})+B_r(\boldsymbol{x})=M,\ j_n=j_r=0}} (i_n + i_r)\lambda_F \pi(\boldsymbol{x}) \tag{C.6}$$

and $\Lambda_P = (1 - P_{PU}^B)\lambda_P$ respectively. Note that the corresponding expression, R_{PU}, which denotes the forced termination rate of PUs due to new user arrivals always equals zero since none of the ongoing PUs can be terminated due to the arrivals of new users. Consequently, R_{PU} does not exist in (C.5).

D. Capacity and Blocking Probability Calculation

The capacity of a service in a CRN can be defined as the average number of service completions per time unit. In the considered network model, SU and PU services can be finished in both N-CRN and R-CRN. Hence the capacity of SU and PU services is given by

$$\rho_{SU} = \sum_{\boldsymbol{x}\in\mathcal{S}}(j_n + j_r)\mu_S\pi(\boldsymbol{x}), \quad \rho_{PU} = \sum_{\boldsymbol{x}\in\mathcal{S}}(i_n + i_r)\mu_P\pi(\boldsymbol{x}) \tag{C.7}$$

respectively. As already pointed out in Sec. II-B, a newly arrived SU service is blocked when all available channels in the N-CRN (excluding the failed channels) are occupied by PUs or/and SUs. On the other hand, when all available channels in the N-CRN are occupied by PUs, a newly arrived PU service is blocked. Consequently, the blocking probability of both services is expressed as follows respectively.

$$P_{SU}^{B} = \sum_{\substack{\boldsymbol{x}\in\mathcal{S} \\ B_n(\boldsymbol{x})=M-R}} \pi(\boldsymbol{x}), \quad P_{PU}^{B} = \sum_{\substack{\boldsymbol{x}\in\mathcal{S} \\ i_n+f_n=M-R}} \pi(\boldsymbol{x}). \tag{C.8}$$

Table C.1: Summary of symbols in the CTMC and value configuration

Symbol	Description	Value
M	Total number of channels in the CRN	8 or 12
R	Number of reserved channels in the R-CRN	2
A	Parameter which limits R	$2 - 10$
λ_P, λ_S	PU and SU arrival rates	5 services per unit of time
μ_P, μ_S	PU and SU service rates per channel	2 services per unit of time
λ_F	Failure rate of channels	$0.01 - 0.2$ failures per unit of time
μ_R	Repair rate per channel	0.1 or 0.12 repairs per unit of time

IV. Numerical Results and Discussions

The numerical results to assess our previous analysis are presented in this section. To derive the numerical results, we coded, in MATLAB, the CTMC model developed in Sec. III. The values of the network parameters are configured as in Table C.1 unless otherwise declared. Note that, it is necessary to configure λ_F and μ_R to small values in comparison with the arrival and service rates since channel failures do not occur frequently.

A. Retainability

Fig. C.2 presents the retainability of PU and SU services under different R values as a function of the PU arrival rate. It is observed that by increasing the number of reserved channels, the retainability of services improves for both PUs and SUs.

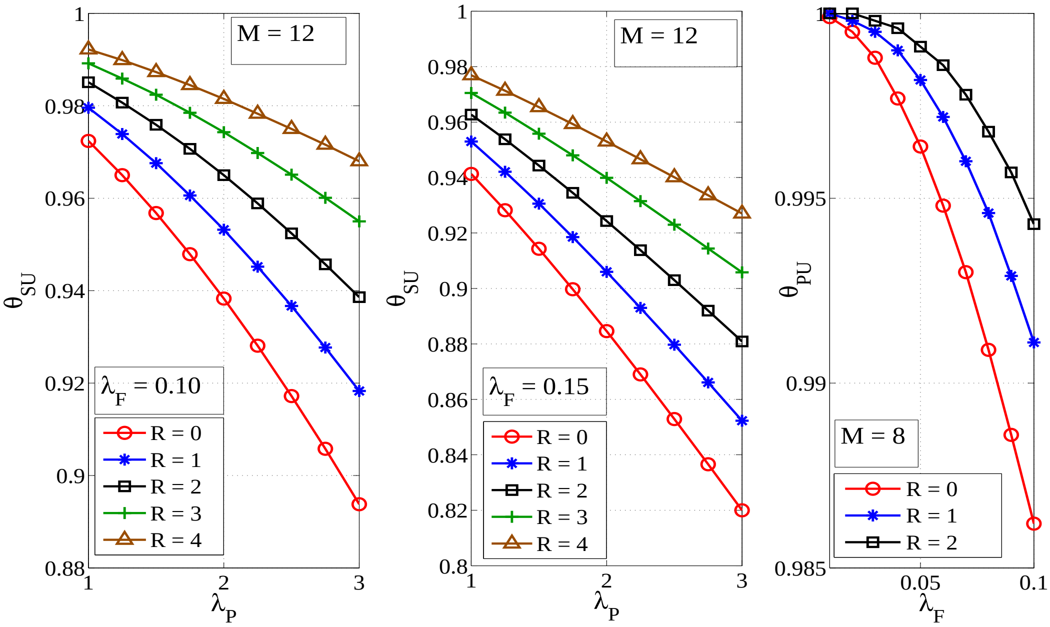


Figure C.2: Retainability of PU and SU services as a function of PU arrival rate and channel failure rate respectively.

With further analysis of the obtained results, we also note that the benefit of channel reservation is prominent when the channel failure rate is higher. For instance, if the system can reserve $R = 4$ out of $M = 12$ channels, the percentage of the SU retainability improvement is approximately 8% *with reference to the network without channel reservation* given that $\lambda_F = 0.1$ and $\lambda_P = 3$. However, this percentage improvement becomes 16% when the failure rate is $\lambda_F = 0.2$. Also, it is worth mentioning that the retainability of PU services is always higher than that of SUs due to its priority on accessing channels in both N-CRN and R-CRN.

B. PU and SU Capacity

From the retainability point of view, we observe that both primary and secondary networks achieve significantly improved performance with reserved channels. However, the price in which the system pays for channel reservation is the decreased capacity and higher blocking probability as shown in Fig. C.3 and Fig. C.4 respectively. In Fig. C.3, we plot the capacity achieved by both PU and SU services as a function of the channel failure rate, λ_F. As can be observed from this figure, when λ_F increases, the capacity of both services decreases as expected. Moreover, capacity degradation occurs when channel reservation is applied. However the system will be able to reduce the capacity loss with an increased channel repair rate if the channel failure rate reaches higher values. For instance, 20% increase of the channel repair rate sufficiently compensates the SU capacity loss when $\lambda_F > 0.04$. On the

C

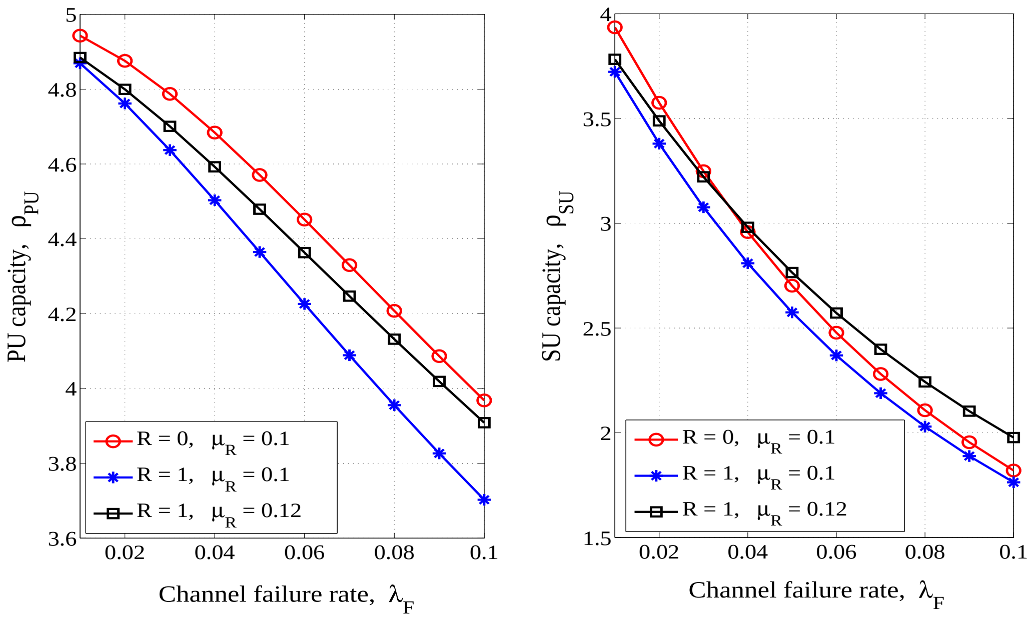


Figure C.3: Capacity of PU and SU services when channel reservation is applied in a CRN with $M = 8$ channels.

other hand, when the channel failure rate is not significant or comparatively small, the increase of repair rate may not help to recompense the capacity loss. Therefore, at a very low λ_F, channel reservation is not recommended. Motivated by this observation, we attempt to discover an optimal channel reservation for the CRN under different QoS requirements as discussed in details in Sec. V.

C. Blocking Probability

Maintaining a high level of service retainability is considered as more fundamental for both voice and data communications since service interruptions lasting over long periods are not acceptable. Therefore, channel reservation up to a certain extent is recommended although the blocking probability of new users increases as depicted in Fig. C.4. The numerical results presented in Fig. C.4 show that the blocking probability for PU and SU services increases as a result of both channel failures and channel reservations. If channel reservation is not adopted, i.e., when $R = 0$, all channels would be otherwise used for accepting new service requests. Again, it is observed that with a higher channel repair rate, fewer PU and SU services are blocked. This is because that, when μ_R is higher, the failed channels are recovered shortly and become idle again, resulting in increased channel access opportunities for new users. Moreover, we have observed that with an increasing channel failure rate, the blocking probability tends to monotonically increase due to lack of idle channels.

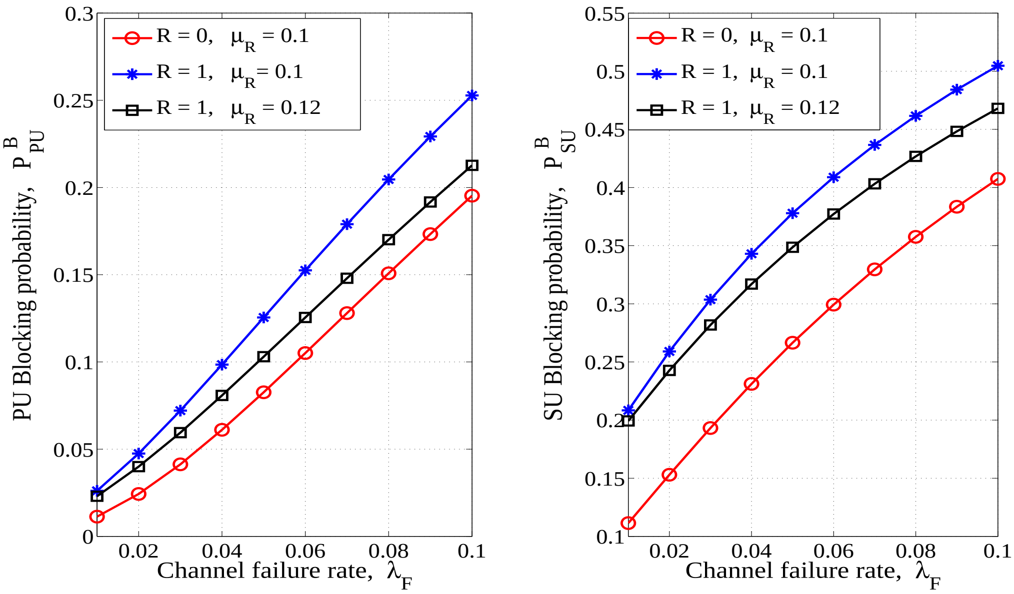


Figure C.4: Blocking probability of PU and SU services when channel reservation is applied in a CRN with $M = 8$ channels.

V. Retainability-based Optimal Channel Reservation

As stated in Subsec. IV-C, once the number of reserved channels is increased, the blocking probability of PU and SU services consequently increases. In other words, channel availability for new users gradually reduces when R increases. On the other hand, if there are no reserved channels, the retainability of services diminishes and therefore the QoS level degrades. Thus, it would be essential to develop a methodology to select the optimal number of reserved channels which can satisfy certain level of QoS requirements. Herein we propose a retainability-based channel reservation method as shown in Algorithm C.1. Algorithm C.1 is able to calculate and obtain the optimal number of reserved channels denoted as R_{opt} in the R-CRN for a given set of conditions. In this algorithm, a CRN consisting of M channels is considered where the system does not allow to reserve more than $\lfloor \frac{M}{A} \rfloor$ channels. Furthermore, the system should guarantee at least the $(\theta_{SU})_{MIN}$ level of retainability for SU services while *maintaining the blocking probability as low as possible*.

The value of R_{opt} can be found considering various traffic conditions. As an example, we determine R_{opt} while investigating the impact of λ_P on the optimal solution. Let the required minimum retainability level be $(\theta_{SU})_{MIN} = 0.92$, $\lambda_F = 0.10$, $M = 12$ and $A = 2.2$. As observed in Fig. C.2, SUs will achieve $\theta_{SU} = 0.92$ of retainability even without channel reservation if $\lambda_P \leq 2.5$. Therefore, in order to

Algorithm C.1: Retainability-based channel reservation algorithm.

Input: M : Total number of channels in the CRN
Input: $\theta_{SU}|_{R=k}, k = 0, 1, 2, \cdots$: Retainability of SU services when k channels are reserved in the R-CRN
Input: A : Parameter to determine the upper bound for R
Input: $(\theta_{SU})_{MIN}$: Required minimum retainability level for SUs
Output: R_{OPT} : The optimal value of R

```
Calculate R' = ⌊M/A⌋
if (θ_SU)_MIN ≤ θ_SU|R=0 then
|   R_OPT = 0
else if θ_SU|R=0 < (θ_SU)_MIN ≤ θ_SU|R=1 then
|   R_OPT = 1
···
else if θ_SU|R=k−1 < (θ_SU)_MIN ≤ θ_SU|R=k then
|   R_OPT = k
···
else if θ_SU|R=R'−1 < (θ_SU)_MIN ≤ θ_SU|R=R' then
|   R_OPT = R'
else
|   R_OPT = R'
end
```

fulfill the requirement of maintaining a minimum blocking probability, R_{opt} should be 0 when $\lambda_P \leq 2.5$. However, if the PU arrival rate appears in the range of $2.5 < \lambda_P < 2.9$, the system should reserve at least 1 channel to achieve the minimum retainability level. The reason is as follows. When λ_P becomes higher, the forced termination probability of SU services rises due to the access priority of PUs leading to reduced SU retainability. In this case, to attain the required retainability level, the system has to reserve more channels to the R-CRN.

Likewise, through the analysis of the statistics obtained from Fig. C.2, the value of R_{opt} which satisfies the given conditions can be obtained for a given range of λ_P. The obtained numerical results are illustrated in the left sub-figure in Fig. C.5 for $1 \leq \lambda_P \leq 3$ for different values of the required minimum θ_{SU}. However, the channel failure rate is kept fixed as $\lambda_F = 0.10$. As observed at a given λ_P, the optimal value for R needs to be increased when the required retainability level rises.

The effect of λ_F on determining R_{opt} is also depicted in the right sub-figure of Fig. C.5, by configuring the required minimum retainability level fixed as $(\theta_{SU})_{MIN} = 0.90$. As illustrated in this figure, with respect to an increasing channel failure rate and at a constant λ_P, the system has to allocate correspondingly a higher number of channels to the R-CRN to meet the $\theta_{SU} = 0.90$ requirement. Owing to frequent channel failures, the SU forced termination probability increases by lowering θ_{SU}. Accordingly, in order to compensate the retainability loss, the system will set a

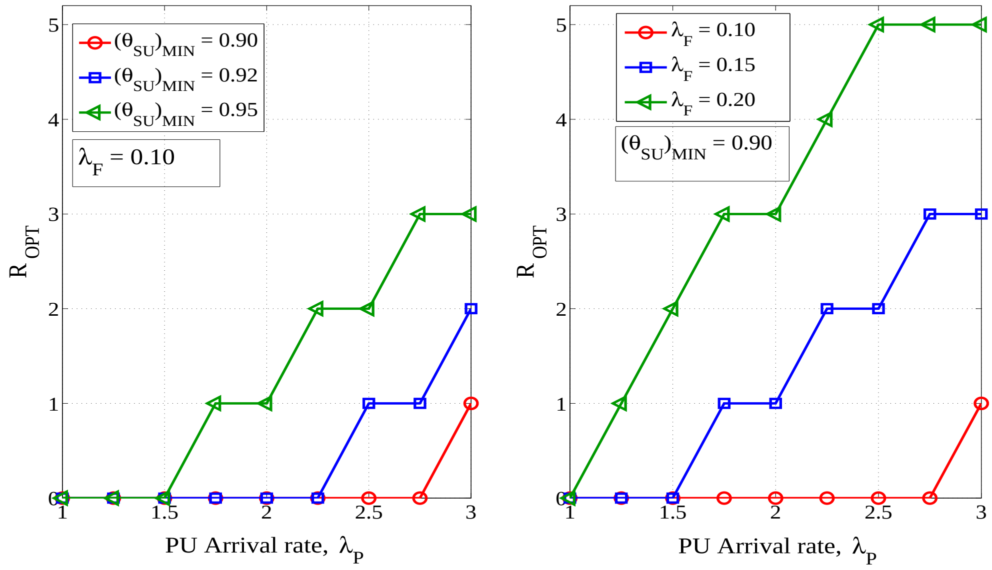


Figure C.5: Optimal number of reserved channels when the required minimum retainability level, θ_{SU} is given in a CRN with $M = 12$ channels.

comparatively larger R_{opt} when λ_F becomes higher.

VI. Conclusions

In this paper, we investigated the impact of random channel failures on the performance of channel access in CRNs. A channel reservation based dynamic spectrum access strategy is proposed with the main objective of improving the retainability level of ongoing transmission sessions. Furthermore, a CTMC model has been developed to evaluate the performance of this strategy. According to the obtained numerical results, a high level of retainability can be achieved through channel reservation. At the same time, keeping the number of reserved channels minimum is beneficial for obtaining high capacity and low blocking probability perspectives. Therefore, based on given QoS requirements and spectrum management constraints, we have proposed an algorithm to determine the optimal number of reserved channels in the CRN.

Paper C: References

[1] R. Sattiraju, P. Chakraborty, and H. D. Schotten, "Reliability modeling and analysis of a wireless transmission as a repairable system," in *Proc. IEEE GLOBECOM Workshops on Ultra-Low Latency and Ultra-High Reliability in Wireless Communications*, pp. 1397-1401, Austin, TX, USA, Dec. 2014.

C

[2] N. Cheng, N. Zhang, N. Lu, X. Shen, J. W. Mark, and F. Liu, “Opportunistic spectrum access for CR-VANETs: A game-theoretic approach,” *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 237-251, Jan. 2014.

[3] I. A. M. Balapuwaduge and F. Y. Li, “System times and channel availability analyses in multi-channel cognitive radio networks,” in *Proc. IEEE ICC*, pp. 1320-1325, Sydney, Australia, 10-14 Jun. 2014.

[4] 3GPP, “Key performance indicators (KPI) for evolved universal terrestrial radio access network (E-UTRAN): definitions,” *3GPP TS 32.450*, pp. 6-10, Jun. 2010. [Online]. Available: http://www.3gpp.org/DynaReport/32450.htm.

[5] A. Osseiran et al, “Scenarios for 5G mobile and wireless communications: the vision of the METIS project,” *IEEE Commun. Magaz.*, vol. 52, no. 5, pp. 26-35, May 2014.

[6] L. Sun, W. Wang, and Z. Lu, “On topology and resilience of large-scale cognitive radio networks under generic failures,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3390-3401, Jun. 2015.

[7] D. Hamza and S. Aïssa, “Enhanced primary and secondary performance through cognitive relaying and leveraging primary feedback,” *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2236-2247, Jun. 2014.

[8] C. B. R.-Estrello, G. H.-Valdez, and F. A. C.-Pérez, “System-level analysis of mobile cellular networks considering link unreliability,” *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 926-940, Feb. 2009.

[9] D. Hong and S. S. Rappaport, “Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures,” *IEEE Trans. Veh. Technol.*, vol. 35, no. 3, pp. 77-92, Aug. 1986.

[10] J. Wang, Q.-A. Zeng, and D. P. Agrawal, “Performance analysis of a preemptive and priority reservation handoff scheme for integrated service-based wireless mobile networks,” *IEEE Trans. Mobile Comput.*, pp. 65-75, vol. 2, no.1, Jan.-Mar. 2003.

[11] Telecommunication Standardization Sector of ITU, “Quality of service and dependability vocabulary,” *Recommendation ITU-T E.800*, 2007. [Online]. Available: https://www.itu.int/rec/T-REC-E.800-198811-S/en.

[12] L. Jiao, F. Y. Li, and V. Pla, “Modeling and performance analysis of channel assembling in multichannel cognitive radio networks with spectrum adaptation,” *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2686-2697, Jul. 2012.

Table C.2: Transitions from a generic state $\boldsymbol{x} = (i_n, j_n, f_n, i_r, j_r, f_r)$ of the channel access strategy upon PU/SU events and channel failures.

Activity	Destination State	Tran. rate	Conditions
1. PU AR. A vacant channel exists in the N-CRN.	$(i_n+1, j_n, f_n, i_r, j_r, f_r)$	λ_P	$B_n(\boldsymbol{x}) < M'$.
2. PU AR. No vacant channel exists. SU_N performs handover to the R-CRN.	$(i_n+1, j_n-1, f_n, i_r, j_r+1, f_r)$	λ_P	$B_n(\boldsymbol{x}) = M', j_n > 0, B_r(\boldsymbol{x}) < R$.
3. PU AR. SU_N is forced to terminate.	$(i_n+1, j_n-1, f_n, i_r, j_r, f_r)$	λ_P	$B_n(\boldsymbol{x}) = M', j_n > 0, B_r(\boldsymbol{x}) = R$.
4. SU AR. A vacant channel exists in the N-CRN.	$(i_n, j_n+1, f_n, i_r, j_r, f_r)$	λ_S	$B_n(\boldsymbol{x}) < M'$.
5. PU DP from the N-CRN. PU_R performs spectrum handover to the N-CRN.	$(i_n, j_n, f_n, i_r-1, j_r, f_r)$	$i_n\mu_P$	$i_n > 0, i_r > 0$.
6. PU DP from the N-CRN. SU_R performs handover to the N-CRN.	$(i_n-1, j_n+1, f_n, i_r, j_r-1, f_r)$	$i_n\mu_P$	$i_n > 0, i_r = 0, j_r > 0$.
7. PU DP from the R-CRN. No spectrum adaptation.	$(i_n, j_n, f_n, i_r-1, j_r, f_r)$	$i_r\mu_P$	$i_r > 0$.
8. SU DP from the N-CRN. SU_R performs handover to the N-CRN.	$(i_n, j_n, f_n, i_r, j_r-1, f_r)$	$j_n\mu_S$	$j_n > 0, i_r = 0, j_r > 0$.
9. SU DP from the R-CRN. No spectrum adaptation.	$(i_n, j_n, f_n, i_r, j_r-1, f_r)$	$j_r\mu_S$	$j_r > 0$.
10. Idle channel failure in the N-CRN.	$(i_n, j_n, f_n+1, i_r, j_r, f_r)$	$(M' - B_n(\boldsymbol{x}))\lambda_F$	$B_n(\boldsymbol{x}) < M', f_n < M'$.
11. Idle channel failure in the R-CRN.	$(i_n, j_n, f_n, i_r, j_r, f_r+1)$	$(R - B_r(\boldsymbol{x}))\lambda_F$	$B_r(\boldsymbol{x}) < R, f_r < R$.
12. A PU occupied channel in the N-CRN fails. A vacant channel exists in the N-CRN.	$(i_n, j_n, f_n+1, i_r, j_r, f_r)$	$i_n\lambda_F$	$i_n > 0, B_n(\boldsymbol{x}) < M'-1$.
13. A PU occupied channel in the N-CRN fails. SU_N performs handover to the R-CRN.	$(i_n, j_n-1, f_n+1, i_r, j_r+1, f_r)$	$i_n\lambda_F$	$i_n > 0, B_n(\boldsymbol{x}) = M', j_n > 0, B_r(\boldsymbol{x}) < R$.
14. A PU occupied channel in the N-CRN fails. SU_N is forced to terminate.	$(i_n, j_n-1, f_n+1, i_r, j_r, f_r)$	$i_n\lambda_F$	$i_n > 0, B_n(\boldsymbol{x}) = M', j_n > 0, B_r(\boldsymbol{x}) = R$.

In this table and Table C.3, the notations AR and DP indicate an arrival event and a departure event respectively.
An SU and a PU service in the N-CRN are denoted as SU_N and PU_N respectively.
An SU and a PU service in the R-CRN are denoted as SU_R and PU_R respectively. Denote $M' = M - R$.

C

Table C.3: Transitions from a generic state $\boldsymbol{x} = (i_n, j_n, f_n, i_r, j_r, f_r)$ of the channel access strategy upon channel failures and repairs.

Activity	Destination State	Tran. rate	Conditions
15. A PU occupied channel in the N-CRN fails. The interrupted PU accesses a channel in R-CRN.	$(i_n - 1, j_n, f_n + 1, i_r + 1, j_r, f_r)$	$i_n\lambda_F$	$i_n > 0, B_n(\boldsymbol{x}) = M - R, j_n = 0, B_r(\boldsymbol{x}) < R$.
16. A PU occupied channel in the N-CRN fails. SU_R is forced to terminate.	$(i_n - 1, j_n, f_n + 1, i_r + 1, j_r - 1, f_r)$	$i_n\lambda_F$	$i_n > 0, B_n(\boldsymbol{x}) = M', j_n = 0, j_r > 0, B_r(\boldsymbol{x}) = R$.
17. A PU occupied channel in the N-CRN fails. The interrupted PU service is forced to terminate.	$(i_n-1, j_n, f_n+1, i_r, j_r, f_r)$	$i_n\lambda_F$	$i_n > 0, B_n(\boldsymbol{x}) = M', j_n = 0, j_r = 0, B_r(\boldsymbol{x}) = R$.
18. An SU occupied channel in the N-CRN fails. A vacant channel exists in the N-CRN.	$(i_n, j_n, f_n + 1, i_r, j_r, f_r)$	$j_n\lambda_F$	$j_n > 0, B_n(\boldsymbol{x}) < M' - 1$.
19. An SU occupied channel in the N-CRN fails. The interrupted SU accesses a channel in R-CRN.	$(i_n, j_n - 1, f_n + 1, i_r, j_r + 1, f_r)$	$j_n\lambda_F$	$j_n > 0, B_n(\boldsymbol{x}) = M', B_r(\boldsymbol{x}) < R$.
20. An SU occupied channel in the N-CRN fails. The interrupted SU service is forced to terminate.	$(i_n, j_n-1, f_n+1, i_r, j_r, f_r)$	$j_n\lambda_F$	$j_n > 0, B_n(\boldsymbol{x}) = M', B_r(\boldsymbol{x}) = R$.
21. A PU occupied channel in the R-CRN fails. The interrupted PU accesses another channel in R-CRN.	$(i_n, j_n, f_n, i_r, j_r, f_r + 1)$	$i_r\lambda_F$	$i_r > 0, j_n = 0, B_n(\boldsymbol{x}) = M', B_r(\boldsymbol{x}) < R$.
22. A PU occupied channel in the R-CRN fails. SU_R is forced to terminate.	$(i_n, j_n, f_n, i_r, j_r-1, f_r+1)$	$i_r\lambda_F$	$i_r > 0, j_r > 0, j_n = 0, B_n(\boldsymbol{x}) = M', B_r(\boldsymbol{x}) = R$.
23. A PU occupied channel in the R-CRN fails. The interrupted PU service is forced to terminate.	$(i_n, j_n, f_n, i_r-1, j_r, f_r+1)$	$i_r\lambda_F$	$i_r > 0, j_r = 0, j_n = 0, B_n(\boldsymbol{x}) = M', B_r(\boldsymbol{x}) = R$.
24. An SU occupied channel in the R-CRN fails. The interrupted SU accesses another channel in the R-CRN.	$(i_n, j_n, f_n, i_r, j_r, f_r + 1)$	$j_r\lambda_F$	$j_r > 0, B_n(\boldsymbol{x}) = M', B_r(\boldsymbol{x}) < R$.
25. An SU occupied channel in the R-CRN fails. The interrupted SU service is forced to terminate.	$(i_n, j_n, f_n, i_r, j_r-1, f_r+1)$	$j_r\lambda_F$	$j_r > 0, B_n(\boldsymbol{x}) = M', B_r(\boldsymbol{x}) = R$.
26. A failed channel in the N-CRN is repaired. PU_R performs spectrum handover.	$(i_n + 1, j_n, f_n - 1, i_r - 1, j_r, f_r)$	$f_n\mu_R$	$i_r > 0, f_n > 0$.
27. A failed channel in the N-CRN is repaired. SU_R performs spectrum handover.	$(i_n, j_n + 1, f_n - 1, i_r, j_r - 1, f_r)$	$f_n\mu_R$	$i_r = 0, j_r > 0, f_n > 0$.
28. A failed channel in the R-CRN is repaired. No spectrum adaptation.	$(i_n, j_n, f_n, i_r, j_r, f_r - 1)$	$f_r\mu_R$	$f_r > 0$.

Paper D

Title: Reliability Analyses of Dynamic Spectrum Reservation Strategies for CR Networks under Random Channel Failure and Recovery

Authors: Indika A. M. Balapuwaduge[†], Frank Y. Li[†], and Vicent Pla[‡]

Affiliation: [†] Dept. of Information and Communication Technology, University of Agder (UiA), N-4898 Grimstad, Norway
[‡] Dept. of Communications, Universitat Politècnica de València (UPV), 46022 València, Spain

Journal: Submitted to *IEEE Transactions on Wireless Communications*, March 2016.

D

Dynamic Spectrum Reservation for CR Networks under Random Channel Failure and Recovery: Medium Access and Reliability Analyses

Indika A. M. Balapuwaduge, Frank Y. Li, and Vicent Pla

Abstract — **Providing channel access opportunities for new service requests and guaranteeing a continuous connection for ongoing flows until their service completion are two essential challenges for providing quality of service (QoS) in wireless networks. Channel state measurements in wireless networks have shown that channel failures, which are typically caused by hardware and software failures or/and by intrinsic instability in radio transmissions, can easily result in network performance degradation. In cognitive radio networks (CRNs), secondary transmissions are inherently vulnerable to connection breaks due to licensed users' arrivals as well as channel failures. To explore the advantages of channel reservation on performance improvement in error-prone channels, we propose and analyze in this paper a dynamic channel reservation (DCR) algorithm and a dynamic spectrum access (DSA) scheme. The key idea of the proposed DCR algorithm is to reserve a number of channels exclusively for the interrupted services in order to avoid forced terminations. We make the number of reserved channels as dynamically adjustable according to the ongoing traffic load in the network to enhance channel availability for new users and to maintain service retainability for ongoing users. Furthermore, the performance of CRNs in the presence of heterogeneous channel failures is also investigated.**

Keywords—Cognitive radio networks, dynamic channel reservation, CTMC, retainability, network unserviceable probability

I. Introduction

The current global research efforts on 5th generation (5G) mobile communications have identified the need for large extent improvements of accessibility and reliability of communication services. With limited bandwidth and static spectrum allocations, current cellular networks cannot achieve this target unless more flexible and dynamic spectrum access is enabled. Cognitive radio (CR) allows future

Manuscript submitted March 21, 2016.

D

wireless networks to dynamically access the licensed spectrum without causing interference to incumbent users [2], [3] and it is considered as a key component in the 5G paradigm which can tackle the challenges such as ultra reliable communication (URC). In the context of 5G, URC refers to the provision of certain level of communication services with high degrees of availability and reliability [4]. Therefore performance measures related to dependability attributes such as reliability and availability of CR networks (CRNs) are of major importance to its successful operation in future 5G wireless networks.

When compared with other wireless networks, CRNs are more prone to channel access failures from the secondary network's point of view. It is therefore less predictable in terms of quality of service (QoS) experienced by users. This feature poses challenges to the provisioning of a dependable service that meets the CR users' requirements on reliability and availability. Traditionally, service level performance evaluation of CRNs has been addressed merely by considering resource insufficiency, where the impact of wireless channel failures and recovery has been largely overlooked partially due to the complexity of analyses. In wireless networks including CRNs, a connection can fail owing to two fundamental reasons. That is, 1) hardware failures or software malfunctioning, and 2) intrinsic features in radio transmissions, such as channel fading and shadowing [5]. Owing to those failures, the network capacity or the number of active subscribers that the network can support may decrease and the response time for a service request may degrade. Therefore traditional performance models that ignore failures and recovery of channels generally overestimate network capacity and other performance measures. To perform realistic reliability and availability analysis, it is required to consider performance changes that are associated with channel failures and recovery processes. Consequently, the goal of this paper is to study the performance of dynamic spectrum access (DSA) together with dynamic channel reservation (DCR) in CRNs with a focus on the impact of channel failure and their recovery.

In a CRN, a primary user (PU) always has priority for channel access while a secondary user (SU) can only opportunistically access the idle channels. When there are no idle channels at a given instant, a new secondary request will simply be blocked. An ongoing SU connection could also be dropped upon PU arrivals or channel failures. Typically, forced termination of PU sessions has not been considered in CRNs since ongoing PU sessions cannot be preempted by new user arrivals. Once channel failures occur, however, an ongoing PU session can also be terminated before service completion [6]. Thus, *channel availability* and *service retainability* which describe the properties of a system from a *dependability perspective* are two important metrics for performance evaluation in a CRN [7]. Note that the *dependability* of a communication system is defined as its ability to deliver services that can justifiably be trusted, when faced with failures of their components [8], [9]. The reason that a system does not behave as it is specified can be, for instance, any fault in its design, or the failure of some of its components when facing with unpredicted changes in the system's environment. The theory to analyze this type of phenomenon is known as *dependability theory* [10] in the research community. Therein, retainability refers to the ability of finishing a service completely without being terminated before its actual end [11] and it is considered as one of the key performance indicators in evolved UMTS terrestrial radio access (E-UTRAN) [12]. Therefore, retainability is a measure of service reliability which reflects how often a user abnormally losses its connection [13].

A. Related Work

The research work on CR has been largely focusing on aspects such as spectrum sensing and dynamic spectrum access [14]. Few studies which take into account channel failure when performing performance modeling exist. A majority of them considered system centric performance analysis rather than analysis from a dependability perspective. By considering a topology in a CRN, the formation of *blackholes*, i.e., explosive spreading of random failures was investigated in a recent paper [15]. Furthermore, network resilience was investigated by assuming the existence of a giant component with surviving nodes that span through the entire network. Even though correlations among failures and cascading effects are studied, the results presented in [15] are limited to the analysis on failure statistics and network percolation. Heterogeneous failures of network functions are considered in [16] when analyzing the service resilience in 5G mobile systems. Although a proactive restoration mechanism is also proposed to ensure service resilience in [16], it targets only on cloud-based mobile networks.

A cognitive relaying scheme was proposed in [17] to enhance the PU and SU performance in the event of transmission failures. However, the failures of secondary and primary links are assumed to be with a constant probability. Although the secondary throughput maximization is mainly targeted in [17], forced termination of connections has not been investigated. Instead, forced termination and blocking probability analyses were conducted in [18] considering both resource insufficiency and link unreliability. However the traffic model considered in [18] is a multi-cellular system and the analysis does not directly apply to CRNs with reserved channels. A model for file transfer over an unreliable channel was considered in [19] by taking into account of interruptions during file transfer due to server failures. Instead of restarting the interrupted file transferring from the beginning, the authors proposed a file fragmentation policy which could bound the expected completion time. However, the analysis in [19] was performed for file fragmentation and no analytical or simulation results were presented.

In addition, a few recent studies considered also channel reservation mechanisms which are applicable to CRNs. Channel reservation policies have been used for improving spectrum utilization in CRNs by considering both PUs and SUs. In [20], a PU based channel reservation policy is applied to a CRN where a suitable number of channels are initially reserved for PUs. As long as the CRN has available reserved channels, the PUs cannot occupy the unreserved channels. However, the probability of forced SU terminations may not decrease substantially since ongoing SU services in the unreserved band can still be terminated if the whole reserved band is occupied. Another DCR scheme was proposed in [21] to reduce forced termination probability while minimizing the increase of blocking probability. Since the reserved channels can also be occupied, forced terminations of SU services in the non-reserved band may not be decreased adequately [21]. As a solution to this problem, we proposed a static channel reservation (SCR) scheme in [1] which solely targeted on retainability. However, static reservation generally leads to lower channel availability for CRs [1].

This paper is motivated by the aforementioned interesting previous studies. However, the work presented in this paper is distinct from those related studies with respect to mainly *four aspects*. First, the proposed channel reservation scheme *dynamically adjusts* the number of reserved channels and it targets at minimizing forced terminations for both ongoing PU and SU services. Second, both PUs and SUs can access the reserved band in the proposed scheme. A common assumption in many previous studies on DCR in CRNs is that the reserved spectrum is occupied exclusively by the SUs or the PUs. This type of reservation would lead to considerably reduced spectrum utilization in the network. In contrast, our approach provides higher flexibility for channel access upon a sudden increase of spectrum demand of the primary network (PN) or the secondary network (SN). Third, random failures of channels are also considered to obtain more realistic results of performance metrics. We study also the heterogeneity of failures which is more realistic in real-life wireless systems. Fourth, the performance of those schemes is investigated from the perspective of dependability theory. This is because the terminology related to QoS of telecommunication services has been defined in the *ITU-T E.800 Recommendation* [22] based on the dependability theory. Correspondingly, the definitions of dependability metrics used in this study conform with the ITU recommendations.

B. Contributions

In brief, the contributions of this paper can be summarized as follows.

- A DSA scheme which enables dynamic channel reservations and spectrum access in a multi-channel CRN is proposed.
- A DCR algorithm which can dynamically adjust the reserved spectrum is proposed and it is incorporated with the DSA scheme. Two working modes are proposed in the DCR algorithm targeting at either maintaining service retainability or enhancing channel availability respectively.
- The combined effect of random channel failures and resource insufficiency is investigated while considering both homogeneous and heterogeneous channel failures when evaluating the performance of the proposed scheme.
- A continuous time Markov chain (CTMC) model is developed to evaluate the system performance of *both primary and secondary networks*. Several dependability metrics including channel availability, service retainability and network unserviceable probability are derived based on the CTMC model. Furthermore, simulations are performed to validate the correctness and the preciseness of the derived analytical models.
- Based on different performance requirements and conditions, another algorithm that can identify the optimal upper bound for the number of reserved channels in a CRN is proposed.

The remainder of this paper is organized as follows. An overview of the network scenario and the modeling assumptions are presented in Section II. In Section III, we introduce an algorithm which performs DCR and incorporate them with a spectrum access scheme for a CRN with error-prone channels. By providing theoretical analysis in Section IV, CTMC models are developed to analyze the performance of the PN and the SN. Numerical results are illustrated in Section V and heterogeneous channel failure analysis is presented in Section VI. An algorithm to find the optimal number of reserved channels is proposed in Section VII. Finally the paper is concluded in Section VIII.

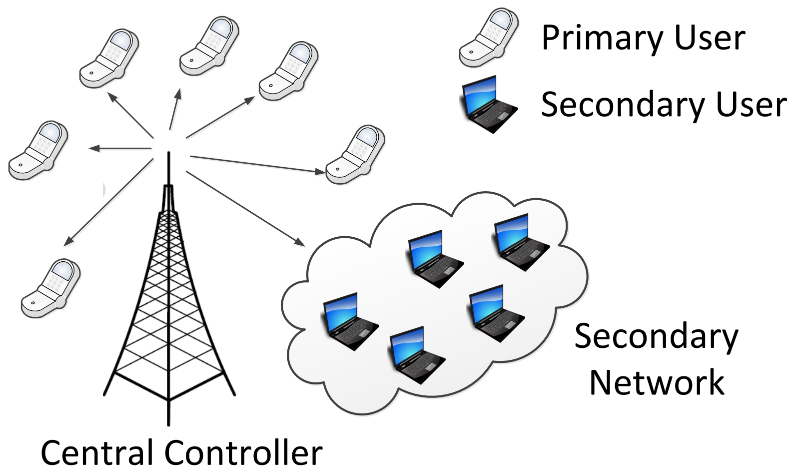


Figure D.1: A CRN with a centralized architecture.

II. Network Scenario and Assumptions

A centralized CRN architecture as illustrated in Fig. D.1 is considered in this study. It comprises two types of networks: the PN and the SN with multiple PUs and SUs. Spectrum allocation and management are performed by a central controller. To do so, the central controller needs to collect information about the spectrum utilization of both PUs and SUs as well as the channel state information about each channel of the network. Based on this information, a decision on DSA is made by the central controller.

Consider a PN operating on $M \in \mathbb{Z}^+$ equal bandwidth channels, where $\mathbb{Z}^+$ is a set of positive integers. The SN opportunistically accesses and reuses the licensed spectrum allocated to the PN. As already mentioned, channels are prone to different kinds of failures which could interrupt ongoing communication sessions. In order to avoid forced terminations, a certain, typically small, number of channels are reserved. *Those reserved channels can only be accessed by the SU and PU services that are interrupted due to channel failures or the preempted SU services upon new PU arrivals*. As depicted in Fig. D.2, $R \in \mathbb{Z}^+$ out of M channels are reserved. The set of reserved channels is denoted as R-CRN whereas the set of non-reserved channels is denoted as N-CRN.

The number of channels that can be reserved to the R-CRN (i.e., R) is dynamically adjustable depending on ongoing channel occupancy status. However, to enforce a certain degree of fairness between the newly arrived users and ongoing users, we define an upper bound R_{max} on the total number of channels that can be reserved, i.e., $R \leq R_{max}$, where $R_{max} \in \mathbb{Z}^+$. It is worth mentioning that the set of channels allocated to the N-CRN or R-CRN does not necessarily have to be con-

D

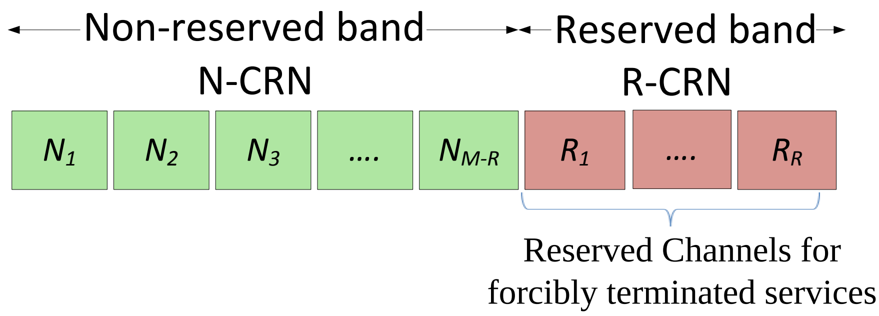


Figure D.2: Reserved (R-CRN) and non-reserved (N-CRN) channel assignment, where N_i, $i \in \{1, 2, \cdots, M-R\}$ represents the channels in the N-CRN and R_j, $j \in \{1, 2, \cdots, R\}$ represents the channels in the R-CRN.

tiguous. For ease of presentation, R-CRN channels are shown as a contiguous band in Fig. D.2. In the next section, we discuss the methodology adopted for adjusting R under different conditions of the system. In our analysis, the following assumptions are made as the basis for developing the analytical model presented in Section IV.

- The arrivals of both PU and SU services are Poisson processes with rates λ_P and λ_S for PU and SU services respectively. Moreover, the service times for PU and SU services are exponentially distributed, with corresponding service rates per channel μ_P and μ_S respectively.
- The on-time duration (during which the channel is working properly until a failure occurs) of a channel is exponentially distributed with a failure rate per channel λ_F in both N-CRN and R-CRN. Note that heterogeneous channel failures which have different failure rates and repair rates are also investigated in Section VI. Moreover, we assume that both occupied and idle channels are prone to failures.
- The repair time of a failed channel in both N-CRN and R-CRN is exponentially distributed, with a repair rate per channel μ_R. Moreover, all failed channels can be repaired simultaneously.

III. The Proposed Dynamic Spectrum Access Scheme and DCR Algorithm

The proposed DSA scheme includes a channel reservation and adjustment procedure and a channel allocation procedure upon six distinct events: PU/SU arrivals, PU/SU departures, channel failures and recovery. For channel reservation, a DCR algorithm which determines the required number of reserved channels in the R-CRN is proposed.

The proposed DCR algorithm works with two working modes. One of these working modes needs to be selected in order to adjust the number of reserved channels before channel allocation to the newly arrived users. Herein, we explain the proposed DCR algorithm before presenting the proposed channel allocation scheme.

Algorithm D.1: Dynamic channel reservation (DCR) algorithm with two working modes.

Input: M : Total number of channels in the whole CRN
Input: i_n : Total number of channels occupied by PUs in the N-CRN
Input: j_n : Total number of channels occupied by SUs in the N-CRN
Input: i_r : Total number of channels occupied by PUs in the R-CRN
Input: j_r : Total number of channels occupied by SUs in the R-CRN
Input: f : Total number of failed channels in the whole CRN
Input: R_{max} : The upper bound of the number of reserved channels
Input: $a_i : a_{k+1} < a_k < \ldots < a_i < \ldots < a_1 < a_0; 0 < a_i < 1,$
$i = 1, \ldots, k; k \in \mathbb{Z}^+; a_0 = 1, a_{k+1} = 0$
Output: R : The allowable number of reserved channels that will be allocated to the R-CRN

Calculate $\Phi = (i_n + j_n + i_r + j_r)/(M - f)$
Calculate $N_{Avail} = M - (i_n + j_n + i_r + j_r + f)$
Calculate $N_R = i_r + j_r$
for $i = 0 : 1 : k$ **do**
    if $(a_{i+1} \leq \Phi < a_i)$ **then**
        traffic_load_level= i
        break;
    end
end
Select a working mode: *Working mode* $= 1$ intends to enhance the retainability of ongoing services while *Working mode* $= 2$ increases channel availability for new users.
if *Working mode* $= 1$ **then**
    $R' = R_{max}-$traffic_load_level
end
if *Working mode* $= 2$ **then**
    $R' = R_{max} - (k-$traffic_load_level)
end
$R' = \max\{R', N_R\}$: $R' \in \mathbb{Z}^+$ *is a local variable which is created to store the selected value of* R *where* $0 \leq R' \leq R_{max}$
$R = \min\{N_{Avail} + N_R, R'\}$

A. Dynamic Channel Reservation Algorithm

The proposed algorithm is designed with two different working modes considering distinct reliability aspects: service retainability and channel availability. The ongoing traffic load is taken as the basis for determining R in Algorithm D.1. In this study, the ongoing traffic load, Φ, is calculated as the ratio between the number of ongoing PU and SU services in the *whole CRN* and the number of operational channels (see Line 1 of Algorithm D.1). Note that an *operational channel* indicates a channel that is not in a failed state and each operational channel can be either occupied or idle.

The design principle for resource reservation in the R-CRN with respect to the two working modes in the proposed algorithm is explained as follows.

1. *Working mode 1* intends to maintain and enhance the *retainability* of ongoing services in the CRN subject to the channel reservation constraint R_{max} and the ongoing traffic load. To improve retainability, *working mode 1* allocates a higher number of channels to the R-CRN when the ongoing traffic load becomes heavier. This is reasonable from the retainability perspective since forced terminations are more likely to occur at a higher traffic load.

2. Conversely, resource reservation in the R-CRN in *working mode 2* is performed for the purpose of reducing blocking probability of new users. In other words, the improvement of channel availability for new users is of most interest in this case. Accordingly when the ongoing traffic load is higher, *the network reserves fewer channels to the R-CRN in working mode 2*.

Hereafter, we refer to *working mode 1* and *working mode 2* in short as *mode 1* and *mode 2* respectively. In both working modes, the ongoing traffic load is characterized by $k+1$ levels, where $k \in \mathbb{Z}^+$. To distinguish these levels, k configurable parameters, $a_1, a_2, \ldots, a_k$ are required as input parameters to Algorithm D.1 such that $0 < a_k < a_{k-1} < \cdots < a_2 < a_1 < 1$. For instance, consider a scenario with $k = 2$. The ongoing traffic load is characterized by three levels as high, medium and low traffic levels. Accordingly, we configure $a_1 = 0.70$ and $a_2 = 0.35$. Therefore, the high, medium and low traffic loads indicate that $\Phi \geq 0.7$, $0.35 \leq \Phi < 0.7$ and $\Phi < 0.35$ respectively.

The basic channel reservation procedure is shown in Algorithm D.1. Recall that Algorithm D.1 is run prior to channel allocation upon PU and SU service arrivals. As the initial step, the ongoing traffic load, Φ, is calculated on Line 1 and then corresponding a traffic load level is determined. If Φ is higher, *mode 1* tends to allocate a larger value of R, i.e., a value closer to R_{max}. However, *once the traffic* *load decreases, the parameter, traffic_load_level increases thus* mode 1 *assigns a smaller value of* R as indicated at Line 12. By allocating narrower spectrum to the R-CRN at low traffic loads, it is able to avoid spectrum under-utilization of the reserved band. Otherwise, the allocation of additional channels to the R-CRN will not really help much to decrease the rate of forced terminations at low traffic loads.

Unlike in *mode 1*, *mode 2* assigns a lower value of R if the the traffic load increases, as indicated in Line 15. This is for the purpose of increasing channel access opportunity for new users. On the other hand, when the current traffic load falls within lower traffic levels, *mode 2* tends to allocate a higher value of R, i.e., closer to R_{max}. With appropriate configurations of parameters a_i, spectrum under-utilization can be minimized.

To avoid spectrum handover of ongoing services from the R-CRN to the N-CRN and to keep the value of R' positive, the algorithm selects the maximum value between R' and N_R as indicated in Line 17. Here N_R is the number of ongoing services in the R-CRN. In this study, spectrum handover from the R-CRN to the N-CRN is not needed according to the reservation criterion. The last line of the algorithm inspects the availability of idle channels in the CRN in order to *adjust the reserved band* according to the value R' obtained in Line 17 and to determine the appropriate value for R.

The selection of an appropriate working mode depends on channel failure and recovery rates, the QoS requirements of the PN and the SN, as well as the arrival rates of users. For instance, when most of the SUs demand low forced termination probability while operating in error-prone channels, *mode 1* would be recommended.

B. Dynamic Spectrum Access Scheme

In the following, we provide an overview of the proposed DSA scheme which is operated along with the DCR algorithm proposed above. It is worth reiterating that in the proposed DSA scheme, the reserved channels could be allocated *only for the interrupted services* to avoid forced terminations. That is, a newly arrived service will not be assigned to the R-CRN no matter it is a PU or an SU. The reason for applying such a restriction is due to the fact that maintaining a higher retainability level for established connections is one main QoS requirement. The proposed channel access scheme is explained below upon the occurrence of six events. The DSA scheme with those six events is partly illustrated in a flow chart in Fig. D.3. The step where the operation of the DCR algorithm is required is shown by the process block *X* in the flow chart.

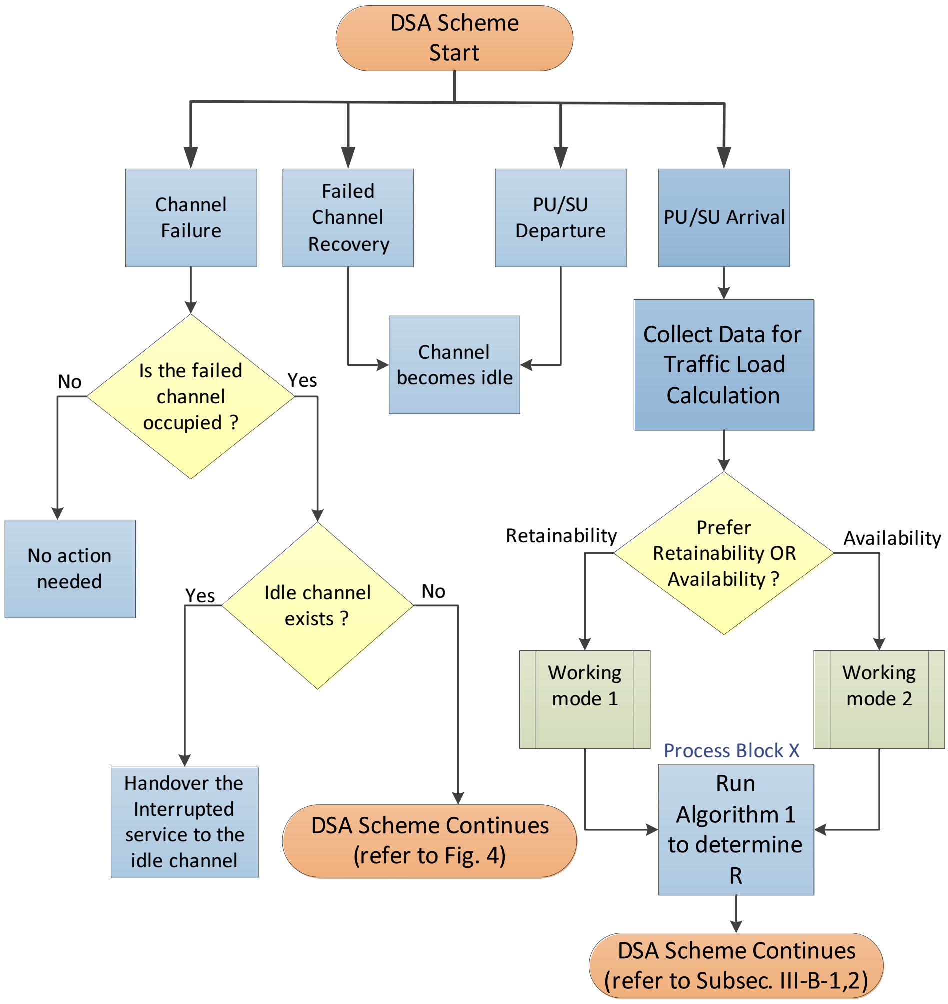


Figure D.3: Flow chart illustration of the proposed DSA scheme in part.

1) PU Arrival: First, the DCR algorithm presented in Subsection III-A is run to determine the corresponding R and adjust the R-CRN accordingly. If there is an idle channel in the N-CRN, a newly arrived PU will commence transmission on that idle channel. In case where the whole N-CRN is occupied, one of the SU services in the N-CRN is interrupted. It performs spectrum handover to a vacant channel in the R-CRN and a channel is released for access by the new PU. This interrupted SU service is allowed to access the R-CRN since it is subject to a forced termination upon the PU arrival. As mentioned earlier, the channels in the R-CRN are exclusively reserved for the ongoing services which are interrupted and cannot find another channel in the N-CRN. If the interrupted SU service cannot find an idle channel in the R-CRN, it is forced to terminate. When all operational channels in the N-CRN are occupied by PUs, a new PU request will be blocked.

D

2) SU Arrival: The same as performed in PU arrivals, the system adjusts the number of channels in the N-CRN and the R-CRN by adopting Algorithm D.1. Upon an SU arrival, the system will allocate an idle channel to the new SU from the N-CRN if it exists. If all operational channels in the N-CRN are occupied either by PUs or/and SUs, the new SU request is blocked without considering the channel availability in the R-CRN. Since PUs have priority over the channels in the N-CRN, a newly arrived SU cannot preempt other ongoing services.

3) PU departure: No handover actions are performed after a departure of PU services from the N-CRN or the R-CRN. Since the R-CRN is reserved for interrupted services, it is not reasonable to perform handover of services in the R-CRN to the N-CRN upon service departures in the N-CRN. The number of available channels in the network is accordingly increased by one.

4) SU departure: Similar to the above mentioned PU departure event, no channel adjustment is required after a departure of an SU service. The network allows ongoing services in the R-CRN to complete their services in the reserved band without performing spectrum handover back to the N-CRN.

5) Channel failure: If an idle channel in the N-CRN or R-CRN failed, the number of available channels in the network is accordingly decreased by one. However, when *an occupied channel in the N-CRN or R-CRN fails*, a spectrum handover procedure is required, as illustrated in Fig. D.4. Upon a failure of an already occupied channel, the handover process is performed based on the priority level assignment for PU and SU services in the N-CRN and the R-CRN. In this study, the ongoing PU services in the R-CRN are considered as the highest priority class while the SU services in the R-CRN are considered as the second highest priority class. The PU and SU services in the N-CRN are regarded as the third and fourth priority classes respectively. Let $PL(s)$ denote the priority level assigned to a service type, s. Therefore, $PL(PU_{R-CRN}) > PL(SU_{R-CRN}) > PL(PU_{N-CRN}) > PL(SU_{N-CRN})$ where $PL(PU_{R-CRN})$ denotes the priority level of an PU service in the R-CRN and so on. Since the ongoing SU services in the reserved band are considered with higher priority than the PU services in the N-CRN, the PU services in the N-CRN cannot preempt an SU service in the R-CRN upon a channel failure in the R-CRN. Moreover, this priority level assignment is independent of the selected working mode of the DCR algorithm.

6) Channel repair: Once a failed channel is re-established to its normal conditions and ready for carrying services, it is said that the channel is repaired or recovered. However, as operational channels, ongoing services do not perform spectrum handover when a failed channel is recovered.

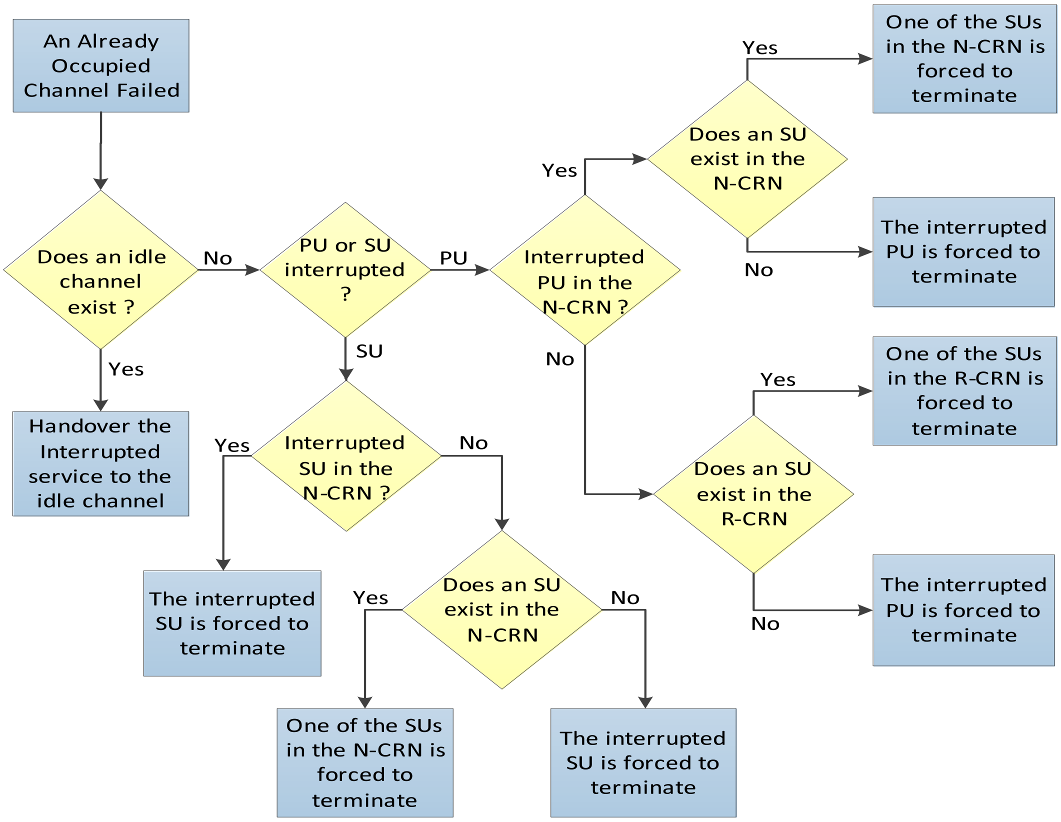


Figure D.4: Flow chart illustration of the proposed DSA scheme when a currently occupied channel failed.

IV. CTMC Modeling and Performance Metrics

We model the considered network and the proposed DSA scheme using a Markov chain, with continuous time and discrete states. Let $\mathcal{S}$ be the set of feasible states in the CTMC model. The states of the CTMC model corresponding to the DSA scheme are represented by $\boldsymbol{x} = (i_n, j_n, i_r, j_r, f)$ where i_n and j_n denote the number of PU and SU services in the N-CRN, i_r and j_r denote the number of those services in the R-CRN respectively and f denotes the number of failed channels in the whole CRN. The total numbers of occupied channels in the N-CRN and the R-CRN for a given state $\boldsymbol{x}$ are denoted by $B_n(\boldsymbol{x})$ and $B_r(\boldsymbol{x})$ respectively, i.e., $B_n(\boldsymbol{x}) = i_n + j_n$ and $B_r(\boldsymbol{x}) = i_r + j_r$. Moreover, denote the sum of occupied plus failed channels in the whole CRN as $B(\boldsymbol{x})$, i.e., $B(\boldsymbol{x}) = B_n(\boldsymbol{x}) + B_r(\boldsymbol{x}) + f$. Therefore, the number of idle channels in state $\boldsymbol{x}$ can be obtained as $M - B(\boldsymbol{x})$. Note that the number of reserved channels in state $\boldsymbol{x}$ is denoted as $R(x)$. The state transition rates associated with different events are summarized with different conditions in Table D.1. The transitions and transition rates mentioned in this table constitute the transition rate matrix $\boldsymbol{Q}$.

D

D

Table D.1: Transitions from a generic state $\boldsymbol{x} = (i_n, j_n, i_r, j_r, f)$ of the channel access scheme upon PU/SU events and channel failures and repairs.

Event	Destination State	Tran. rate	Conditions
1. PU AR. A vacant channel exists in the N-CRN.	$(i_n+1, j_n, i_r, j_r, f)$	λ_P	$B_n(\boldsymbol{x}) < M - R(x)$.
2. PU AR. No vacant channel exists in the N-CRN. SU_N performs handover to the R-CRN.	$(i_n+1, j_n-1, i_r, j_r+1, f)$	λ_P	$B_n(\boldsymbol{x}) = M - R(x); j_n > 0; B_r(\boldsymbol{x}) < R(x)$.
3. PU AR. An SU_N is forced to terminate.	$(i_n+1, j_n-1, i_r, j_r, f)$	λ_P	$B_n(\boldsymbol{x}) = M - R(x); j_n > 0; B(\boldsymbol{x}) = M$.
4. SU AR. A vacant channel exists in the N-CRN.	$(i_n, j_n+1, i_r, j_r, f)$	λ_S	$B_n(\boldsymbol{x}) < M - R(x); B(\boldsymbol{x}) < M$.
5. PU DP from the N-CRN.	$(i_n-1, j_n, i_r, j_r, f)$	$i_n\mu_P$	$i_n > 0$.
6. PU DP from the R-CRN.	$(i_n, j_n, i_r-1, j_r, f)$	$i_r\mu_P$	$i_r > 0$.
7. SU DP from the N-CRN.	$(i_n, j_n-1, i_r, j_r, f)$	$j_n\mu_S$	$j_n > 0$.
8. SU DP from the R-CRN.	$(i_n, j_n, i_r, j_r-1, f)$	$j_r\mu_S$	$j_r > 0$.
9. Idle channel failure.	$(i_n, j_n, i_r, j_r, f+1)$	$(M - B(\boldsymbol{x}))\lambda_F$	$B(\boldsymbol{x}) < M$.
10. An occupied channel fails. An idle channel exists in the CRN.	$(i_n, j_n, i_r, j_r, f+1)$	$(B(\boldsymbol{x}) - f)\lambda_F$	$f < B(\boldsymbol{x}) < M$.
11. An occupied channel fails. No idle channels exist in the CRN. An SU_N is forced to terminate.	$(i_n, j_n-1, i_r, j_r, f+1)$	$(M-f)\lambda_F$	$B(\boldsymbol{x}) = M; j_n > 0$
12. An occupied channel fails. No idle channels exist in the CRN. A PU_N is forced to terminate.	$(i_n-1, j_n, i_r, j_r, f+1)$	$(M-f)\lambda_F$	$B(\boldsymbol{x}) = M; j_n = 0; i_n > 0$.
13. An occupied channel fails. No idle channels exist in the CRN. An SU_R is forced to terminate.	$(i_n, j_n, i_r, j_r-1, f+1)$	$(M-f)\lambda_F$	$B(\boldsymbol{x}) = M; B_n(\boldsymbol{x}) = 0; j_r > 0$.
14. An occupied channel fails. No idle channels exist in the CRN. A PU_R is forced to terminate.	$(i_n, j_n, i_r-1, j_r, f+1)$	$(M-f)\lambda_F$	$B(\boldsymbol{x}) = M; B_n(\boldsymbol{x}) = j_r = 0; i_r > 0$.
15. A failed channel is repaired.	$(i_n, j_n, i_r, j_r, f-1)$	$f\mu_R$	$f > 0$.

The notations AR and DP indicate an arrival event and a departure event respectively. An SU service and a PU service in the N-CRN are denoted as SU_N and PU_N respectively. An SU service and a PU service in the R-CRN are denoted as SU_R and PU_R respectively.

Let $\pi(\boldsymbol{x})$ denote the steady state probability of being in state $\boldsymbol{x}$. The steady state probabilities of each state can be calculated according to

$$\boldsymbol{\pi Q} = \mathbf{0}, \ \sum_{\boldsymbol{x}\in\mathcal{S}} \pi(\boldsymbol{x}) = 1, \tag{D.1}$$

where $\boldsymbol{\pi}$ is the steady state probability vector and $\mathbf{0}$ is a row vector of all 0's. In the following, we derive mathematical expressions to analyze performance metrics of the CRN.

A. Capacity

In this study, *capacity* is defined as the rate of service completions, i.e., the average number of service completions per time unit. Let ρ_P and ρ_S be the capacity of PU and SU services respectively. Correspondingly, we obtain

$$\rho_P = \sum_{\boldsymbol{x}\in\mathcal{S}} (i_n + i_r)\mu_P \pi(\boldsymbol{x}), \tag{D.2}$$

$$\rho_S = \sum_{\boldsymbol{x}\in\mathcal{S}} (j_n + j_r)\mu_S \pi(\boldsymbol{x}). \tag{D.3}$$

B. Channel Availability

In a CRN, PUs have priority for channel access while SUs can only opportunistically access the idle channels. Once all the channels are occupied at a given instant, a new user request will simply be blocked and the CRN is said to be unavailable for new users. Thus, channel availability based measurements are of importance for both PUs and SUs when performing channel access in a CRN [23]. In this paper, channel availability for PU or SU services is defined as *the probability that the CRN will allocate a channel to a new PU or SU arrival* without blocking the request. Blocking of a new PU service occurs when all the operational channels in the N-CRN are occupied by PUs. Let A_P denote the channel availability of PU services. We obtain

$$A_P = 1 - \sum_{\substack{\boldsymbol{x}\in\mathcal{S} \\ (B(\boldsymbol{x})=M \text{ or } B_n(\boldsymbol{x})=M-R(x));\ \ j_n=0}} \pi(\boldsymbol{x}). \tag{D.4}$$

Moreover, an SU service will be blocked if all the operational channels in the N-CRN are occupied by PUs or/and SUs. Therefore, the channel availability of SU services can be obtained by

$$A_S = 1 - \sum_{\substack{\boldsymbol{x}\in\mathcal{S} \\ B(\boldsymbol{x})=M \text{ or } B_n(\boldsymbol{x})=M-R(x)}} \pi(\boldsymbol{x}). \tag{D.5}$$

Accordingly, the blocking probabilities of PU and SU services, denoted as P_P^B and P_S^B respectively, are obtained as

$$P_P^B = 1 - A_P \text{ and} \tag{D.6}$$

$$P_S^B = 1 - A_S. \tag{D.7}$$

C. Retainability

A session-oriented service often explicitly measures the probability that a session delivers service with an acceptable level of quality until the session terminates [24].

1) Retainability definition: According to [142], retainability is defined as the probability that a connection, once established, will operate within specified transmission tolerance without interruption for a given time interval. Mathematically, the retainability of a service, θ, is expressed as

$$\theta = 1 - P_F \tag{D.8}$$

where P_F is the forced termination probability of that service.

2) Retainability of the SN: The probability of forced termination represents the probability that an ongoing SU service in the network is forced to terminate before its communication is finished [25]. Note that when considering random channel failures, forced terminations could happen to both SU and PU services. The forced termination probability of SUs, P_F^S, can be expressed as the ratio between the mean forced termination rate of SU services, P_R^{SE}, and the effective rate in which a new SU service is assigned a channel, Λ_S [25]. Denote the rate of forced terminations of SUs due to PU arrivals as R_S. Then we have

$$R_S = \lambda_P \sum_{\substack{\boldsymbol{x} \in \mathcal{S} \\ B(\boldsymbol{x})=M;\ j_n>0}} \pi(\boldsymbol{x}). \tag{D.9}$$

In addition, ongoing SU services can also be terminated upon a channel failure when all other channels in the CRN are busy. Denote the rate of forced terminations of SUs due to channel failures as $R_S^{'}$. It is obtained by

$$R_S^{'} = \lambda_F \sum_{\substack{\boldsymbol{x} \in \mathcal{S} \\ B(\boldsymbol{x})=M; \\ ((j_n>0) \text{ or } (B_n(\boldsymbol{x})=0;\ j_r>0))}} (M - f)\pi(\boldsymbol{x}). \tag{D.10}$$

D

Since the effective rate in which a new SU service is assigned a channel is $\Lambda_S = A_S\lambda_S$, we have $P_S^F = \frac{R_S+R_S^{'}}{\Lambda_S}$. Correspondingly, the retainability of SU services, θ_S, can be expressed as

$$\theta_S = 1 - \left(\frac{R_S + R_S^{'}}{\Lambda_S}\right). \tag{D.11}$$

3) Retainability of the PN: Similarly, the forced termination probability of PU services due to channel failures, P_P^F, can be expressed as

$$P_P^F = \frac{R_P + R_P^{'}}{\Lambda_P}, \tag{D.12}$$

where $R_P^{'}$ and Λ_P are given by

$$R_P^{'} = \lambda_F \sum_{\substack{\boldsymbol{x}\in\mathcal{S} \\ B(\boldsymbol{x})=M; \\ ((j_n=0;\ i_n>0)\ \text{or}\ (B_n(\boldsymbol{x})=0;\ j_r=0;\ i_r>0))}} (M-f)\pi(\boldsymbol{x}) \tag{D.13}$$

and $\Lambda_P = A_P\lambda_P$ respectively. Note that R_P, which denotes the forced termination rate of PUs due to new user arrivals, always equals zero since none of the ongoing PUs can be terminated due to the arrivals of new users. Therefore, the retainability of PU services, θ_P, is given by

$$\theta_P = 1 - \frac{R_P^{'}}{\Lambda_P}. \tag{D.14}$$

D. Network Unserviceable Probability (NUP)

Network unserviceable probability is an uncharted item in CRN research. From the viewpoint of users in the network, the service can be accomplished only if the network provides a connection *when it is required and as long as it is required* [26]. Thus the CRN is unserviceable if either of the following conditions happens:

1. No channel is available upon a service arrival due to failed channels and/or traffic congestion.

2. The channel allocated for the connection failed before the session is finished due to channel failures (for both PUs and SUs) or PU arrivals (for SUs only).

Accordingly, the NUP for SU services, Q_S, can be defined as the probability that an SU service cannot be completed successfully. It is obtained by calculating the ratio between the rate of service completions and the rate of arrivals as follows

$$
\begin{aligned}
Q_S &= 1 - (\text{Prob. of successfully finishing an SU service}), \\
&= 1 - \frac{\lambda_S(1 - P_S^B)(1 - P_S^F)}{\lambda_S}, \\
&= P_S^B + P_S^F - P_S^B P_S^F. \qquad \text{(D.15)}
\end{aligned}
$$

From (D.15), it is clear that both blocking probability and forced termination probability are reflected when evaluating NUP. Similarly, the NUP for PU services, denoted as Q_P, can be derived as follows.

$$Q_P = P_P^B + P_P^F - P_P^B P_P^F. \qquad \text{(D.16)}$$

Generally speaking, the lower NUP, the better from a service provisioning's perspective in a network. Moreover, there is a tradeoff between channel availability and service retainability when performing the proposed DCR algorithm. That is, a larger R reduces channel availability for newly arrived users but leads to higher retainability for ongoing services. Therefore, deriving a metric which can evaluate the joint effect of those performance measures will help us assess the overall performance of the proposed DCR algorithm. According to (D.15) and (D.16), the effects of both channel availability and service retainability are considered when evaluating the NUP.

Note that, the expressions for performance measures deduced above are derived as general expressions and therefore they are applicable to any scenario that employs the proposed schemes given that $M > 1$ and $R_{max} < M$.

V. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we first provide numerical results to investigate the impact of channel failures and traffic conditions on the above derived performance metrics based on the proposed channel reservation algorithm. Subsequently, we study optimal channel reservation given a set of conditions in the next section. The default configurations of the network parameters are listed in Table D.2 unless otherwise stated. Four traffic load levels are considered when calculating R. Correspondingly, we configure $a_1 = 0.75, a_2 = 0.50$ and $a_3 = 0.25$ in Algorithm D.1. Note that the number of channels allocated to the R-CRN in the static channel reservation is denoted as R in the presented numerical results. The curves representing *Without channel reservation* in this section are obtained by setting $R = 0$.

Furthermore, the channel failure rate is configured as 0.05 failures per time unit which guarantees a considerably lower rate compared with the arrival rates of PU and SU services. For instance, when the PU arrival rate is 10 flows per time unit,

Table D.2: Summary of symbols in the CTMC and value configuration

Symbol	Description	Value
M	Total number of channels in the CRN	8
R_{max}	Upper bound of the number of reserved channels in DCR	2 or 4
R	Number of reserved channels in SCR	2
λ_P, λ_S	PU and SU arrival rates	5 services per unit of time
μ_P, μ_S	PU and SU service rates per channel	2 services per unit of time
λ_F	Failure rate of channels	0.05 failures per unit of time
μ_R	Repair rate per channel	1.0 repairs per unit of time

we can expect that the possibility of occurring a single channel failure is after every 200 arrivals of PU services. On the other hand, in our configurations, the average session duration for a service is 0.5 time units whereas the mean time between two consecutive failures is 20 time units. Therefore, a termination of a service upon a channel failure does not frequently occur. On average, it is less likely to have frequent channel failures in a communication network unless there is a severe breakdown in a situation such as a natural disaster or a highly distorted channel environment. Therefore, it is necessary to be cautious when selecting values for λ_F and μ_R.

A. Capacity of the PN and the SN

In Fig. D.5 and Fig. D.6, we plot the achieved capacity for the SN and the PN respectively as a function of the PU arrival rate considering both working modes in the DCR algorithm with different configurations of R_{max}. Note that the channel failure rate is kept constant in this scenario, i.e., $\lambda_F = 0.05$. From Fig. D.5, we notice that the capacity of the SN decreases when channel reservation is adopted, however, only when the PU arrival rate is lower than approximately 10 arrivals per time unit. Without employing channel reservation, the SN achieves the highest capacity given that λ_P is small.

At a lower λ_P, an SU service is unlikely to be terminated forcibly and consequently the reserved channels are rarely accessed. On the other hand, the incoming SU services cannot access the R-CRN although there are vacant channels in the reserved band, regardless of the working mode adopted. Therefore, most of the reserved channels are under-utilized and this fact leads to capacity degradation compared with the case without channel reservation. Furthermore, at a higher λ_P, the forced termination probability of SUs increases and some of the interrupted services will get channel access opportunities in the reserved band to continue their services

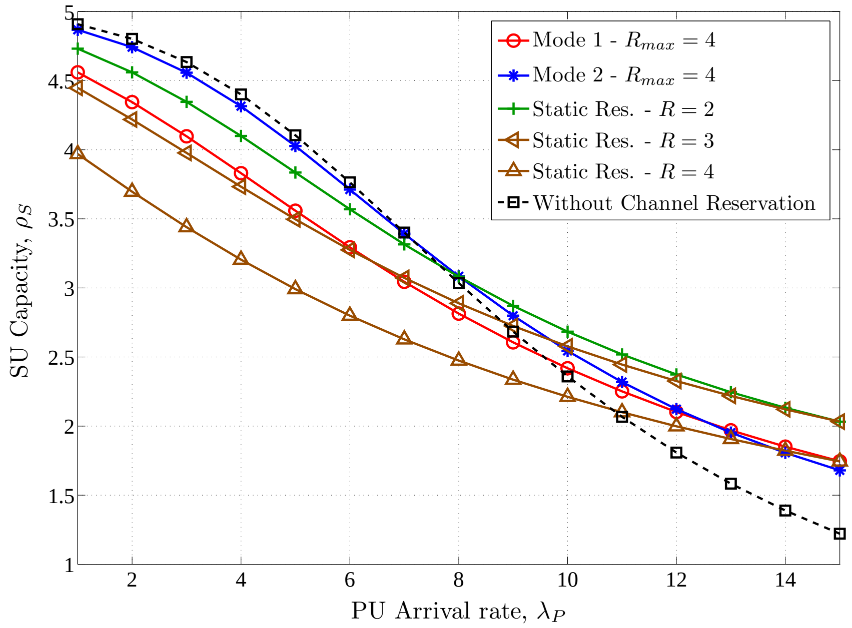


Figure D.5: Capacity achieved by SUs as a function of the PU arrival rate.

until completion. Consequently the capacity of the SN with DCR is increased compared with the case without channel reservation. Accordingly, we can conclude that the advantage of DCR over the scheme without channel reservation in terms of the achieved SU capacity is evident when λ_P is higher, i.e., $\lambda_P > 9.5$ for mode 1 and $\lambda_P > 7$ for mode 2, as shown in Fig. D.5.

Fig. D.6(a) shows the capacity of the PN as λ_P varies. The performance diversity of the proposed working modes is shown more clearly in this figure. It is observed that the capacity achieved in the PN becomes higher in mode 2 than in mode 1 due to the following reasons. The number of reserved channels, R, in mode 1 is increased when the ongoing traffic load in the network rises as illustrated in Fig. D.6(b). In contrast, mode 2 decreases R with an increasing traffic load. This is because, for a given high value of λ_P, the system allocates a comparatively larger number of channels to the R-CRN in mode 1 than in mode 2. Consequently, new PUs lose channel access opportunities in mode 1, leading to a capacity reduction when compared with mode 2. The advantage of forced termination reduction is not significant for the PN since PU services have priority over SU services for channel access in the N-CRN.

Let us further analyze the SU capacity in Fig. D.5. It is shown that mode 2 can outperform mode 1 only if $\lambda_P < 13$. Again, this is due to the priority consideration for channel access opportunities between new SU requests and ongoing SU

D

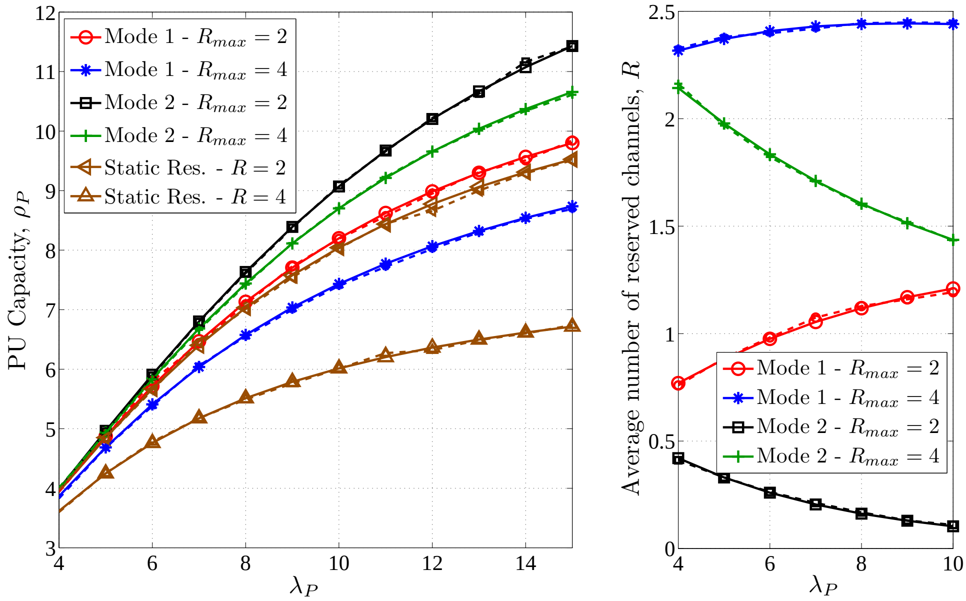


Figure D.6: Capacity achieved by PUs as a function of the PU arrival rate. The dashed lines represent the results obtained by simulations.

services. In other words, when the traffic arrival rate is low, mode 2 serves better since it opens more opportunities for new services. At higher arrival rates, mode 1 is preferred since it protects ongoing services. Moreover, at higher arrival rates, both static and dynamic channel reservations show better performance compared with the scheme without channel reservation in terms of SU capacity. As already mentioned, the provision of channel access opportunities for interrupted SU services in both DCR and SCR at higher λ_P leads to an uncreased capacity. On the other hand, without channel reservation, those channel access opportunities would be obtained by PUs, leading to a lower SU capacity.

Furthermore, extensive simulations have been conducted by utilizing MATLAB to validate the correctness and the preciseness of the obtained analytical results. For the sake of illustration clarity, we represent simulation results by dashed lines only in Fig. D.6 and Fig. D.7. The simulations were performed in a similar way as presented in our previous study [2]. The results clearly demonstrate that the results obtained from the theoretical derivations match closely the simulation results.

B. Retainability

Now let us observe the achieved retainability by SU services as the PU arrival rate varies for both working modes as plotted in Fig. D.7. As shown in this figure, the retainability of services can be increased significantly when one of the working

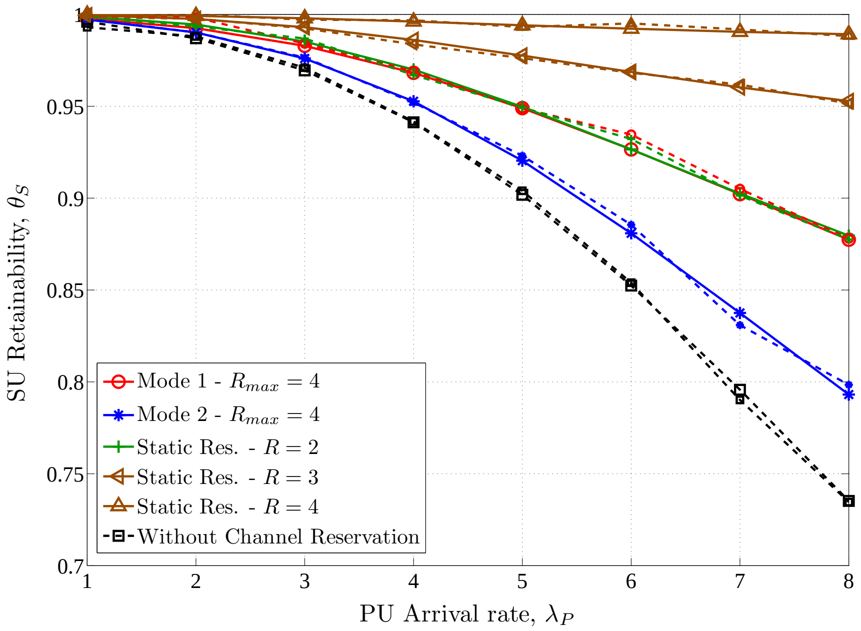


Figure D.7: Retainability of SU services as a function of the PU arrival rate. The dashed lines represent the results obtained by simulations.

modes is employed in the DCR algorithm, compared with a CRN without channel reservation. For instance, when $\lambda_P = 8$, mode 1 or mode 2 based DCR improves retainability by approximately 20% and 10% respectively in comparison with the CRN without channel reservation. This result confirms that better retainability is achieved by selecting mode 2.

To achieve higher retainability, more channels should be reserved in the R-CRN, according to the number of ongoing sessions in the CRN. An ongoing session in the N-CRN needs channel access in the R-CRN when it is interrupted due to PU arrivals or channel failures if all other operational channels are busy. Mode 1 follows this principle. This is most likely to happen when the traffic load is high. Conversely, mode 2 may allocate channels to the R-CRN even when there is a low demand for channel reservation. At a higher traffic load, the number of channels in the R-CRN is reduced in mode 2 and therefore the possibility of improving retainability diminishes. For this reason, the retainability exhibits always a higher value with mode 1 than with mode 2. Moreover, via SCR with a higher number of reserved channels, it is able to provide higher service retainability than in DCR since the number of channels in the R-CRN is kept fixed and they are exclusively allocated for ongoing services [1]. However, SCR is not flexible and it shows comparatively poor performance in terms of channel availability, as to be pointed out in the next

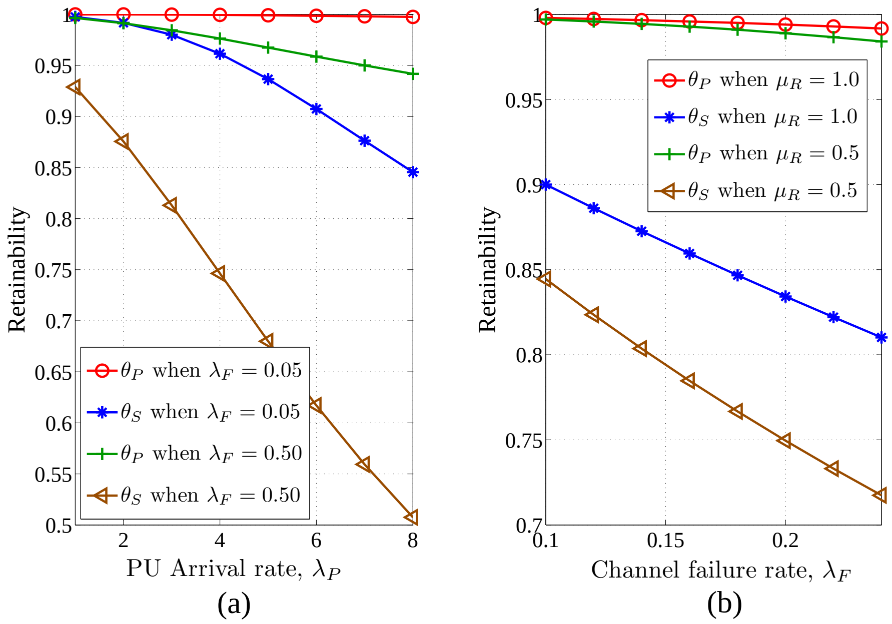


Figure D.8: Retainability of SU and PU services as a function of the PU arrival rate and channel failure rate when mode 1 is selected in Algorithm D.1.

subsection.

In Fig. D.8, we investigate the retainability level of PU and SU services with respect to different channel failure rates and repair rates. As observed in the figure, higher channel failure rates and lower channel repair rates lead to a decreased retainability level. With a large λ_F, channels are more likely to fail. With a lower μ_R, the channels have to stay for a longer period of time as an unavailable (or failed) channel. Consequently, the number of idle channels in the network diminishes and thus the retainability degrades. However, the retainability level of PU services is always higher than that of SU services owing to its channel access priority in both N-CRN and R-CRN.

C. Channel Availability

In Fig. D.9, we evaluate the channel availability of PU and SU services respectively considering different configurations. In the presence of random channel failures, channel availability reveals access opportunities upon PU and SU arrivals at different traffic load levels. From this figure, it is clear that the channel availability decreases due to channel reservation and the degree of degradation is proportional to the number of reserved channels. If channel reservation is not adopted, i.e., when $R = 0$, all channels would be available for new services. Therefore, the scheme without channel reservation always exhibits the highest availability for both networks.

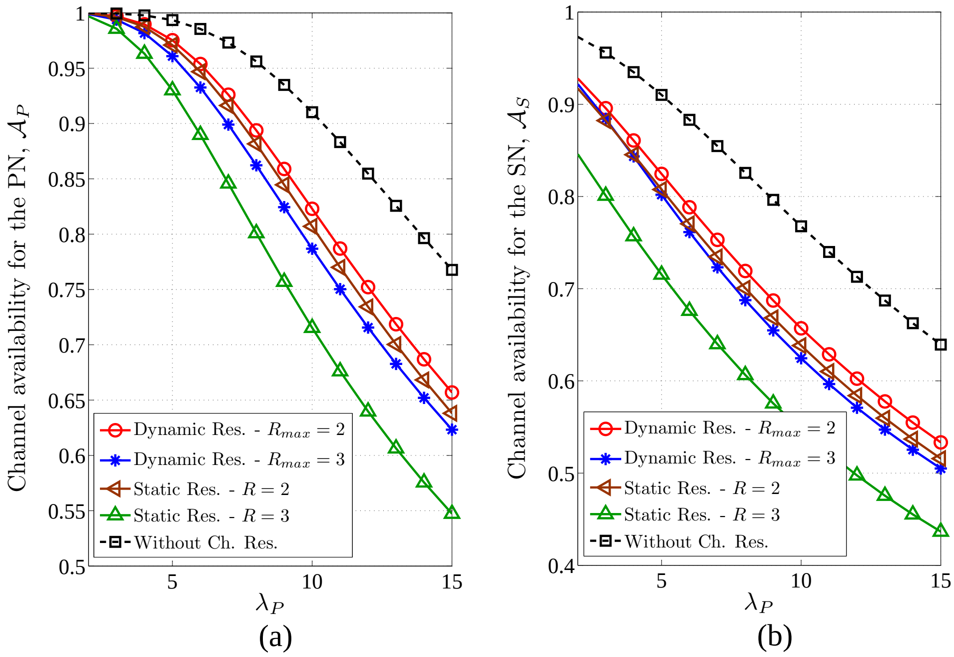


Figure D.9: Channel availability for PN and SN as a function of the PU arrival rate. DCR indicates mode 1 in Algorithm D.1.

Fig. D.9 shows also that the channel availability is significantly influenced by the number of reserved channels in static reservation more than in the DCR scheme. For instance, consider SCR at $\lambda_P = 15$. The percentage of channel availability reduction for SUs with $R = 3$ is 15.4% with respect to the case of $R = 2$. On the other hand, the DCR scheme with mode 1 achieves a percentage of channel availability reduction for SUs as 5.3% for the case of $R_{max} = 2$ with the same configuration. The reason is as follows. Mode 1 intends to assign a larger value for R when the ongoing traffic load in the network becomes higher than a certain threshold, resulting in better spectrum utilization. Fig. D.9 illustrates further that the PUs' channel availability is higher than that of the SN. This is due to the priority access of the PN, i.e., the ongoing SU services can be preempted by an incoming PU, but not the other way round.

Furthermore, the achievable channel availability for the SN is reported in Fig. D.10 considering different channel failure and recovery rates. As expected, with a higher channel failure rate, channel availability tends to monotonically decrease due to lack of idle channels. On the other hand, with a higher channel repair rate, fewer PU and SU services are blocked. This is because that, when μ_R is higher, the failed channels are recovered shortly and become idle again, resulting in increased channel access opportunities for new users.

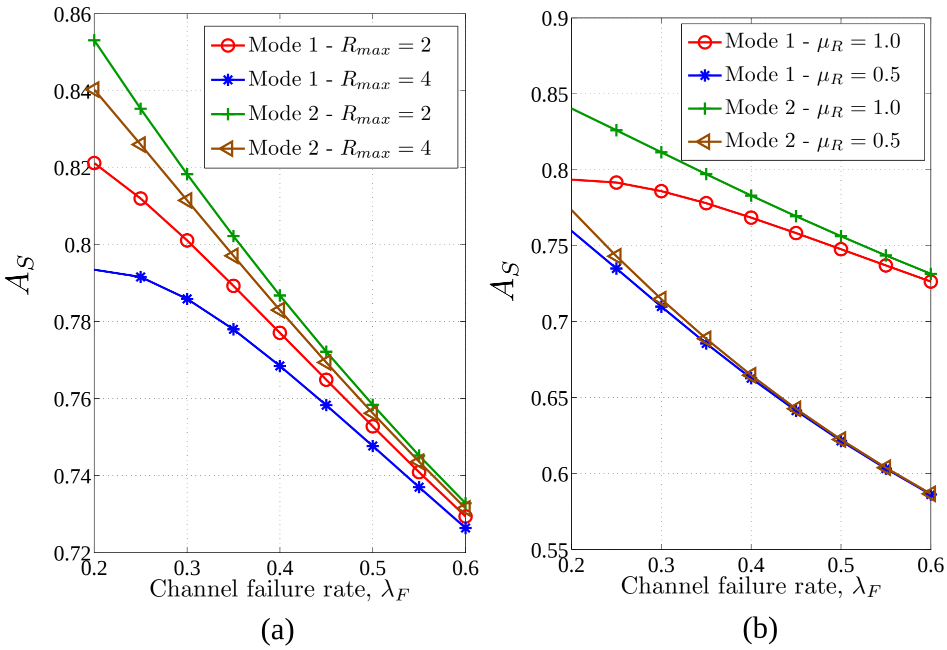


Figure D.10: Channel availability for the SN as a function of the channel failure rate.

As expected, the achieved channel availability for new SU services becomes higher if mode 2 is selected. On the other hand, selecting mode 2 leads to a lower retainability level in the DCR algorithm. This result indicates a need for an assessment of the tradeoff between the new users' channel availability and the ongoing services' retainability. The NUP-based channel reservation algorithm presented in Section VII provides insight regarding this tradeoff.

D. Network Unserviceable Probability

Fig. D.11 confirms that NUP increases with a higher channel failure rate. Moreover, DCR schemes generally outperform the SCR scheme in terms of the NUP. To illustrate this, let us consider the scheme "Mode 2 with $R_{max} = 4$". At $\lambda_F = 0.1$, the network's NUP is only 0.23. Under the same configuration, the network experiences a higher NUP with static reservation when $R \geq 2$. Compared with mode 2, the NUP of SUs increases by 13% if the static scheme with $R = 2$ is employed. This percentage becomes 39% and 73% for static channel reservation with $R = 3$ and $R = 4$ respectively. Even though DCR reduces NUP considerably, the channel access schemes without channel reservations cannot be outperformed in terms of NUP for the arrival rates of PU and SU services used in Fig. D.11. However, by employing mode 2, SUs experience lower NUP compared with DCR with mode 2.

Another important observation regarding the NUP performance of the SN can be found in Fig. D.12. This figure depicts the NUP of SU services, as a function of the

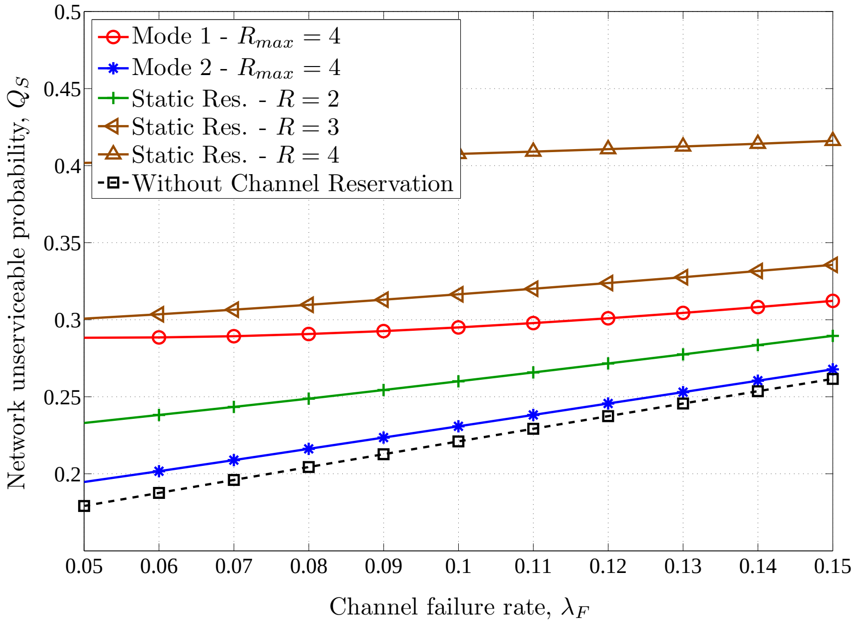


Figure D.11: Network unserviceable probability as channel failure rate varies.

PU arrival rate. In the DCR algorithm presented in Subsection III-A, channel allocation to the R-CRN is controlled based on the ongoing traffic load. Therefore, the NUP performance of SUs is sensitive to the traffic arrival rate. The results presented in Fig. D.12 indicate that, under certain traffic loads, channel reservation provides better performance than the CRN without channel reservations. For instance, when $\lambda_P < 7$, the best NUP performance is observed under the DCR with $R = 0$, i.e., without any channel reservation. However, once $\lambda_P > 11$, the network experiences the highest NUP without channel reservation. The reason for this result is explained below considering the joint effect of channel availability and service retainability.

When the primary traffic load is low, forced terminations of SU services do not happen with higher probabilities. However, once some channels are reserved, a new user could be blocked even though there are vacant channels available in the R-CRN. This leads to a significant increase of NUP in the channel reservation scheme because of high blocking probabilities. Contrary, at larger PU traffic loads, the ongoing SU services are interrupted by PU arrivals and thus the forced termination probability of SUs increases. With reserved channels, those forced terminations can be alleviated and therefore the network could avoid potential capacity losses, leading to a low NUP. Unlike in the channel reservation enabled networks, the CRNs without channel reservation cannot protect ongoing SU services at higher PU traffic loads. Consequently, the NUP for those systems exhibits a large value. In brief, it

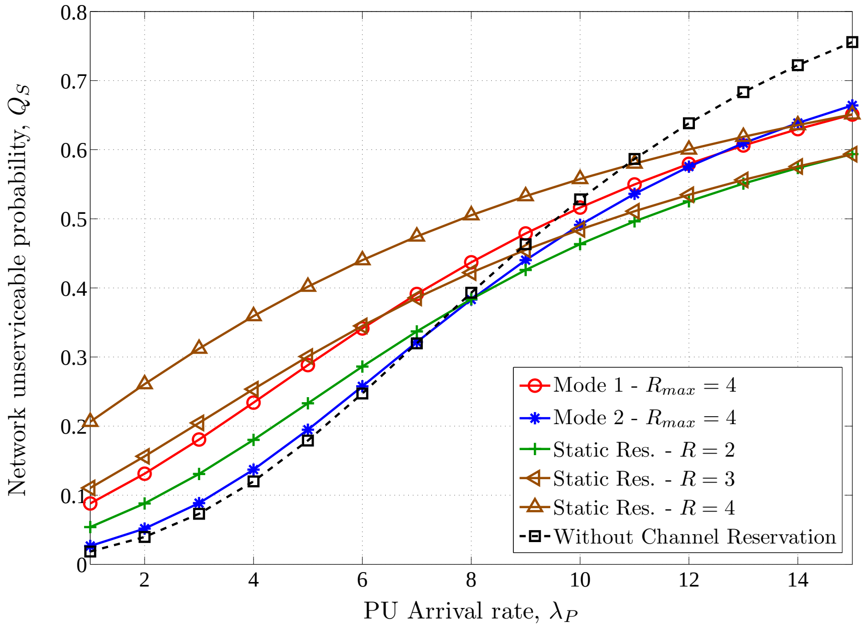


Figure D.12: Network unserviceable probability as a function of the PU arrival rate.

can be concluded that the proposed DCR helps to reduce the NUP in a CRN when λ_P is higher. However, it is recommended to limit the number of reserved channels for low arrival rates.

VI. Modeling and Performance Analysis under Heterogeneous Channel Failures

As wireless networks may interconnect with each other under diverse environments, different types of failures may happen. Consequently, some failures may occur more frequently than others and some failed channels may be recovered faster than others. By considering these factors, we design a failure model which consists of two types of channel failures in order to further investigate the effect of heterogeneous failures on the performance of CRNs.

A. CTMC Model for Heterogeneous Failures in CRNs with Reserved Spectrum

The CTMC model presented in Section IV is extended to model the heterogeneous failure types. Consider two types of failures which have distinct failure rates and repair rates. Let the failure rate and the repair rate of a channel for *Type 1 failure* be λ_{F_1} and μ_{R_1} respectively. Similarly those values for *Type 2 failure* will be λ_{F_2} and μ_{R_2}. Let $\mathcal{S}$ be the set of feasible states in the CTMC model. The states of the CTMC model corresponding to the proposed DCR based spectrum access scheme is then represented by $\boldsymbol{x} = (i_n, j_n, i_r, j_r, f_1, f_2)$ where the terms f_1 and f_2 represent

D

the number of failed channels in the CRN due to the first and second failure types respectively. Term $B(\boldsymbol{x})$ which represents the sum of the occupied and failed channels in the whole CRN becomes $B(\boldsymbol{x}) = B_n(\boldsymbol{x}) + B_r(\boldsymbol{x}) + F$ where $F = f_1 + f_2$, $B_n(\boldsymbol{x}) = i_n + j_n$ and $B_r(\boldsymbol{x}) = i_r + j_r$ as mentioned in the homogeneous failure case.

The corresponding state transition table is obtained by updating Table D.1 with necessary modifications. In Table D.3, we list the updated state transitions numbered **9** to **15** for *Type 1* failures. Note that the first 8 events are the same as listed in Table D.1, but we need to replace the final element, f, of each state by f_1, f_2. Similarly, the transitions for *Type 2* failures can be obtained, however, they are not listed here. In order to accommodate heterogeneous failure in Algorithm D.1, the only update that needs to be done is located in *Line 1* and *Line 2* of Algorithm D.1.

Accordingly, the expressions in *Line 1* and *Line 2* should be changed as

Line 1: $\Phi = (i_n + j_n + i_r + j_r)/(M - (f_1 + f_2))$ and

Line 2: $N_{Avail} = M - (i_n + j_n + i_r + j_r + (f_1 + f_2))$

respectively, in Algorithm D.1. Furthermore, the corresponding equations which represent the forced termination rates of SU and PU services due to channel failures in the heterogeneous case are obtained as

$$R_S' = (\lambda_{F_1} + \lambda_{F_2}) \sum_{\substack{\boldsymbol{x}\in\mathcal{S} \\ B(\boldsymbol{x})=M;\ ((j_n>0) \text{ or} \\ (B_n(\boldsymbol{x})=0;\ j_r>0))}} (M - F)\pi(\boldsymbol{x}) \tag{D.17}$$

and

$$R_P' = (\lambda_{F_1} + \lambda_{F_2}) \sum_{\substack{\boldsymbol{x}\in\mathcal{S} \\ B(\boldsymbol{x})=M;\ ((j_n=0;\ i_n>0) \text{ or} \\ (B_n(\boldsymbol{x})=0;\ j_r=0;\ i_r>0))}} (M - F)\pi(\boldsymbol{x}). \tag{D.18}$$

The other expressions which represent the PU/SU capacity, blocking probability, NUP and retainability are the same as in the homogeneous failure case.

B. Numerical Results with Heterogeneous Failure Rates

The numerical results presented in this subsection are obtained for a CRN with $M = 6$ channels and $R_{max} = 2$.

Table D.3: Transitions from a generic state $\boldsymbol{x} = (i_n, j_n, i_r, j_r, f)$ of the channel access scheme upon PU/SU events and channel failures and repairs.

Event	Destination State	Tran. rate	Conditions
9. Idle channel fails due to a first type failure.	$(i_n, j_n, i_r, j_r, f_1+1, f_2)$	$(M - B(\boldsymbol{x}))\lambda_{F_1}$	$B(\boldsymbol{x}) < M$.
10. An occupied channel fails due to a *Type 1* failure. An idle channel exists in the CRN.	$(i_n, j_n, i_r, j_r, f_1+1, f_2)$	$(B(\boldsymbol{x}) - F)\lambda_{F_1}$	$F < B(\boldsymbol{x}) < M$.
11. An occupied channel fails due to a *Type 1* failure. No idle channels exist in the CRN. An SU_N is forced to terminate.	$(i_n, j_n - 1, i_r, j_r, f_1 + 1, f_2)$	$(M - F)\lambda_{F_1}$	$B(\boldsymbol{x}) = M; F < M; j_n > 0$
12. An occupied channel fails due to a *Type 1* failure. No idle channels exist in the CRN. A PU_N is forced to terminate.	$(i_n - 1, j_n, i_r, j_r, f_1 + 1, f_2)$	$(M - F)\lambda_{F_1}$	$B(\boldsymbol{x}) = M; F < M; j_n = 0; i_n > 0$.
13. An occupied channel fails due to a *Type 1* failure. No idle channels exist in the CRN. An SU_R is forced to terminate.	$(i_n, j_n, i_r, j_r - 1, f_1 + 1, f_2)$	$(M - F)\lambda_{F_1}$	$B(\boldsymbol{x}) = M; F < M; B_n(\boldsymbol{x}) = 0; j_r > 0$.
14. An occupied channel fails due to a *Type 1* failure. No idle channels exist in the CRN. A PU_R is forced to terminate.	$(i_n, j_n, i_r - 1, j_r, f_1 + 1, f_2)$	$(M - F)\lambda_{F_1}$	$B(\boldsymbol{x}) = M; F < M; B_n(\boldsymbol{x}) = j_r = 0; i_r > 0$.
15. A failed channel due to a *Type 1* failure is repaired.	$(i_n, j_n, i_r, j_r, f_1 - 1, f_2)$	$f_1\mu_{R_1}$	$f_1 > 0$.

For the activities corresponding to user arrivals and departures, refer to the events numbered **1** to **8** in Table D.1. The only modification required in those events is the replacement of the final element, f, in each state by f_1, f_2.

The notations AR and DP indicate an arrival event and a departure event respectively. An SU service and a PU service in the N-CRN are denoted as SU_N and PU_N respectively. An SU service and a PU service in the R-CRN are denoted as SU_R and PU_R respectively. Denote $B_n(\boldsymbol{x}) = i_n + j_n$, $B_r(\boldsymbol{x}) = i_r + j_r$, $B(\boldsymbol{x}) = B_n(\boldsymbol{x}) + B_r(\boldsymbol{x}) + F$, where $F = f_1 + f_2$.

D

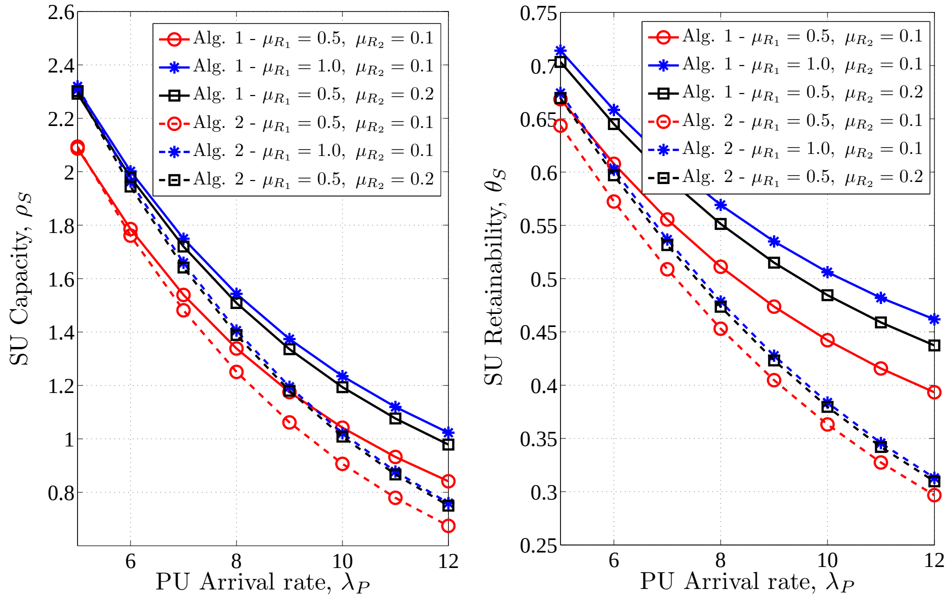


Figure D.13: Capacity and service retainability of the SN in the presence of heterogeneous channel failures when $\lambda_{F_1} = 0.1$ and $\lambda_{F_2} = 0.01$. The solid lines and the dashed lines represent the numerical results corresponding to mode 1 and mode 2 respectively.

1) Capacity and Retainability: In Fig. D.13, the capacity and the retainabilty of SUs are studied when two types of channel failures exist, with the failure rates, $\lambda_{F_1} = 0.1$ and $\lambda_{F_2} = 0.01$ respectively. Here we assume that a *Type 2 failure* occurs 10 times slower than a *Type 1 failure*. In other words, channel failures occur at a rate 0.11 failures per time unit and that 0.1/0.11= 10/11 is the fraction of all failures that is Type 1 (i.e., 10 out of 11). Relevant to this assumption, a *Type 1 failure* could happen due to path loss, multi-path fading or shadowing, while hardware or power failures, which are unlikely to happen often, could be represented by *Type 2 failures*.

The performance of SUs is shown in Fig. D.13 under different configurations of channel repair rates. Consider a reference configuration when $\mu_{R_1} = 0.5, \mu_{R_2} = 0.1$. As illustrated in this figure, the achieved capacity is increased with a higher repair rate in both working modes. The percentage of increments is approximately 18% and 12% for mode 1 and mode 2 respectively, given that $\lambda_P = 12$ when the repair rate of a failed channel due to *Type 1* failure is doubled. If we now double the repair rate of *Type 2* failed channels instead of *Type 1*, the above mentioned percentage increments become 14% and 10% for mode 1 and mode 2 respectively. Therefore, it is observed that the effect of an increased repair rate is more evident for failures with higher failure rates. From a retainability point of view, this observation

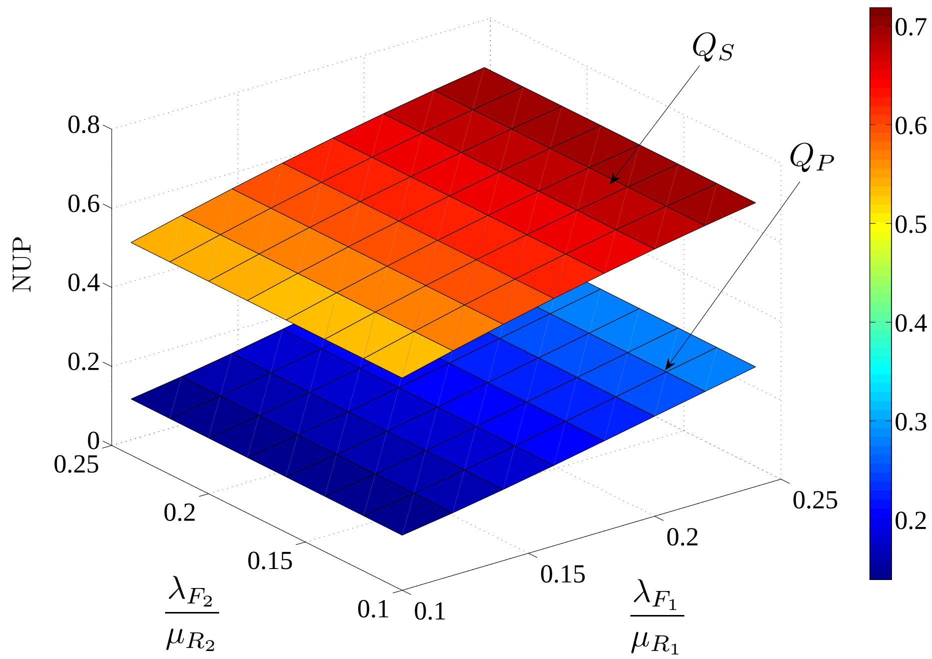


Figure D.14: The NUP as a function of $\frac{\lambda_{F_1}}{\mu_{R_1}}$ and $\frac{\lambda_{F_2}}{\mu_{R_2}}$. The numerical results are corresponding to mode 1 in Algorithm D.1 when $\lambda_P = \lambda_S = 2$ and $\mu_P = \mu_S = 1$, given that $\mu_{R_1} = 0.5$ and $\frac{\lambda_{F_1}}{\mu_{R_1}} = \frac{\lambda_{F_2}}{\mu_{R_2}}$ at each point on the mesh grid. Moreover, the ranges of parameters are $0.05 \leq \lambda_{F_1} \leq 0.12$, $0.01 \leq \lambda_{F_2} \leq 0.045$ and $0.04 \leq \mu_{R_2} \leq 0.45$.

appears to be valid.

As we observed with the homogeneous channel failures, the retainability of SU services shows higher values in mode 1. Furthermore, the figure depicts that the retainability of services can be improved with higher channel repair rates. When μ_{R_1} or μ_{R_2} is raised, the frequency of recovering failed channels becomes higher and consequently there exists a comparatively larger number of available channels in the network. Then the interrupted ongoing SU services can find channel access opportunities more easily, leading to higher retainability.

2) Network unserviceable probability: In order to investigate the effect of heterogeneous channel failures on the NUP performance of CRNs, we illustrate the numerical results for the NUP for both PU and SU services in Fig. D.14. As expected, the NUP of the PN is always lower than that of the SN, owing to the channel access priority of the PN. We note also that the increase of channel failure rates leads to a raise of the NUP for both services.

However the rate of the increase of the NUP highly depends on the failure type characteristics. The two failure types considered in Fig. D.14 have different failure and repair rates. However, those configurations are set to satisfy, $\alpha = \frac{\lambda_{F_1}}{\mu_{R_1}} = \frac{\lambda_{F_2}}{\mu_{R_2}}$

D

where α denotes the ratio between the failure rate to the repair rate of a channel. Note that, at each point on the mesh grid, $\lambda_{F_1} > \lambda_{F_2}$ and $\mu_{R_1} > \mu_{R_2}$. In generally, a network that has a higher channel repair rate would be available for users for a longer period than a network with a lower channel repair rate.

However, the advantage of higher repair rates would be dominated by higher failure rates, as we observed in Fig. D.14. Interestingly, although the repair rate of *Type 1* failures is higher than that of *Type 2* failures, the rate of the increase of NUP is more significant in *Type 1* failures when the ratio, α, is increased. This fact can be clearly observed from the gradient of the NUP mesh plot with respect to each axis representing the ratio, α. The reason is explained as follows. The rate of occurring *Type 1* failures is two times (or more) as higher as that of the second type. Therefore, the time between failures tends to be shorter in *Type 1* failures. That is, *the network with Type 1 failures cannot be serviceable for a longer period of time even though the failed channels are recovered shortly*. Accordingly, we conclude that, the failure type with higher repair rate does not necessarily provide better performance in terms of the NUP since the failure rates would also influence the duration of channel available period.

VII. Identifying the Optimal R_{max} for DCR

Identifying an optimized upper bound for the number of reserved channels in the DCR algorithm is also important since there is a tradeoff between channel availability and service retainability as pointed out in the previous sections. In this section, we propose an NUP-based algorithm to identify the optimal R_{max} which guarantees a bounded NUP for the network.

Algorithm D.2 is proposed to obtain the optimal upper bound for the number of reserved channels denoted as R_{max}^{OPT} in the R-CRN for a given NUP constraint. Note that, the algorithm can be extended to calculate R_{max}^{OPT} also for any other performance metric mentioned in Section IV. The NUP constraint in this algorithm means that the system should guarantee the NUP of SU services always being below a certain threshold level given by $(Q_S)_{max}$ while *achieving the highest possible retainability for SU services*. In this algorithm, a CRN consisting of M channels is considered where no more than $\lfloor \frac{M}{A} \rfloor$ channels could be reserved. Here $A > 1$ is a parameter which constraints the amount of spectrum to be reserved. Herein, we determine R_{max}^{OPT} while investigating the impact of λ_P on the optimal solution. This optimization problem may also be formulated as

D

$$\begin{aligned} &\text{maximize}_{R_{max}=0,1,\cdots,\lfloor \frac{M}{A} \rfloor} && \theta_S \\ &\text{subject to:} && Q_S \leq (Q_S)_{max}. \end{aligned} \tag{D.19}$$

Configure respectively the maximum tolerable NUP for SUs as $(Q_S)_{max} = 0.40$, $\lambda_{F_1} = 0.01$, $\lambda_{F_2} = 0.005$, $\mu_{R_1} = 0.5$, $\mu_{R_2} = 0.1$, $M = 8$ and $A = 2.5$. As observed in Fig. D.15(a), Q_{SU} can be maintained below 40% even when $R_{max} = 3$ if $\lambda_P \leq 5.4$. Therefore, in order to fulfill the requirement of maintaining a higher retainability level, R_{max}^{OPT} should be 3 when $\lambda_P \leq 5.4$. However, if the PU arrival rate appears in the range of $5.4 < \lambda_P < 5.9$, the system should not allocate $R_{max} = 3$ since SUs experience a NUP greater than 40%. Instead $R_{max} = 2$ can guarantee the required NUP in that range. In general, when λ_P becomes higher, the NUP of SU services rises owing to the access priority of PUs. To keep the NUP below the threshold level, the network has to reduce R_{max}. Similarly, through the analysis of the curves obtained from Fig. D.15(a), the value of R_{max}^{OPT} which satisfies the given conditions can be obtained for a given range of λ_P. The obtained numerical results are illustrated in Fig. D.15(b) for different values of $(Q_S)_{max}$.

Algorithm D.2: NUP-based algorithm for identifying an optimal R_{max}.

Input: M : Total number of channels in the CRN
Input: $Q_S|_{R_{max}=k}, k = 0, 1, 2, \cdots$: NUP of SU services when $R_{max} = k$ is configured in Algorithm D.1.
Input: A : Parameter which constraints the amount of spectrum to be reserved where $A > 1$.
Input: $(Q_S)_{max}$: Maximum tolerable NUP for SUs
Output: R_{max}^{OPT} : The optimal value of R_{max}
Calculate $R_a = \lfloor \frac{M}{A} \rfloor$
if $(Q_S)_{max} \leq Q_S|_{R_{max}=0}$ **then**
| $R_{max}^{OPT} = 0$
else if $Q_S|_{R_{max}=0} < (Q_S)_{max} \leq Q_S|_{R_{max}=1}$ **then**
| $R_{max}^{OPT} = 1$
else if $Q_S|_{R_{max}=1} < (Q_S)_{max} \leq Q_S|_{R_{max}=2}$ **then**
| $R_{max}^{OPT} = 2$
$\cdots$
else if $Q_S|_{R_{max}=k-1} < (Q_S)_{max} \leq Q_S|_{R_{max}=k}$ **then**
| $R_{max}^{OPT} = k$
$\cdots$
else if $Q_S|_{R_{max}=R_a-1} < (Q_S)_{max} \leq Q_S|_{R_{max}=R_a}$ **then**
| $R_{max}^{OPT} = R_a$
else
| $R_{max}^{OPT} = R_a$
end

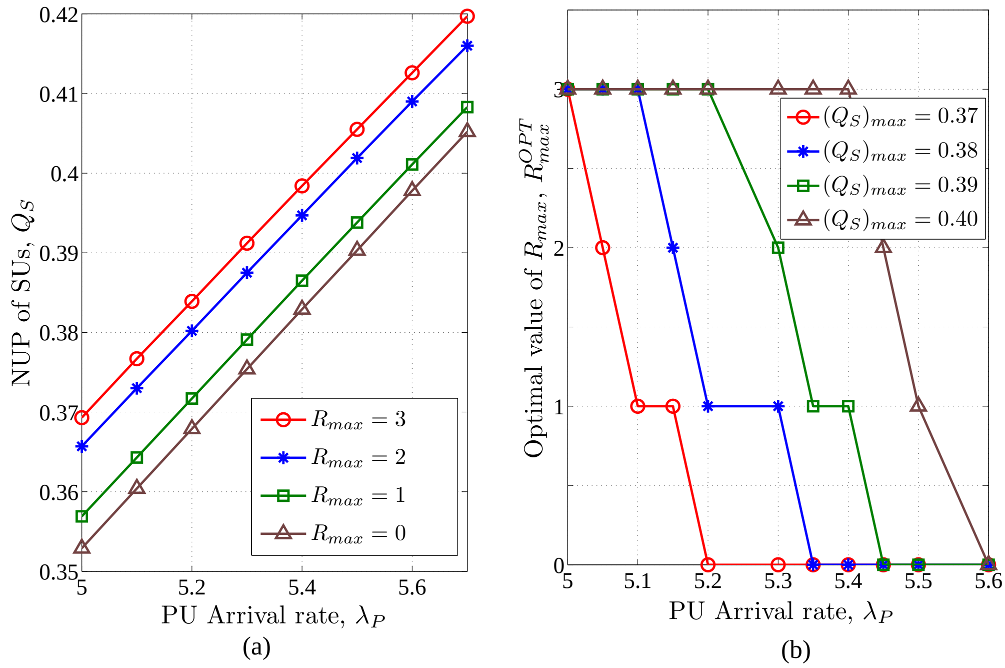


Figure D.15: The optimal value of R_{max} that can be allocated in Algorithm D.1 when mode 1 is selected in order to have a guaranteed NUP for the SN given that $M = 8$ and $A = 2.5$. The results shown in sub-figure (b) are obtained based on the NUP performance shown in sub-figure (a).

VII. Conclusions

In this paper we proposed a dynamic spectrum access scheme with an embedded dynamic channel reservation algorithm which can enhance users' performance in terms of service retainability or channel availability. The proposed DCR algorithm dynamically configure the number of reserved channels depending on the channel occupancy and channel failure status. A CTMC-based analytical model for performance evaluation of the proposed dynamic spectrum access scheme is developed. Based on the CTMC model, mathematical expressions for retainability and network unserviceable probability that capture both resource insufficiency and channel failures are derived for both primary and secondary networks. The numerical results reveal that the major drawbacks of static channel reservation such as drop of capacity and increase of blocking probability can be alleviated significantly through the employment of dynamic channel reservation. Furthermore, the proposed analytical model for a CRN is extended to heterogeneous failures with different failure and repair rates. The performance of SU and PU services in the presence of two failure types is also evaluated. Finally, the optimal upper bound for the number of reserved channels which satisfies a specific QoS requirement is obtained.

D

PAPER D: REFERENCES

[1] I. A. M. Balapuwaduge, F. Y. Li, and V. Pla, "Significance of channel failures on network performance in CRNs with reserved spectrum," in *Proc. IEEE ICC*, pp. 1-6, Kuala Lumpur, Malaysia, May 2016.

[2] Y. Xiao, Z. Han, C. Yuen, and L. A. DaSilva, "Carrier aggregation between operators in next generation cellular networks: A stable roommate market," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 633-650, Jan. 2016.

[3] A. Kumar, A. Sengupta, R. Tandon, and T. C. Clancy, "Dynamic resource allocation for cooperative spectrum sharing in LTE networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5232-5245, Nov. 2015.

[4] H. Shariatmadari, S. Iraji, and R. Jäntti, "Analysis of transmission methods for ultra-reliable communications," in *Proc. IEEE PIMRC*, Hong Kong, China, Aug./Sep. 2015, pp. 2303-2308.

[5] A. Azarfar, J.-F. Frigon, and B. Sansò "Improving the reliability of wireless networks using cognitive radios," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 2, pp. 338-354, Second Quarter, 2012.

[6] K. Y. Wong, Y. Keung, C. Li, and B. Li, "Dynamic call admission control with connection level service guarantee in multi-service wireless cellular networks," in *Proc. IEEE ICC*, Paris, France, Jun. 2004, pp. 4248-4252.

[7] D. Kreutz, F. M. V. Ramos, P. E. Veríssimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proc. IEEE*, vol. 103, no. 1, pp. 14-76, Jan. 2015.

[8] S. Ni, Y. Zhuang, Z. Cao, and X. Kong, "Modeling dependability features for real-time embedded systems," *IEEE Trans. Dependable Secure Comput.*, vol. 12, no. 2, pp. 190-203, Mar./Apr. 2015.

[9] A. Banchs, D. Qiao, and D. Westhoff, "Dependability issues with ubiquitous wireless access [Guest Editorial]," *IEEE Wireless Commun. Mag.*, vol. 15, no. 5. pp. 8-9, Oct. 2008.

[10] G. Rubino and B. Sericola, *Markov Chains and Dependability Theory*, Cambridge University Press, pp. 1-2, New York, NY, USA, 2014.

[11] A. Osseiran et al, "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26-35, May 2014.

[12] 3GPP, "Key performance indicators (KPI) for evolved universal terrestrial radio access network (E-UTRAN): Definitions," *3GPP TS 32.450*, pp. 6-10, Jun. 2010. [Online]. Available: http://www.3gpp.org/DynaReport/32450.htm.

[13] A. G.-Andrades, P. Muñoz, E. J. Khatib, I. d.-l.-Bandera, I. Serrano, and R. Barco, "Methodology for the design and evaluation of self-healing LTE networks," *IEEE Trans. Veh. Technol.*, Early access available in IEEE Xplore, DOI: 10.1109/TVT.2015.2477945, Feb. 2016.

[14] Z. He, S. Mao, and S. Kompella, "A decomposition approach to quality-driven multiuser video streaming in cellular cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 728-739, Jan. 2015.

[15] L. Sun, W. Wang, and Z. Lu, "On topology and resilience of large-scale cognitive radio networks under generic failures," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3390-3401, Jun. 2015.

[16] T. Taleb, A. Ksentini, and B. Sericola "On service resilience in cloud-native 5G mobile systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 483-496, Mar. 2016.

[17] D. Hamza and S. Aïssa, "Enhanced primary and secondary performance through cognitive relaying and leveraging primary feedback," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2236-2247, Jun. 2014.

[18] C. B. R.-Estrello, G. H.-Valdez, and F. A. C.-Pérez, "System-level analysis of mobile cellular networks considering link unreliability," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 926-940, Feb. 2009.

[19] J. Nair, M. Andreasson, L. L. Andrew, S. H. Low, and J. C. Doyle, "On channel failures, file fragmentation policies, and heavy-tailed completion times," *IEEE/ACM Trans. Networking*, vol. 24, no. 1, pp. 529-541, Feb. 2016.

[20] T. Chakraborty and I. S. Misra, "Design and analysis of channel reservation scheme in Cognitive Radio Networks," *Computers & Electrical Engineering*, vol. 42, pp. 148-167, Elsevier, Dec. 2014.

[21] S. Deng, B. Wang, Y. Zheng, and K. K. Chai, "Dynamic channel reservation based on forecast in cognitive radio," in *Proc. IEEE VTC Spring*, Dresden, Germany, Jun. 2013, pp. 1-6.

[22] Telecommunication Standardization Sector of ITU, "Quality of Service and Dependability Vocabulary," *Recommendation ITU-T E.800*, 2007. [Online]. Available: https://www.itu.int/rec/T-REC-E.800-198811-S/en.

[23] A. Amich, M. A. Imran, R. Tafazolli, and P. Cheraghi, "Accurate and efficient algorithms for cognitive radio modeling applications under the i.n.i.d. paradigm," *IEEE Trans. Veh. Technol.*, vol. 64, no. 5, pp. 1750-1765, May 2015.

[24] E. Bauer, R. Adams, and D. Eustace, *Beyond Redundancy: How Geographic Redundancy can Improve Service Availability and Reliability of Computer-based Systems*, John Wiley & Sons Inc. IEEE Press, pp. 20-23, New Jersey, NJ, USA, Dec. 2011.

[25] L. Jiao, I. A. M. Balapuwaduge, F. Y. Li, and V. Pla, "On the performance of channel assembling and fragmentation in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5661-5675, Oct. 2014.

[26] L. Lee, "The unserviceable probability of a class of telecommunications networks," *IEEE Trans. Rel.*, vol. R-20, no. 3, pp. 132-135, Aug. 1971.