Optimizations in Heterogeneous Mobile Networks

Popovska Avramova, Andrijana; Dittmann, Lars; Ruepp, Sarah Renée; Yan, Ying

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Popovska Avramova, A., Dittmann, L., Ruepp, S. R., & Yan, Y. (2016). Optimizations in Heterogeneous Mobile Networks. Technical University of Denmark (DTU).

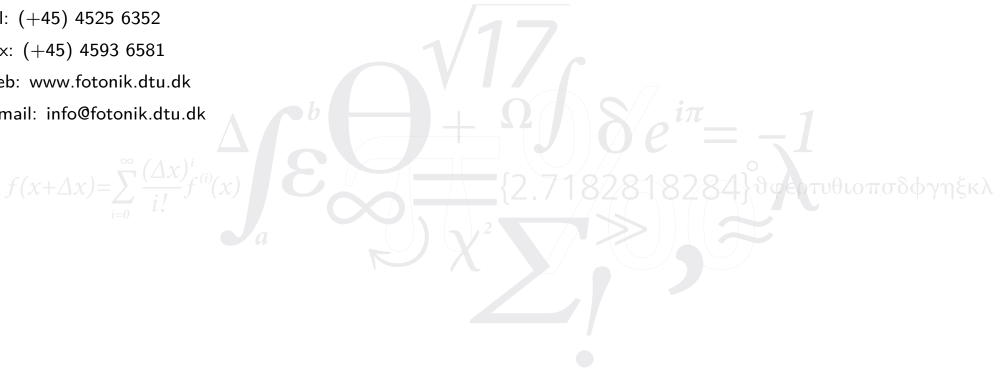
Optimizations in Heterogeneous Mobile Networks

Andrijana Popovska Avramova

PhD Thesis

Technical University of Denmark
Kgs. Lyngby, Denmark
February 2016

Technical University of Denmark
Department of Photonics Enginieering
Networks Technologies and Service Platforms
Ørsteds Plads 343
2800 Kgs. Lyngby
DENMARK
Tel: (+45) 4525 6352
Fax: (+45) 4593 6581
Web: www.fotonik.dtu.dk
E-mail: info@fotonik.dtu.dk



Preface

This thesis presents a selection of the research work conducted during my Ph.D. study from August, 2010 until February, 2016. The project was mainly done in the Networks Technology and Service Platforms group at the Department of Photonics Engineering - Technical University of Denmark under supervision of Professor Lars Dittmann, Associate Professor Sarah Ruepp and Senior Researcher Ying Yan.

A six months research stay took place at Nokia Networks (Nokia - Bell Labs) in Aalborg, Denmark, under supervision of Professor Klaus I. Pedersen and Associated Professor Hua Wang.

This Ph.D. study, the external research stay and participation in international conferences were financed by the Technical University of Denmark, The Danish Advanced Technology Foundation through the project SAIRS - "Standard Agnostic Intelligent Radio Systems for the High Capacity Wireless Internet", as well as by travel grants from Otto Mønsteds Fond and Oticon Fond.

Abstract

Heterogeneous Mobile Networks bring advantages over homogeneous deployments in achieving the demand for mobile network capacity and coverage not just outdoor rural and urban areas, but also to homes and enterprises where the large portion of the mobile traffic is generated. However, the heterogeneity in the mobile networks bring many challenges that are discusses with this dissertation. More focus is placed on specific issues within different areas of heterogeneity by proposing optimizations in order to overcome the considered problems.

The heterogeneity of mobile networks, together with the densification of the base stations, bring into a very complex network management and operation control for the mobile operators. Furthermore, the need to provide always best connection and service with high quality demands for a joint overall network resource management. This thesis addresses this challenge by proposing a universal hierarchical framework that enables flexible and effective management of diverse resources, namely spectral, optical and computational.

Dual Connectivity (DC) is an emerging architecture, which allows for simplified and flexible mobility management and enhanced load balancing among nodes. The independent control of the user's transmit power at each node may cause degradation of the overall performance. In this line, a dedicated study of power distribution among the carriers is performed. An optimization of the power allocation is proposed and evaluated. The results show significant performance improvement to the achieved user throughput in low as well as in high loads in the cell. The flow control of the data between the nodes is another challenge for effective aggregation of the resources in case of dual connectivity. As such, this thesis discusses the challenges in providing efficient flow control, and investigates an optimal traffic rate allocation method.

Cloud Radio Access Network (C-RAN) designates a leading technology for the Radio Access Network (RAN) architecture that is able to support dense deployments, while ensuring network level energy and cost efficiency for the operator. This thesis thoroughly investigates the achievable multiplexing gains under C-RAN through a mathematical model based on the teletraffic theory. The work allows for evaluation of the key parameters and conditions for optimized cell deployment. The model can be applied to dynamically re-assign cells to a pool of baseband units. Furthermore, an evaluation of the various functional splits in the baseband processing is introduced. The proposed mathematical model quantifies the multiplexing gains and the trade-offs between centralization and decentralization concerning the cost of the pool, fronthaul network capacity and resource utilization. Among the benefits that C-RAN brings is the possibility for sharing of the radio spectrum and the resources required for baseband processing among operators. This thesis investigates strategies for active sharing of radio access among multiple operators and analyses the individual benefits depending on the sharing degree.

Resumé

(Summary in Danish)

Heterogene Mobilnet har væsentlige fordele i forhold til homogene implementeringer i at dække behovet til mobil netværkskapacitet og dækning ikke bare udendørs i landdistrikter og byområder, men også til private hjem og virksomheder, hvor en stor del af den mobile trafik bliver genereret. Men den heterogenitet i de mobile netværk giver mange udfordringer, der bliver diskuteret i denne afhandling. Afhandlingen har stor fokus på specifikke problemstillinger inden for forskellige områder af heterogenitet ved at foreslå optimeringer for at overvinde de nævnte problemer.

Den heterogenitet af mobilnet sammen med tættere placering af basestationer, giver en meget kompleks netværks vedligeholdelse og drift kontrol for mobiloperatørerne. Behovet for altid at give den bedste forbindelse og tjeneste med høj kvalitet kræver en fælles, overordnet netværks ressource kontrol. Denne afhandling omhandler denne udfordring ved at foreslå et universelt hierarkisk framework, der muliggør fleksibel og effektiv styring af forskellige ressourcer, nemlig spektrale, optiske og processeringsmæssige.

Dual Connectivity (DC) er en ny arkitektur, som giver mulighed for forenklet og fleksibel mobilitetsstyring og forbedret ressourcebalancering mellem knudepunkter. Den uafhængige kontrol af brugerens transmissionseffekt ved hvert knudepunkt kan medføre forringelse af den samlede ydelse. Med hensyn til denne udfordring, er der inkluderet en dedikeret undersøgelse af effektfordeling mellem carriers. En optimering af effektallokering foreslås og evalueres. Resultaterne viser signifikant forbedring i den maksimale hastighed for lave samt høje belastninger i cellen. Kontrol af data flow mellem knuderne er endnu en udfordring for en effektiv sammenlægning af ressourcerne i tilfælde af dual connectivity. Denne afhandling diskuterer udfordringerne i at udføre effektiv flow kontrol, og undersøger en optimal trafik fordelingsmetode.

Cloud Radio Access Network (C-RAN) er betegnelsen for en førende teknologi til Radio Access Network (RAN) arkitektur, der er i stand til at understøtte tætte udrulning og samtidig sikre netværks energi- og omkostningseffektivitet for operatøren. Denne afhandling undersøger detaljeret de opnåelige multiplexing gevinster ved CRAN gennem en matematisk model baseret på tele-trafik teori. Arbejdet giver mulighed for evaluering af de vigtigste parametre og betingelser for optimeret celle design. Modellen kan anvendes til dynamisk at tildele celler til en pulje af baseband enheder. Endvidere bliver den givet evaluering af de forskellige funktionelle splits i basebands processering. Den foreslåede matematiske model kvantificerer de opnåelige multiplexing gevinster og kompromisser mellem centralisering og decentralisering vedrørende pool omkostninger, fronthaul netkapacitet og ressourceudnyttelse. Blandt de fordele, som C-RAN giver er muligheden for at dele radiofrekvenser og nødvendige ressourcer til baseband behandling mellem operatørerne. Denne afhandling undersøger strategier for aktiv deling af radio adgang mellem flere operatører, og analyserer de individuelle fordele afhængigt af delings grad.

Acknowledgements

I would not be able to complete the work that is presented in this thesis without help of many people who were around me during my Ph.D. studies.

First and foremost, I would like to thank my supervisors: Professor Lars Dittmann, Associate Professor Sarah Ruepp and Senior Researcher Ying Yan that have steered me through the studies with their knowledge and encouragements. I am grateful for the given opportunity and their trust in me.

I would like to thank all the members of Networks Technology & Service Platforms group that have created a pleasant working environment. I acknowledge my colleagues Matteo Artuso and Aleksandra Checko who have been supportive in every way.

There are no words to express my gratitude to Dr. Villy Bæk Iversen for showing me the world of teletraffic and teaching me the theory and different analytical models. He has been a great inspiration for many research directions in this study.

I deeply admire Associate Professor Henrik L. Christiansen, who always had time for discussing various challenges in my study. His intriguing questions have been inspiring and have helped me to improve my work.

I am grateful to Klaus Pedersen and Dr. Hua Wang from Nokia Networks for their help and encouragements. Working as a visiting scholar in their research group was a pleasant and rewarding experience.

Most importantly of all, I show extensive gratitude to my family that has always been there when I needed them most. I thank my mother who taught me to be ambitious and never to give up. I thank my father who has been my tutor my whole life, showed me the love towards science and opened the doors of the world of telecommunication. I would also like to thank my sister - her strength and willingness have supported me and encouraged me ever since my first years at the university. I thank my husband whose optimisms have always brought brightness and confidence in my work. And last but not least, I thank my daughters whose love, laugh and joy made this journey much more enjoyable.

Ph.D Publications

Publications on the topic Optimizations in Heterogeneous Mobile Networks

This dissertation only includes work for the topic on Optimizations in Heterogeneous Mobile Networks. The PhD project resulted in 14 peer reviewed papers and they are listed here in chronological order:

[1] A. Checko^{1st}, **A. Popovska Avramova**^{1st}, M. Berger, and H.L. Christiansen. "Evaluating C-RAN fronthaul functional splits in terms of network level energy and cost savings". In: Journal of Communications and Networks, accepted for publication (2016)

[2] **A. Popovska Avramova**, H. Wang, L. Dittmann, and K. I. Pedersen. "Optimal Uplink Power Control for Dual Connected Users in LTE Heterogeneous Networks". In: IEEE 83rd Vehicular Technology Conference: VTC2016-Spring, accepted for publication, May 2016.

[3] **A. Popovska Avramova**, and V.B. Iversen. "Optimal Traffic Allocation for Multi-Stream Aggregation in Heterogeneous Networks". In: IEEE 7th Int. Workshop on Heterogeneous and Small Cell Networks (HetSNets), IEEE GLOBECOM 2015, Dec. 2015.

[4] **A. Popovska Avramova**, H. L. Christiansen, and V.B. Iversen. "Cell Deployment Optimization for Cloud Radio Access Networks using Teletraffic Theory", In: The 11th Advanced International Conference on Telecommunications, AICT 2015. July 2015.

[5] **A. Popovska Avramova**, and V.B. Iversen, "Radio Access Sharing Strategies for Multiple Operators in Cellular Networks". In:The First IEEE International Workshop on 5G & Beyond - Enabling Technologies and Application, IEEE ICC 2015, June 2015.

[6] A. Zakrzewska, **A. Popovska Avramova**, H. Christiansen, Y. Yan, A. Checko, A. Dogadaev, S. Ruepp, M.S. Berger, and L. Dittmann. “A Framework for Joint Optical-Wireless Resource Management in Multi-RAT, Heterogeneous Mobile Networks”. In:Workshop on Optical-Wireless Integrated Technology for Systems and Networks (OWITSN) 2013, IEEE ICC 2013, June 2013, pp. 895–899.

[7] A. Zakrzewska, **A. Popovska Avramova**, S. Ruepp, M.S. Berger, and L. Dittmann. “CSMA-based SON Mechanism for Greening Heterogeneous Networks”. In: INFOCOM Student Session, INFOCOM 2013, April 2013

[8] E. Kisielius, A. Zakrzewska, **A. Popovska Avramova** and S. Ruepp. “Energy Efficiency in Self Organising Networks”. In: OPNETWORK 2013, Aug. 2013

[9] **A. Popovska Avramova**, Y. Ying, and L. Dittmann. “Evaluation of a Cross Layer Scheduling Algorithm for LTE Downlink.” In: Telfor Journal 5.1 (2013), pp. 26–31.

[10] **A. Popovska Avramova**, Y. Yan, S. Ruepp, and L. Dittmann. “Resource Allocation: Current Issues and Future Directions”. In: 4th Nordic Workshop On System And Network Optimization For Wireless (SNOW 2013). Apr. 2013.

[11] A. Dogadaev, A. Checko, **A. Popovska Avramova**, A. Zakrzewska, Y. Yan, S. Ruepp, M.S. Berger, L. Dittmann, and H. Christiansen, H. “Trafic Steering Framework for Mobile-Assisted Resource Management in Heterogeneous Networks”. In: Ninth International Conference on Wireless and Mobile Communications, ICWMC. July 2013.

[12] **A. Popovska Avramova**, Y. Yan, and L. Dittmann. “Cross layer scheduling algorithm for LTE downlink”. In: 20th Telecommunications Forum (TELFOR), Nov. 2012, pp. 307–310.

[13] **A. Popovska Avramova**, Y. Yan, and L. Dittmann. “Modeling and Simulation of Downlink Subcarrier Allocation Schemes in LTE”. In: OPNETWORK, Aug. 2012.

[14] **A. Popovska Avramova**, Y. Yan, and L. Dittmann. “Modeling of Bandwidth Aggregation over Heterogeneous Wireless Access Networks”. In: OPNETWORK, Aug. 2012

Publications on the topic Wireless Sensor Networks

[15] B. Stojkoska and **A. Popovska Avramova**. “Wireless Sensor Networks Framework for Indoor Temperature Regulation”. In: 8th Annual South-East European Doctoral Student Conference. Sept. 2013, pp. 354–364.

[16] B. Stojkoska, **A. Popovska Avramova**, and P. Chatzimisios. “Application of Wireless Sensor Networks for Indoor Temperature Regulation”. In: International Journal of Distributed Sensor Networks (2014). Article ID 502419.

Publications on the topic Critical Communications and Public Safety Networks

[17] S. Michail, **A. Popovska Avramova**, and L. Dittmann. “MIH Based Mobility for TETRA-LTE Network”. In: ICT Innovations 2013. Sept. 2013.

[18] **A. Popovska Avramova**, S. Ruepp, and L. Dittmann. “Towards Future Broadband Public Safety Systems: Current Issues and Future Directions”. In: International Conference on Information and Communication Technology Convergence (ICTC). Oct. 2015, pp. 74–79.

Contents

List of Figures

List of Tables

Acronyms

3GPP Third Generation Partnership Project

AI Artificial Intelligence

BBU Baseband Unit

BS Base Station

C-RAN Cloud Radio Access Network

CA Carrier Aggregation

CAPEX CAPital EXpenditure

CC Component Carrier

CCO Coverage and Capacity Optimization

CN Core Network

CoMP Coordinated Multi-Point

CPRI Common Public Radio Interface

CQRs Channel Quality Reports

CRE Cell Range Expansion

CRRM Common Radio Resource Management

CSMA Carrier Sense Multiple Access

DC Dual Connectivity

DL downlink

eGAN enhanced Generic Access Network
eICIC enhanced Inter-cell Interference Coordination
EPC Evolved Packet Core
ePDG Evolved Packet Data Gateway

HetNets Heterogeneous Networks

I-WLAN Interworking WLAN
IEEE Institute of Electrical and Electronics Engineers
IETF Internet Engineering Task Force
IMS IP Multimedia Subsystem
IQ In-phase/Quadrature

JRRM Joint Radio Resource Management

KPI Key Performance Indicators

LAA License Assisted Access
LPN Low Power Node
LPNs Low Power Nodes

MAC Media Access Control
MIH Media Independent Handover
MIMO Multiple Input Multiple Output
MLB Mobility Load Balancing
mmWave milimeter-wave
MRO Mobility Robusteness Optimization
MRRM Multi-access Radio Resource Management

OBSAI Open Base Station Architecture Initiative
OPEX OPerating EXpenditure

ORI	Open Radio equipment Interface
P-GW	Packet Data Network Gateway
PDCP	Packet Data Convergence Protocol
PHY	Physical
QoS	Quality of Service
RAN	Radio Access Network
RATs	Radio Access Technologies
RIWCoS	Reconfigurable Interoperability of Wireless Communications System
RRH	Remote Radio Head
RRM	Radio Resource Management
S-GW	Serving Gateway
SDR	Software Defined Radio
SINR	Signal to Interference plus Noise Ratio
SLA	Service Level Agreement
SON	Self-Organizing Networks
UE	User Equipment
UL	uplink

CHAPTER 1

Introduction

Ever since the volume of mobile data traffic has surpassed the voice traffic, access to mobile Internet has become fundamental/integrated part of every day and professional life. The mobile devices have become more affordable, as well as more capable of delivering variety of services to the end user. These developments, together with great diversity of captivating mobile applications, have driven the increase of mobile data traffic demand with exponential rate, that is expected to surpass the traffic from wired devices by 2019 [19].

This immense increase of the mobile data traffic demand has been extensively addressed by the research community. The standardization bodies such as Institute of Electrical and Electronics Engineers (IEEE), Third Generation Partnership Project (3GPP) together with the equipment vendors, mobile operators and the research centers from the academy, focus their research activities towards mobile networks that will provide ubiquitous access and high quality of service to the end users. The existence of various wireless standards results in highly heterogeneous mobile networks where multiple Radio Access Technologies (RATs) coexist and cells (macro cell, pico cell, femto cell) with different size overlap. Furthermore, various Base Station (BS) designs and architectures for connecting the access nodes towards the core network bring additional heterogeneity in the mobile networks. The vision of the upcoming wireless communicating systems is to provide integration of these technologies and enable further convergence with new revolutionary technologies. However, there are many challenges resulting from the heterogeneity in mobile networks that are still open and need to be resolved before having a truly converged network that will provide seamless experience to the end users.

1.1 Background on Heterogeneous Mobile Networks

The demand for increased bandwidth, high Quality of Service (QoS) and ubiquitous connectivity is growing continuously. Different access technologies can be used in order to provide access to services, therefore efficient and cost-effective solutions for interworking is required. In this context, a heterogeneous network is regarded as a compound of multiple access technologies and transmission solutions, cells with different radius coverage and various access network architectures.

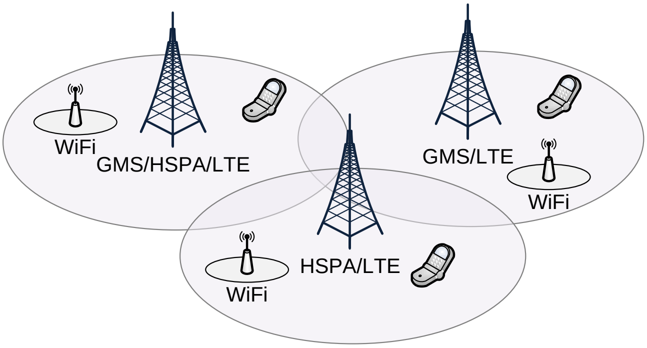


Figure 1.1: Multi-RAT heterogeneous network.

This thesis addresses these three domains of heterogeneity in mobile networks. The subsequent subsections present each of them separately, discussing their key features and research directions. The main contributions are briefly stated in the following subsections along for each introduced area of heterogeneity.

1.1.1 Multi-RAT Heterogeneous Networks

The First Generation (1G) of mobile communication was introduced in 1981, known as the Nordic Mobile Telephony (NMT). The NMT is the first fully automatic cellular phone system based on analogue technology. The Second Generation (2G) comprises the Global System for Mobile Communications (GSM) which introduced digital encryption of the phone conversations and the Short Message Service (SMS). The subsequent enhancements – General Packet Radio Service (GPRS) known as 2.5G and Enhanced GPRS (EDGE) known as 2.75G provided IP functionality and data transfer rates up to 1.3Mb/s in downlink and 635kb/s in uplink. The Third Generation (3G) representative, the Universal Mobile Telecommunications System (UMTS) and its enhancements – the High Speed Packet Access (HSPA) enabled new applications including Voice over IP (VoIP), uploading pictures and sending large e-mail messages. And finally, the LTE-Advanced (LTE-A) and the Wireless Interoperability for Microwave Access (WiMAX) Release 2 are recognized as the technologies that fulfill the International Mobile Telecommunications Advanced (IMT-Advanced) specifications for the Fourth Generation (4G) of mobile systems.

The coexistence of these technologies has resulted in highly heterogeneous mobile networks as depicted in Fig. 1.1. Today, Wireless Fidelity (WiFi), UMTS and Long Term Evolution (LTE) are not seen as competing but rather as complementary technologies. For example, cellular operators are looking at WiFi as a cost effective opportunity for offloading traffic from the cellular networks. In order to provide truly ubiquitous service to the end users, many challenges exist in the integration of these technologies. In this thesis substantial research has been done in addressing the challenges in providing a unified framework for network resource management and achieving performance gains in resource aggregation from heterogeneous access nodes.

Interworking Requirements and Architectures

Interworking between different RATs leads to improved overall performance benefits to the end users as well as to the network operators. The integration between RATs can be achieved on different levels, where some requirements [20, 21] may or may not be fulfilled completely. The main vision is to provide such flexibility that will maintain the user's experience across different access networks, such that the charging the access to the service is consistent and in accordance with the user's Service Level Agreement (SLA). Furthermore, the service should be provisioned in seamless fashion, meaning that the handover process should have minimal or no impact to the service quality. Another important aspect is the network selection procedure that will enable the user to be attached to a network that provides better service at a certain cost (according to user profile).

There are many different approaches in achieving integration between the existing RATs, each with specific network architecture and support for the above mentioned requirements for interworking which are detailed in [22, 23, 24, 25]. The two most basic levels of coupling are the loose and tight coupling. With the loose coupling the different access networks are seen as complementary access to the Internet and the services. Installation of new networks is simplified and operators can deploy, maintain and perform traffic engineering separately, as there is no need for high cooperation among networks. Each separate network can utilize different mechanisms for authentication and billing. The subscriber database can be shared, and Mobile IP [26] can be used as mobility management for handover between networks. On the other hand, with the tight coupling the networks are deployed as alternative RAN. This architecture is more complex as more modifications are required in the networks - namely WiFi gateways need to implement 3GPP authentication, mobility management etc. As the non-3GPP networks need to implement 3GPP mobility management, the tight coupling has better handover performance compared to the loose coupling.

The first interworking architectures standardized by 3GPP where the Interworking WLAN (I-WLAN) [27] and the enhanced Generic Access Network (eGAN) architecture [28, 29]. The I-WLAN is an example of loose coupling, where the standard Internet Engineering Task Force (IETF) protocols for authentication, accounting and mobility can be used. The GAN technology is an example of tight coupling interworking. The interconnection is achieved through new network element that hides the type of access network towards the core network. Thus this architecture does not impose any modification to the underlying radio access technology however the terminal needs to be enhanced with the eGAN capability. Interworking among different RAT can be also achieved with IP Multimedia Subsystem (IMS) as it allows for a variety of multimedia services to be extended over both, fixed and mobile heterogeneous networks. IMS architecture is designed with the aim of providing ubiquitous cellular access, convergence of services (with strong accent on IP multimedia services), mobility of users (roaming, handover), support for different levels of QoS and policy-based delivery of services [30].

Since Release 8, the Evolved Packet Core (EPC), as an IP based multi-access core network, integrates both 3GPP (GSM, UMTS, LTE) and non-3GPP (including both fixed and wireless) access technologies. With the EPC, a user can utilize one or more access networks to connect to the core network. The 3GPP like access networks are integrated via the Serving Gateway (S-GW). Depending on the trust level between the core network operator and the access network operators, different point of integration has been defined. The trusted non-3GGP access technologies are connected through the Packet Data Network Gateway (P-GW), while the untrusted are connected through the Evolved Packet Data Gateway (ePDG).

Network Resource Management

Providing an integrated resource management that will enable efficiency in the resource management functionalities over multiple technologies is of crucial importance for both the network operators and end users. The convergence of the existing and emerging technologies requires common framework for control procedures that will manage the diverse resources jointly for enhanced user experience.

Several management platforms have been proposed in the literature [31, 32, 33, 34, 35, 36]. The concepts proposed by these in the literature deal with enhancements for seamless operation, improving the RAT selection schemes and handover performance, as well as load balancing. However, the proposed frameworks for Radio Resource Management (RRM) deal with resource diversity coming from multi -standard technologies. Yet, these management frameworks lack functionalities that manage the overall network resources such as optical and computational resources.

This thesis addresses the above issue by proposing a novel framework for design of network resource management that considers other areas of heterogeneity resulting from the evolution of the mobile networks. Furthermore, the framework is enhanced with the Self-Organizing Networks (SON) concept and the cognitive principles, improving the abilities for automation and the learning abilities into the individual functionalities of the management tasks.

Resource Aggregation

From a user's point of view, maintaining seamless connection to Wi-Fi networks is not as straightforward as connecting to cellular networks. Often users do not recognize the available Wi-Fi networks, and, even more, they get discouraged to use Wi-Fi when credentials are necessary to connect to open or public networks. In order to provide a cellular-like experience and the Global Wi-Fi implementation, the Wi-Fi Alliance Hotspot 2.0 Specification, which references the IEEE 802.11u amendment, has been presented [8]. The IEEE 802.11u [9] aims to provide an overall end-to-end solution for interworking with external networks. It defines Layer 2 transport for a query-response protocol that allows users to effectively query the network for the information relevant to the network selection prior to performing the authentication procedures.

However, lately more studies are focused on the RAN layer integration between LTE and WiFi networks. In this line, the License Assisted Access (LAA) is emerging as a candidate technology for telecommunication companies to utilize unlicensed spectrum[1], particularly the $5GHz$ band, for wireless data traffic offloading [37]. More specifically, the LAA is based on the method of Carrier Aggregation (CA) between a primary cell, operating in licensed spectrum to deliver critical information and guaranteed QoS, and a secondary cell, operating in unlicensed spectrum to opportunistically boost the data rate. Such a coexistence between LTE-LAA and WiFi requires fairness as to ensure that the LTE operation will not impact the existing WiFi users more than an additional WiFi network on the same carrier [38]. Another type of tight integration can be achieved at the Packet Data Convergence Protocol (PDCP) layer from the LTE protocols stack, such that the users are simultaneously connected to both, the WiFi and the LTE network. This way, link aggregation is achieved such that the scheduler at the PDCP layer needs to classify packets into two flows - one with destination characterized by the LTE access node and the other one characterized by the WiFi access node.

This thesis addresses the challenge of resource aggregation at higher layers, that can be achieved for example at the PDCP layer. This type of aggregation is later referred to as multi-stream aggregation in order to provide the generalized term for resource aggregation across heterogeneous access nodes. The motivation for this study comes form the potential benefits seen in improved data throughput and enhanced quality of experience for the users. The main challenges for such

---

[1] license-exempt spectrum.

cooperation is the optimal split of the traffic, for which variety of optimization procedure can be found [39, 40, 41, 42]. An overview of this methods is provided, and more depth study of the important factors that influence the performance is provided. Furthermore, a new optimization of the methods for flow control considering additional degree of heterogeneity is proposed.

1.1.2 Multi-tier Heterogeneous Networks

Dense and heterogeneous network deployments, where Low Power Nodes (LPNs) are overlaid by macro cells (Fig. 1.2), significantly improve the capacity of a cellular network in a cost-effective and pragmatic way [43], [44]. The LPNs can be classified as remote radio heads, pico, femto and relay nodes, and their transmission power varies between 250 mW to 2W for outdoor purpose, while for indoor use the transmission power is of 100mW or less. The relay nodes differ from the other low power nodes as the same spectrum is used for communication with the mobile users as well as for the wireless backhaul to the macro BS. A group of Low Power Node (LPN) with the same transmission power can be seen as a separate tier in the network, and therefore this type of heterogeneity is referred to as multi-tier Heterogeneous Networks (HetNets)[2]. The cost-effectiveness in the heterogeneous topologies comes from the significantly decreased cost of the low power nodes compared to the high power macro nodes, which reduces the overall CAPital EXpenditure (CAPEX). The heterogeneous networks are more pragmatic as the LPNs target locations where the traffic demand is increased (stadiums, shopping malls, airports, etc.) and are improving the data rates at the cell edge of the macro cells. The small cells effectively increase the capacity through a better spatial reuse and offload the macro cell. By having reduced load, the macro cell can provide wider and more reliable coverage for mobile users at medium and high velocity.

However, HetNets impose changes in the theoretical modeling, analysis and performance evaluation, and at the deployment, operation and management stages of the cellular networks [45], [46]. These implications challenge the state of the art resource management and require re-design and improvements. Since there are large disparities between the transmission powers of the nodes in multi-tier HetNets, a BS that offers the highest Signal to Interference plus Noise Ratio (SINR) may not always offer the best data rates to the user. The available resources to a given user become dependent on the cell load as well. Cell Range Expansion (CRE) overcomes this problem, as the SINR is positively biased in favor of the small cells and the offload from the macro towards the small cells is controlled. Therefore, CRE improves the load balancing between the tiers and users throughput[44]. But, as the name implies, with CRE, the coverage size of a cell is expanded, hence the

---

[2] The term *heterogeneous networks* considers all types of heterogeneity, whereas *HetNet* relates to a network with different cell sizes.

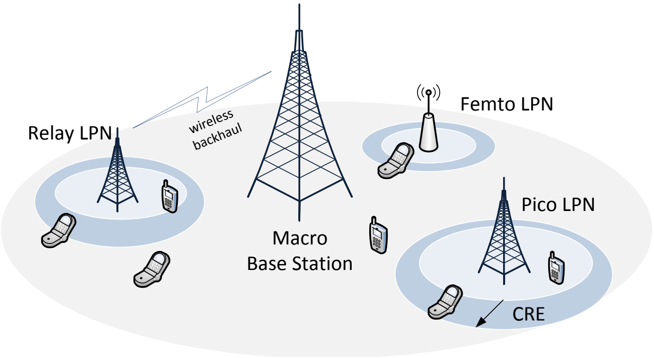


Figure 1.2: An example of heterogeneous network: mixture of high and low power nodes.

handover and cell (re)selection decisions need to be adapted as well. The increased number and variety of cell sizes may degrade the mobility performance due to frequent ping-pongs, handover and radio link failures. In order to optimize the mobility performance, cell specific handover parameters are required in multi-tier heterogeneous networks [47]. Additionally, the significant difference in the received power and interference levels in the uplink (UL) and downlink (DL) makes a single BS association not optimal for both UL and DL data transmission. Decoupling of uplink and downlink cell association, where uplink and downlink transmission are threated independently, is highly beneficial as it improves the spectral and energy efficiency [48].

Another, very important challenge is how all low power nodes will be connected to the core network. The increased data rates on the air interface need to be supported by suitable capacity in the backhaul network. On the other hand the backhaul network need to be cost-effective, meaning the cost will scale down proportionally with the expected decrease of the equipment in the RAN and leasing costs. This question is further addressed in .

Dual Connectivity architecture as a small cell enhancement has been introduced within 3GPP Release 12 [49], [50]. With DC, a User Equipment (UE) with multiple transceivers is able to simultaneously connect to two power nodes, and as such the user can benefit from inter-node resource aggregation. Unlike inter-site CA, DC is characterized with more flexibility as it does not depend on backhaul network with strict requirements on capacity and latency.

Network densification through increased number and often uncoordinated operation of low power nodes, leads to substantial technical challenges with respect to operation and management of the network. With increased complexity and heterogeneity, automated operation of the network is regarded as imperative for optimized utilization of network resources. In order to ensure proper integration of

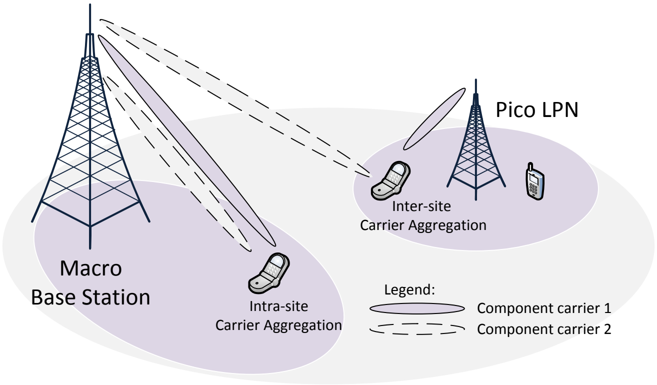


Figure 1.3: Examples on Carrier Aggregation in heterogeneous network.

automated features into the existing processes and overall architecture for operation and management, the 3GPP has started the standardization of SON use-cases and procedures since Release 8. Such automotive procedures are important for improving the energy and spectral efficiency in the network, and ensure lower OPerating EXpenditure (OPEX).

Energy efficiency

Dense deployments of LPN lead to increased overall energy consumption in the networks. Green network deployments have become very popular among the telecommunication industries, not just for cost reduction, but also for reaching green environmental targets. As such, network energy efficiency becomes an important performance indicator for cellular network design. In order to minimize the energy consumption and CO_2 emission, the transmission power and overall operation of the nodes needs to be controlled in an automatic way. High fluctuations in the traffic demand on space and time opens the possibility for adjustment of the transmission power of the nodes according to the load per area and time, while ensuring proper performance in the networks in terms of coverage and capacity. By placing the nodes into sleep mode (through on/off cycles) when the traffic demand is notably low, significant gains in energy efficiency can be achieved [51, 52, 53].

Spectrum efficiency

In underlay channel deployment strategy of HetNets, the frequency spectrum is fully reused at each power node. While this strategy provides a better spatial reuse, severe interference can degrade the performance or even create coverage holes. Therefore, advanced interference techniques are required to mitigate both inter-tier and intra-tier interference. Time-domain enhanced Inter-cell Interference Coordination (eICIC) has been standardized in LTE Rel-10 in order to coordinate the interference between tiers. With eICIC, the interfering cells coordinate transmission by configuring subframes with almost no transmission. These subframes are then used by the opposite tier to serve the UEs that have experienced harsh interference conditions. Beside coordinating the transmissions, interference suppression can be achieved through Coordinated Multi-Point (CoMP) transmissions. In CoMP, the main idea is to make the interference signal to become useful, which is achieved such that the nodes within a COMP cluster jointly encode/decode the data traffic. Additionally, recent techniques for interference rejection at the receiver have shown improvements by suppressing the interference and mitigating multipath fading of the desired signal [54].

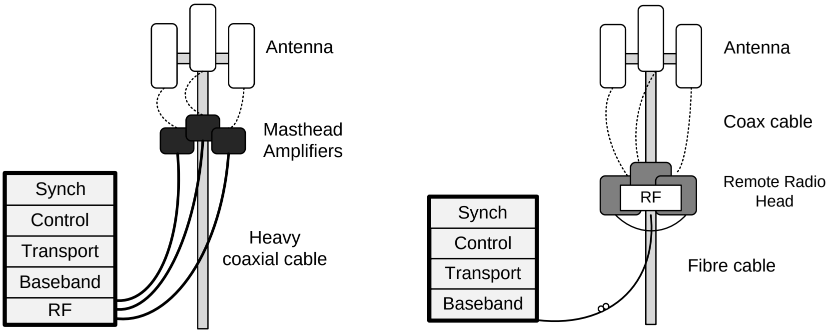
In overlay channel deployments, the low power nodes are deployed on different carrier frequency then the macro cell. Each tier becomes homogeneous, and inter-tier interference problem is avoided. While this is beneficial in case of closed femto cells (nodes with restricted policy for association), it is highly inefficient from spectral point of view due to fragmented spectrum. Another possibility is to dedicate one carrier frequency for macro coverage, while the other one to be shared by the macro and small cell tier. An example of such a deployment is illustrated in Fig. 1.3, where Component Carrier (CC) 1 is used by the macro cell, while CC 2 is shared. Hence, for terminals that can aggregate the reception of both carriers, the entire spectrum bandwidth becomes available. CA was first introduced in Rel-10, to efficiently use fragmented spectrum in order to achieve the data rates required by IMT-Advanced. Afterwards, CA was adopted for non-collocated sites, such as in case of remote radio heads where ideal, high capacity and low latency backhaul exists. With the introduction of DC architecture for small cells, it was natural to extend the inter-site CA for small cells.

One of the goals of this thesis is to study the implications of CA for DC, where the scheduling of radio resources on each CC is independent. The lack of joint scheduling raises challenges in the power allocation for uplink transmissions. Analysis of these challenges and enhanced scheme for power allocation is provided in Chapter 3 that takes in consideration the load of the cell and the cell effective interference. The achieved gains by the model have been compared to the results from other methods [55].

1.1.3 Radio Access Network Architecture

Network densification through low power nodes significantly improves the users' experience providing high data rates on the link between the mobile station and the access point. Nevertheless, the enhancements on the air interface must be complemented with equally efficient backhaul connectivity. In the last years, the backhaul technologies have evolved in several directions. One of them is C-RAN architecture that has been addressed by this PhD dissertation. This subsection introduces the evolution of the radio access networks, while Chapter 5 discusses the benefits and challenges of C-RAN and studies the optimal deployments that lead to increased transport efficiency in the network part between the access nodes and the Core Network (CN). Additionally, Chapter 6 provides analysis of the benefits form resources sharing between mobile operators, that is enabled by the centralization and virtualization features of the C-RAN.

In the last years, the mobile network operators have continuously upgraded their network in order to improve the network capacity, provide extended coverage and most importantly ensure the quality in the service delivered to the users. These network upgrades, mostly related to the RAN, require additional CAPEX and increased OPEX, of which most is related to the radio access network. A large portion of the total cost is associated with the site acquisition, cooling, power supply, equipment and their installation. There has been great focus on research efforts on finding innovative ways for managing the deployment and operation costs. The research has resulted in several base station architectures that are illustrated in Fig. 1.4 and Fig. 1.5.



(a) Conventional Base Station (b) Distributed Base Station

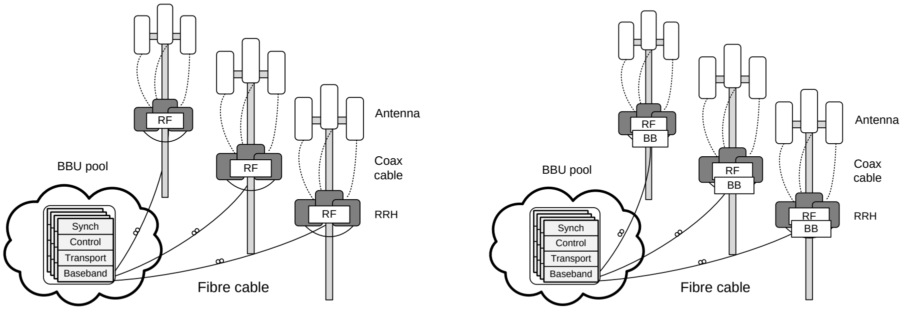
Figure 1.4: Base Station Evolution.

The traditional base station (Fig. 1.4(a)) was mostly utilized for 1^{st} and 2^{nd} generation of mobile networks. The main goal of this antenna is to provide wide area coverage, and therefore, the antenna is usually mounted on a mast or a rooftop. The antenna is connected to the radio module (hardware related to radio and baseband processing) via coaxial cable. As these cables exhibit high power losses, the radio module needs to be located in proximity (few meters). The energy consumption and the equipment cost related to this design is to high, which makes the traditional base station unattainable for dense deployments.

In order to reduce the time-to-market, as well as the cost of deployment and operation, the industry have seen potentials in separating the radio unit and the baseband processing. This has resulted in a distributed design of the BS, illustrated in Fig. 1.4(b). The radio unit part, in form of Remote Radio Head (RRH), is placed close to the antenna, which alleviates the power dissipation due to the coaxial cables. The RRH incorporates an amplifier, frequency filter, up/down converter and performs digital to analogue and analog to digital conversion [56]. The baseband digital processing (PHY/MAC) is performed at the Baseband Unit (BBU), while the interconnection to the corresponding RRH is done via optical fiber or microwave technology. Several standards have been defined for the interface between the RRH and the BBU. The most common protocol used for Inphase/Quadrature (IQ) data transmission is the Common Public Radio Interface (CPRI) specified in [57]. The CPRI protocol is bidirectional constant bit rate and implies very strict synchronization and latency control between the BBU and the RRH. The Open Radio equipment Interface (ORI) protocol complements the CPRI protocol by specifying the control and management plane [58]. Another protocol that can be used is the Open Base Station Architecture Initiative (OBSAI) [59].

The functionality performed by the RRH makes it independent on the nature of baseband data that are to be transmitted thus providing support for multi-mode communications [56]. Beside this advantage, the distributed antenna allows for more flexible deployments as the distance between the RRH and BBU can be extended and the BBU equipment can be located in more feasible place thus allowing reduction in the cost for site rental.

However, with the distributed base station there is still static mapping between the RRHs and the BBUs, and the resource utilization is still inefficient as the dimensioning needs to be done according to peak load. However, as the load at the cell sites varies across time and space, the BBU resources can be shared among heavy and less loaded base stations. This is achieved by C-RAN (Fig. 1.5) where the BBUs from several cell sites are aggregated at one location which is refereed to as BBU pool. Beside the centralization, another important feature of C-RAN is the virtualization/cloudification which allows for general purpose processors to be utilized for the baseband processing [60]. The statistical multiplexing of computational resources allowed by C-RAN [61], improves the resource utilization and effectively results in lowering the required infrastructural resources. As such



(a) Full centralization (b) Split in Baseband Processing

Figure 1.5: C-RAN Architectures.

the costs associated with cite rental, energy consumption as well as network upgrades and maintenance are reduced [62].

Besides the multiplexing gains that leads to energy and cost savings for the operator, C-RAN improves the cooperative radio resource management and enables better quality of service delivered to the user. As the dense deployment is challenged by high interference, mechanisms such as eICIC and CoMP are required. However, their performance greatly depends on the latency in the backhaul network interconnecting the base stations. This challenge is mitigated with C-RAN as the baseband processing from many cell sites may be done over one BBU pool [63]. Another advantage is the service deployment at the edge. A C-RAN network covers a larger area and thus serves more users compared to traditional RAN, thus it is possible to move/deploy new services on the RAN side [64]. Furthermore, the C-RAN architecture enables the active sharing among operators. A novel architecture that was introduced with C-RAN and received great attention is the notion of offering the RAN as a service [65].

With the introduction of the BBU pool, the radio access network is divided into a fronthaul and a backhaul network. The fronthaul network spans between the RRHs and the BBU pool, while the bakchaul connects the BBU pool to the CN. Fig. 1.5 shows two different definitions for the fronthaul network interface in C-RAN. In Fig. 1.5(a) the baseband processing is fully centralized, which results in IQ data transmissions that require high capacity and low latency network with very strict jitter. The data rates at the fronthaul link is expected to be the 12 to 55 times higher compared to the user data rates on the air interface [62], which challenges the economic benefits of the C-RAN due to the costly transport. One way of reducing the required capacity can be achieved with I/Q data compression [66], however, the data rates still scale with the number of antennas used and is independent of the user traffic. Recently, different split points in the baseband processing have received attention in order to relax the burden on the bandwidth. The idea is illustrated by Fig. 1.5(b) such that only part of the baseband processing is centralized [67, 68, 69]. Each split point designates separate interface definition and present particular operational impact on CoMP gains, pooling gains as well as bandwidth and latency requirements [70].

Related to the physical medium used in the RAN, it is important to state that due to the cost of the fiber and the sometimes infeasible locations of the low power nodes, placement of fiber cannot always be justified. In such cases, alternative solutions based on wireless technologies become more advantageous. Solutions based on microwave can provide capacity from 10 Mbps-100 Mbps, and even up to 1 Gbps for short distance (1.5 km) [71]. As the channels between low power nodes and the macro BS is fairly static, wireless backhauling using milimeter-wave (mmWave) frequency has received great attention. Several studies have investigated the feasibility and have shown the benefits of mmWave based fronthaul [72], [73]. Additionally, massive Multiple Input Multiple Output (MIMO) (both for low and high frequency bands) is another capacity improvement techniques that can be adopted for wireless backhauling [74].

1.2 Thesis Structure

The overall scope of this thesis is illustrated in Fig. 1.6. This figure presents the core chapters with their respective area of research as well as the publications that these chapters are based on.

This work performed with the PhD study is presented through seven chapters. This chapter provides introduction to the challenges and defines the problems addressed in this thesis. Chapter 2 reviews the existing RRM frameworks that have been intended for multi-RAT heterogeneous mobile networks. The chapter proposes a novel and universal approach in designing architecture for managing network resources. The proposed framework takes into consideration the heterogeneity in the mobile networks in several areas as elaborated in the introductory chapter. It further applies the SON and cognitive principles in order to provide flexible and effective management over diverse resources such as spectrum, optical and computational resources. The control procedures are organized in hierarchical manner, achieving the benefits from both the centralized and the distributed approach for management.

Chapter 3 moves the focus towards the DC architecture for HetNets, where the connection between the access nodes is based on traditional backhaul networks characterized by certain latency. One of the challenges for uplink resource aggregation in case of dual connectivity is to define the coordinated distribution

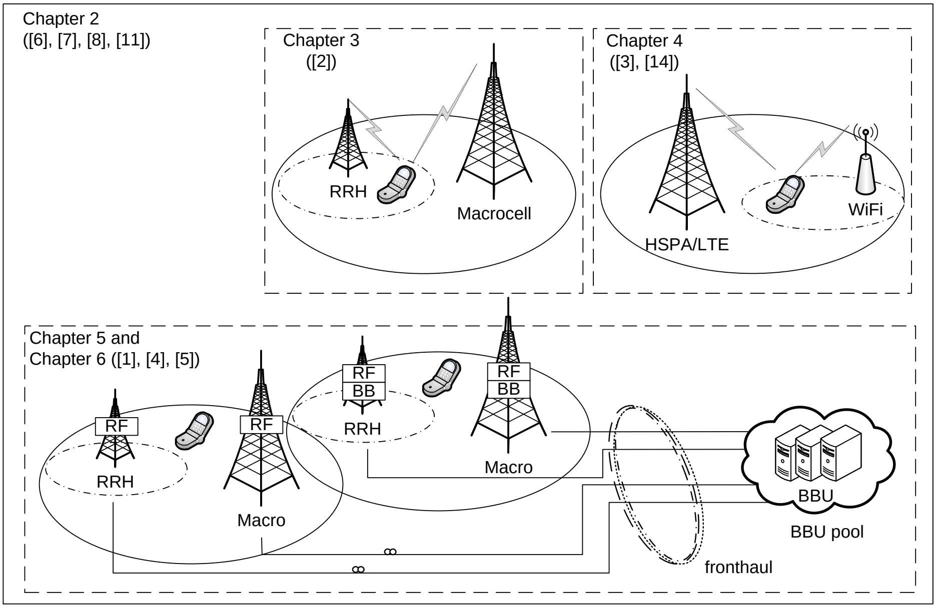


Figure 1.6: Thesis structure.

of the user's power over the individual cells. This issue has been addressed in this chapter by proposing an enhanced uplink power control model for dual connected users. The proposed solution is generalized and could be applicable for future LTE release and/or upcoming 5G standards supporting UE transmissions towards multiple cells.

Chapter 4 enhances the previous chapter by addressing the challenge of adaptive flow control in case of resource aggregation over heterogeneous access nodes. The heterogeneity and the dynamic nature of radio access networks are considered as important factors that determine the performance improvement by multi-stream aggregation. The networks with various qualities of service provisioning, capacity and delay variations are represented by different queuing systems. Services with different traffic characteristics in terms of QoS requirements are considered. The simulation results show the advantages of the proposed model with respect to efficient increase in data rate while ensuring accepted delay compared to traditional schemes.

Chapter 5 drives the topic towards the RAN, by analyzing the challenges and benefits of C-RAN. In the first part, the chapter addresses the importance of proper dimensioning of the BBU pools while considering the QoS and achievable gains by multiplexing the traffic from multiple cells. The challenges in the fronthaul networks in C-RAN are addressed in the second part of the chapter, where the importance in redesigning the baseband processing chain has been indicated. Heterogeneous networks with cells that are served with different splits are evaluated by quantifying the trade-offs between centralization and decentralization concerning the cost of the pool (gains), fronthaul network capacity and resource utilization.

Chapter 6 analyses networks with dynamic sharing of resources, where it is important to carefully design the sharing strategies in order to ensure performance gains in the entire system. Moreover, the effects and implications on each partner that enters a sharing agreement need to be analyzed. As such, strategies for active resource sharing of the radio access network between operators are investigated, identifying the individual benefits depending on the sharing degree.

Finally, Chapter 7 concludes this thesis by summarizing the main results and contributions of this dissertation. Future research directions are indicated as well.

CHAPTER 2

Network Resource Management

This chapter discusses the fundamental principles for network resource management resulting from the evolution of the mobile communication systems. It proposes a novel and universal approach in designing an architecture for managing network resources. The proposed framework takes into consideration the heterogeneity in the mobile networks resulting from diverse radio access technologies, cells formed by a range of low powered nodes and base station architecture. The traditional mobile network features and support for the new advanced multi-cell functionalities are included as well.

This chapter is organized as follows. A brief introduction of the problem and goals of the network resource management is given in Section 2.1. Afterwards, in Section 2.2, the related work is presented where the most relevant cooperative resource management systems are summarized. Section 2.3 elaborates the proposed management framework and demonstrates its flexibility and applicability for resource management in heterogeneous mobile networks. Enhancements to the proposed framework with the self-optimization and cognitive principles are presented in Section 2.4, while Section 2.5 describes the internal architecture and the communications of modules. Section 2.6 discusses challenges in the distributed management and provides mechanisms for improved coordination and decision distribution. The main contributions in this sections are summarized in Section 2.7.

Note: This chapter is based on publication [6], [7], [8] and [14]. The text has been revised and appropriately modified for the purpose of this chapter. A copyright permission is enclosed.

2.1 Introduction

An efficient provisioning of services in heterogeneous access network environments for users with multiple access network interfaces and QoS-based applications requires a universal resource management framework that is independent of the deployed network technologies and architectures. It is also crucial that the network resources are intelligently managed and adapted in strong correlation with the network capabilities and the particular goals of users, service providers and network operators. Furthermore, the resource management should embrace all phases of mobile network development including planning, deployment, operation, maintenance, migration to new technologies, etc.. As such, the network management must also optimize the costs that are related to the mobile network and provide improved revenue to the operators.

The basic radio resource management functionalities are related to call admission and control, power control, spectrum allocation and packet scheduling. However, with the introduction of different access technologies, cells with different sizes as well as mobile terminals with different capabilities, these rather simple mechanisms became complex, and new mechanisms need to be introduced in order to avoid congestion and to balance the load over multiple cells, for example. Each of these RRM functionalities often lead to an NP hard optimization problem, or to an exhaustive search over a large set of possible values for different configurations. As such, each of them represents an important research area in mobile communication. The cooperation and joint optimization over different technologies and architecture facilitate the management of the network as an integral entity. Naturally, this leads to improved performance compared to the performance of each individual entity.

2.2 Related Work

Most of the existing systems for radio resource management in the literature deal with only one type of heterogeneity - that is the multi-RAT. Their architectures are designed to deal with various access technologies and provide interoperability for enhanced service personalization and transparent communicating environment [31]. The specifics of these resource managements differs in the way the resources are managed (centralized vs. decentralized) and in the level of coupling between the networks.

Integration and cooperation among coexisting multi-RAT is prevalent for seamless connectivity anytime, anywhere, anyhow, while not interrupting the active sessions. The Media Independent Handover (MIH) (IEEE 802.11) standard strongly facilitates the handover procedures for transparent handover in heterogeneous environments. With the MIH an intermediate abstraction layer is designed between the Physical (PHY) and the Media Access Control (MAC) layers, that provides services to upper layers and helps the handover information flow through the entities that take part in the handover. The Reconfigurable Interoperability of Wireless Communications System (RIWCoS) [32] architecture for RRM relies on the concept of MIH, and it aims at providing highly reconfigurable heterogeneous environment.

The Common Radio Resource Management (CRRM) [33] developed within

EVEREST project [75] and the Joint Radio Resource Management (JRRM) [34] follow centralized approach for resource management. The centralized way of resource management provides complete view of resource availability. However, with the increased number of entities, the signaling load becomes excessive and it takes longer time to acquire the relevant information. Thus high latency is introduced and as such these systems are more difficult to be realized, especially for dense deployed small cells. The CRRM introduces the notion of a pool of radio resources belonging to different RAT that are managed in a common and flexible way. With the CRRM, different levels of interworking can be defined: low and high degree. The low level coupling is a policy based, the functionalities of the RRM per RAT remain local, and the centralized CRRM enforces policies definitions across the individual RRM entities. On the other hand with the high degree of coupling, the central server is responsible for RAT selection and handover decisions, in which case more frequent exchange of measurements from the terminals and across RAT is required.

The JRRM is based on the device capability for multi-homing and service splitting across different RAT. As each access technology has its own strengths, multiple simultaneous connections allow for integrations of their benefits. For example, increased coverage and reliable communication can be established through 3G, while the data rates can be enhanced with WiFi. The central server in the JRRM manages the overall capacity of sub-networks and is responsible for service prioritization, splitting and aggregation of the traffic streams. This type of management allows for joint radio resource scheduling and admission control.

A distributed approach for radio resource management based on multi-agents is proposed by the Multi-access Radio Resource Management (MRRM) [35]. The agents cooperate with each other in order to achieve global optimization. The agents are characterized with cognitive capabilities, for control, adjustments and monitoring of the assigned radio network and the other agents. Beside cognition, the agents have also a memory layer for intelligent control and fast adjustment based on the history of successful configurations. Another architecture for radio resource management that makes use of the cognitive concept was introduced in the ARAGORN project [36]. With the ARAGON project the different radio access technologies are transparently integrated with the concept of Software Defined Radio (SDR), making the architecture independent of the specific hardware and radio technology.

2.3 Network Management Framework

With the evolution of the mobile networks, the management of network resources has become more complex. The new directions within the mobile network developments, have introduced new challenges for the network management and they

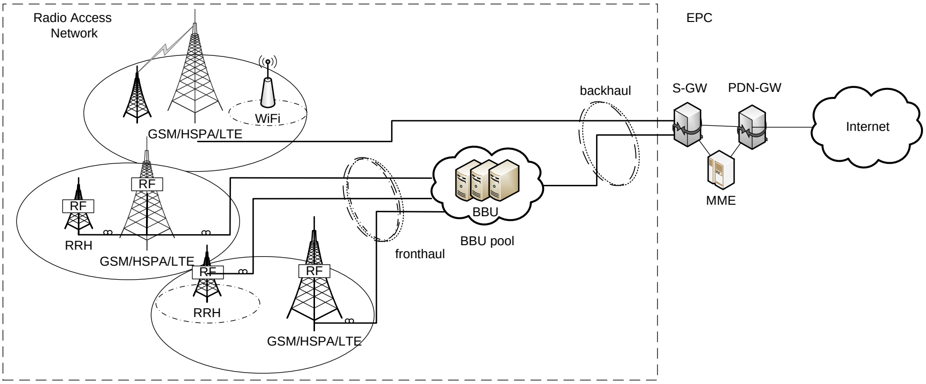


Figure 2.1: Scope of RAN resource management.

can be summarized as following:

- Resource Diversity: spectrum, optical and computational resources. Most of the resource managements mentioned in the related work consider the management of the radio resources on the air interface towards the users. However, as mentioned in Chapter 1 the high data rates on the air interface must be supported with sufficient fronthaul/backhaul capacity. Different medium and technology can be used for the fronhaul/backhaul network, as such diverse resources are introduced: optical, wireless. Many of the functionalities introduced by LTE-A require high-speed signal processing, therefore the computational resources need to be managed in an efficient way. These resources are tightly interconnected and need to be managed jointly in order to ensure the guarantee in the quality of service.

- Functional Challenges: This is related to the introduction of new features in LTE-A such as CoMP, CA, eICIC, DC, and MIMO. These functionalities have specific requirements (for example fronhaul/backhaul data rate and latency), that need to be realized in order to ensure performance gains.

- Architectural/deployment challenges. The introduction of small cells and their dense deployments for increased capacity has lead to new transformation in the management of mobile networks. The new architectures in the RAN due to baseband separation, together with virtualization and cloud computing have resulted in essential changes in the way the mobile networks operate.

With respect to the above challenges, Fig. 2.1 illustrates the scope of the framework for network management proposed in this chapter. This figure indicates

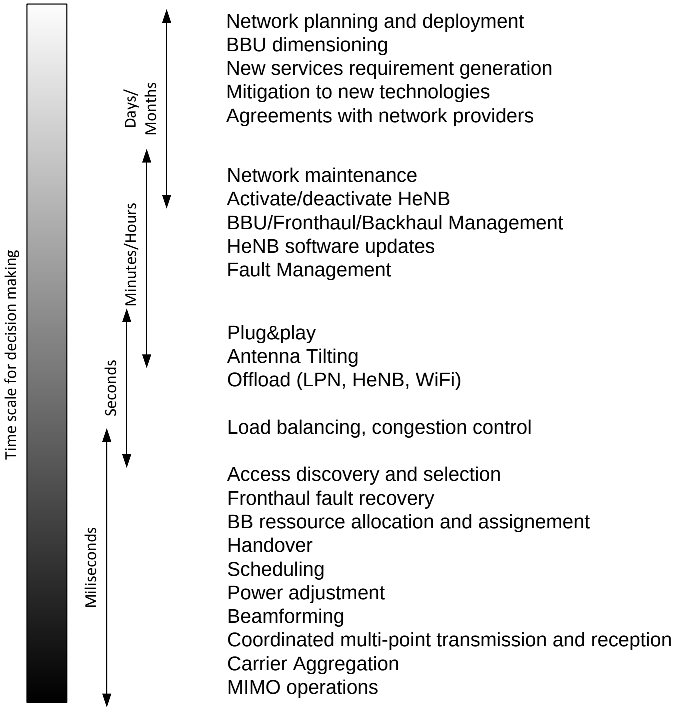


Figure 2.2: Time scale of management functionalities.

the heterogeneity of the mobile networks due to multiple RAT, multiple tiers resulting from cells with different coverage area as well as backhaul connectivity. The discussed resource management platforms in the previous sections do not fully cover all requirements resulting from such heterogeneity. However, some of the concepts are still valid and can be included in the new management framework. The advantages of the cognitive agents for local optimization (as in MRRM) are multi-fold and it is valuable to adopt the concept. Furthermore, the cognitive principles can be adopted on a more global level. SON mechanisms are paramount for automated management of dense and often uncoordinated deployment of small cells (such as the case of HeNBs), and as such should be employed in particular RRM functionalities.

2.3.1 Network Management Functionalities

Fig. 2.2 presents the core functionalities for network management. Depending on the management task that each of these functionalities is responsible for, the time it takes to reach/execute a certain decision can vary in duration from milliseconds to months. As such they can be organized in time interval groups as shown on the figure, which is based on the findings in [76, 77, 78, 79, 80, 81]. It should be noted that the idea is to group the management tasks, and that the figure does not indicate a specific order of time required for a task. Furthermore, some of the tasks are highly flexible - meaning that time interval in which the action is performed depends on a specific scenario/case. For example, antenna tilting can be done electronically, which means it can be controlled very fast. On the other hand, if manual operation is required it may take longer time for this functionality. Another, maybe more representative task is the fault management. Depending on the importance and severity, the task may take from 100ms to days for the fault to be recovered [78]. The decision cycle for load balancing needs to be carefully chosen in order to avoid more extensive signaling than needed to ensure optimal solution [82]. This cycle needs to adapt to the dynamic nature of the traffic, due to which it can change from seconds to minutes (ex a sport event) and hours (during night)[76]. Therefore, the figure does not give a definite grouping of the tasks.

Beside the time scale, these management tasks differ in the area of responsibility or the level of decision and control of resources. For example, BBU dimensioning refers to the task of planning the resources at the BBU pool in C-RAN and more global knowledge of the traffic demand is required. As such the decisions made by this task have global scope and influence larger portion of network resource. On the other hand, beamforming and power adjustments, control the resources with a cell site, hence the decision and actions have local scope.

2.3.2 Hierarchical Management Framework

The management framework proposed by this work is motivated by the time span and decision scope that the different management tasks have. Consequently, the network management tasks can be organized by the following modules:

- Ultra-fast modules: actions that are performed at millisecond time scale.
- Fast modules: require seconds to ensure that the decision takes effect.
- Slow modules: less frequent actions with minutes to hours time span.
- Ultra-Slow modules: tasks that require careful planing and consequently longer time to be executed.

The granularity of the modules is not fixed, and it can be adopted to the needs and preferences of the network operator as well as the type of deployment and equipment used. New modules can be defined, that are in-between the ones defined here, or even define super fast modules.

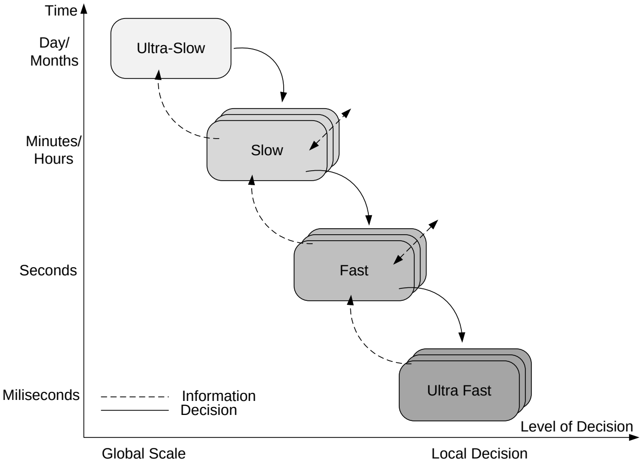


Figure 2.3: Hierarchical Network Management framework.

These modules introduce division of the management task depending on the time it takes for certain control functions to be propagated and executed. Furthermore, in this novel notion of control action division, a hierarchical organization illustrated in Fig. 2.3 is assumed. The main purpose for such an approach is to solve the dilemma between the centralization and distributed resource management. The functionalities that need to make fast effect need to be close to the resources/entities they control, hence such tasks need to be distributed. As time scale and decision scope is increased, the more centralized the control of the resources can be performed. For example the task of load balancing needs to be centralized over multiple cells that cover certain area.

In this hierarchical approach, 1 to N mapping is assumed. The modules have one superior (except the ultra-slow module) and may have several inferior modules (except the ultra-fast module). Each module is controlled by the superior module and delegates tasks to the inferior modules. The modules are able to exchange information to these layers such that the slow and fast modules are also able to exchange information horizontally with modules of the same layer. In the following the modules are described in more detail, and examples for each module are presented.

ULTRA-SLOW MODULES

In the proposed hierarchy of the modules, the ultra-slow module tasks are fully centralized and have the most global view and control of the resources in the network. The decisions/actions that are propagated within the entire network are long-term and performed very rarely. For example, an action that requires global orchestration and need to be triggered within the entire management tree is when software updates are performed. Network planning and dimensioning (for example of the BBU pools) and migration to new technology require careful preparation and have higher effect on the entire network (or on a large portion of it).

Slow modules

The slow modules centralize the information received from the inferior fast modules and use this information when reaching a decision. Furthermore, it provides information of the network resources towards the ultra-slow modules, as well to the peer slow modules. In order to reach a decision and execute certain functionality, information from the peer slow modules may be used. Activation/deactivation of HeNBs may be controlled by the slow modules, which should be based on information such as load, utilization of the resources that are controlled by the underlying layers, as well as neighboring peer modules related to a specific area of interest.

Fast modules

The fast modules need to manage the network resources in a fast manner, within few seconds intervals. The control of the resources is delegated by slow module, and information exchange may be communicated with other fast modules as well as underlying modules. The fast modules control management procedures such as congestion control, which may trigger actions that require fast handover to another access point or change in the traffic flow of barear split in case of dual connectivity.

Ultra-fast modules

The ultra-fast modules are placed on the bottom of the hierarchy in a highly distributed manner. They are very close to the actual network resources, such as the radio spectrum. They control the management procedures that need to make very fast and often decision like scheduling and other baseband operations related to MIMO, CoMP and beamforming. The reason for not including any information exchange among ultra-fast modules is to indicate that such communication requires ideal connections. Furthermore, the ultra-fast modules of several entities may be collocated such that they orchestrate the baseband processing for several cells. This could be the case of C-RAN or distributed RRH with joint baseband processing that is required in order to achieve gains from CoMP or CA for example.

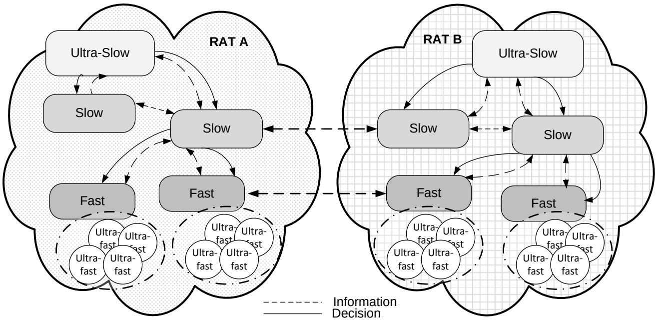


Figure 2.4: Inter-RAT management cooperation.

2.3.3 Management of Heterogeneous Networks Resources

The proposed framework for network resource management, based on the modules described above, is highly flexible and is technology independent. It can be applied for any type of heterogeneous network, including multi-RAT and multi-tier heterogeneous networks, as well as BS with various architectures that were presented in Chapter 1. It provides support for existing and new functionalities of LTE-A, such as CoMP, MIMO, resource aggregation in terms of LAA, CA, DC and LTE-WiFi aggregation. Furthermore, it considers different stages of network evolution, such as dimensioning, migration to new technologies, planning and implementation of new services, software upgrade etc.

Fig. 2.4 illustrates the relationship among the modules within a single RAT and in case of inter-RAT cooperation. It is most appropriate to enable information exchange among the slow and fast modules, and it is implementation specific. The communication related to a certain functionality depends on the network operator preferences, the type of integration (tight vs. loose) as well as on the requirements of the functionality. The information exchange between the ultra-fast modules may require extensive signalling and it may be time consuming. However, different types of cooperation can be established. For example in case of LAA, the implementation of the resource scheduler over the unlicensed spectrum needs to ensure time access fairness with Wi-Fi and conform to regulations in various regions.

The proposed framework provides support for management of diverse network resources: spectrum, optical and computational resources. The management tasks that directly control these resources are distributed and placed close to physical resources, while the higher layers in the framework provide centralized and joint orchestration of these resources. This way, the management framework allows for ensuring services by allocating resource in a flexible and dynamic way in order to respond to the demand of different use cases and specific implementation of a service. This demand can be expressed for example in data rate, latency or coverage. The resources from the higher layers are seen more in logical rather than physical instances. This is of great relevance to the future way of building and operating networks in order to accommodate various vertical industries. This is the main idea behind the concept of logical network slices that will enable operators to provide networks on an as-a-service basis and meet the wide range of use cases [83].

2.4 SON and congnitive aspects

The introduction of self-optimization for automation of the network operation is an advanced and appropriate mechanism to deal with the rising complexity of the mobile networks due to heterogeneity as well as to provide optimization in the utilization of the diverse network resources [76]. The SON mechanisms are becoming crucial for the mobile operators due to the enormous potentials in minimization of the costs related to network operation and maintenance. SON standardization was introduced to LTE with Release 8 [84] by defining use cases procedures and open interfaces in order to ensure interoperability among different vendors.

The SON use cases are categorized in three distinctive functionalities along the operation, administration and maintenance of mobile networks:

- Self-Configuration,
- Self-Optimization, and
- Self-Healing.

These processes can be adopted in the proposed framework for resource management. The self-configuration and self-healing can be done at the fast and slow level. The specific layer where the actual functionality needs to be implemented depends on its complexity and importance. Self-healing relates to detection of degradation, root cause analysis and compensation. Dealing with cell outage can be detected for example, by monitoring system variables and performance indicators, which can instantly trigger an action such as increasing the transmission power of a small cell. On the other hand, self-configuration deals with the auto-configuration of new and existing network elements and automatic cell configuration which has more global scope of information and decision spread. Self-optimization is highly beneficial for the ultra-fast and fast modules, where the resource are managed in more local manner and the decisions/actions need to be planned and performed immediately.

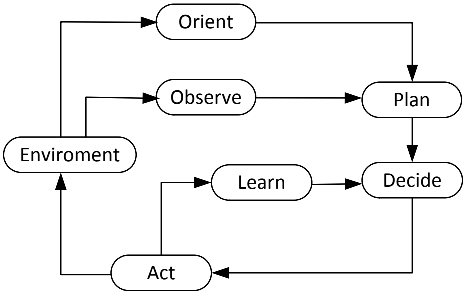


Figure 2.5: The cognitive loop based on [85]

The self-optimization refers to optimization of the network operation where it is important that the individual entities exchange information and alert the neighboring and upper/lower layers in case of violation of Key Performance Indicators (KPI). Different algorithms can be implemented to reflect self-optimization, that is related to functions like Mobility Robusteness Optimization (MRO), Mobility Load Balancing (MLB), Coverage and Capacity Optimization (CCO) and energy efficiency. Information related to the load of the cells and the neighboring cells, the availability of different tiers and cells based on different RAT, the capabilities of the terminals and the requested service by the user may not be available within one module. The self-optimization functionality may be further spread in the sense that some modules will act as detection point, Policy Decision Point or Policy Enforcement Point. These categories define if the module will monitor for violation of the KPI, will plan and forward the decision or execute the action based on the given commands.

Further development of SON can be achieved with the cognitive concepts [86] and relevant directions from Artificial Intelligence (AI) through machine planning and machine learning. A cognitive cycle contains the following phases: observe, orient, plan, decide, learn and act while their interaction is illustrated in Fig. 2.5. The cognitive concept can be adopted to individual modules, in order to intelligently adapt to the changes in the environment. By sensing the environment and monitoring the state of different management entities, the necessary adjustments and actions can be planned and afterwards these changes can be acted upon. The knowledge from previous states and decisions and the learning ability represent significant part of the cognitive loop in order to accelerate the decision making and avoid conflicting actions with other entities.

However in highly complex multi-RAT, multi-tier environment with various base station architectures, there is a multitude of network parameters that need to be optimized. Many of the parameters are related to several management

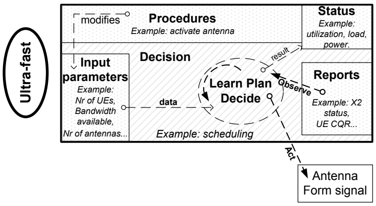


Figure 2.6: Inner architecture for ultra-fast module

functionalities, hence the interactions and conflicting decisions need to be resolved in an efficient manner. For example, the adjustment of the transmission power on a small cell can be directly related to different functionalities such as: CCO, MLB and energy optimization. Furthermore, the modules need to be aware of the decisions of the other modules and in timely manner take their configuration updates into consideration when making their own decision. Some of the decisions can be made simultaneously, and therefore it is important to coordinate the decisions in order to reach a stable system. So, in such a distributed management, it is important to design mechanisms for avoiding potential conflicts. Such a method for coordination among small cells is proposed in Section 2.6.

2.5 Internal architecture and communication of modules

The internal architecture of different modules (levels) of the management framework is generalized such that it adapts to the cognitive concept. Each module consists of the following basic elements (units): input parameters, decision, procedures, status and reports. Fig. 2.6 illustrates the units for the ultra fast module. Although on the figure the different units are placed under one block, they actually may be located on different physical locations. The text below explains the role of each unit and an example is given again for the ultra fast management module.

- Input parameters: All the parameters that can be configured / modified from the modules of the higher level are considered in this unit. They can be considered as arguments for the unit responsible for reaching a decision and can include different thresholds and goals that need to be reached by the module. For the ultra-fast module, the assigned bandwidth (number of resource blocks per TTI) and the number of UEs that are assigned to the scheduler could be considered as input parameters.

- Status: All the performance measurements that need to be monitored by different modules are considered in this unit. In case of the ultra-fast module, this could be the utilization of physical channels, e.g. downlink, uplink or control channels.

- Reports: The reports are used by the units in order to monitor/sense the environment. This information is to be used in the decision process. For example, the scheduler in the ultra fast module uses Channel Quality Reports (CQRs) to reach a decision on the distribution of the resource blocks.
- Decision: This unit contains the algorithm/tool for reaching a decision based on the reports (the observation) and the status updates from the different modules (it orients to other modules). Before reaching a certain decision, this unit also includes the planning process as well as the learning (remembering previous decisions) process. Each module has a defined range of parameters that define the area of responsibility. The decision can trigger certain procedure and act upon a parameter and/or propagate the decision towards the lower modules. In case of the ultra-fast module, this unit is represented by the scheduler design.
- Procedures: Based on the decision reached, a certain procedure (action) needs to be executed by the module. For example the called procedures can set certain configuration parameter. In the case of the ultra-fast module, the power settings can be adjusted by the fast module.

Fig. 2.7 and Fig. 2.8 illustrate the communication among different modules. Two different communication directions are defined: layer and peer communication. Layer communication takes place between modules that belong to two different management levels. It is represented by the actions (procedures) and the reporting unit (observation). Peer communication is among modules that belong to the same levels and it is represented by the orient state. In order to reach a decision, all modules need to consult (orient) the status of different modules. For instance, in case of the load balancing, different performance measurements need to be constantly monitored, such as utilization, load and energy consumption [87]. In this type of communication, there is no master control of the information exchange, nor centralized database. Therefore it is of great importance to solve the problem of conflicting decisions/actions that need to be performed by these modules. This problem is further elaborated in the following section.

2.6 Control Mechanism

In case of centralized management there is a single entity that performs optimization tasks over a global area and enforces the decisions across each entity. This approach implies significant control and information exchange and moreover, for the case of dense deployed small cells, information aging and control delay

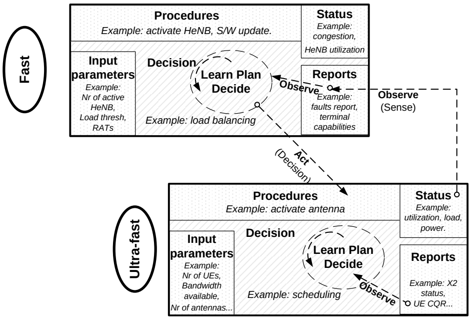


Figure 2.7: Example of module communication between different levels

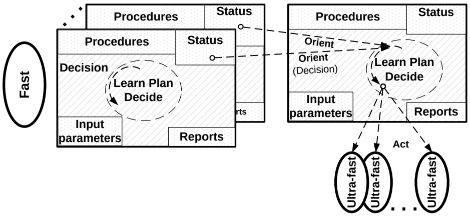


Figure 2.8: Example of module communication among modules in one level

will degrade the overall performance. Therefore, distributed management is more appealing but, as discussed in Section 2.4, the decision making process at modules need to be coordinated. This section proposes a method for communicating the decision among the neighboring entities, which is illustrated in Fig. 2.9.

In Fig. 2.9, the optimization procedure is considered to have the following time periods: sensing and reporting (orienting), decision, action and frozen period. During the overall operation, each cell needs to continuously monitor the environment and inform the neighboring cells of its current state. The actual intervals for reporting and alerting depend on many parameters such as current load, the time of the day, expected traffic demands, power consumption and so on. If the measured KPIs are getting closer to the actual thresholds, the alerting interval could be reduced. If the KPI thresholds are violated then the optimization procedure needs to reach a decision that will eventually result into an action (e.g. change of transmission power, antenna tilting).

The question of who should perform an action among the neighboring cell is resolved as in Carrier Sense Multiple Access (CSMA) scheme. Before an action is conducted, a notification is sent to all neighboring cells. If a conflict occurs (two cells receive their notification for action), it is resolved by applying a random

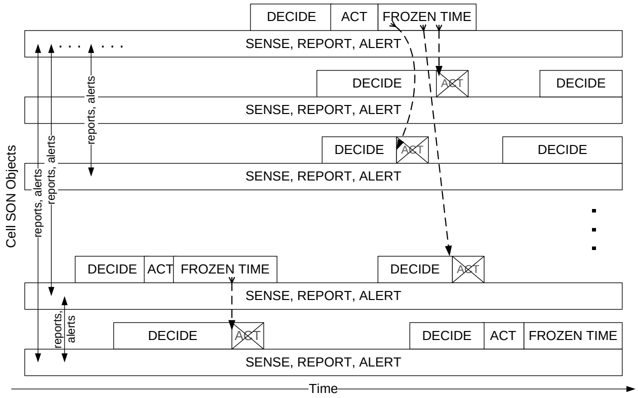


Figure 2.9: Communication time frame

backoff period for the action. This means that a frozen period (similar to busy channel in CSMA) occurs for all neighboring cells. During the frozen period the neighboring cells need to monitor the environment again before reaching a decision and possibly notifying and performing an action. This period is required in order to make sure that the small cells have taken into account the change in the environment as a result of the performed action. The frozen period can be predetermined or agreed among neighboring cells. The duration of the periods is dynamic and depends on the implemented optimization algorithm as well as on the current state of the cell and neighboring cells.

2.7 Summary

This chapter has elaborated on the essentials for resource management in a heterogeneous network, where the joint management of radio, optical and computational resource is essential for performance optimization. A novel universal approach for resource management that facilitates the emerging technologies for spectrum, energy and cost efficiency was presented. The proposed framework is generic and can be applied for various scenarios and network deployments and different functionalities related to resource management in a multi-RAT network, (dense) small cell deployments and BS architectures were addressed.

The structure of the framework is hierarchical where information exchange is foreseen among neighboring modules. This approach provides balance between centralized management characterized with global knowledge but excessive signaling and distributed management with local control of resources. The proposed hierarchical structure requires thorough arrangement of the management functionalities in order to ensure synchronization of the individual decisions and joint management of wireless, optical and computational resources. Each of the individual modules is enhanced with cognitive behavior and learning abilities in order to assist the decision process and timely adapt to the changes in the network. The communication towards the higher layers allows for establishing global goals, while the communication with the lower and peer layers allows for more general and detailed knowledge of the system performance. The cognitive concept and the SON principles augment the proposed framework allowing for more efficient and automated resource management.

CHAPTER 3

Uplink Power Control for Dual Connected Users

The focus of this chapter is on the implications of dual connectivity architecture for HetNets, where the connection between the access nodes is based on traditional backhaul networks characterized by a certain latency. As the users can be configured with two radio access nodes, one from the macro cell tier and the other from the small cell tier, the maximum data rates from the configured cells can be aggregated. However, as non-ideal backhaul is assumed, the radio resource management at each node is independent which impacts the gains from resource aggregation. In uplink direction, the maximum power of the user can be easily exceeded and even more the interference towards the other cells can be increased. Therefore, one of the challenges for uplink dual connectivity is to define coordinated distribution of the user's power over the individual cells. This issue has been addressed in this chapter by proposing an enhanced uplink power control model for dual connected users with 3GPP Rel-12 LTE dual connectivity as main reference. The proposed model takes in consideration the load of the cell and the cell effective interference. The solution is generalized and could be applicable for future LTE release and/or upcoming 5G standards supporting UE transmissions towards multiple cells. Furthermore, a criteria for dual connectivity configuration is determined. The performance assessment of the proposed model is done via advanced system-level simulations, where important mechanisms that influence the performance have been accurately modeled. The achieved gains by the model have been compared to the results from other methods.

This chapter is organized as follows. Introductory motivation behind this work is presented in Section 3.1. Section 3.2 provides an overview of dual connectivity and the challenges for resource aggregations with this architecture. The tradition power control for uplink and the implications due to dual configuration are are discussed in Section 3.3. Section 3.4 elaborates the proposed optimization for users' power allocation. Section 3.5 presents the simulation assumptions under which the proposed optimization model is evaluated and discusses the obtained results. The last section concludes this chapter and remarks future directions.

Note: This chapter is based on publication [2]. The text has been revised and appropriately modified for the purpose of this chapter.

3.1 Introduction

HetNets, where low power nodes co-exist with macro BS, have dominant role in addressing the surge of mobile data traffic increase, the uneven traffic distribution and limited spectrum availability [88]. The low power nodes are characterized with smaller size/weight and lower cost of deployment, which make them an attractive solution for fast deployment at hot spot areas both for indoor and outdoor purpose. Heterogeneous networks with small cells bring many potential advantages compared to homogeneous networks. However, as already discussed in Chapter 1 there are fundamental and technical issues that need to be addressed in order to fully benefit from such deployments.

As additional spectrum becomes available at higher frequency such as 3.5 GHz and above, frequency separated small cell deployment has received great attention both by the wireless industry and research community [89]. Due to the propagation characteristics, higher frequency bands are more suitable for small cells, where the distance between the small cell node and the users is shorter. Nevertheless, improving per user-throughput by utilizing radio resource at one cell is difficult [90], and resource aggregation over multiple nodes motivates the research towards better integration of small cells, especially in case of non-ideal backhaul. In order to bring the interworking between the small cell and macro cell tier closer, and allow for improved and aggregated resource utilization at both tiers, novel HetNets architecture, known as DC, has been proposed and gained much momentum by the 3GPP forum.

In dual connectivity the control plane used for connection establishment and the user (data) plane are separated and may no longer transmitted by the same physical node. For each user a single control connection is maintained which makes this architecture an attractive solution that meets the challenges related to the energy efficiency and mobility robustness. Furthermore, the user plane can be split and transmitted through different nodes, such that the split can be dynamically adapted to the radio conditions and load of the multiple configured cells. Significant gains in users' throughput can be achieved by such inter-node radio resource aggregation. The cell-edge users, for example, can further benefit from additional capacity by having multiple transmitting/receiving streams. While downlink inter-node resource aggregation in DC is supported with 3GPP Release 12 [49, 50], there are still technical challenges for supporting uplink resource aggregation and further studies are required to demonstrate the potential benefits. This chapter direct the attention towards this potential by studying the challenges and requirements for improved uplink data rates. The core topic of this chapter is the uplink power control as it plays a significant role in providing the required SINR while controlling the interference caused to the surrounding cells.

3.2 Dual Connectivity

Under the assumption of non-ideal backhaul, the dual connectivity provides an architecture for higher-layer interworking between macro and small cell tier with the aim to optimize resource utilization. This type of interworking between the macro and small cell tier allows for distributed deployment and operation of the access nodes [91].

The architecture in DC is based on the split concept between the Control (C)-plane and the User (U)-plane and was first introduced to the open literature as soft-cell scheme [92] and Phantom or Macro Assisted (MA) small cell solution [93]. Common to both is the fact that the small cell no longer transmits cell specific signals and channels, while their visibility and connectivity relies on the signaling from the macro cell. This way the macro cell becomes responsible for UE channel establishment and release, while the MA cell transmits UE specific data. A more detailed proposal for architectural framework for DC is given in [94], and it relies on Channel State Information-reference signals (CSI-RS) to differentiate among the small cells and to assist small cell (re)selection procedures.

3.2.1 Dual Connectivity Protocol Architecture

A DC capable UE can be configured with two radio access nodes so that it is able to consume radio resources from both nodes in more efficient manner. The concept of DC for a considered scenario by 3GPP is illustrated in Fig. 3.1. The two nodes involved in DC are referred to as Master eNB (MeNB) and Secondary eNB (SeNB). Multiple serving cells can be associated with each evolved Node B (eNB) and they are referred to as Master Cell Group (MCG) and Secondary Cell Group (SCG). One primary cell (referred to as PCell from MCG and PSCell from SCG) in each group is configured to carry the physical uplink/downlink control channel and is used for radio link monitoring. The MeNB terminates the signaling (S1-MME) interface dedicated for exchange of signaling messages between the Mobility Management Entity (MME) and the eNB. There is only one S1-MME per UE, and the Signaling Radio Bearer (SRB) cannot be split or offloaded to the SeNB. The MeNB becomes responsible for the Control (C)-plane of the UE, such that the Radio Resource Control (RRC) signaling towards the UE required for connection establishment, modification and release is handled by the MeNB (Fig. 3.2). In case of radio link failure towards the SeNB, the UE stops sending and receiving and needs to inform the MeNB via RRC message. The measurement reporting and configuration is also controlled by the MeNB. The SeNB serves to provide additional radio resources in order to improve the user's throughput. The coordination between the MeNB and SeNB is done via the X2-Control (X2-C) interface. This interface is used for SeNB addition, SeNB

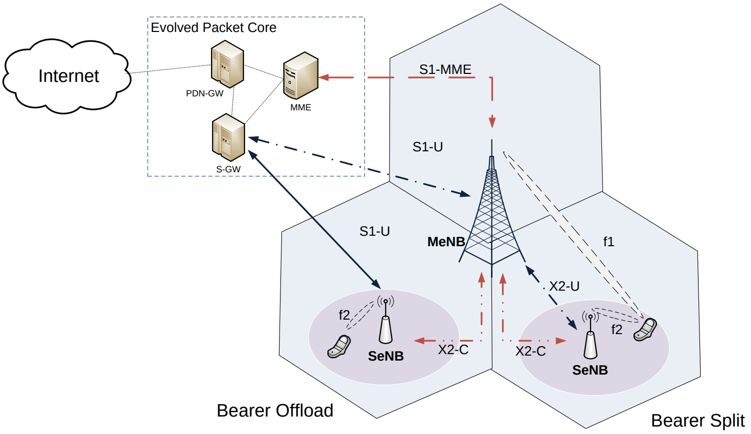


Figure 3.1: Deployment scenario for Dual Connectivity

modification (modification of UE bearer context), SeNB change (with respect to mobility) and SeNB release.

While the (C)-plane is fixed, two alternatives exist for the User (U)-plane that handles user data and they are illustrated in Fig. 3.2. The S1-U interface for user plane tunneling can either be terminated at the SeNB or at the MeNB. When terminated at the SeNB, the data and the signaling radio bearers go through different eNBs and this type of bearer option is also known as *bearer offload*. This alternative can improve the load balancing between tiers, but it does not allow for resources utilization across the involved eNBs for the same bearer. The other alternative, referred to as *split bearer*, allows for resource utilization across eNBs. As seen from Fig. 3.2, the eNBs have independent Radio Link Control (RLC) and MAC layers, while the PDCP layer at the MeNB becomes responsible for transmission of data units towards the SeNB and packet reordering at reception. The user data is transmitted over the X2-User (X2-U) interface, hence this split requires a good backhaul between the eNBs. As the data plane is split at the MeNB, mobility robustness is further improved as the small cell mobility is hidden towards the core network and no data forwarding is required between the involved SeNBs.

The bearer split in DC can significantly increase per-user throughput by resource aggregation and enable more dynamic load balancing between eNBs. The benefits for cell-edge users in case of DC under realistic and irregular HetNets deployment studied in [95] are evident both for low and high load in the network. In order to fully harvest the possible gains from bearer split, several technical challenges need to be addressed. These challenges are discussed in the following

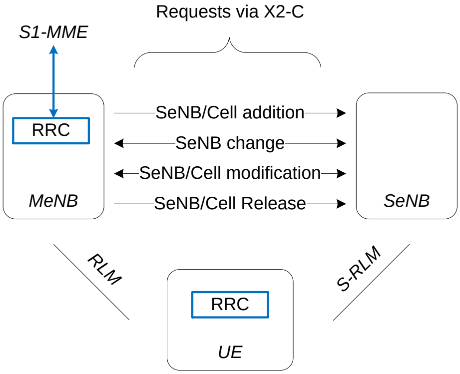


Figure 3.2: Control Plane Architecture for Dual Connectivity and SeNB management

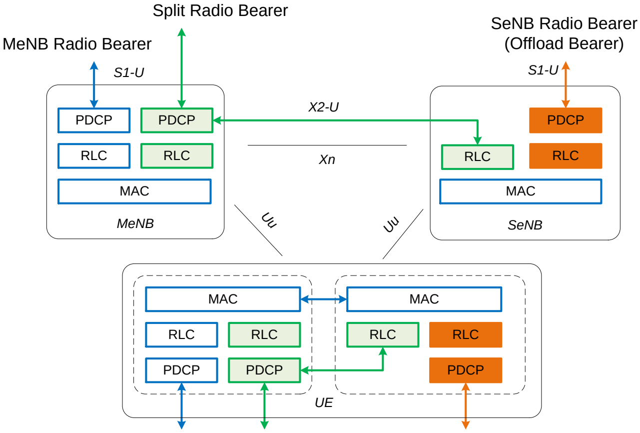


Figure 3.3: User Plane Architecture for Dual Connectivity

text, such that more emphasis is given to uplink transmission.

3.2.2 Challenges for Split Bearer

Split bearer in DC extends the inter-site CA functionality for small cells. The main difference to inter-site CA is that in DC each eNB owns its radio resources and their scheduling is independent (without tight coordination in form of cross-carrier scheduling) due to the non-ideal backhaul. Packet scheduling together with cell association play an important role in efficient resource utilization and load balancing. With DC, the cell association problem is defined as selection of the optimal pair of MeNB and SeNB for a given UE and the configuration of the data bearer type. The optimal pair of eNBs should consider the signal strength of available cells, the cell load as well as the backhaul connectivity [96]. The sum-rate maximization problem for configuring UEs with either bearer split or bearer offload addressed in [97], shows that if all UEs are configured with DC with bearer split, the network capacity may be compromised. Hence, not all UEs can benefit from split bearer and therefore additional criteria need to be considered when configuring the UE with DC. A simple analysis in [98] shows that the throughput gains per UE are increased if the UE experiences similar channel conditions and similar bandwidth availability at the two layers.

In the downlink direction, the data has to be forwarded by the PDCP layer at the MeNB towards the SeNB over the X2-U interface. Therefore, the data transmission from the small cells will be influenced not just by the buffering time at the SeNB but as well by the backhaul latency. In order to avoid buffer overflow and packet loss or data starvation, efficient flow control of the data stream between the eNBs is required. Exchange of flow control information in form of buffer size and user's throughput towards the MeNB is required in order to assist the MeNB in deciding how much of the data needs to be forwarded towards the SeNB [99]. The flow control should target the data rates experienced at the SeNB as well as be robust to backhaul and traffic conditions [98]. Flow control mechanisms are required in the uplink direction, where the PDCP layer at the UE should effectively split the flow in order to avoid out-of-order packet arrival at the MeNB. As the split bearer in both directions depends on the X2 link conditions, it implies that packet reordering at PDCP is always needed, unlike in the legacy LTE where it is required upon handover and RRC re-establishment [90].

In downlink, the packet scheduling at the SeNB is influenced by the data flow from the MeNB. However, in uplink, due to the independent resource and power allocation, packet scheduling becomes more complex. As there is no information exchange about the number of resources and assigned Modulation and Coding Scheme (MCS) for each UE, the total power required by the allocated grants from both eNBs can exceed the power limitation at the UE. In this case power scaling would be required, which may lead to lower MCS to be used by the UE, or even cause a drop of UL transmission. The lack of centralized processing for power allocation can further lead to increased uplink effective interference in each layer, nullifying the benefits of dual connectivity. This is investigated in [55] where the analysis shows that as the load in the network is increased, the decision for configuring users with bearer split should take into consideration the additional interference that the user will create towards the secondary cell tier. Power scaling and increased interference lead to decreased spectrum efficiency and in order to ensure optimal resource utilization, coordination between eNBs is required. Furthermore, a guarantied maximum power used for uplink per eNB should be determined.

In traditional eNB, the scheduler derives a scheduling metrics for each UE in order to prioritize among the users. The number of assigned resource blocks (time/frequency pair) is determined by the available power at the UE and the MCS chosen to achieve certain value for the BLock Error Rate (BLER). In order to ensure benefits from bearer split in uplink at least the following technical issues should to be re-thought:

1. *Fairness among users.* The Proportional Fair (PF) based scheduler design ensures fairness among all users within the cell. Hence, users that are located at the cell edge will benefit by more frequent resource allocation, in order to achieve the required data rates. However, having a cell-edge user configured with bearer split can create imbalance in fairness towards users that are not configured with bearer split. Therefore, when calculating the scheduling metric the average user throughput aggregated from all configured cell needs to be considered instead of the users' per-cell throughput. This means that it is important that the eNBs exchange information regarding achieved data rate per user.

2. *Buffer status reporting.* With the bearer split illustrated in Fig. 3.3, the RLC entities at the UE are independent, and therefore the buffer status in each RLC entity can be reported to the corresponding eNB. However, the complexity in Buffer Status Report (BSR) comes from the fact that the PDCP entity is shared. By reporting the full PDPC buffer status to both eNBs may result in resource allocation that can be wasted, especially if the data portion to be transmitted is small. Therefore, a split ratio need to be introduced in BSR for the PDCP layer, such that a virtual status is created for each logical link [100]. The split ratio should reflect the channel conditions and the load of the cells in order to improve the load balancing and resource utilization. As the split ratio will influence the amount of resources allocated per eNB, it is important to consider the backhaul latency

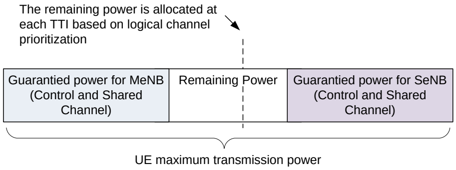


Figure 3.4: Distribution of UE maximum power

when defining the split ratio as it will have impact on the packet reordering at the MeNB.

3. *Power distribution between MeNB and SeNB.* As discussed above, if the UE receives scheduling grants from the two eNBs within one Transmission Time Interval (TTI), there is possibility that the total power allocated by the eNBs will exceed the power limitation at the UE. In order to avoid power scaling, it is important to define a guarantied power that each eNB can use for the transmission towards that eNB [101]. This is illustrated in Fig. 3.3, which shows that the sum of the guarantied power per eNB may be lower than the total UE power. Then, the remaining power can be dynamically allocated by either eNB, such that in general, a higher priority is given to transmissions towards the MeNB and it depends on the Uplink Control Information (UCI) type. Several semi-static power scaling schemes have been investigated in [102]. The distribution of the maximum power between layers have been investigated in [103] and [104], where the configuration is done per UE, and require additional signaling of UE channel state information among the cells.

4. *Power Headroom Report (PHR).* The PHR is used by the scheduler in order to determine how much additional power can be allocated to the user before reaching the power limit. By reporting the power headroom to all configured eNBs, each eNB will be able to evaluate the pathloss and resource utilization of the other eNB. This will help the schedulers to make more optimized decisions on resource allocation by avoiding too aggressive or too pessimistic resource allocations. The PHR may be reported over the air, or the information can be exchanged via the X2 interface. In both cases the schedulers need to agree on the maximum power that the UE can use per configured cell.

5. *Logical Channel Prioritization (LCP).* The LCP procedure in legacy LTE is based on token bucket algorithm, such that for each logical channel a priority level is assigned by the eNB. Additionally, a Prioritized Bit Rate (PBR) and

Bucket Size Duration (BSD)[1] are defined in order to avoid starvation of lower priority channels and are related to the QoS defined for the bearer. As the (C)-plane is carried over the UE-MeNB link, more control elements will exist in this link compared to UE-SeNB link, leading to non-uniform prioritization for the same bearer between the two MAC entities. Additionally, if there is only one PBR defined for the split bearer, it may be enforced twice by the two MAC entities, leading to unfair resource scheduling. Hence, the LCP between the two MAC entities at the UE need to be in tight coordination. This can be done through a common token bucket, where a single bucket is defined for the split bearer and the number of tokens that are removed is derived from the sum of the data units generated at both MAC layers. The second option is to define a share of the PBR towards the bucket of each MAC entity.

6. *Sounding Reference Signal (SRS).* If the transmit power limit is exceeded, then the power is distributed such that higher priority is given to the MeNB and control channels (for example the Physical Uplink Control Channel (PUCCH) that contains HARQ/ACK has higher priority than Physical Uplink Shared Channel (PUSCH) without UCI). The lowest priority is given to the SRS [101]. The SRS is used by the eNB to extract the frequency selective uplink channel quality, which is used by the Link Adaptation (LA) to derive the most appropriate MCS as well as the scheduling in the frequency domain. Hence, if there is no sufficient power, there are two possibilities: the SRS transmission can either be dropped due to power limitation and scaling avoidance, or different Power Spectral Density (PSD) can be used for SRS and data transmission on PUSCH. Both cases impact the LA, as it may fail to correctly estimate the channel. An error need to be accounted due to the difference between estimated SINR based on the uplink SRS strength measurements and the experienced SINR on the data channel.

In this chapter, the main objective is to provide a simplified uplink power allocation for DC that reduces the information exchange among eNBs, while the gains with respect to achieved user throughput are maximized. The assumptions for these technical issues are also elaborated in more details.

3.2.3 Advantages and Considerations in Dual Connectivity Architecture

Interworking between the macro cell and small cell tier through DC brings several other advantages. As the small cell no longer need to transmit acquisition and

---

[1] the BSD determines the maximum bucket size of a logical link

reference signals, the energy efficiency in the network can be significantly improved in conjunction with more dynamic on/off schemes and power transmission adjustments. Longer sleep cycles when the small cell is neither sending nor receiving data bring potential energy efficiency gains in the network [105]. The small cell can be awaken by the macro cell whenever capacity enhancement and traffic offload is required. During sleep mode the low power nodes do not transmit any pilot signals, which challenges their discovery by the UEs. A database-aided connection procedures is proposed in [106] for this purpose. Additionally, the energy consumption of the (ideal) backhaul networks can be considered in the cell activation criteria and offload to the small cells [107].

As the macro cell handles UE connection establishment, modification and release, DC allows for enhanced mobility performance in terms of low probabilities of experiencing handover failures or ping-pong events [108]. The mobility robustness comes at a cost of increased workload at the macro cell [109]. Although the macro cell hides the small cell mobility towards the CN, signaling between the UE and macro node in case of addition/removal of secondary cell from the SeNB is still required. In case of dense deployment of small cell, measurements reports from the UEs sent in uplink RRC signaling can cause significant overhead. The overhead becomes more evident as the velocity of the UE is increased. This issue has been addressed in [110] in form of inter-site CA and a solution is proposed such that the small cell mobility management is based on UE autonomous decisions with a certain degree of network control.

3.3 Uplink Power Control

3.3.1 Traditional Power Control

The Fractional Power Control (FPC) scheme [111] used for uplink transmission in LTE has two components: Open Loop Power Control (OLPC) and Closed Loop Power Control (CLPC). The OLPC serves to compensate the slow varying path gain and shadowing. Further optimization of the system performance can be achieved via the CLPC, which provides adjustment of the transmission power in order to compensate for rapid variations and errors in path loss measurements.

In this work, the focus is on the OLPC, according to which, the power allocation at transmission time t, for UE i associated with cell k in dB scale is defined as:

$$P_{i,k}^{t} = min\{P_{max,k}, P_{0,k}^{t} + 10log_{10}M_{i,k}^{t} + \alpha_k \cdot PL_{i,k}^{t}\} \tag{3.1}$$

In eq. (3.1), $P_{i,k}^{t}$ represent the transmission power at time t for UE i limited by the maximum power of $P_{max,k}$ that can be defined per cell such that $P_{max,k} \leq P_{max}$. $M_{i,k}$ represent the number of Physical Resource Blocks (PRBs) allocated to the user by cell k.

The parameters that determine the performance of the OLPC [112] are the normalized power density ($P_{0,k}$) and path loss compensation factor (α_k). The purpose of α_k, which is a cell specific parameters is to define the fraction of the path loss that will be compensated. With $\alpha_k < 1$, the transmission power of the cell-edge users is reduced compared to the case of full compensation ($\alpha_k = 1$). This way the OLPC regulates the generated interference to the neighbouring cells. The value of α_k should be chosen such that the overall performance in the cell is improved, while minimizing the impact on the cell coverage.

There are different methods for adjusting the normalized power density, which can be user or cell specific parameter. In this work the Load Adaptive Power Control (LAPC) detailed in [113] has been used. The LAPC dynamically adjusts the user's power spectral density depending on the bandwidth variations. According to the LAPC, $P_{0,k}^{t}$ is adjusted periodically by the following equation:

$$P_{0,k}^{t} = P_{max} - 10log_{10}M_{i,k}^{avg} - \alpha_k \cdot PL_{95\%,k} \tag{3.2}$$

where $M_{i,k}^{avg}$ denotes the estimated average number of Physical Resource Block (PRB)s allocated per UE in cell k, approximated as total number of available PRBs, divided by the number of UEs served by the cell. In order to ensure minimal power limitation at the cell edge users, an estimate of the 95%-ile of user path loss in the cell ($PL_{95\%,k}$) has been considered in eq. (3.2).

3.3.2 Power control for Dual Connectivity with Bearer Split

The power allocation defined by eq. (3.1) and (3.2) is adequate for single connections. If the same power allocation is used for UEs that are configured with DC with bearer split, the total power allocated at each CC, can exceed the maximum power limitation at the UE. As discussed in Section 3.2, if that happens, power scaling and reduction of the MCS initially set will be required, which lead to reduced spectral efficiency.

Due to the non-ideal backhaul, the RLC and MAC layers at each eNB are independent. The UEs are configured with independent power control and link adaptation settings per link, meaning different settings are defined per macro and small cell link. Furthermore, the scheduling decision are uncoordinated, meaning that macro cell scheduler is unaware of the scheduling grants sent by the small cell, and vice versa. The overall problem can be illustrated with the three example cases shown in Figure 1:

Case (1): If the UE is only scheduled for transmission towards the small cell in a given TTI, the optimal transmit power setting and MCS are defined by the small cell. The traditional power control defined by eq. (3.1) can be applied.

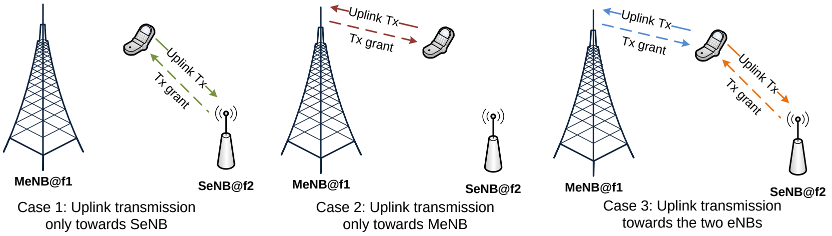


Figure 3.5: Cases for uplink transmissions in Dual Connectivity

Case (2): If the UE is only scheduled for transmission towards the macro cell in a given TTI, the optimal transmit power setting and MCS are defined by the macro cell. Again the power control defined by eq. (3.1) can be applied, while the OLPC parameters may be different from these in Case (1).

Case (3): If both the macro and small cell schedule the UE in the same TTI, the maximum power (P_{max}) of the UE will have to be shared between the two links, meaning that optimal power allocation and MCS selection for each of the two links is different from Case (1) and Case (2) respectively.

However, the current LTE DC operates with one power control parametrization per UE per link where the MCS selection is conducted independently per link. As such the current solution is sub-optimal. Therefore, in this work an enhanced solution is proposed where both the power settings and MCS selection are accurately adjusted depending on whether the UE is scheduled for macro-only, small cell only, or both. This procedure, detailed in Section 3.4, allows for more efficient uplink link adaptation in case of a UE configured with bearer split and is relevant for both future LTE releases and upcoming 5G releases.

3.4 Enhanced Uplink Power Control

In the considered HetNets deployment in this work, the macro cell tier and the small cell tier are deployed at two non-overlapping Component Carriers (CCs) and that there is only one cell configured per sector/LPN. It is assumed that the traditional high-power macro cell operate at a carrier frequency $F1$ and provide wide area coverage. The low-power small cells layer operate at a carrier frequency $F2$ and they enhance the capacity at certain areas. When the users are configured with DC, it is assumed that the control channels are always associated with the macro cell, for mobility purposes. However, since in this work the main concern is throughput improvements with bearer split, herein the notation of *first* and *aggregation* tier is used, denoted as (f) and (a) perspectively. The *first* cell, denoted by $k^{(f)}$, is the cell that the UE would initially associate with if that UE is not configured with DC. The *aggregation*, denoted by $k^{(a)}$, is the second best cell that belongs to the opposite tier than the *first* cell.

As shown in [55], configuration of DC will cause the uplink interference to increase, leading to performance degradation compared to the case where none of the UEs is configured with DC. Therefore, the maximum power needs to be distributed between the two layers in an effective way, considering the interference and load at the *first* and *aggregated* cell.

3.4.1 Maximum Power Distribution

Let k denotes the best candidate cell for the i-th user, and let c_i be a channel with bandwidth W_i assigned to the user. Then the effective interference on link towards the cell for the user at transmission time it in linear scale is given by:

$$\tilde{E}_{i,t} = \frac{n_{i,t} + \sum_{j \in J} g_{j,k} \tilde{P}_j}{g_{i,k}} \tag{3.3}$$

where n_i is the noise power at the receiver, while $g_{x,k}$ defines the channel power gain between the transmitter at the x-th UE and the receiver at the k-th cell. The set J defines the UEs that transmit over the same or a portion of the channel c_i and create interference to the signal transmitted by the i-th UE. The transmit power of a user $j \in J$ is defined by $\tilde{P}_j$. Then the achievable data rate for the i-th user is defined thorough the Shannon capacity formulation:

$$R_{i,t} = W_{i,t} \log_2(1 + \frac{\tilde{P}_{i,t}}{\tilde{E}_{i,t}}) \tag{3.4}$$

For users configured with DC, the capacity of each link can be defined using eq.(3.4) as $R_i^{(f)}$ and $R_i^{(a)}$, respectively for the *first* and *aggregation* layer. Assuming an optimal cell selection at each tier, and fixed bandwidth assigned from each cell to the user, the problem of maximizing the data rate for a UE that is configured on two non-overlapping CC can be formulated as:

$$\begin{aligned} \max_{\tilde{P}_{i,t}^{(f)}, \tilde{P}_{i,t}^{(a)}} \Big\{ W_{i,t}^{(f)} \log_2(1 + \frac{\tilde{P}_{i,t}^{(f)}}{\tilde{E}_{i,t}^{(f)}}) + W_{i,t}^{(a)} \log_2(1 + \frac{\tilde{P}_{i,t}^{(a)}}{\tilde{E}_{i,t}^{(a)}}) \Big\} \\ \text{Subject to:} \\ \tilde{P}_{i,t}^{(f)} + \tilde{P}_{i,t}^{(a)} \le \tilde{P}_{max} \\ \tilde{P}_{i,t}^{(f)}, \tilde{P}_{i,t}^{(a)} \ge 0 \end{aligned} \tag{3.5}$$

where the notation in Table 3.1 defines the symbols used.

Table 3.1: Notation

Symbol	Definition
$\tilde{P}_{max} = 10^{\frac{P_{max}}{10}}$	Maximum UE transmit power (200mW)
$\tilde{P}_{i,t}^{(f)}$	The i-th UE transmit power towards *first* eNB.
$\tilde{P}_{i,t}^{(a)}$	The i-th UE transmit power towards *aggregation* eNB.
$c_{i,t}^{(f)}$	Channel assigned by *first* cell to the i-th UE.
$c_{i,t}^{(a)}$	Channel assigned by *aggregation* cell to the i-th UE.
$W_{i,t}^{(f)}$	Bandwidth (Hz) allocated by *first* cell to the i-th UE.
$W_{i,t}^{(a)}$	Bandwidth (Hz) allocated by *aggregation* cell to the i-th UE.
$\tilde{E}_{i,t}^{(f)}$	Effective interference on channel $c_{i,t}^{(f)}$
$\tilde{E}_{i,t}^{(a)}$	Effective interference on channel $c_{i,t}^{(a)}$

The optimal power for each channel can be then formulated through Lagrange multipliers as shown in [104]:

$$\begin{aligned}
\tilde{P}_{i,t}^{(f)} &= \min\{\tilde{P}_{max}, \frac{W_{i,t}^{(f)}\tilde{P}_{max} - W_{i,t}^{(a)}\tilde{E}_{i,t}^{(f)} + W_{i,t}^{(f)}\tilde{E}_{i,t}^{(a)}}{W_{i,t}^{(f)} + W_{i,t}^{(a)}}\} \\
\tilde{P}_{i,t}^{(a)} &= \min\{\tilde{P}_{max}, \frac{W_{i,t}^{(a)}\tilde{P}_{max} - W_{i,t}^{(f)}\tilde{E}_{i,t}^{(a)} + W_{i,t}^{(a)}\tilde{E}_{i,t}^{(f)}}{W_{i,t}^{(f)} + W_{i,t}^{(a)}}\}
\end{aligned} \tag{3.6}$$

For TTI where the UE has scheduling grant from both cells, eq. (3.6) can be used to define the transmission power per CC that will maximize the user's throughput. However, the UE will always transmit with maximum power, since $\tilde{P}_{i,t}^{(f)} + \tilde{P}_{i,t}^{(a)} = \tilde{P}_{max}$. Thus, having multiple UEs configured with DC, would potentially increase the uplink interference level on the frequency layer that they would not transmit to, if not configured with uplink DC. Additionally, this scheme requires that the eNB will inform the UE of the effective interference on the PRBs allocated to the user, which will increase the signaling overhead over the air interface.

3.4.2 Power Control Framework for Bearer Split

The work in this chapter proposes fundamental modifications in order to overcome the issues discussed above and they are illustrated in Fig. 3.6:

1. At a given period $t = T$, each pair of macro and small cell eNB exchange the following information: $\tilde{W}^t$ and $\tilde{E}^t$, where $\tilde{W}^t$ is defined as the total

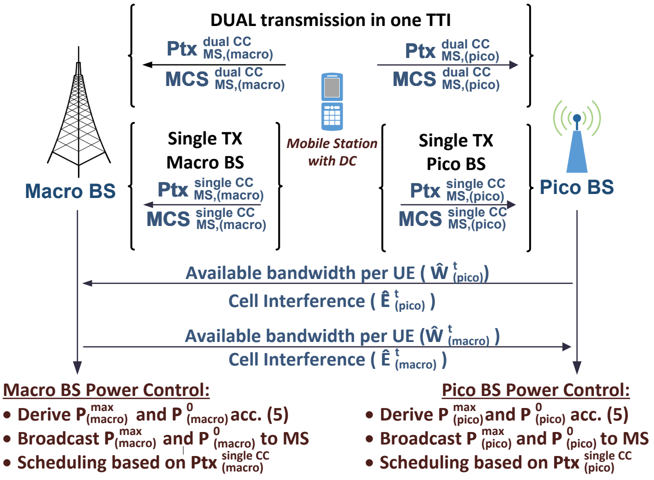


Figure 3.6: Illustration of the proposed power control for Dual Connectivity

available bandwidth divided by the number of UEs associated with the BS. The equation for the effective interference is modified as to capture the interference on the total bandwidth:

$$\tilde{E}^t = \frac{\tilde{N} + \tilde{I}}{\tilde{PG}_{95\%}} \tag{3.7}$$

Where $\tilde{N}$ is the thermal noise and $\tilde{I}$ represents the transmission power received from the UEs not associated with the cell, but transmitting on the corresponding CC. The path gain is related to the $PL_{95\%}$, s.t. $\tilde{PG}_{95\%} = 10^{\frac{-PL_{95\%}}{10}}$.

2. Based on the information exchanged in (1), the first and the aggregation layer will use the eq. (3.6) to define the distribution ot the maximum power across the CCs: $\tilde{P}_t^{(cc)}$, for $c \in \{a, f\}$. Hence, the maximum power that a UE can transmit on a given tier, if it is scheduled for dual transmission in a TTI, is defined as $P_{max}^{(f)} = 10log_{10}\tilde{P}_t^{(f)}$ and $P_{max}^{(a)} = 10log_{10}\tilde{P}_t^{(a)}$ in dB scale.

3. Based on the maximum power distribution in (2), the first and the aggregation layer will use eq. (3.2) to derive the normalized power density at each CC for DC, defined as: $P_{0,dc}^{t,(cc)}$, for $c \in \{a, f\}$. These values are broadcasted to the UE configured with DC, together with the maximum power defined in (2) that the UE can use at each CC in case of double transmission within a TTI. Based on the broadcasted normalized power densities for dual connectivity,

the i-th UE will be able to derive the transmission power $P_{i,dc}^{t,(cc)}, cc \in \{a, f\}$, using eq. (3.1).

4. The derivation of the $P_{max}^{(cc)}$ and $P_{0,dc}^{t,(cc)}$, for $cc \in \{a, f\}$ is performed at each eNB, for a defined pair of a macro and a small cell. This means that the distribution of the maximum UE transmission power will depend to which pair of macro and small cell the UE is associated with. For example a UE that is associated with Macro Cell 1 and Small Cell 1, may not have the same $P_{max}^{(f)}$ and $P_{max}^{(a)}$ as the UE that is associated to Macro Cell 1 and Small Cell 2. Each UE configured with DC will have four values of P_0: $P_{0,dc}^{t,(cc)}$, for $cc \in \{a, f\}$ used by the UE when scheduled in one TTI by both layers, and $P_{0,k}^{t}$, for $k \in \{m, s\}$ when the UE is scheduled by one eNB (macro or small cell) at a given TTI.

The fundamental idea is that each UE maintains two pairs of transmission power and MCS as indicated in Fig. 3.6. The network configures the UE with different uplink transmit power control settings: one for transmission towards the macro only, one for transmission towards the small cell only, and individual configuration for each cell in case of simultaneous transmission towards the macro and small cell in the same TTI. In each TTI, the UE checks if it has received scheduling grant at both configured CCs. If there is only one scheduling grant on a CC (cc), the UE will use power $P_{i,k=(cc)}^{t}$ and MCS corresponding to a given BLER target. If there are two scheduling grants, then the UE will use the dual transmission power towards each CC, defined as $P_{i,dc}^{t,(cc)}$ and corresponding MCS. As such, the scheduling grants that are sent from the eNBs should contain two MCS suggestions: one for single transmission and one for dual transmission. As the macro and small cell does not know on an instantaneous basis if the UE is scheduled from both cells, the uplink UE transmission could include a header to indicate the used MCS. The solution could also work without such a header, and instead rely on blind MCS detection between the MCS for single and dual transmission.

3.4.3 Dual Connectivity Configuration

The UEs configured with uplink DC, have an additional limitation of the maximum UE transmission power due to increased peak-to-average power ratio (PAPR) and inter-modulation effects in case of bandwidth aggregation across non-contiguous carriers [114]. The effect is modelled by power back-off ($P_{backoff}$) such that the maximum power reduction mask equation by 3GPP [115] is used to determine an upper bound of the required power back-off. The value of $P_{backoff}$ (in dB) is determined based on the allocated and the available bandwidth aggregated from the CCs. The exact procedure is further detailed in [116]. As such for a UE to be configured with DC, the following power limitation checks can be defined:

$$\begin{aligned} P_{(a)}^{t,dc} &\leq P_{max} \\ 10log_{10}(10^{P_{i,dc}^{t,(f)}/10} + 10^{P_{i,dc}^{t,(a)}/10}) &\leq (P_{max} - P_{backoff}) \end{aligned} \tag{3.8}$$

This limitation will not configure with DC the UEs that are very far from the aggregation layer, since these UEs could increase the uplink interference on the aggregation layer, especially when being scheduled only by the aggregation cell in a given TTI. Additionally, by having the power limit for dual transmissions on two CCs, the need for power scaling will be avoided.

3.5 Performance Analysis

A system level simulator has been used to evaluate the performance of the proposed power control. The evaluation guidelines for heterogeneous networks provided by 3GPP in [117] have been followed. LTE specifications has been followed as well, and major radio resource management functionalities like packet scheduling, link adaptation, Hybrid ARQ (HARQ), power control etc. have been included. The algorithms used and the main parameters are summarized in Table 3.2.

The simulated network topology consists of seven regular 3-sector hexagonal macro sites, and 4 small cells randomly dropped per macro sector area according to spatial uniform point process. Two CC are configured, each with 10 MHz bandwidth. The macro cell are deployed with CC at 1.8 GHz carrier frequency, while the small cells are deployed with CC at 2.6 GHz carrier frequency. Two separate transceivers are assumed at each UE, in order to be configured with two CC simultaneously. The user distribution over the simulation area assumes hotspot dropping, such that 2/3 of the total UEs have been dropped uniformly within $40m$ radius to the center of the small cells, while the remaining UEs have been uniformly placed within the entire simulation area. The applied traffic model is finite buffer with fixed payload size. The calls arrive according to Poisson distribution with intensity of λ calls per second at each macro sector area, while the call is terminated when the total buffer is successfully transmitted. The product of the call arrival rate and payload size defines the offered load per macro sector area.

3.5.1 Simulation Assumption

As already discussed previously, there are several technical challenges that need to be considered for DC with bearer split. This works threats the problem of power distribution, while for the other challenges the following assumptions have been considered.

Table 3.2: Simulation Parameters

Parameters	**Settings**
Network layout	7 macro sites (21 macrocells), wrap-around 4 small cells randomly placed per macrocell
Channel profile	Spatial Channel Model (SCM) channel model with 3D antenna
ISD / cell radius	Macrocell: 500 m / small cell: 40 m
Pathloss to macro[2]	$40(1-0.004H)\log_{10}(D)-18\log_{10}(H)$ $+21\log_{10}(F)+21\log_{10}(F)+80$
Pathloss to pico	$140.7+21\log_{10}(F/2000)+36.7\log_{10}(D)$
eNB transmit power	Macro eNB: 46 dBm; small cell: 30 dBm
Carrier frequency	10MHz@1.8GHz & 10MHz@2.6GHz
eNode-B receiver	2-Rx Maximal Ratio Combining (MRC)
UE transceiver	2 separate transceivers
UE bandwidth allocation	Adaptive Transmission Bandwidth [118]
Packet scheduling	Throughput based joint proportional fair [116]
Bursty traffic model	Poisson arrival with hotspot UE distribution Fixed payload size of 1 Mbits per UE
Cell association metric	RSRQ
Available MCSs	BPSK (R=1/5,1/3) QPSK (R=1/4,1/3,1/2,2/3,3/4) 16QAM (R=1/2,2/3,3/4,5/6)
Max UE power	200 mW [23 dBm]
Average Power backoff	4.0 dB
HARQ	Synchronous and adaptive with max 4 trans.
BLER target	20%
Link adaptation	Fast adaptive modulation and coding (AMC)
α	0.8
Power spectral density	Independent Load Adaptive Power Control [113]
95%-ile user pathloss	$L_{95\%,M}=123$ dB, $L_{95\%,S}=116$ dB
Average power back-off	f 4.0 dB
Update period T	10 ms

CELL ASSOCIATION

Cell selection is based on the downlink Reference Signal Received Quality (RSRQ) since it captures the load and interference conditions in the considered cell-layer. A UE selects a cell from each layer that has the highest value of the measured RSRQ. The two layers are identified as follows:

$$
\begin{aligned}
k^{(f)} &= arg\max_{k}\left\{\max_{k\in M}\{RSRQ_k\}, \max_{k\in S}\{RSRQ_k + RE_k\}\right\} \\
k^{(a)} &= arg\min_{k}\left\{\max_{k\in M}\{RSRQ_k\}, \max_{k\in S}\{RSRQ_k + RE_k\}\right\}
\end{aligned}
\tag{3.9}
$$

where RE_k is the range extension offset applied to the small cell for load balancing purposes.

MAC Scheduling

DC is intended for deployment where ideal backhaul connectivity is not present. Due to absence of tight scheduling synchronization between the macro and small cell layer, independent radio resource allocation at the MAC is performed at each cell. This means that *first* layer will not be aware if the *aggregation* layer has scheduled the UE within the same TTI and wise versa. In this work, PF MAC scheduling based on adaptive transmission bandwidth as proposed in [118] is implemented at each cell. In order to ensure fairness of the UE across the layers, joint PF scheduling defined in [116] is used. In the joint scheduling, the sum of the average scheduled throughput over all CCs at which the UE has been scheduled is considered when calculating the PF metric. This means that on a certain period of time, the MAC schedulers exchange information on the average UE throughput.

Link Adaptation

Independent MCS adaptation is performed by the MAC entity at each eNB in order to match the quality of the two links. The link adaptation determines the most suitable MCS based on frequency selective channel state information such that a certain BLER target is achieved. An outer loop link adaptation algorithm described in [119] is used in order to compensate measurement errors. In this work, for each DC configured UE, at both eNBs, there are two separate link adaptation, one for single and the other for dual transmission. This is needed in order to determine the most appropriate MCS for both cases when the user is scheduled by a single eNB or by both eNBs at the same TTI. As discussed previously, the link adaptation unit require that the UEs transmit SRS over the whole or a fraction of the scheduling bandwidth such that the power spectral density used for SRS is the same as the power spectral density used for data transmissions. With the proposed mechanisms in this work, the UEs need to send SRSs for both single and dual transmissions. However, due to UE power capabilities, there can be limit on the sounding bandwidth or the power spectral density used for the SRS. However in this work, power limitations for the SRS are not taken into account.

3.5.2 Simulation Results

In order to show the improvements due to the proposed power allocation scheme, we have used as a reference the model detailed in [55] where the UE has power allocation defined by eq. (3.1) independently on whether it is schedule on a single or both CCs. Additionally, we show the performance in case all UEs are configured with DC (Prop. Alg) and when power limitation check according to eq. (3.8) is performed (Power Limit) such that both follow the proposed power allocation mechanism.

The Cumulative Distribution Function (CDF) of the average transmission power is depicted in Fig. 3.7. In case of no DC configuration, the UE transmission power will not vary much as the power control algorithm is load dependent. In case of the proposed DC power allocation, the UEs will use more power as they need to transmit towards the two layers. Almost 85% of the time, both eNBs will send scheduling grants in the same TTI, meaning that the UEs will accumulate more transmission power. Additionally, the average transmission power is higher with the proposed algorithm as more UEs are configured with DC compared to 80 % with the reference model [55] at 10Mbps load. As the load is increased, less users will be configured with DC with (Power Limit) configuration, leading to lowered average transmission power. Due to the power backoff, the transmission power of the cell edge UEs is reduced in case of UEs configured with DC, compared to case without DC.

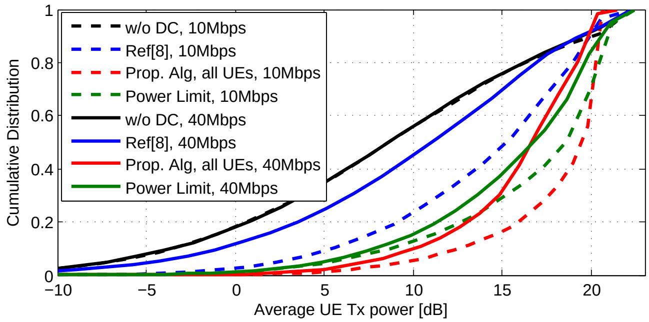


Figure 3.7: Average UE transmission power for different traffic loads

Fig. 3.8 shows the CDF of the aggregated interference over thermal noise in both layers for 10Mbps and 40Mbps offered traffic load. A comparison is made with the reference model and the case where no UEs is configured with DC. The figure shows that by applying the proposed power allocation scheme, the overall interference in the cell is reduced. The UEs will have higher energy and spectral efficiency as more power is distributed towards a cell that is less loaded and has less interference. Thus, on average the UE will leave the system faster and therefore cause less interference. As the probability for dual transmission is very high at 10Mbps load, by configuring all UEs with DC, the maximum transmission power per layer will be reduced, causing less interference compared to the case of (Power Limit) configuration. In (Power Limit) configuration, 80% of the UEs will be configured with DC, leaving 20% transmitting at maximum P_{max} per layer. At high load, the probability for dual transmission is reduced to less than 60% in case of (Prop. Alg.) compared to 75% in case of (Power Limit), while the number of UEs configured with DC is comparable to the case of 10Mbps. Therefore at high load, the (Prop. Alg.) will lead to higher interference over thermal noise than (Power Limit) configuration.

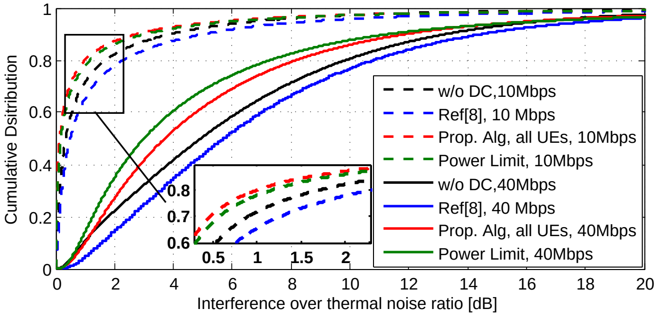


Figure 3.8: Interference over thermal noise for different traffic loads

Fig. 3.9 and Fig. 3.10 show the 50%-ile and the 5%-ile of the user throughput. The performance of the involved mechanisms are analysed against increased traffic load in a macro cell area. For all algorithms it is common that the performances gains are decreasing as the load is increased. At very low load, there are very few UEs in the system. With the proposed algorithm, the UEs will distribute more power towards the cell with less interference and less load, hence higher spectral and energy efficiency can be achieved. The UEs will highly benefit from bandwidth aggregation and leave the system very fast, hence create less interference towards each layer. As the load is increased, the number of UEs per macro cell area will increase, and so will the interference.

By allowing all UEs to have aggregation layer configured, high gains in the 5%-ile user throughput can be observed compared to (Power Limit) and [55]. This is because with (Power Limit) and [55], cell edge users will not be configured, and not benefit from bandwidth aggregation. As the load is increased, the gains are reduced since cell edge users will increase the interference and influence the overall user throughput. The scheduler will further reduce the transmission power due to the power backoff, and they will not be able to increase their transmission power to combat the interference. Therefore, not configuring the cell edge UEs with

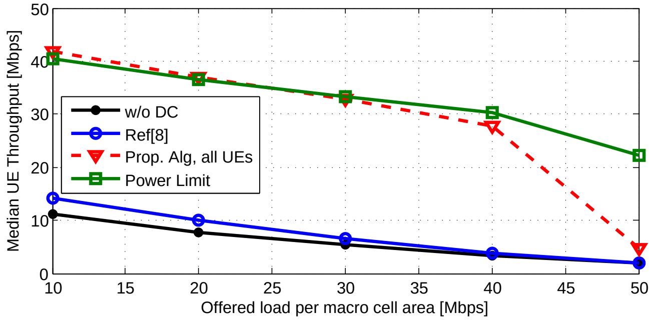


Figure 3.9: 50%-ile user throughput for different traffic loads

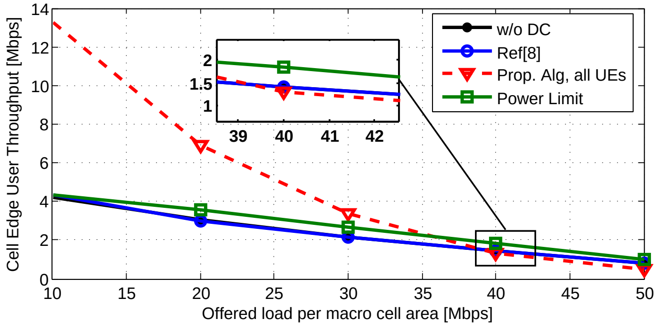


Figure 3.10: 5%-ile user throughput for different traffic loads

DC will be beneficial at high load, as the interference will be reduced as shown in Fig. 3.8. In overall the reason for the decreasing performance in the cell edge throughput at high load comes from the probability of having dual transmissions during TTI. As the load is increased, and there are more users to be scheduled, the probability of being scheduled in the same TTI is reduced. Now the UEs will transmit with higher power most of the time, creating much more interference in the corresponding layer.

3.6 Conclusion

With dual connectivity a new form of multi-cell cooperation is established which becomes very important with respect to the outlook of small cell development - further densification and spectrum extension through exploration of higher spectrum frequencies for the small cells. The DC provides enhancements to sleep mechanisms for reduction of the overall power consumption as well as efficient mobility management where the small cell handover is no longer necessary.

Most importantly, the dual connectivity allows for resource aggregation across carries without the need of ideal backhaul connection. However, several technical challenges need to be addressed first. This chapter provides a comprehensive overview and discussion of the main issues that need to be considered for efficient resource aggregation for dual connected users. The work described treats the problem of user power allocation scheme for dual connectivity with bearer split. It proposes that two separate configurations per component carrier are maintained: one for single transmission and the other in case of dual transmissions within TTI. The design on the power control has been simplified as to ensure minimal coordination among base stations as well as minimal air interface signaling. The performance of the scheme has been assessed with system level simulations and comparison with the previous case study has been evaluated. An important outcome of the study is that with bearer split high gains can be achieved for the median and cell edge user throughput. For cell edge user throughput almost threefold gain is accomplished for low load, while approximately 30% gain is achieved at high load (compared the case when DC is not configured). The results indicate that the distribution of the power is important in order to reduce the possible interference increase in the aggregation layer. At very low load, the decreased number of UE configured with DC can limit the gains from dual connectivity, while at high load it is important to reduce the number of UEs configured with DC, due to the cell edge performances. The analysis also reveals that the probability of a UE being scheduled in one TTI by two eNBs has high influence on the cell interference. This work also indicates that further study of coordination among eNBs scheduling is required in order to reduce the degradation of the coverage performance. This could be in a form of coordination for the scheduling grants towards cell edge users such that the probability for dual transmission is kept low.

It should be underlined that dual connectivity is a very relevant architecture for the future mobile networks, as it eliminates the need for stringent requirements in the backhaul networks. Bearer split as part of dual connectivity, extends carrier aggregation for small cells, and is able to achieve high gains for the user. Therefore, bearer split becomes an important consideration for the future LTE release and upcoming 5G standards. The benefits from DC are multifold, however deeper study through investigations on the implications of dual connectivity is required before any real deployment takes place. In this line, further study on the impact on higher layers and other mechanism is also important. It is worth stating that the dual connectivity influences the UE sleep mechanisms such is Discontinuous Reception (DRX). The decisions on setting the related DRX parameters need to be agreed among the eNBs in order for the UE to benefit from sleep cycles while not compromising the achievable throughput. Additionally, integration of the dual transmission with multi-path TCP should be studied as well.

CHAPTER 4

Optimal Traffic Allocation for Multi-Stream Aggregation

Motivated by the multi-mode capability of the mobile devices and the fact that the heterogeneous wireless access networks overlap in coverage, mobile operators are looking for solutions that will benefit by simultaneous use of the available multiple access interfaces. The integration of different technologies through aggregation of two or more heterogeneous links has therefore received great attention. The benefit of such cooperation can be seen in improved data throughput and enhanced quality of experience. As such, this chapter investigates optimal traffic rate allocation method for multi-stream aggregation over heterogeneous networks.

The heterogeneity and the dynamic nature of radio access networks are considered as important factors that determine the performance improvement by multi-stream aggregation. The optimization method proposed in this work, models the networks by different queuing systems in order to indicate networks with different quality of service provisioning, capacity and delay variations. Furthermore, services with different traffic characteristics in terms of QoS requirements are considered. The simulation results show the advantages of the proposed model with respect to efficient increase in data rate while ensuring accepted delay compared to traditional schemes.

This chapter is organized as follows. Section 4.1 provides the motivation for this work, while Section 4.2 provides an overview of the related work. Section 4.3 introduces the network and traffic models, while Section 4.4 describes the optimization problem. Section 4.5 relaxes the optimization problem described in the previous section, and presents a linear optimization analytical model. Section 4.6 presents and analyses the results under different case studies for multi-stream aggregation. Conclusions and remarks for future work are given in the last section.

Note: This chapter is based on publication [3]. The text has been revised and appropriately modified for the purpose of this chapter. A copyright permission is enclosed.

4.1 Introduction

Ever since the smart devices have entered the market, the demand for mobile data traffic has resulted in stunning growth. Both, the development of smart device capabilities and improved processing power (screen resolution and pixel density), as well as vast range of diverse applications and services, impact the future forecast of the mobile data volumes [120]. On the other hand, the existing wireless technologies and providers differ in their capabilities to offer wider coverage, sufficient capacity and consistent QoS. Furthermore, most of the mobile stations today have an ability to simultaneously connect to different overlapping access networks (for example an LTE and WiFi networks). In order to fully satisfy the needs of mobile users, these three facts inevitably lead toward a need of an integration framework where advantages and disadvantages of each network, device, or application will compensate from each other and the user will get the most/best of the available resources offered by the actual momentum of the heterogeneous environment. The level of cooperation and the integration play a significant role in the network performance achieved by aggregation of the unique strengths of the each mobile technology.

By optimizing the traffic allocation at each node, the inter-RAT aggregation deals with effective utilization of the available bandwidth across multiple radio access technologies. The previous chapter (Chapter 3) elaborated on the problem of bearer split in case of dual connectivity which is a specific case of RAT aggregation. One of the challenges for such aggregation is the flow rate control between the access nodes, which defines how to effectively split the bearer. This chapter sets the focus to an adaptive traffic rate allocation that can be used independently of the underlying technology and can be applied for both, intra-RAT and inter-RAT aggregation. In the proposed model the abstraction of the access nodes is done through a queuing model and both, the heterogeneity (capacity, quality of service (QoS) provisioning) and the dynamic nature (load, heterogeneous traffic requirements, delay variations) of the access networks are considered. The optimal aggregation is challenged by the heterogeneity and the dynamics of access networks. The flow rate control should improve the achievable data rate and the packet delay, while minimizing the out-of-order packet delivery.

4.2 Related work

In a multi-RAT environment, multi-mode capable devices can simultaneously utilize the overlapping areas of various access networks for increased data rates, reduced handover rates and improved service. Achieving improved users' quality of experience by aggregation of multiple access networks has been studied on different levels:

1. *Physical layer:* Carrier Aggregation. Multi-carrier aggregation between 3GPP like technologies is covered by Release 12. Next step is the aggregation of licenced and unlicenced spectrum (LTE Assisted Access, LAA) [121], while in the future, aggregation of any type of communication is expected for converged 5G system.

2. *Link Layer:* Dual Connectivity, LTE+WiFi Link Aggregation. It allows for technology aggregation at the RAN layer. The enhanced performance with this approach has been validated by Qualcomm [122], Alcatel-Lucent [123] and KT [124] at the Mobile World Congress 2015. The focus in this work in on this approach as it relies on software updates at the mobile terminals and features in the RAN. Furthermore, while achieving similar results to LAA, it does not require spectrum sensing or interference control mechanisms as in the previous approach.

3. *TCP/IP Layer:* Network Aggregation, Multi-path TCP. The concept is similar to the previous one with the main difference that the traffic is combined at a network proxy in the Internet, rather than at the cellular RAN [125], [126]. This type of approach require additional implementation/new hardware and the impact on the end-to-end characteristics of the individual paths have high influence on the achieved performance.

The existing studies with respect to bandwidth aggregation provide extensive analysis of various static and dynamic policies for traffic allocations over heterogeneous networks with different technologies and cell sizes ([39]-[40]). In [39] the cells are modeled with processor sharing queue, and only elastic traffic is considered. An optimization model with objective to minimize the mean sojourn time and balance the load among overlapping cells is considered. The study shows that optimal performance can be achieved with knowledge of the average service rates and number of incoming flows to the cells. The works in [41] and [42] show the advantages of bandwidth aggregation in multi-hop networks. The numerical results in [41] indicate the energy efficiency of the proposed flow distribution over multiple interfaces and robustness to link outages. The work in [42] is based on frame aggregation in multi-hop networks, and the proposed optimization model gives improvements of end-to-end delay and packet loss compared to a proactive routing protocol.

In [127] the authors address the problem of joint subnet selection and optimal traffic allocation to each subnet in order to maximize the data rates under QoS constraints and subnet capacity. The available networks have been modeled as single server queuing system (M/M/1) with static subnet delays. In [128], network cooperation of wireless personal and wide area networks has been considered, while the analytical model deals with minimization of transmission delay. The same network cooperation has been considered in [129], where the optimization problem is formulated as minimization of packet delay probability and cost for queuing packets.

Rate allocation for multi-user video streaming has been addressed in [130], [131] and [40]. In [130] distortion-aware concurrent multipath transfer is proposed, while [131] minimizes the expected distortion of the video streams based on the available bit rate and round trip time. In [40] a layered video streaming allocation scheme is given, such that the video flow is divided into basic and enhance layers. The analytical model optimizes the transmission ratio of each layer and achieves improved data rates under acceptable system delays. It also considers modifications to the M/M/1 model for the transmission delay, such that the service rate is adapted based on the size of the queue.

The related work shows the importance and advantages of aggregation by analyzing different proposed models under various study cases. In this work, an additional degree of network heterogeneity is introduced by modeling different degrees of QoS provisioning. Thus, a novel traffic allocation scheme is proposed that additionally considers how the services are provisioned and furthermore considers buffering requirements due to packet reordering. A linear programming model is defined and analysis is conducted by comparison with traditional models for traffic allocation. The numerical results show that the proposed algorithm outperforms the reference schemes in terms of achieved data rate, inter-packet delay and buffering time.

4.3 System Model

The heterogeneous network considered in this work is consisted of two types of access nodes: for example LTE and WiFi as illustrated in Figure 1. A mobile terminal in dual mode can connect to two (or more) access nodes at the same time, such that the traffic from (UL) or to (DL) of the mobile terminal is divided into two separate streams. The delay and the data rate of each flow depend on the system characteristics of each node, the volume of the incoming traffic for access points and the QoS requirements of the service. In this section, the traffic characteristics of the services are described first and then the queuing models used to present the different access networks.

4.3.1 Service class and traffic characteristics

In the analysis we focus on two classes of services. The services can be either real-time that require guaranteed bit rate (GBR) or elastic for which the bit rate is

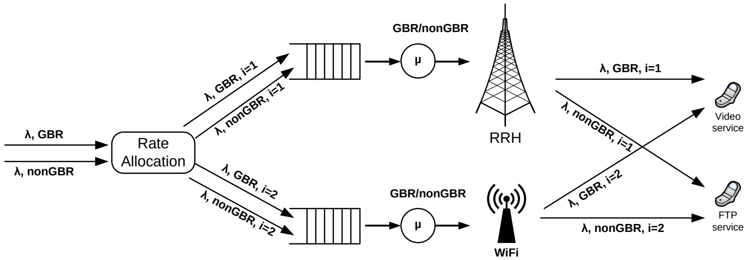


Figure 4.1: System for multi-stream aggregation over heterogeneous networks

not guaranteed (nonGBR). Each has defined data rate traffic of R_{gbr} and R_{nongbr} bps respectively. We consider a Poisson distribution for packet arrival with rate of λ_{gbr} and λ_{nongbr}. The packets have length of L_{gbr} and L_{nongbr} bits, such that $R_s = \lambda_s \cdot L_s$, where $s \in \{gbr, nongbr\}$. As shown on Fig. 4.1, the traffic of each service is split to two sub-streams, each with packet arrival of λ_s^{RAT1} and λ_s^{RAT2}, such that $R_s = \sum_{i=1}^{2}(\lambda_s^i \cdot L_s)$. The set $\{\lambda_s^{RAT1}, \lambda_s^{RAT2}\}$ for each user defines the traffic rate allocation under multi-stream aggregation. The offered traffic at each interface for a service s can be defined as: $A_s^{RAT} = \frac{\lambda_s^{RAT}}{\mu_s^{RAT}}$, where μ_s^{RAT} is the service rate at the corresponding interface. The service rate relates to the available capacity of the RAN defined as ${R_c}^{RAT}$, such that $\mu_s^{RAT} = {R_c}^{RAT}/L_s$.

4.3.2 Heterogeneous access networks

The considered access networks differ in the QoS provisioning and available bandwidth. The first network has higher QoS provisioning such that it prioritize the GBR type of service over the nonGBR. Therefore, it is modeled as priority queuing system (M/G/1) with non pre-emptive discipline [132], such that the waiting time can be distributed to the services according to pre-defined preferences. In case of $|S|$ levels of service prioritization, the mean sojourn (service plus queuing time) time for service of class s, is given by:

$$\overline{W_s} = \frac{V_{1,N}}{(1 - \sum_{i=0}^{s-1} A_i) \cdot (1 - \sum_{i=0}^{s} A_i)} + \frac{1}{\mu_s} \tag{4.1}$$

In case all type of services have exponentially distributed service time with intensity μ_i, the remaining mean service time at a random point of time is defined as:

$$V_{1,N} = \sum_{i=1}^{|S|} \frac{\lambda_i}{(\mu_i)^2} \tag{4.2}$$

The second network does not provide any QoS, but distributes the resources equally among all services in the system. Therefore, such network is modeled as M/M/1 queuing system with processor sharing [132]. As such, this access node serves all services simultaneously by time sharing of the available resources. In such a system, there is no real queue as all services are assigned an equal fraction of the available resources (capacity). In case of multi-service queuing system with processor sharing and $|S|$ different streams, the mean sojourn time for service of type s, is given by:

$$\overline{W_s} = \frac{\frac{1}{\mu_s}}{1 - (\sum_{s=1}^{|S|} A_s)} \tag{4.3}$$

The complete notation of the parameters used in modeling the optimization problem is summarized in Table 4.1.

Table 4.1: Notation of parameters used

Parameter	**Description**
$N = 2$	Number of access points
K	Number of users
$i \in I = [1, N]$	Interface index
$s \in S$	Service index, (1 for GBR and 2 for nonGBR)
α_s	Weight (importance) of service s ($\sum_{p=1}^{|S|} \alpha_s = 1$)
λ_s^i	Packet arrival rate of service s at interface i
$\lambda_{k,s}^i$	Packet arrival rate of user k with s type of service at interface i
$\lambda_{k,s} = \{\lambda_{k,s}^i \mid i \in I\}$	Traffic allocation for user k with service type s
$\lambda_{k,s}$	Packet arrival rate of user k with service type s
L_s^i	Packet data size in bits of service type s, at interface i
$1/\mu_s^i$	Mean service time for service type s, at interface i
Lb_s	Minimum bit rate for service s
Ub_s	Maximum bit rate for service s
R_s	Data rate of service s
R_c^i	Total capacity at interface i
$W_{k,s}^i$	Sojourn time for user k and service s on interface i
$W_{max,s}^i$	Maximum sojourn time for service s on interface i
$maxD_s$	Maximum difference in packet delay controlled by the buffer size

4.4 Optimal Traffic Rate Allocation

In this work the objective is to minimize the maximum delay per application as to reduce the variance in the delay on each interface. The maximum delay is of importance as indicated in [125], while the difference between the delay over the links has influence on packet reordering [133]. For bandwidth aggregation over two heterogeneous networks, the main challenge is estimating the characteristics of the available networks, such as delay and throughput. Since the links have different propagation delays and queuing delays the packet can be delivered out of order. There are two ways to overcome out-of-order packets: 1° by sending the packets in predefined order so that they are delivered in order at the receiver (thus minimizing the out-of-order delivery) or 2° by using larger buffers at the receiver side. Out-of-order packets are less problematic for elastic services, but much critical for delay sensitive applications, such as gaming, video streaming, etc.

Using the notation in Table 4.1 and the eq.(4.1) and eq.(4.3), the total sojourn delay per each interface for every type of service can be summarized as:

$$\overline{W^{i}_{k,s}} = \begin{cases} \dfrac{\sum_{p=1}^{|S|} \frac{(\lambda^{i}_{k,p}+\lambda^{i}_{p})}{(\mu^{i}_{p})^{2}}}{1-\frac{\lambda^{i}_{k,s}+\lambda^{i}_{s}}{\mu^{i}_{s}}} + \dfrac{1}{\mu^{i}_{s}}, \quad (i=1, s=1) \\ \dfrac{\sum_{p=1}^{|S|} \frac{(\lambda^{i}_{k,p}+\lambda^{i}_{p})}{(\mu^{i}_{p})^{2}}}{(1-\sum_{p=1}^{s-1} \frac{\lambda^{i}_{k,p}+\lambda^{i}_{p}}{\mu^{i}_{p}})\cdot(1-\sum_{p=1}^{s} \frac{\lambda^{i}_{k,p}+\lambda^{i}_{p}}{\mu^{i}_{p}})} + \dfrac{1}{\mu^{i}_{s}}, (i=1, s \geq 2) \\ \dfrac{\frac{1}{\mu^{i}_{s}}}{1-\sum_{p=1}^{|S|} \frac{\lambda^{i}_{p}+\lambda^{i}_{k,p}}{\mu^{i}_{p}}}, \quad (i=2, s \in S) \end{cases} \tag{4.4}$$

In order to increase overall QoS, under the given constraint of data rate for each service, the objective of the traffic allocation in case of multi-RAT aggregation can be defined as minimization of the maximum transmission delay. Hence, the optimization problem can be summarized as:

$$\min_{s} \max_{i} \overline{W^i_{k,s}} \tag{4.5}$$

subject to:

$$\sum_{s=1}^{|S|} (\lambda^i_{k,s} + \lambda^i_s) \cdot L^i_s \leq R^i_c, \quad i \in I \tag{4.6}$$

$$0 \leq \lambda^i_{k,s} + \lambda^i_s < \mu^i_s, \quad s \in S, \quad i \in I \tag{4.7}$$

$$R_s = \sum_{i=1}^{N} (\lambda^i_{k,s} \cdot L_s), \quad s \in S \tag{4.8}$$

$$|\overline{W^1_{k,s}} - \overline{W^2_{k,s}}| \leq maxD_s, \quad s \in S \tag{4.9}$$

The set $\lambda_{k,s} = \{\lambda^i_{k,s} | i \in I\}$ defines the optimal traffic allocation for each user of service type $s \in S$. The objective function (4.5) of the optimization problem is defined as minimization of the maximum delay. For two type of services, two unknowns ($W_{k,s}$) are added to the optimization problem in order to balance minimization of the maximum delay for GBR compared to the maximum delay at the nonGBR service. Constraint (4.6) ensures that the sum of the traffic that is accepted by the radio access node does not exceed the available capacity. Constraint (4.7) must be satisfied in order to ensure statistical equilibrium and steady state of the queues. Constraint (4.8) reflects the data rate requirements of each service, while constraint (4.9) reflects the buffer size at the receiver that is responsible for packet reordering. The number of packets being buffered depends on the traffic allocation, while the time that they are buffered depends on the delay difference. Delaying the packets at the sender side by buffering at each interface with time $d_{maximum} - d_{interface}$, reduces out-of-order packet delivery. Therefore, by controlling the delay difference, the buffering requirements are controlled as well. Furthermore, this is also important because the transmission buffers may introduce delays that are not tolerant for some delay sensitive applications [126]. The value of the $maxD_s$ should be chosen reasonable, that is to consider the packet inter-arrival rate.

The optimization problem defined through (4.5)-(4.9) is non-linear and there is no evidence that the objective function is convex. The common methods for solving convex optimizations are not applicable in this case. Alternative, numerical search methods can be used, but the search space due to four unknown is quite large. Therefore, in the following section transformation of the oprimization problem is done in order to lead to a linear and simpler optimization problem.

4.5 Tractable Problem Formulation

Instead of minimizing the maximum delay, the objective of the optimization is reformulated such that the data rates are maximized under constraints of maximum allowed delay for each service and interface. Again, using the notation in Table 4.1, a more tractable optimization problem can be stated as:

$$\sum_{s=1}^{|S|} \alpha_s \cdot \max\left(\frac{\sum_{i=1}^{|N|} \lambda_{k,s}^i \cdot L_s^i}{Ub_s}\right) \tag{4.10}$$

subject to:

$$\overline{W_{k,s}^i} \leq W_{max,s}^i, \quad i \in I, \quad s \in S \tag{4.11}$$

$$Lb_s \leq \sum_{i=1}^{N} (\lambda_{k,s}^i \cdot L_s^i) \leq Ub_s, \quad s \in S \tag{4.12}$$

$$\sum_{s=1}^{|S|} (\lambda_{k,s}^i + \lambda_s^i) \cdot L_s^i \leq R_c^i, \quad i \in I \tag{4.13}$$

$$0 \leq \lambda_{k,s}^i + \lambda_s^i < \mu_s^i, \quad s \in S, \quad i \in I \tag{4.14}$$

With this optimization problem the multiple-objective function (4.10) is formulated to maximize the data rate of each service class. The parameter α_p, where $\sum_{s=1}^{|S|} \alpha_s = 1$, defines the relative importance of the service. The higher α_p is, more weight is assigned to the service. Constraint (4.11) defines the maximum delay that can be experienced on an interface for each service class. With (4.11) the waiting time on each interface for the individual service type is controlled. By setting an appropriate value of $W_{max,s}^i$, the constrain (4.9) can be avoided. Constraint (4.8) reflects the data rate requirements of each service, such that for constant data rate applications, the GBR, Lb_s and Ub_s will be same or very similar, compared to nonGBR traffic. Constraint (4.13) ensures that the available capacity at the access node is not exceeded, while constraint (4.14) must be satisfied in order to ensure statistical equilibrium and steady state of the queues.

In this optimization problem all equations are linear, except the equations under (4.11). By considering eq.(4.4), for all cases except the case of $(i = 1, s = 2)$ eq.(4.11) can be converted to linear by simple algebra. The case of $(i = 1, s = 2)$ can be also converted to linear under the assumption that

$$\left(1 - \frac{\lambda_{k,s=1}^i + \lambda_{s=1}^{i=1}}{\mu_{s=1}^{i=1}}\right) \sim 1 \tag{4.15}$$

By combining the equations for $(i = 1, s = 1)$ and for $(i = 1, s = 2)$, the following can be concluded:

$$(W^{i=1}_{max,s=1} - \frac{1}{\mu^{i=1}_{s=1}}) < (W^{i=1}_{max,s=2} - \frac{1}{\mu^{i=1}_{s=2}}) \tag{4.16}$$

Then for the case of $(i = 1, s = 2)$ the following equation must hold as well:

$$(1 - \sum_{s=1}^{|S|} \frac{\lambda^i_{k,s} + \lambda^i_s}{\mu^i_s}) \leq 1 \tag{4.17}$$

4.6 Numerical Analysis and Performance Evaluation

This section provides analysis of the results obtained by numerical simulation of the proposed traffic allocation scheme. In order to better evaluate the performance of the proposed scheme, a comparison is provided by considering the commonly used traffic allocation schemes: single interface allocation and load balancing.

With the single allocation (*Single Interface*), the user is assigned with only one access point. It is considered that the GBR traffic is only offered to the first interface, while nonGBR is offered only to the second interface. This way the priority handling at the first access node will not influence the quality of the non-GBR traffic.

With the load balancing traffic allocation (*Load Balancing*) the traffic allocation is done according to the following equation:

$$R^i_s = R_s \frac{R^i_c}{\sum_{i=1}^{N} R^i_c} \tag{4.18}$$

With eq.(4.18) the rate allocation is defined according to the capacity available at each interface. A modification of the load balancing algorithm has been considered as well (*Load Balancing with Limitation*), such that the traffic on an interface is limited in order to keep the total transmission time lower then the maximum one as defined by $W^i_{max,s}$. In the numerical analysis it is assumed a maximum delay of $13ms$ for the GBR and $50ms$ for nonGBR service at each interface.

Throughout the numerical analysis the weighting parameter α_s is set to 0.8 for GBR and 0.2 for nonGBR service. The available capacity at the different interfaces are set to $2Mbps$ for RAT1 and $3Mbps$ for RAT2. The numerical analysis is organized in two case studies described in the following subsections. At each case study, for service type s, the following key indicators are evaluated:

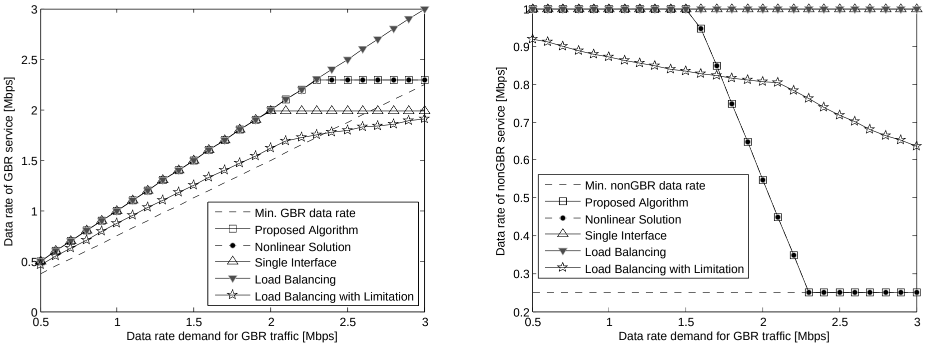
1. achieved data rate $\sum_{i=1}^{|N|} \lambda^i_{k,s} \cdot L^i_s$,
2. maximum delay among $\{W^i_{k,s} | i \in I\}$ and

3. buffer requirements as $abs(W^1_{k,s} - W^2_{k,s})$.

In the analysis a *Nonlinear Solution* that threats the proposed allocation scheme in Section 4.5 is included. This way it is demonstrated that the assumptions considered by eq. 4.17 are justified.

4.6.1 Increase of Data Rate for GBR traffic

In this case study we set the data rate of nonGBR with fixed minimum and maximum value, such that $Lb_{nonGBR} = 0.25Mbps$, while $Ub_{nonGBR} = 1Mbps$. The data rate demand of the GBR traffic is increasing such that the maximum value varies from $0.5Mbps$ to $3Mbps$. The minimum required bit rate is related to the maximum rate Ub_{GBR} by: $Lb_{GBR} = 0.75 \cdot Ub_{GBR}$. The performance of the proposed algorithm is given by Fig. 4.2 to Fig. 4.4.



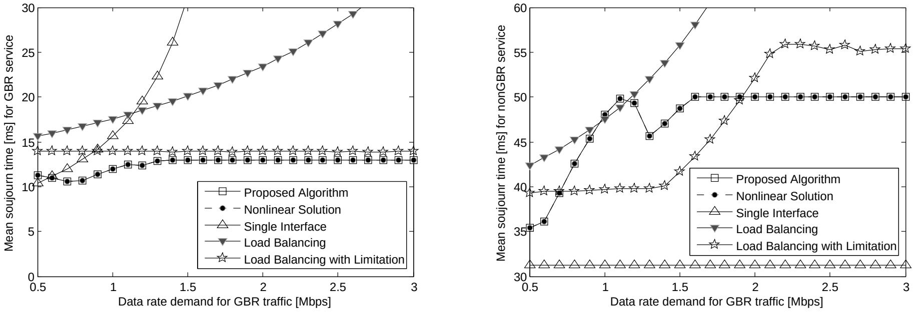
(a) GBR traffic

(b) nonGBR traffic

Figure 4.2: Data rate performance under increased demand for GBR traffic

Fig. 4.2 show the data rate variation for both services. For both services, the considered algorithms keep the data rates above the minimum limitation. The highest reduction of the nonGBR is with the proposed algorithm. For the GBR service, at high load, the *Single Interface* and *Load Balancing with Limitation* reduce the data rate below the minimal value. The *Load Balancing* approach keeps the data rate with the demand, however that comes at a price of increased sojourn time.

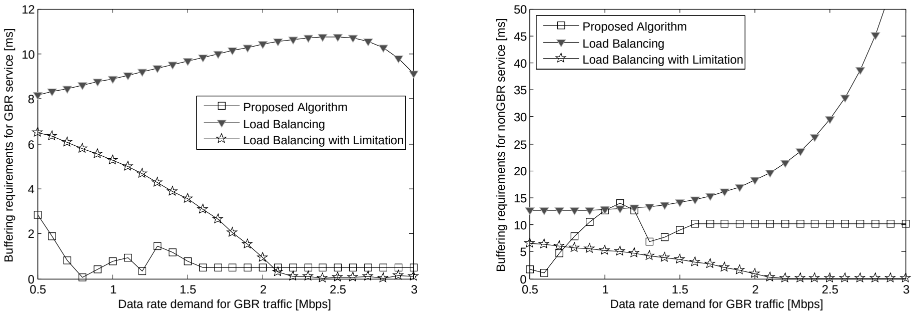
The sojourn time for the two services in illustrated in Fig. 4.3. It can be stated that the *Load Balancing* algorithm performs the worst, as the sojourn time is increasing for both services even for low load. While for low load the



(a) GBR traffic

(b) nonGBR traffic

Figure 4.3: Transmission delay variation under increased demand for GBR traffic



(a) GBR traffic

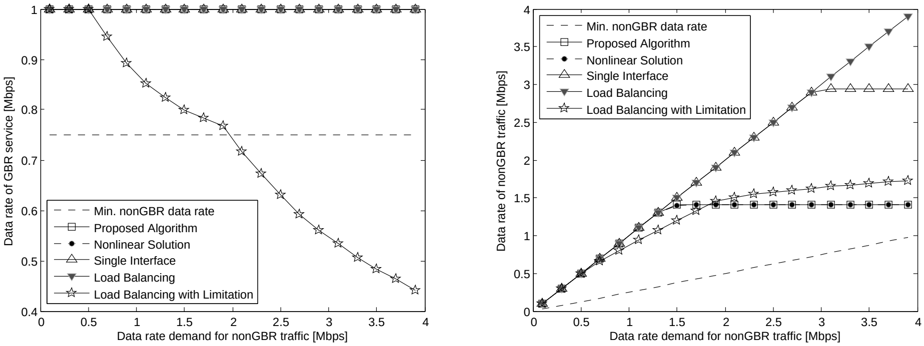
(b) nonGBR traffic

Figure 4.4: Buffering time variation under increased demand for GBR traffic

Single Interface manages lower sojourn time than the maximum, for higher load, the sojourn time is worst and the increase is higher then the *Load Balancing*. This is of course expected as the load towards the RAT1 is increased which leads to long queuing delay for the GBR service. The proposed algorithm and the *Load Balancing with Limitation* approach keep the maximum sojourn time at a reasonable (below the maximum) for both services. However at lower load, the sojourn time for the nonGBR is better with the *Load Balancing with Limitation*, but as the load is increased, the sojourn time is improved with the proposed algorithm.

Until now the proposed algorithm and the *Load Balancing with Limitation* approach have comparable performance. Another metric that is considered in this study is the buffering requirements due to packet reordering, which is presented by Fig. 4.4. For high load, again the two have similar behavior for high load. For low load, the buffering time is higher for the *Load Balancing with Limitation* than the proposed algorithm. The requirements for the data rate and sojourn time are managed better with the proposed algorithm, both for low and higher load. By having higher weight to the GBR service, the proposed algorithm allocates the resources by ensuring the performance of this service first.

4.6.2 Increase of Data Rate for nonGBR traffic



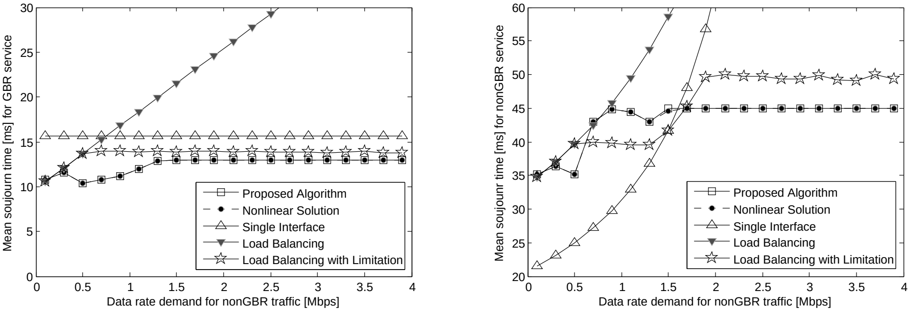
(a) GBR traffic

(b) nonGBR traffic

Figure 4.5: Data rate performance under increased demand for nonGBR traffic

In this case study the data rate of GBR is fixed, such that $Lb_{GBR} = 0.75Mbps$, while $Ub_{GBR} = 1Mbps$. The data rate of the nonGBR traffic is increasing from $0.1Mbps$ to $4Mbps$. The minimum required bit rate of the nonGBR service is related to the maximum rate Ub_{nonGBR} by: $Lb_{nonGBR} = 0.25 \cdot Ub_{nonGBR}$. The performance of the proposed algorithm for this case study is given by Fig. 4.5 to Fig. 4.7.

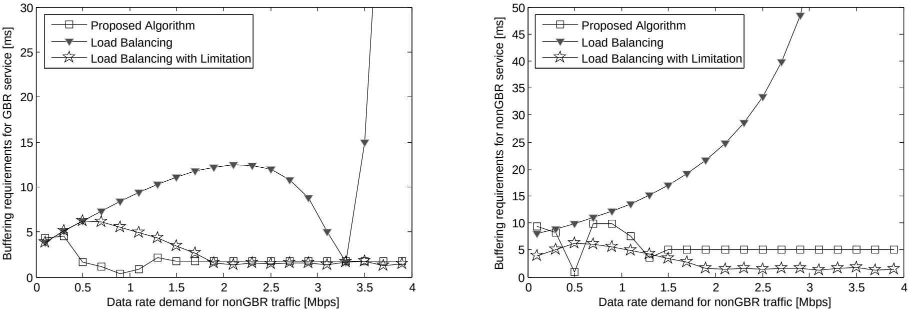
Fig. 4.5 illustrates the data rates achieved for the two services. For the GBR service, all algorithms except *Load Balancing with Limitation* achieve the maximum data demand for GBR service. For the nonGBR service, the *Load Balancing* algorithm achieves the maximum data rate, however all the other manage a data rate that is higher then the minimum one.



(a) GBR traffic (b) nonGBR traffic

Figure 4.6: Transmission delay variation under increased demand for nonGBR traffic

The maximum sojourn time for each service is presented in Fig. 4.6. As the *Load Balancing* does not consider the limitation on the sojourn time, it does not reduce the data rate, and the sojourn time is increasing do to larger queuing. The same is valid for the *Single Interface*, that can not even manage the required sojourn time for the GRB service, although there is low load at the considered interface. The lowest sojourn time for both services is achieved by *Load Balancing with Limitation* and the proposed algorithm at a cost of reduced data rate for the nonGBR.



(a) GBR traffic (b) nonGBR traffic

Figure 4.7: Buffering time variation under increased demand for nonGBR traffic

The buffering requirements are presented in Fig. 4.7. Again the proposed algorithm and the *Load Balancing with Limitation* approach have similar performance. However at low load, the proposed algorithm manages the buffering requirements better for GBR service by giving higher priority to this service. The performance for the nonGBR in case of higher load is achieved by reducing the data rate. However, the proposed algorithm achieves better sojourn time than the *Load Balancing with Limitation* approach.

4.7 Conclusion

Research institutions are currently working on establishment of standards that will allow for carrier aggregation on different levels as discussed in this chapter. As such, stream aggregation among heterogeneous networks is an important research direction and requires investigations of the benefits and challenges under various conditions.

This chapter has discussed the problem of adaptive traffic allocation in cooperative and heterogeneous network. A novel optimization problem has been proposed with the objective of improving the traffic transmission quality under QoS constraints for different services. Compared to the previous research, this work considers three levels of heterogeneity among the available access networks: capacity, delay variations and QoS provisioning. In the model, the heterogeneity of each interface has been represented with an appropriate queuing process. Furthermore, services with different quality requirements have been considered. The numerical results for different case studies confirmed the benefits of multi-stream aggregation over heterogeneous networks.

The performance and superiority of the proposed allocation is demonstrated by assessment with traditional allocation methods known in the literature. The performance is evaluated with respect to the achieved throughput, maximum transmission delay among the available interfaces as well as the buffering requirements needed to reduce the packet reordering at the receiver. It is worth to mention that the buffering requirement has not been explicitly addressed in the previous literature.

The proposed algorithm can be used to provide useful insights on how to allocate resources and how to establish the cooperation among heterogeneous networks. It indicates that the mechanisms for simultaneous transmission over two access nodes with the purpose of improved transmission quality need to consider how the service is handled at each node, and not just the available capacity. Therefore, when the access points are selected for stream aggregation, the priority handling at each of them needs to be considered.

The proposed traffic allocation problem can be easily adopted in case of aggregation across two similar access nodes. Hence, the proposed solution provides flexibility and can be applied for different deployments. The stream aggregation can also be performed over network belonging to different operators in order to improve the offered service quality. As such the numerical results can be used by the mobile operators as to enter into cooperative agreements, and by that the results will effectively influence the operator's business model.

CHAPTER 5

Cloud Radio Access Network

This chapter pays due attention to the optimizations and enhancements required for a RAN that will be able to complement the technological advances on the air interface. The core topic of this chapter is the C-RAN, a leading technology that ensures low operational and capital expenditures associated with the provision of the telecommunication services. C-RAN is based on pooling and centralization of baseband processing, exploiting the advantages of cloud computing concept and technology.

Proper resource placement and dimensioning of the pools is required in order to provide the desired QoS in C-RAN under acceptable cost and energy conditions. The dimensioning of the pool should exploit the so-called tidal effect due to user mobility in cellular networks, in order to maximize the cost savings on baseband processing. The fist part of this chapter analyzes the pooling gains in heterogeneous networks with various cell traffic profiles. Based on the analysis the key parameters and conditions for cell deployment optimization are identified.

However, the multiplexing gains achieved by full centralization of the baseband processing can not always justify the expenditure in the fronthaul network. In case of full centralization, the low transport efficiency in the fronthaul challenges the cost-effectiveness of C-RAN. Redesign of the C-RAN through functional split in the baseband processing chain has been proposed to overcome these challenges. Therefore, the second part of this chapter considers heterogeneous networks with cells that are served with different splits. The study quantifies the trade-offs between centralization and decentralization concerning the cost of the pool (gains), fronthaul network capacity and resource utilization.

This chapter is organized as follows. After the introductory motivation in Section 5.1, the related work is presented in Section 5.2. The analytical model used in this chapter is presented in Section 5.3. Section 5.4 discusses the approach taken in this work to evaluate the multiplexing gains by pooling resources, as well as principles for pool dimensioning. The multiplexing gains due to the tidal effect are evaluated in Section 5.5, while Section 5.6 analyses heterogeneous deployments with respect to fronthaul splits. Concluding remarks are offered in Section 5.7.

Note: This chapter is based on publications [1] and [4]. The text has been revised and appropriately modified for the purpose of this chapter. A copyright permission is enclosed.

5.1 Introduction

The strong advances on the processing abilities at the mobile devices as well as innovative and popular applications, have pushed the focus of mobile networks from voice to data dominated. In the past few years, there has been exponential growth of the mobile data traffic, that is expected to continue even more rapidly in the next years, especially in the busy hours [19]. The forecast for high capacity and coverage demand, together with the mobile subscriber proliferation and increased user expectations, has lead to fast development of cellular networks, where densification and advanced air interface technologies have prevailing importance.

The current cellular networks with the traditional design of base station are highly inefficient. The operator need to dimension the BS according to the peak hours, thus leading to over-provisioning of resources. In fact measurement reports show that the offered traffic varies both geographically and temporarily, leading to only 15-20% of the sites to carry 50% of the total traffic[134]. Secondly, the traditional design of the BS limits the tight coordination among the resource management at each cell, thus prohibits advanced cooperation technologies such as CoMP, CA and (massive) MIMO. Thirdly, most of operators costs are associated with the RAN, that is 60% of CAPEX and 80% of OPEX [135], and they will continue to increase drastically with network densification. These costs together with the flat subscription rates, decreases the operators revenue and challenges the economical profitability of the cellular networks.

Striving towards cost and energy efficient mobile networks, C-RAN designates a leading technology for the RAN architecture [136] that addresses the exponential growth of the mobile traffic. C-RAN inherits the design of a distributed BS, where the RRH is separated and connected via fiber to the baseband processing server, called BBU. The baseband processing is gathered in a centralized pool, which facilitates advanced multi-cell cooperation techniques. Thus the radio resource management is more efficient as the network is flexible and scalable to the temporally and spatial fluctuations of the mobile traffic. The centralization is further enhanced with cloud computing [62], providing elasticity and virtualization with possibility for multitenancy among operators. The computational resources can be pooled and dynamically allocated to a virtual BS, which brings cost-effective hardware and software design [137]. Thus the statistical multiplexing gains that can be achieved by C-RAN, scale with the mobile operator costs: with C-RAN less equipment is required and as such the CAPEX of the network is to be reduced. Less energy will be required to operate the pool of BBUs, thus energy savings can be achieved which directly lowers the OPEX of the network [64].

The multiplexing gains achieved by pooling the baseband resources, have been quantified by several contributions [138, 139, 140, 141, 142, 143, 144]. Their observations of the cost and energy savings by multiplexing are obtained based on simulations. While the simulation based analysis gives better insight on the protocol interaction, the mathematical approach allows to create simpler models that provides the basis for analysis, design and optimization of fronthaul and backhaul deployments. In the work presented by this chapter, traffic engineering approach is used in order to perform a quantitative study of C-RAN, and evaluate the cost and energy savings. The focus is placed on the multiplexing gains coming from traffic depended sources, that is user data and heterogeneous traffic characteristics. The models used are based on multi-dimensional Markov model and direct routing network model, and are able to capture the session level dynamics in the network as well as the relation between baseband processing and radio resource requirements. Furthermore, the derived mathematical model can be used to initially investigate and indicate the conditions for optimal dimensioning of the BBU pools (Section 5.5).

With the initial proposal for C-RAN, where the complete baseband processing is centralized, radio samples are exchanged between the RRH and BBU pool. This architecture yields a fronthaul that requires high bandwidth and low latency interconnect network. Therefore the fronthaul network presents cost burden to the C-RAN architecture, especially when high number of cells are associated with the BBU pool. In order to alleviate this challenge, alternative interfacing between the RRH and BBU pool have been proposed [70]. Furthermore, the concept of Heterogeneous C-RAN has emerged, where the resources from the low power nodes are pooled, while the high power nodes are interconnected with the traditional S1/X2 link towards the BBU pool [145]. The main purpose of such heterogeneous C-RAN is to provide improved signaling and resource management through coordinated cross cell scheduling and hybrid backhauling [146, 147].

The optimal solution on how to split the baseband functionality represents an open challenge that has been addressed in this chapter as well. Based on the direct routing network model, a mathematical model to analyze different split points in terms of multiplexing gains is derived. The results show that in heterogeneous deployments, when cells are served by different splits, different multiplexing gains can be achieved at the fronthaul links and at the BBU pool. In this line, Section 5.6 considers flexible and heterogeneous design on the fronthaul network and evaluates the costs due to processing requirements, fronthaul capacity and overall resource efficiency.

5.2 Related Work

The possible multiplexing gains achieved with C-RAN have been evaluated by the research and industry community. In [138] Werthmann *et al.* show multiplexing gains can be achieved if multiple sectors are aggregated into one single cloud base station and the spatial user distribution has high influence on the compute resource load. A framework for partitioning base stations into groups and their scheduling on a multi-core compute platform is proposed by Bhaumik *et al.* in [139]. They conclude that the variation in the processing load across base stations leads to savings of compute resources. The statistical multiplexing gain as a function of cell layout has been analyzed by Namba *et al.* in [140]. In [141] Madhavan *et al.* quantify the multiplexing gain of consolidating WiMAX base stations under different traffic conditions. They show that the gain increases linearly with network size and it is higher when base stations are experiencing higher traffic intensity. However, in [61] Liu *et al.* show the opposite, that the multiplexing gain reaches significant level for the medium-size pools and the increase in gain for larger pools in negligible.

The multiplexing gains under different cell traffic profiles has been analyzed in [142] based on system level simulator. Based on the simulation results, the paper concludes the set of parameters that maximizes the potential cost savings with C-RAN. Furthermore, a packet based fronthaul architecture is proposed in order to enable flexible assignments between RRH and BBU pools. The benefit of dynamic assignment of baseband processing to RRHs has been further analyzed in [143] and [144], where it has been emphasized that one to one mapping of BBU and RRH is sub-optimal. Their work shows that the configuration in the network must be flexible in order to provide high performance and energy efficiency. Semi-static and dynamic RRH-BBU switching schemes have been proposed and analyzed with respect to efficiency in the BBU pool. The results show that a percentage of BBUs can be reduced, depending on the traffic load and the applied scheme for assignment.

5.3 Direct Routing Network Model

Teletraffic theory has been used for planning and administration of real telecommunications systems even when the first telephone system became commercially available. The models are technology independent, and are based on the relationship between the system capacity and its characteristic, the strategy of operation and the statistical properties of the traffic. The main purpose is to derive the dependency towards the grade of service and assist the decision making related to the optimization of telecommunication systems. An overview of different mathematical models, their characteristics and relevance is provided in [148], while a broad application of different models for modeling and dimensioning of mobile networks is given in [149].

The direct routing network model used in this work provides an engineering tool to study the multiplexing gains in hierarchical cellular networks [150]. The numerical results from the analysis influence the decisions for cellular systems dimensioning. As the network model itself is complex, it is worth to first introduce the model for a single link which the network is composed of.

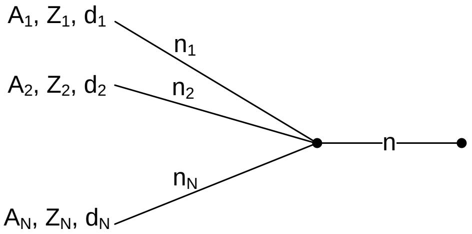


Figure 5.1: Multiplexing in single link

However, before describing the model, it is important to state the considered system, strategy and the traffic characterization. The system is assumed to be homogeneous, meaning that the system is comprised of identical basic units that can be defined as channels, servers, trunks or slots. In case of the considered computational resources, the basic units may be defined as Central Processing Unit (CPU) cycles defined as cycles-per-second (CPS) of CPU time [141] or as compute resource efforts in Giga Operations Per Second (GOPS) [138]. The operational strategy assumed is blocked-calls-cleared, meaning that the calls are rejected if there is not sufficient resources available. The traffic is characterized with call arrival rate (λ), and the service time which is assumed to be exponentially distributed. The offered traffic to the system (A) is defined as a product form of the arrival rate and the mean service time ($\frac{1}{\mu}$), hence $A = \frac{\lambda}{\mu}$. The model is insensitive to service time distribution and only the mean value influence the state probabilities. Each traffic flow may have individual mean service time, however in this work, without loss of generality it is assumed the service times are exponentially distributed with mean value of unit time.

5.3.1 Single Link Model

A single link with capacity of n basic units (BUs) is shared among N statistically independent (uncorrelated) flows of Binomial, Poisson, or Pascal (BPP) traffic, as illustrated in Fig. 5.1. The j-th traffic flow is characterized by it's mean value A_j, and standard deviation std_j. The maximum number of BUs that the j-th flow can occupy is denoted by n_j, while d_j is the number of BUs that the flow requires for the entire duration of the connection. This system can be described by N-dimensional Markovian process with state space defined by the vector $x = (x_1, x_2, ..., x_N)$ where $x_j = i_j \cdot d_j$ and i_j represent the number of connections of a flow j. Then the restrictions that lead to truncation of the state space can be formulated as:

$$0 \leq x_j \leq n_j, \quad \sum_{j=1}^{N} x_j \leq n, \quad where \quad \sum_{j=1}^{N} n_j \geq n \tag{5.1}$$

In the case n is sufficiently large such that there is no global restriction, the system corresponds to N independent one-dimensional loss systems (classical BPP loss system), that are represented by state probabilities $p_j(x_j)$.

The system described above is reversible and has product form. Due to the product form, the algorithm based on convolution [151] can be applied to obtain the individual performance metrics of each stream. By successive convolution of one flow at a time, the state probabilities can be aggregated and one-dimensional vector can be used to describe the system (* denotes the operation of convolution):

$$p(x) = p_1(x_1) * p_2(x_2) * ... * p_N(x_N), \tag{5.2}$$

The probability $p(x)$ of observing the system in state x is equal to the mean proportion of time where exactly $(i_1, i_2, ..., i_N)$ connections of the corresponding traffic flow are established (accepted by the system). The convolution is done such that first two flows j and k are convolved with limitation $\min(n_j + n_k, n)$. Then the third flow is added to the previous convolution and so on. Due to the truncation, normalization at each step needs to be performed in order to get the true state probabilities. To calculate the time, call, and traffic congestion for a flow j, all flows except j need to be convolved into $p_{N/j}$. The derivation of the three types of congestion is given in [151]. Here for the purpose of this chapter, the calculation for the carried traffic (in number of BUs) is only presented:

$$Y_j^n = \sum_{x=0}^{n} \sum_{x_j=0}^{x} x_j \cdot p_{N/j}(x - xj) \cdot p_j(x_j) \tag{5.3}$$

and $C_j^n = (A_j - Y_j^n)/A_j$ represents the traffic congestion. By applying the above method, the performance measures for each flow can be derived.

5.3.2 Network Model

A network model is defined as a group of the above described single links. A network with direct routing [150] is characterized by N routes and K links. The routes (R_j) represent the different traffic flows, while the links (L_k) define the capacity l_k as the maximum number of basic unit that the traffic flows can use on that link. The number of BUs that the j-th traffic flow can use in link k is

Table 5.1: Direct Routing Network Representation

	Routes				
Links	R_1	R_2	...	R_N	*Capacity*
L_1	$d_{1,1}$	$d_{1,2}$	...	$d_{N,1}$	l_1
L_2	$d_{1,2}$	$d_{2,2}$	...	$d_{N,2}$	l_2
...	...	...	...	...	...
L_K	$d_{1,K}$	$d_{2,K}$	...	$d_{N,K}$	l_K

defined as $d_{j,k}$. The restriction on each link can be expressed as:

$$\sum_{j=1}^{N} x_{j,k} = \sum_{j=1}^{N} i_j \cdot d_{j,k} \leq l_k, \quad k = 1, 2, ...K \tag{5.4}$$

Table 5.1 provides an illustration of how a network with direct routing can be defined. All the routes are independent therefore the convolution algorithm can be applied to aggregate the state probabilities of any two routes to one route, until one route remains for which the performance metrics are calculated. Now, during convolution, each link has to be considered one by one, as a restriction to the state space. Because each link can restrict one or more routes, the number of busy channels at each link, or the number of connections at each routes need to be tracked. The algorithm becomes more complex since multi-dimensional vectors need to be convolved, where the number of links defines the dimension. The state number increases to maximum $\prod\limits_{k=1}^{K} (l_k + 1)$, which requires large memory for calculation. However, in this work, due to symmetry in the definition of the equivalent direct routing model to the cellular network studied, the number of states can be significantly reduced.

5.3.3 Network layout mapping to a C-RAN deployment

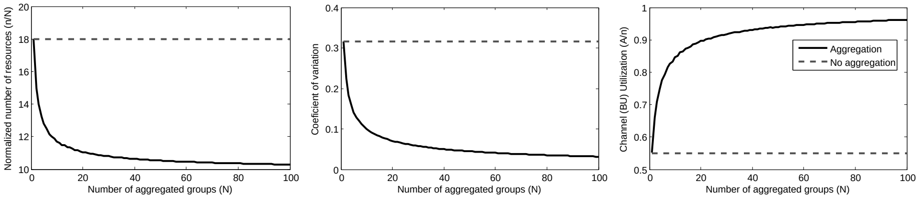
Based on the analytical model presented in Section 5.3, the following notation will be used throughout this chapter to describe a C-RAN network. A BBU pool is associated with N RRHs, where a RRH j can use up to n_j radio resources. The number of baseband processing power (or computational resources) in a BBU pool is given by n, where $n \leq \sum_1^N n_j$. The traffic at RRH j is represented through the mean value of offered traffic A_j and standard deviation std_j. A call j requires d_j radio and computational resources (required for baseband processing) for the entire duration of a connection. In the multi-dimensional Markov model there will be two types of truncations. The truncation due to the limited radio resources is referred to as blocking probability due to radio resources, while the truncation that results from computational resource limitation in the BBU pool, defined through n, is referred to as blocking probability due to computational resources (BBU pool limitation). Hence, for each traffic flow, the call blocking probability depends on the blocking probability due to radio resources and blocking probability due to computational resources.

5.4 Discussion on multiplexing gain and BBU pool dimensioning

This section outlines the approach considered for evaluation of the multiplexing gain and the conditions for optimal dimensioning of the pool. The rationals for the considered performance metrics are discussed as well.

5.4.1 Multiplexing Gain

The fundamental problem of network dimensioning corresponds to the task of determining the network capacity that will provide the required QoS and, at the same time, demand for a reasonable cost associated to the deployment and operation of the network. When multiple traffic sources (can be seen as users, base stations) share the available capacity, multiplexing gains can be achieved, taking advantage over the fact that peak data requests do not occur exactly at the same time. The assumption that the traffic streams are statistically independent is important. Multiplexing gains result from traffic randomness. However, the best multiplexing gains are achieved with combination of bursts of traffic streams (irregular, large peaks to average) and/or traffic groups with different characteristics in terms of daily loads variations. The latter refers to the tidal effect, where the different groups are associated with residential and office sites for example.



(a) Required Capacity

(b) Carried Traffic Characteristics

(c) Resource Utilization

Figure 5.2: Analysis of multiplexing gain with aggregation.

The multiplexing gain comes from the principle of group conservation [149], according to which the highest blocking probability of a given group among certain number of groups is reduced when they are combined in one group. The group conservation principle is here explained through a simple example by comparing the n number of resources (BUs) required to achieve a blocking probability of 1% in case of serving an individual stream (one group) and in case of an aggregation of the N streams (combination of groups). In this example a single group has capacity of 18 BUs, and the offered traffic is characterized by its mean value of $A = 10$ (offered traffic is 10 Erlang) and standard deviation $std = \sqrt{\sigma^2} = \sqrt{10}$ (Poisson arrivals). Fig. 5.2 shows different parameters for the case when N traffic streams are aggregated ($N \in [1, 100]$), and the for case when each of the N stream is alone. In Fig. 5.2(a) the dashed line shows a constant value of the normalized number of BUs (n/N) when the traffic streams are served independently. However, the full line shows the normalized number of BUs required to serve the same traffic load in an aggregated group. As it can be seen the required number of BUs in the group is decreasing as N is increased but only until a certain point after which the increment is very slow (almost constant). The reason for this comes from the manner the mean value and standard deviation for the aggregated stream are derived. Since each stream is independent of the others, the mean value and the standard deviation of the aggregated traffic are calculated as:

$$A_{agg} = \sum_{1}^{N} (A_j), \quad std_{agg} = \sqrt{\sum_{j=1}^{N} {std_j}^2} \tag{5.5}$$

These equations indicate that the mean value of the total traffic is the same in case of individual streams and stream aggregation. The difference is in the standard deviation, which influences the Coefficient of Variation (CV). The CV is a measure for the irregularity of a distribution and is defined as $CV = \frac{std}{A}$. For the given example the change of the CV by aggregation is illustrated in Figure 5.2(b). The CV is reduced as the number of groups is increased, but already after $N = 30$ the reduction is slow and the traffic becomes more regular. Therefore, any additional increase of N will not lead to significant reduction of the number of required BUs, and will not bring significant multiplexing gains. As such, larger groups will not lead to significant increase of the gain compared to medium size groups.

An important metric is the resource (in terms of BUs) utilization, that is defined as A/n. Fig. 5.2(c) shows the resource utilization, and it can be seen that the utilization will not be significantly improved with large group size. Already at $N = 30$ the channel utilization is above 0.9 which is very close to the maximum of one Erlang. Due to high utilization, very large groups are even more sensitive to overload, and therefore large groups are not recommended. For that

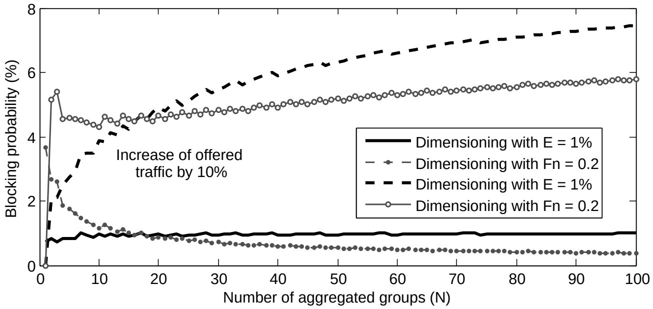


Figure 5.3: Blocking probability dependency on group dimensioning.

reason, the tradeoff between utilization and sensitivity should be considered when dimensioning.

5.4.2 BBU pool dimensioning and multiplexing gains

The discussion above gives insights in the possible gains in pooling the baseband processing resources from each BS into a centralized pool. Moreover, it gives indications on the size of such pools and the impact on the gains (in terms of required BUs and their utilization) as well as on the sensitivity to traffic variations. Two approaches for BBU pool dimensioning can be considered: dimensioning with fixed blocking probability and dimensioning with fixed improvement function. With fixed blocking probability, the dimensioning of the BBU pools is done by restricting the time congestion to a threshold such that the number of calls that need to re-attempt the connection will be low. This principle was used in the example above. As seen, this type of dimensioning can easily lead to a system with high utilization for large group size, but in the same time to a very sensitive system because it does not consider the channel utilization.

The dimensioning principle based on improvement function is also known as Moe's principle for dimensioning. The improvement function is defined as the increase in carried traffic when the number of channels (n) is increased by 1, $F_n(A) = Y^{n+1}(A) - Y^n(A)$, where $Y^n(A) = \sum_{j=1}^{N} Y_j^n(A_j)$. As one BU can carry at most one Erlang, the following is correct $0 \leq F_n(A) \leq 1$. The improvement function can be set to a fixed improvement value F_{target}, such that a desired balance between high utilization and sensitivity is be ensured. A cost requirement can also be included in determining the optimal number of resources in the pool. Then, the improvement value depends on the cost of adding a RRH to a BBU pool - for example the cost of fiber required for the fronthaul network, computational/energy resources etc. The benefits from the increase of the carried traffic should also be included as income, such that $F_{target} = \frac{cost}{income}$. In this work however, the cost and the income are not considered in order to define the value of F_{target}.

Given the previous example, Fig. 5.3 illustrates how the blocking probability is influenced when the offered traffic to the aggregated group is increased by 10%. As it can be seen for small size groups, dimensioning based on blocking probability leads to better sensitivity in case of increase of offered traffic. The increase in the blocking probability is lower if the group is dimensioned based on blocking probability. However, compared to the case when the group is dimensioned based on blocking probability, dimensioning for medium to large group size based on improvement functions leads to increase of n (group capacity) and therefore to improved sensitivity to traffic increase.

Therefore, if medium to large groups are considered in a BBU pool, dimensioning based on improvement function is more appropriate. Based on the improvement function, the required resources $n_{BBUpool}$ can be determined such that $F_n(A)|_{n=n_{BBUpool}} \leq F_{target}$. If there are N individual groups of capacity n_{RRH} then the multiplexing gain achieved by pooling resources can be quantified by:

$$\mathrm{MG}_{c-ran} = \frac{\sum_1^N n_{RRH}}{n_{BBUpool}} \tag{5.6}$$

Equivalent to eq. (5.6) is the percentage of savings that can be achieved by the pooling and can be expressed as:

$$\mathrm{BBUsave}_{c-ran} = \frac{\sum_1^N (n_{RRH}) - n_{BBUpool}}{\sum_1^N n_{RRH}} \tag{5.7}$$

5.5 Multiplexing Gains for Centralized Baseband Processing

In [61], the authors model the dynamics of the BBU pool with a multi-dimensional Markov model. The work shows that the system parameters such as pool size, QoS requirements at the radio part, and the traffic load have impact on the system design. In their analysis, all the cells that are associated in a common pool of BBUs, have the same characteristics: size (BS transmission power), type of traffic and QoS demand. Therefore, the proposed model cannot be directly used if HetNets are analyzed. By extending the single link model with the network model as described in Section 5.3, a direct routing equivalent to a C-RAN can be derived. The model obtained can be used to estimate the performance metrics for a C-RAN architecture that can include cells with different size (as number of radio resources), cells with different traffic profiles (smooth, bursty and random), services that have different QoS requirements (as a minimum number of resources

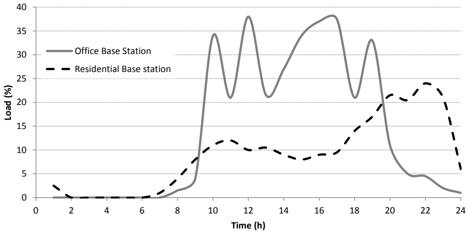


Figure 5.4: Dynamic base station load within daytime [135].

that need to be allocated), as well as multi-layer HetNets deployments. The different models are described in the publication [4]. In this chapter a network based on a mixture of residential and office cells is considered as a baseline. Before presenting the analytical model for such a network, the motivation for this case study is outlined.

5.5.1 Motivation

As indicated in [135] the main multiplexing gain in C-RAN comes from the fact that the cells have diverse traffic load during day hours depending on the area they serve. Fig. 5.4 illustrates how the load at a base station varies during one day. This is the so-called "tidal effect" since the load in the mobile network moves according to the daily routine of the users. During the working hours more users are located in the office areas, hence the BSs associated to those cells are busiest. After working hours, the users move towards the entertainment and residential areas, increasing the traffic demand on the BSs associated to these cells. In case of traditional deployment, the residential cells during working hours and the office cells during evening hours will be underutilized. This figure motivates the research for sharing processing and thus computational/power resources among cells that have variable load. The main objective is to quantify the potential in more efficient utilization of these resources and thus provide a measure of the cost and energy savings of centralizing the baseband processing.

5.5.2 Analytical Model

The direct routing equivalent for a network where the BBU pool aggregates a proportion of cells that serve office and residential area is presented in Table 5.2. The number of office RRHs is O, where each RRH has n_o radio resources. The number of residential cells is $N - O$ where each RRH has n_r radio resources. The office cells are offered bursty traffic model (Pascal distribution) with equal mean and standard deviation across office cells. The traffic at the home cell is modeled using smooth model (Engset distribution) and has equal characteristics among all

Table 5.2: Direct routing equivalent to C-RAN that covers a mixture of office and home cells

<table>
<tr><th rowspan="2">Link</th><th colspan="7">Routes</th><th rowspan="2">Capacity</th></tr>
<tr><th>R_1</th><th>R_2</th><th>...</th><th>R_O</th><th>R_{O+1}</th><th>..</th><th>R_N</th></tr>
<tr><td>L_1</td><td colspan="4" rowspan="3">Identity matrix
of size O</td><td colspan="3" rowspan="3">Zero matrix of
size $[R, O]$</td><td>n_o</td></tr>
<tr><td>...</td><td>...</td></tr>
<tr><td>L_O</td><td>n_o</td></tr>
<tr><td>L_{O+1}</td><td colspan="4" rowspan="3">Zero matrix
of size $[O$ x $R]$</td><td colspan="3" rowspan="3">Identity matrix
of size R</td><td>n_r</td></tr>
<tr><td>...</td><td>...</td></tr>
<tr><td>L_N</td><td>n_r</td></tr>
<tr><td>L_{N+1}</td><td colspan="7">all ones vector of size $[1, N]$</td><td>n</td></tr>
</table>

residential cells. The table consists of an identity matrix of dimension N. Hence, the complexity of the method described in Section 5.3 is highly reduced: the number of the convolutions required to get the performance metrics of one traffic stream is reduced to N. Since there are no dependencies among cells, except the last row, the aggregation of the streams can be done into one-dimensional vectors, and only the global state needs to be remembered. Thus, the number of the states and the required memory is of complexity $O\{n\}$.

5.5.3 Numerical Results and Discussion

The chosen parameters for the analysis follow the example presented in Fig. 5.4 and [142]. The total number of cells is $N = 100$, while the percentage of office cells is varied between 1% and 99% with 1% as step. Each cell has $n_r = n_o = 28$ radio resources, which limits the maximum number of computational resources at the BBU pool at $N \cdot n_r = 2800$. The offered traffic, and standard deviation of the office and residential cells are summarized in Table 5.5.3. The traffic streams will result in very low radio resource blocking probability. The overall blocking probability will be influenced due to the diagonal truncation which results from the limitation of the resources in the BBU pool. Two sub-cases have been considered as two different time snapshots, each for the peak traffic load per cell type. One is from daytime when the traffic of the office cell is higher than the traffic from the home cells. The other is in evening time, when the traffic of the residential cells is higher. By considering these two snapshots, the dynamic of the traffic during one day that influences the dimensioning of the network can be captured.

5.5.3.1 DIMENSIONING THE BBU POOL AND MULTIPLEXING GAINS

As the BBU pool size is high, the Moe's principle for dimensioning the BBU pool in terms of computational resources is used. The improvement value of $F_{target} = 0.2$ is used, and the required amount of resources is defined as n

Table 5.3: Network related parameters

	Daytime		Evening time	
Cell type	Office	Home	Office	Home
Load	30%	10 %	5%	15%
Traffic type	bursty (Pascal dist.)	smooth (Engset dist.)	smooth (Engset dist.)	bursty (Pascal dist.)
A	8.8	2.25	0.8	4.75
std	5.6	0.75	0.4	2.44
CV	0.63	0.33	0.2	0.51

such that $F_{n-1}(A) > F_{target} \geq F_n(A)$. The analysis has been done for the two considered sub-cases: daytime and night. The variation of the required resources in case of different percentage of office and residential cells is given in Fig. 5.5. In daytime, the percentage of required computational resources is increased with the increase of the number of the office cells. The reason for this is that the number of the computation resources scales with the mean value of aggregated traffic. As the mean value of the office cells traffic is larger than the mean value of the home cell traffic, by increasing the number of the office cells, the mean value of the aggregated traffic is increased. During evening time the opposite trend is observed: the increase of the percentage of the office cell reduces the mean value of the aggregated stream, and therefore less computational resources are required.

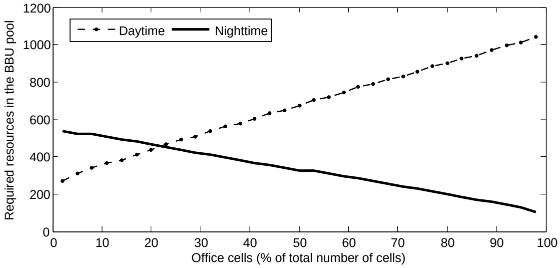


Figure 5.5: Optimal dimensioning of BBU pool.

The two lines cross at 23% of the office cells, meaning that with this ratio of office and residential cells, the same amount of resources will be required during day time and night time. Hence, if the ratio of the office and residential cell is 23 office and 77 residential cells, almost 85% of the maximum resources $(N \cdot n_r)$ in the pool can be saved. However, Fig. 5.5 considers the required resources in the BBU pool assuming that each BS is dimensioned with n_r baseband resources. If we apply dimensioning based on blocking probability with time congestion (E)

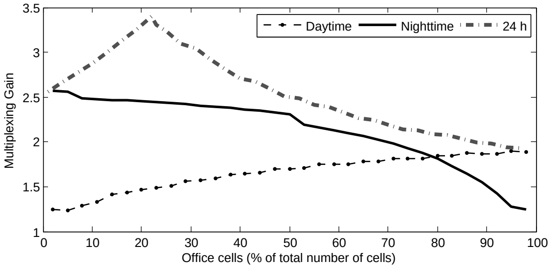


Figure 5.6: Multiplexing Gains in BBU pool.

equal to 1% for each cell, then eq.(5.6) can be used to quantify the multiplexing gains due to aggregation of resources in a centralized pool. Fig. 5.6 shows the multiplexing gains that can be achieved during day and night time if the traffic is aggregated. In order to capture the multiplexing gains during one day, eq.(5.6) was modified to eq.(5.8), such that n^{max}_{RRH} is the maximum resources between the day and night time required at the RRH, while the $n^{max}_{BBUpool}$ defines the maximum resources at the BBU pool required between the day and night time.

$$\mathrm{MG}^{24h}_{c-ran} = \frac{\sum_{1}^{N} n^{max}_{RRH}}{n^{max}_{BBUpool}} \tag{5.8}$$

Fig. 5.6 shows the multiplexing gains achieved with aggregation. As expected they follow the trend of the coefficient of variation, while the maximum multiplexing gains are achieved at 23 % of office cells. The three different approaches considered, show the same results. The analysis based on coefficient of variation is strait forward and requires knowledge of the traffic characteristics. However, in order to define the dimensioning on the BBU pool, the analysis based on multiplexing gains gives better insight. Furthermore, the conclusion of this case study is comparable with the simulation based analysis in [142], which gives confidence in the application of described model.

5.5.4 RRH-BBU pool dynamic mapping

The challenge of the fronthaul design is not just limited to high capacity requirement, but also to the ability to provide flexible and adaptive deployments with respect to RRH-BBU pool assignment. Fiber solutions are capable of supporting high data rates, but are lacking the ability for flexible re-assignment due to the need of manual configurations or very costly optical switches. Adopting any other transport solutions (ex. packet based: wired or wireless) is challenged with strict

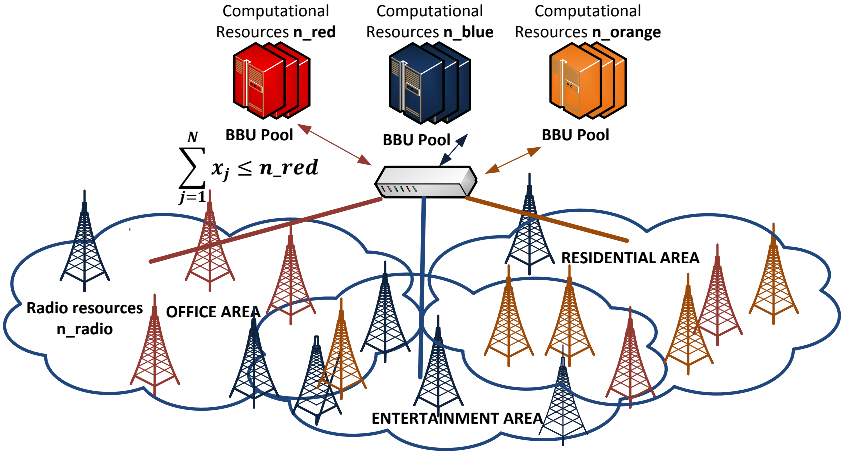
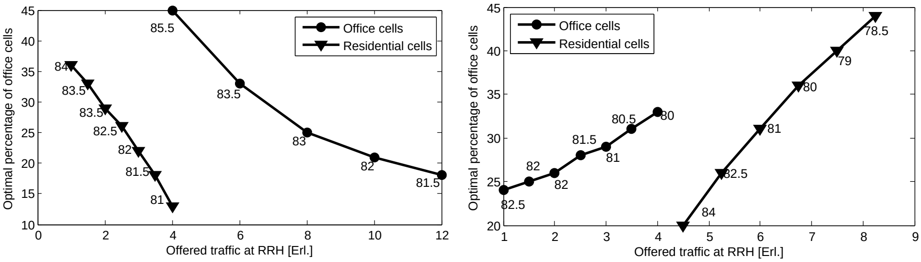


Figure 5.7: Dynamic allocation of RRHs to a BBU pool. The assignment is defined by different colors.

jitter and synchronization requirements but are capable of flexible reconfigurations. The fronthaul network in C-RAN can be enhanced by the Software Defined Networking (SDN), which optimizes network design and operation [152]. The proposed model in this chapter can be implemented at an SDN controller. The SDN controller will be responsible for RRH to BBU pool re-assignment due to traffic distribution change and/or addition of new cells in the network. Thus, the SDN controller can instruct and manage all virtual network components in order to maximize the multiplexing gain and dimension the BBU pools optimally. Figure 5.7 illustrates the dynamic assignment of RRH to BBU pools, where not only the location, but also the traffic load and the type determine the assignment.

The optimal percentage of office cells for different mean value of the traffic offered at a RRH associated with an office and/or a residential cell during day time and during night time is summarized in Figure 5.8. For each optimal deployment these figures present the potential savings by dimensioning the size of the pool using the Moe's principle. The results show that in case of a change of the traffic characteristics, the model can be used for flexible and dynamic re-assignment of RRH to BBU pools. For example, if the mean value of the traffic stream for residential cells during night time is increased, the number of office cells per BBU pool need to be increased. On the other hand, if the mean value of the residential traffic stream during day time is increased, then the number of office cells need to be reduced.

The radio resource blocking probability is low as the load of the cell is not high (Table 5.5.3) and the overall blocking probability is influenced from the blocking probability due to computational resources. This is important as the model complexity is further reduced and only the global state needs to be remembered, which can be described with one dimensional vector of length $n_{BBUpool}$. This simple analysis enables re-configuration in the system in order to optimize the operation by adapting to the dynamic changes. If a certain cell needs to be added to the BBU pool, a convolution needs to be performed in order to aggregate the new cell traffic. If one cell needs to be removed, deconvolution needs to be done.



(a) Variation in Daytime

(b) Variation in Nighttime

Figure 5.8: Optimal deployment for variable load.

5.6 Evaluation of Split Base Station Processing

Seen from mobile operator's perspective, C-RAN provides cost-effective solution, as it enables traffic steering, higher spectrum efficiency and multiplexing gains through pooling of virtualized resources. However, it is important to encounter the implications of C-RAN to the requirements on the fronthaul network. For example, it is worth of notice that point-to-point fiber connections are required in the fronthaul, meaning that the operators that are able to support these connections with marginal cost will be able to fully benefit from cost savings promised by C-RAN. As such this section provides discussion and analysis of different fronthaul architectures that are currently considered to relax the capacity and latency requirements. The analysis is analytical and it provides quantitative indications on the multiplexing gains, which as discussed previously, relates to the cost and energy efficiency of the transport network.

5.6.1 Motivation

In the fully centralized RAN, only the radio frequency unit remains at the RRH, thus IQ data need to be transported between RRHs and BBU pool. The current interface definition, (CPRI or OBSAI), yields for high bandwidth and low latency requirements in the fronthaul infrastructure which challenges the overall benefits of C-RAN. For example in case of CPRI fronthaul interface, for a 20MHz bandwidth and 2x2 MIMO antenna system, the bandwidth required by the fronthaul link between the RRH and BBU pool in one sector is 2457.6 Mbps, where only 150Mbps is offered to a single cell users on the air interface [67]. The maximum allowed latency is in order of few microseconds. Additionally, the CPRI interface assumes point-to-point connections, which makes it inflexible with respect to assignment of RRH to BBU pool.

As the complete baseband processing is placed at the centralized pool, the fronthaul interface is independent of the traffic load and dependent of the number of antennas used, which together highly reduce the transport efficiency. With respect to the technologies on the road towards the 5^{th} generation, such as ultra dense network deployments and massive MIMO, even with compression in place, the capacity needed in the fronthaul will further obstruct the advantages of the C-RAN architecture. The Small Cell Forum [67], iJoin Project [68], and China Mobile [69] have identified that support for flexible functional split in the LTE processing is required in order to relax the requirements on the transport network but still allow cost-effective deployments through centralization and cloudification.

5.6.2 Functional Splits in Baseband Processing

The main goal of introducing the functional split in the baseband processing is to relax the requirements on the fronthaul network such that alternative, cost-effective transport network, such as Ethernet/Internet Protocol, Multiprotocol Label Switching logical networks over microwave, shared fiber and Gigabit passive OPtical networks (GPON) will be able to support the requirements of the fronthaul infrastructure [70]. The upcoming fronthaul interface based on such a split should aim at realizing the benefits of centralization and cloud computation at highest possible level. At the same time the statistical multiplexing on the links should be increased, such that the bandwidth required should be dependent on the load as much as possible. Extensive analysis of most promising fronthaul functional splits is given in: [70], [69], [67]. In the work presented by this section, the focus is placed on the two splits, refereed to as: UE-Cell split and PDCP-RLC split. These two splits are illustrated in Fig. 5.9. The BB-RF split is included in the figure as a fully centralized C-RAN, where the complete baseband functionality is placed in the pool. The main characteristics of the UE-Cell and PDCP-RLC splits are:

UE-Cell split: With this split, the physical layer is separated just before (after) resource (de)mapping in downlink (uplnik) direction. This means that the user data (shared data channels) together with the control channels can be sent

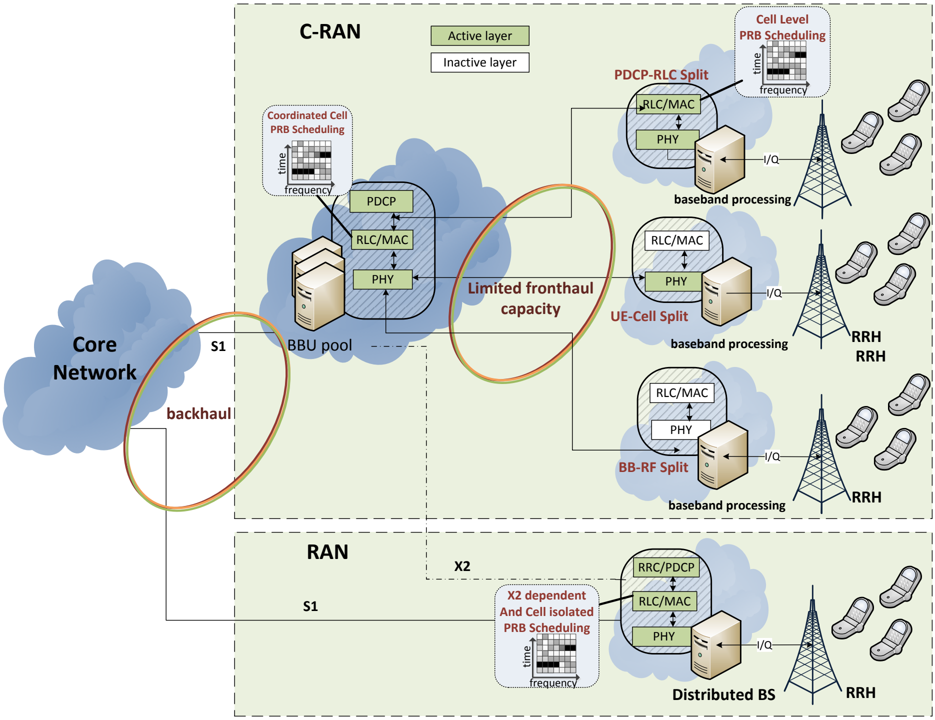


Figure 5.9: RAN Architecture for different functional splits.

towards/from the signal processing unit at the RRH, while all cell related signals that are independent from the traffic (such as cyclic prefix and pilot signals in the downlink) are generated and inserted in the mapping unit at the RRH. As such, the user specific processing remains at the pool, allowing for realization of the traffic steering and CoMP gains for example. In this way, the benefits of the centralization are kept at high level. With this split the cell specific processing remains at RRH, which further alleviates the burden on the bandwidth. However, the antenna dependency is still present. The bandwidth required with this split is reduced to 933Mbps compared to almost 2,5Gbps in case of CPRI (BB-RF split). Yet, the (one way) latency requirements are still tight and are in order of 250 μs [67].

PDCP-RLC split: This split leaves the RLC/MAC scheduling and PHY functionality to reside at the RRH, while the PDCP layer in the LTE-A protocol stack is virtualized at the centralized pool. Since the PDCP functions do not require tight synchronization with the lower layers, this split allows for higher latencies in the order of 30ms [67]. Furthermore, the functionality of the PDCP layer can be generalized as to allow multi-standard support, especially for any upcoming air interface technology [153]. The bandwidth requirements are reduced to 151Mbps [67]. However, this split gives marginal gains with respect to cloud computing and virtualization compared to fully integrated eNB. Furthermore, this splits limits the possibility for joint PHY and MAC processing, which decreases the possible gains in spectral efficiency.

A discussion whether the functional split should be defined statically, per cell, or dynamically per bearer is also important decision that impact the design on the fronthaul network [154]. If the flexibility is introduced per user/bearer within the same cell, the scheduler design will be impacted, especially if PDCP-RLC split is considered. In such a case the decisions on scheduling need to be made partially at the remote end and partially at the pool. In general, the criteria that influence the split decision can be summarized as follows:

- Overall traffic load in the network and quality of service (bandwidth and latency requirements) being provided to the end user together,
- The need for joint processing on the physical or MAC layer,
- Status of the fronthaul links (capacity, utilization and latency),
- Computational/energy resources occupied at the RRH and the BBU pool.

5.6.3 Case study on Fronthaul Network

For heterogeneous cases, when some of the cells are backhauled as traditional RAN, or with C-RAN with flexible processing split, it is important to analyze the possible multiplexing gains and how the total costs of such a network depend on the split decision. In that line, the following work provides a quantitative analysis for such a heterogeneous network. In particular, heterogeneous functional split in the fronthaul interface is considered with dynamic split per traffic flow. With this assumption, a portion of traffic per RRH that is carried with a specific split can be defined, which introduces symmetry in the RRH definition of the analytical model and hence it simplifies the analysis[1].

The case study of heterogeneous case with two splits, UE-Cell and PDCP-RLC, is illustrated in Fig. 5.10. A network of N cells is considered, and each cell has limited radio and baseband processing (computational) resources, denoted as n_r and n_p. The pooled (radio and processing) resources are defined by n_{rBBU} and n_{pBBU}, while n_{rRRH} and n_{pRRH} define the resources that are not centralized and

---

[1] Alternatively, the split can be defined per cell, where a portion of RRHs that are connected via specific split is defined.

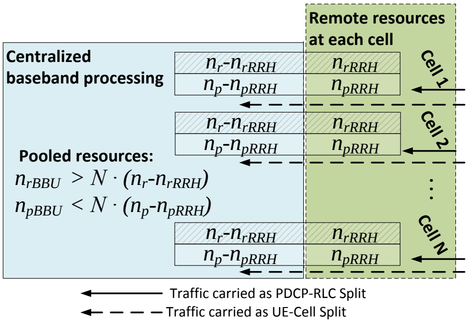


Figure 5.10: Model of C-RAN with N cells, resource definition and offered traffic as PDCP-RLC and UE-Cell split.

Table 5.4: Direct routing equivalent to a functional split in C-RAN

	Route (Stream)							
	$Cell_1$		$Cell_2$		...	$Cell_N$		
Link	R_P	R_U	R_P	R_U		R_P	R_U	Capacity
L_A	d_p^{pdcp}	0	0	0		0	0	n_{pRRH}
L_B	d_r^{pdcp}	0	0	0		0	0	n_{rRRH}
L_A	0	0	d_p^{pdcp}	0		0	0	n_{pRRH}
L_B	0	0	d_r^{pdcp}	0		0	0	n_{rRRH}
L_A	0	0	0	0		d_p^{pdcp}	0	n_{pRRH}
L_B	0	0	0	0		d_r^{pdcp}	0	n_{rRRH}
L_C	0	$d_p^{ue-cell}$	0	$d_p^{ue-cell}$		0	$d_p^{ue-cell}$	n_{pBBU}
L_D	0	$d_r^{ue-cell}$	0	$d_r^{ue-cell}$		0	$d_r^{ue-cell}$	n_{rBBU}

reserved. When the overall traffic is carried as PDCP-RLC split, $n_{pBBU} = 0$, and $n_{pBBU} = 0$ while $n_{rRRH} = n_r$, and $n_{pRRH} = n_p$. Alternatively, when the overall traffic is carried as UE-Cell split, the traffic dependent baseband processing is pooled, hence $n_{pBBU} < \sum_{i=1}^{N} n_p$, where the inequality defines the pooling gain. In order to indicate the gains from joint PHY/MAC processing due to centralized radio resource allocation, the following inequality is defined: $n_{rBBU} > \sum_{i=1}^{N} n_r$. Since we are looking into heterogeneous deployment, the task is to define the portion of n_{rRRH} and n_{pRRH} as well as n_{rBBU} and n_{rBBU}, such that the total traffic is carried with low blocking probability. With this definition of the case study, the equivalent direct routing network model for the fronthaul network with heterogeneous and flexible functional splits can be defined as in Table 5.4.

In Table 5.4 the routes (columns) define the stream(s) j of BPP traffic associated with a RRH. The radio resources are either being scheduled at the MAC layer at the RRHs, or at the BBU pool, depending on the split. The baseband processing power required for these radio resources is reserved in the RRH, or the BBU pool, correspondingly to the assigned radio resources. The limited amount of radio resources and processing power available need to be considered during each convolution step, and in this case they are represented by the link-restrictions. As the performance metrics in case of heterogeneous functional splits are evaluated, the resources (both radio and processing) required at the BBU pool and the RRH for each stream (route) have to be determined. The route R_P defines the mean offered traffic A_j^{pdcp} and the standard deviation std_j^{pdcp} of the traffic that is carried in the fronthaul network through the PDCP-RLC interface at the j-th cell. The route R_U defines the same characteristics ($A_j^{ue-cell}$ and $std_j^{ue-cell}$) for the traffic that is carried as UE-Cell split at the j-th cell. The capacity defines total available resources, both radio and processing, which yields the following restrictions. Link L_A define the restrictions due to the available PRBs at each RRH (n_{rRRH}), while link L_B defines the limited processing possibilities at the RRH, defined through n_{pRRH}. Link L_C, and L_D define the restrictions due to the total available radio and processing resources at the BBU pool, defined as n_{rBBU} and n_{pBBU} respectively.

5.6.4 Case Study Numerical Results and Discussion

For the numerical analysis $N = 100$ cells are considered, each with total offered traffic of $A = 8Erl.$ and arrival rates with Poisson distribution. In order to keep the studies manageable Poisson arrivals with single-slot traffic ($d_p^x = d_r^x = 1, x \in pdcp, ue - cell$) are assumed. At a fixed offered traffic, the percentage of total traffic that is carried as PDCP-RLC split is varied from 0 to 100%, with a step of 10%. This way the network becomes heterogeneous, where the deployment is partially implemented as PDCP-RLC split and partially as UE-Cell split. For each split percentage, we first derive the minimum amount of radio (n_{rRRH}) and processing resources (n_{pRRH}) required at the RRH, such that the carried traffic is equal to the offered traffic (traffic congestion below 1%). Then the resources at the BBU pool (n_{pBBU}) are determined such that the total offered traffic is carried by the network. At each split the following key performance indicators are quantified: multiplexing gains at the BBU pool, resource saving at the fronthaul links and radio resource utilizations. Based on eq. 5.6, the multiplexing gain for the computational resources at the BBU pool is defined as:

$$MG_{BBU-H} = \frac{\sum_1^N n_r}{n_{pBBU} + \sum_1^N n_{pRRH}} \tag{5.9}$$

where n_r represent the maximum processing units available at a RRH in a (traditional) RAN architecture. Only the multiplexing gains that come from the layers below PDCP are considered, as all the gains of centralizing the PDCP layer are present for both functional splits.

On the fronhaul network, the traffic from the UE-Cell split will require larger bandwidth as $933/151Mbps \sim 6$. It is considered that the traffic on the links with the UE-Cell split still remain Poisson, as it should follows the users demand, but it requires six time larger bandwidth on the links than for the case with PDCP-RLC split. Therefore, the offered traffic on the fronthaul link can be dimensioned as $A_{j,FH}^{ue-cell} = A_j^{ue-cell} \cdot d_{j,FH}$, where $d_{j,FH} = 6$. The traffic coming from the PDCP-RLC split will be the same, $A_{j,FH}^{pdcp} = A_j^{pdcp} \cdot d_{j,FH}^{pdcp}$, where $d_{j,FH}^{pdcp} = 1$. Using the same approach as in the BBU pool, the multiplexing gains in the fronthaul network can be defined as:

$$BG_{gain-H} = \frac{n_{FH}^{max}}{n_{FH}^{split\%}} \tag{5.10}$$

where $n_{FH}^{split\%}$ represent the required amount of resources (bandwidth units) for a certain split percentage, and in case all the traffic is multiplexed within the fronthaul network. The maximum number of resources in the fronthaul is required in case all the traffic is carried as UE-Cell split, and that is denoted by n_{FH}^{max}. Alternatively, the multiplexing gains due to shared capacity in the fronhaul network by all RRHs can be defined as (similar to eq.(5.6)):

$$MG_{FH-H} = \frac{\sum_1^N n_{FH,cell}^{split\%}}{n_{FH}^{split\%}} \tag{5.11}$$

where $n_{FH,cell}^{split\%}$ denotes the capacity required by a single cell in order to carry the traffic with very low blocking probability.

Implementing multi-cell cooperation, such as CoMP, leads to increased spectral efficiency, especially at the cell edges, where interference is reduced and even more used as complementary signal in case of joint transmission. In order to indicate the effects of multi-cell cooperation, a percentage of radio resource sharing among cells (*sharing_pct*) is introduced. Then, the radio resource utilization can be defined through the following equation:

$$PRB_{util} = \frac{carried\_traffic}{n_{rCell} + n_{rBBU}} \tag{5.12}$$

where $n_{rBBU} = (n_{rCell} - n_{rRRH}) \cdot sharing\_pct$, and n_{rCell} are the radio resource available at the RRH in case all traffic is carried as PDCP-RLC split. The more possibility there exist for multi-cell cooperation, the PRB are better utilized and hence, less PRBs are needed in order to carry the same traffic volume. Since we consider that the total offered traffic is carried, $carried\_traffic = A = 8Erl.$, while $sharing\_pct = 30\%$ to reflect for example CoMP gains [155].

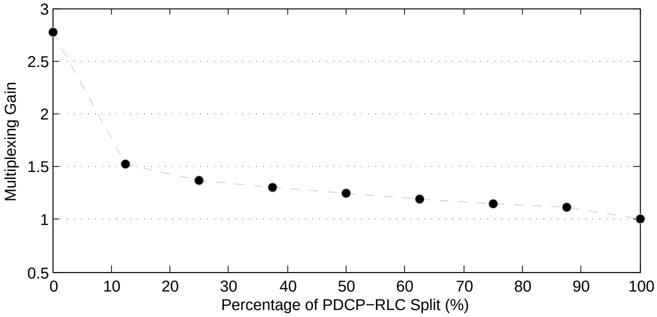


Figure 5.11: Multiplexing gain in BBU pool: PDCP-RLC versus BB-RF split.

The results for each key indicator are given in Fig. 5.11 to Fig. 5.13. As expected, the multiplexing gains in the BBU pool are increasing as more functionality is virtualized at the BBU pool. When the total traffic is carried through the PDCP-RLC split, there is no multiplexing gain from virtualization in the pool as all resource are required at the RRHs and cannot be shared.

On the other hand, the higher bandwidth demand in the fronthaul is when there is only UE-Cell split as seen from Fig. 5.12. As the percentage of PDCP-RLC is increased, gains in the fronthaul link can be achieved as less bandwidth is required. Fig. 5.12 also illustrates the gains achieved in the fronthaul network due to multiplexing (acc to eq. 5.11) when the capacity in the fronthaul network is shared among all cells. It can be noticed that higher multiplexing gains can be achieved with increased PDCP-RLC split percentage, but here it is assumed that the traffic characteristics from both splits follow Poisson distribution. If we consider that the traffic coming from the UE-Cell split is more regular, then higher multiplexing gains can be achieved with the increase of PDCP-RLC split. An other important metric is the physical resource block utilization shown in Fig. 5.13. This metric indicates that the resource efficiency is reduced with lower centralization. This is an important metric to be considered as either additional effort need to be placed in order to provide multi-cell cooperation, or the operator needs to be aware that channel utilization will be increased, in order to carry the same amount of traffic. By increasing the channel utilization, the sensitivity of the overall system to overload is reduced.

In order to evaluate the split percentage based on all three criteria described above, the following cost weight function can be defined:

$$cost_{value} = BBU \cdot c_{bbu} + FH \cdot c_{link} + PRB \cdot c_{prb} \tag{5.13}$$

In equation (5.13), BBU indicates the required amount of processing power

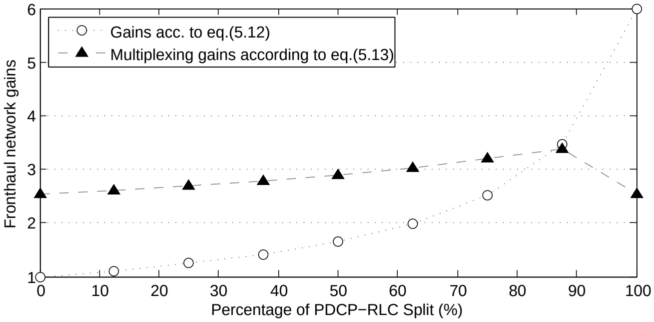


Figure 5.12: Fronthaul network dimensioning.

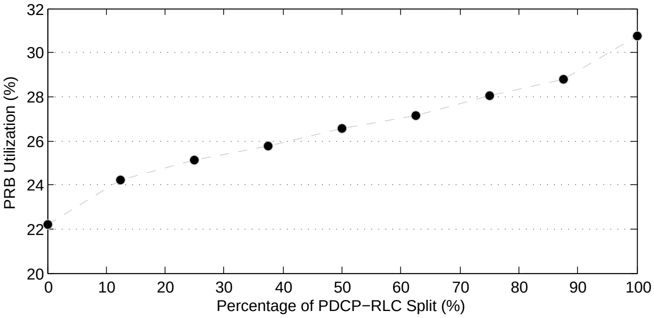


Figure 5.13: PRB Utilization.

(computational resource required for baseband processing), FH indicates the required capacity on the fronthaul links (infrastructure cost), and PRB indicates the percentage of radio resource utilization (spectrum gains). The values of c_{link}, c_{prb} and c_{bbu} represent the associated cost and are normalized such that their sum is equal to unit value. Fig. 5.14 illustrates how the normalized cost of the RAN depends on the weight that is placed on each of the individual gains as considered in eq. 5.13. Real values for the cost are not indicated, as the cost of the equipment depends on the current deployment and hence influences the individual cost for the operator. This figure also shows that heterogeneous deployments should also be considered by the operators depending of the goal and the cost area where the operator needs to save the most. It can help to define the portion of traffic that the operator will route on a particular functional split, depending on the load of the network and the available resources.

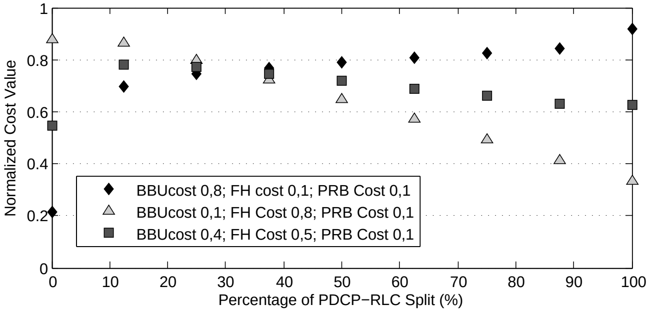


Figure 5.14: Weighted cost evaluation based on gains

5.7 Conclusion

This chapter has focused on C-RAN as an important architecture that offers cost benefits and improved radio resource utilization. As energy and cost savings are related to multiplexing gain, this chapter has evaluated such gains under different networks deployments and traffic profiles. Analytical approach is used, and the used model allows for generalization such that various case studies can be studied. For each case study, traffic dependent requirements have been considered.

In the first part, the benefits of full centralization were analyzed. The work concluded the optimal conditions for dense cell deployments under which the multiplexing gain is maximized. For the studied case, this was defined as the optimal ratio of the two types of cells: serving office and residential areas. The model has been compared with simulation based analysis, which confirms the correctness of the model. The numerical analysis indicated that not only cost, but also sensitivity to traffic variations need to be investigated when dimensioning the pool of baseband units as the pool size is increased. The need for dynamic re-assignment of RRHs to BBU pool has been addressed as well, such that it was elaborated how the method can be adopted in order to optimize the gains in case a new cell needs to be added or in case of the traffic profile change at the existing cells.

Highest multiplexing gains on computational resources can be achieved in case of full baseband centralization. However, such architecture requires the highest fronthaul capacity and may only be viable for operators that can have access to such network at a reasonable cost. Therefore, analysis on the cost and energy savings is important for different functional splits in the baseband processing. Under random traffic definition, traffic dependent multiplexing gains have been quantified for different C-RAN functional splits: separating user and cell specific functions, and PDCP-RLC.

The multiplexing gains are decreased as more baseband functionality is moved towards the cell site. But at the same time, the traffic in the fronthaul network becomes more variable and dependent to the user traffic. As the traffic starts to have variable bit rate, a multiplexing gain on fronthaul links can be achieved, and even more the required capacity will be reduced. Hence, for low traffic load, and even more for bursty traffic, the BBU pool should only have higher layer processing and then the requirements on the fronthaul link can be relaxed. Deployments with heterogeneous functional splits have been evaluated in order to give indications on how much of the traffic can be carried on specific interface. The numerical results suggested that heterogeneous deployments should be considered in order to optimally balance the multiplexing gains, spectral efficiency and cost of the fronthaul infrastructure. As such the observations obtained with presented work, lead to valuable recommendation to the operator in the planning process of the radio access network.

CHAPTER 6

Resource Sharing Strategies for Mobile Operators

C-RAN, SDN and Network Function Virtualisation (NFV) are becoming paramount technologies, each bringing significant benefits for the future mobile networks. Seen from a broader view, these technologies are enablers for dynamic resource sharing both, at the spectrum and the infrastructure level. The opportunities from dynamic sharing have been identified by both, the academia and the industry as an important step towards reduction of the capital and operational expenditures related to the mobile networks. However, dynamic resource sharing requires careful analysis of the benefits for the entire system and, moreover, the effects and implications on each partner that enters a sharing agreement. With respect to this challenge, the work presented in this chapter investigates strategies for active resource sharing of the radio access network between operators, and introduces an analysis the individual benefits depending on the sharing degree. The model used to assess the sharing strategies is based on multidimensional loss systems and networks with direct routing (presented in the previous chapter), while blocking probability is considered as a performance metric.

This chapter starts with a introduction that outlines the motivations and defines the objectives of the study. The up to date work related to sharing of resources in mobile networks is presented in Section 6.2 .The implications of sharing radio resources locally, within a single cell, are analyzed in Section 6.3, while Section 6.4 studies resource sharing strategies for mobile operators on a network level. The conclusions drawn from this study are summarized in Secton 6.5.

Note: This chapter is based on publication [5]. The text has been revised and appropriately modified for the purpose of this chapter. A copyright permission from IEEE is enclosed.

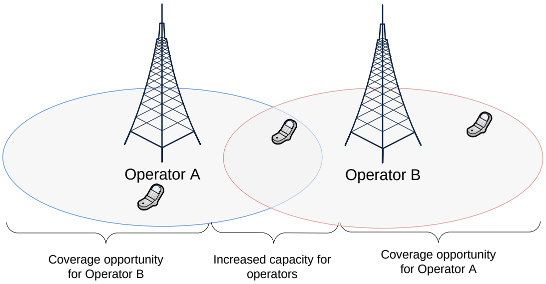


Figure 6.1: Example of resource sharing opportunities.

6.1 Introduction

Contemporary mobile operators are challenged by the users' expectations for a high data service rates, and the demand for always best connection and service - anywhere, at any time. The forecast for high capacity and coverage demand, together with the mobile subscriber proliferation, has resulted in rapid densification of the access network through deployment of cells with different size and overlapping areas served by diverse RATs. These changes and developments in the RAN for the operator translate into increased overall expenditure. Furthermore, the pressure by the state regulations and competition on the voice and data service cost, and flat service rates have negative influence on the operator's revenue.

Among the developed strategies for cost reduction, network sharing has been identified as a promising strategy for substantial improvement of operators costs [156]. Sharing the network capacity is an attractive way of reducing CAPEX and OPEX, while meeting the increasing demand for broadband data services. By introducing sharing in the radio access network, the number of base station deployments (required for coverage enhancement) can be significantly reduced, the resource utilization can be improved, and the overall power consumption can be minimized. Resource sharing among operators is an effective way of extending the coverage and improvement of the network capacity as illustrated in the Fig. 6.1 and enables better load balancing.

Today, network sharing in the radio access part is passive and limited to cell sites. Roaming agreements (national and international) and Mobile Virtual Network Operator (MVNO) concept allow for a way of sharing the OPEX and are suitable in case of well established mobile networks that have an excess in capacity. In case of developing markets, or lack of capacity when dense small cell deployments are required, passive and active infrastructure sharing brings additional reduction of both CAPEX and OPEX. In the close future we may expect a dynamic sharing of radio resources (spectrum) and infrastructure resources (backhaul, access points, routers and switches, computational) as well. Even, on a network level where the spectrum and the infrastructure belonging to different operators are fully integrated and shared through abstraction into network slices [157].

6.1.1 Motivations

Passive infrastructure sharing allows for sharing of non-electronic devices at cell sites such as sites, masts, antennas, power supply, backhaul. It is already an established practice in most European countries [158] and has been proven to reduce the costs and efforts in network deployment and operation. On the other hand, active sharing allows for sharing the active elements in the RAN and it is more difficult to establish and implement, but the operators are looking into the potential of 40% additional reduction of the total costs as indicated in [159]. These facts indicate that active sharing through virtualization of radio resources will be a major part of the roadmap of the fifth generation of cellular networks. The summary of existing sharing agreements between operators in different countries given in [160], indicates that active sharing is already being initiated.

The currently allocated spectrum bands are insufficient to meet the exponential growth of mobile data traffic volumes. Spectrum allocation in the high frequency range has been considered for small cells deployments, as they are appropriate for short distance links due to the propagation characteristics. Additionally, the unlicensed spectrum in the 2.4 GHz and 5 GHz has also received great attention for complementing the license band for LTE operation, and new research item under the term of LAA is undergoing by 3GPP. However, despite these possibilities, spectrum remains a scarce resource and higher spectral efficiency through multi-antenna techniques and flexible spectrum sharing becomes equally important for achieving the needed capacity enhancement.

Dynamic Spectrum Access (DSA) alleviates the problem of spectrum scarcity and underutilization such that the spectrum is shared between primary (license holder) users and secondary (cognitive radio) users. The allocation is done hierarchically, giving priority to the primary over the secondary users. Cognitive radio techniques allow the secondary users to sense the license spectrum band for spectrum holes or white space in the temporal, spatial, and frequency domains [161], while SDR enables adjustments to the transmission parameters in real time [162]. The secondary users need to sense the spectrum at all time and use the spectrum, provided that they do not interfere with the primary users. This means that upon arrival of a primary user, the secondary user needs to abandon the licensed spectrum by adjusting its operation.

Independently of the underlying sharing mechanisms, dynamic and flexible spectrum management among operators is inevitable and remains an important consideration for the future mobile networks. Nevertheless, the spectrum is not the only resource that need to be optimally utilized and as it was emphasized

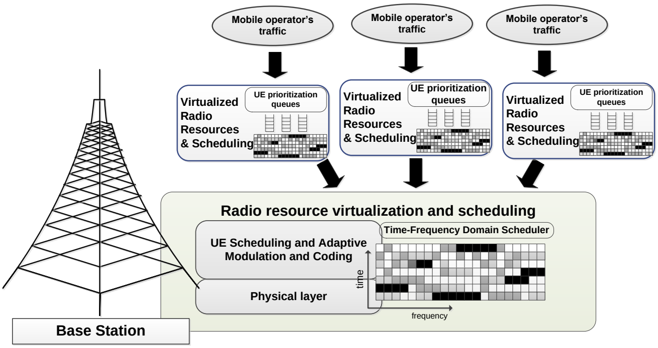


Figure 6.2: Radio resource sharing based on virtualization.

in the previous chapter, high cost are associated with the backhaul networks. Therefore, resource sharing should not only be limited to radio resource, but also consider computational and optical resources in the backhaul network.

This discussion highly motivates the research on resource sharing in the RAN as well as investigations on different sharing strategies for the mobile operators. Before going into the theoretical analysis of such business cooperation opportunities for the operators, it is important to highlight the fundamental technologies that enable resource sharing in the RAN and how they facilitate resource sharing.

6.1.2 Enabling Technologies

Dynamic sharing in the RAN is mainly challenged by the traditional design of the base station. Independently of the deployed RAT, each base station relies on local resource availability that need to be over-provisioned for peak hours. Furthermore, the radio frequency channels are managed locally, within a physical cell. Such distributed spectrum management limits the gains from advanced multi-antenna techniques for higher spectrum efficiency. The cooperation in spectrum management among cells is improved by C-RAN as the baseband processing is pooled and centralized which facilitate the exchange of control information. Having in mind that the current and future mobile networks are highly heterogeneous, both in the deployment scenario and radio access technology, a C-RAN supporting multi-RAT enhances inter-RAT resource management and centralization for joint resource allocation and traffic steering [60]. Such multi-RAT C-RAN allows for improved cooperation among cells with different RAT, such as LTE, HSPA, and even WiFi.

NFV allows for virtualization of the network functions such that they are implemented in software under virtualized computing environment that run on a general purpose computing/storage platforms. RAN virtualization is an important

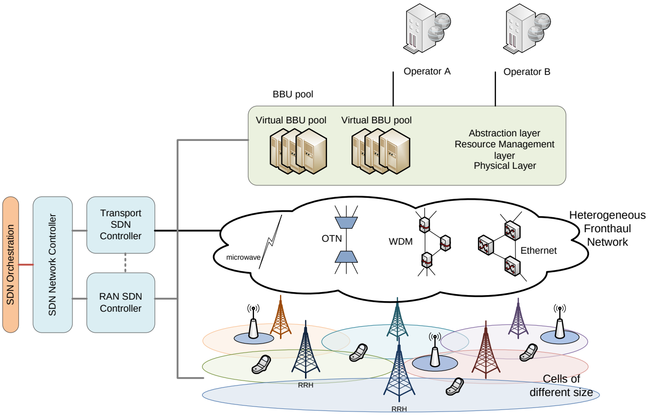


Figure 6.3: Network sharing based on SDN and C-RAN.

use case of NFV, which has the initiative to consolidate and virtualize many network equipment types [163]. The implementation of the baseband processing in software modules and further virtualization of the computational resources that support these software defined base stations, provides a platform for active sharing such that each mobile operator can have a virtual BS operating within one physical BS (Fig. 6.2). Further aspects of virtualization of the RAN for the purpose of sharing spectrum and infrastructure resources among mobile operators is given in [164]. The C-RAN architecture integrates both software defined and virtualization technologies and, as such, C-RAN simplifies the implementation of active sharing among mobile network operators. Furthermore, it fulfills the architectural requirements for full isolation among network operators while insuring cost and efficient operation and the end user connectivity, transparency, and seamless service with agreed level of quality [60]. An example of such architecture, based on C-RAN and NFV, is RAN-as-a-Service (RANaaS) proposed in [65]. The core idea behind RANaaS is to create a RAN as a service, where the operators are isolated from each other, but have the ability to configure and manage the allocated components in the network. The RANaaS architecture allows for dynamic and on-demand provisioning of resources (computational, storage and networking).

Software Defined Networking (SDN) is another important concept that enables dynamic resource management on an infrastructure level. With SDN the control plane is abstracted and fully separated from the forwarding plane. The centralization of the control management allows for improved orchestration of the network control. An architecture with SDN central control for software defined RAN is proposed in [165] as OpenRAN and in [166] as SoftRAN. With OpenRAN *four levels of virtualization* are proposed as follows. On *application level* a network operator/service controls its own virtual space of flow. On *cloud level*, the SDN controller manages virtual BBUs by allocating computing and storage resources. *Spectrum virtualization* is also assumed which further allows for several virtual RRHs to exist on one physical RRH. *Cooperation level virtualization* allows for enhanced communication among virtual entities in the network, with the purpose of creating virtual networks. Within each virtual element, an SDN agent is assumed to resolve the control flow.

Decoupling of the control plane in the RAN through a two-tier model is proposed by Software Defined centralized control plane for Radio Access Network (SoftRAN) [166]. The core idea is abstraction of the base stations in a central controller ("virtual big base station") that is responsible for less frequent decisions, but anyway requires global knowledge of the RAN. Part of the control still remains at the nodes, that need to reach very frequent decisions. A testbed that evaluates the benefits from a policy-centric management framework based on SDN architecture for sharing the physical infrastructure between operators is demonstrated in [167]. Based on [168], Fig. 6.3 illustrates the integration of SDN and C-RAN for integrated radio and infrastructure sharing in the RAN.

6.1.3 Business Model considerations for RAN Sharing

The concept of active sharing brings disturbance of the traditional business model: instead of a single mobile network operator and owner, the network becomes open and mobile network operators contribute to an overall shared network ([65], [169], [170]). Fig. 6.4 illustrates an architecture and the interaction among key entities required to support physical RAN sharing. The owners of the network contribute to an overall physical RAN, that is managed by Mobile RAN Provider (MRP), an independent entity or a consortium constructed by the mobile operators. The MRP is responsible for dynamic allocation of resources to the operators based on agreements and traffic conditions, while the independent support providers ensure monitoring tools for verification of the resources utilization. Each operator has a virtual network assigned, such that it is able to manage and configure the network according to its own policies which ensures operator's independence based on the agreement. The MRP may also be responsible for determining the level of sharing or alternatively, each operator can make a decision of its own, based on the resource utilization level, and give feedback to the MRP.

In case of RAN sharing as depicted in Fig. 6.4, the MRP is responsible for decisions on the network such as upgrades, which results in consolidated physical network and implementation. An important aspect is the commercial independence among operators, which is still present as the operators continue to compete based on offered services and applications.

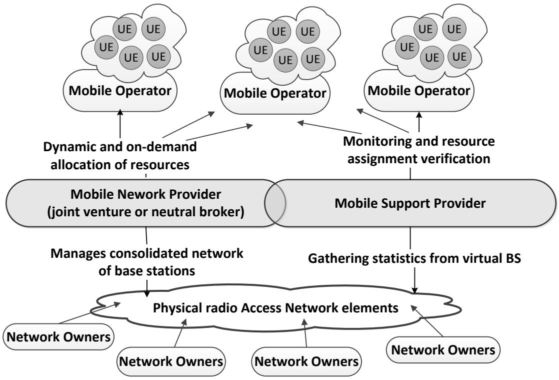


Figure 6.4: Mobile operator architecture by virtualization

6.1.4 Study Objectives and Assumptions

The main objective of the study is to investigate dynamic sharing of resources in the radio access network, where operators decide to share a certain degree of their resources with other operators. This chapter offers a quantitative study on the overall network performance and characterization of the individual benefits with the relation of the sharing degree. In the assessment it is assumed full virtualization of the resources in a RAN. The allocation units in the spectrum can be defined as channels in WiFi or PRB in LTE, or even portions of white space. Beside the time/frequency allocation, space, code and power can also be defined. The analysis in this chapter generalizes the resources in the RAN, and they can be defined as computational resources and/or spectral resources.

Global knowledge of the resource arrangement is assumed, such that their allocation is seamless and can be rearranged among the requesting entities. For example if all the radio resources within one cell are occupied, rearrangement of the calls is possible in order to release the required resources. Such global intelligence is an important assumption from the theoretical point, and allows to model the network as a network with direct routing which was introduced in the previous chapter.

It is assumed that each operator owns certain percentage of the total resources and enters a sharing agreement. The agreement can define the portion of resources that are shared among all operators and can be accessed on a first come first served basis. The performance metrics considered in order to evaluate the benefits from sharing are the individual blocking probability and the carried traffic. The results from studies indicate the optimal degree of sharing such that the agreed performance for each operator is not compromised. The model can be applied for multiple operators, while in the numerical analysis we focus on two operators where we evaluate the sharing degree under different traffic loads, traffic variation and network dimensioning. We consider the impact of sharing strategies in case of a resource sharing within one cell and sharing the resources within a geographical area covered by a heterogeneous network consisting of cells with different sizes.

6.2 Related Work

Resource sharing among self-centered and competitive operators has been proven to be beneficial for all involved entities ([8]-[17]). The degree of benefit is dependent on the type of access to the shared resources that can be based on the principles such as first come first served or some other predefined priority models. The sharing can be regulated according to a utility function, game theory, auctions, etc., where capacity, efficiency, energy consumption, price of spectrum, revenue, resource utilization, and load influence the decision. These principles define the agreement for sharing resources among the involved entities and depend on many parameters such as cost, load, level of quality, etc. The benefits of sharing resources among operators via the Cloud RAN has been analyzed in [171]. The paper presents a study of the resource sharing influence on the network performance, which could be useful when making agreement between operators as well as strategic decisions in network dimensioning.

In [172], the authors investigate a type of sharing where the users connect to the closest BS (form of national roaming), rather than to the home operator it belongs to. The spectrum is negotiated among the different operators and idealistic centralized scheduler is assumed. The results show that this type of sharing leads to better performance in terms of total average frame delay. Game theory approach for spectrum sharing has been considered in [173], [174] and [175]. In the first two paper, a utility function is defined based on the required service rate, spectrum price, and blocking probability according to which the users decide on how to share (trade) a common pool of spectral resources. In [175] a detailed microeconomic analysis based on game theory is given such that the goal of the participants (operators) is to minimize the energy consumption by switching off BSs and redirecting the traffic to a common BS with shared capacity. The trade off between reaching a global optimum and performance optimization for individual operator is considered in [176]. The work in [176] indicates that by deviating from the sharing guarantee, the overall spectral efficiency is increased. Furthermore, it shows the limitation of the spectral efficiency, meaning that there is a pivotal point of the deviation, after which no improvement is observed.

In [177] several types of sharing among competitive operators are compared: capacity sharing, spectrum sharing, virtual spectrum sharing, and virtual PRB sharing. The capacity sharing outperforms the other sharing methods and has been identified as most appropriate for the traditional network design. Yet, capacity sharing is a form of national roaming and requires load based traffic steering to be implemented (involving UE handover). The paper also shows the importance of virtualization of the radio resources, as the virtual PRB sharing performs very close to capacity sharing. As implemented through two tier scheduler (one responsible for prioritization of flows within operator, and the second one for scheduling among UEs), the multiuser diversity is exploited and the spectral efficiency is maximized.

The need for two tier scheduler is also emphasized in [178], [179] and [180]. These papers show the advances in sharing resources through slicing of wireless resources. The first two papers consider a single base station sharing analysis, while the last one concludes that the sharing agreement should be on a larger geographical area as the traffic dynamics varies in time and area. The sharing decision is based on a minimum reserved resources for each operator, while the remaining resources are shared on a first come first served basis. The sum of all resources that are assigned to the involved entities is larger than the total amount of resources in order to exploit statistical multiplexing.

6.3 Resource sharing at a single cell

In this case study the benefits of resource sharing are investigated. The decision on the sharing degree is made local, within a single cell area that is served by two virtual base stations. The main purpose of the case study is to see how the variations of the traffic belonging to one operator influence the performance of the other, depending on the sharing degree. This study can be used for example to grant the MRP to decide on the sharing degree dynamically, by following the traffic load of the operators.

6.3.1 Analytical model

The model is based on Fig. 6.2, such that it is assumed that n RAN resources (RR) are available within a cell area. These resources are shared among N individual mobile operators. The subscribers of each operator generate a flow of Binomial, Poisson or Pascal (BPP) traffic, characterized with mean value A_i, peakedness (ratio of variance and mean value) Z_i. Each connection requires d_i resources for the duration of the connection. The BS operates as a full accessible lost-calls cleared system so that a call from operator i is blocked when less than d_i resources are available. Each operator owns n_i resources, such that $\sum_{i=1}^{N} n_i = n$. The operators are encouraged to enter a sharing agreement in order to increase the individual and overall performance. The operators decide to share a certain degree $s \in [0, 1]$ of the resources. Then operator i has $((1 - s) \cdot n_i)$ resources reserved, but can use at most $n_i^s = (n_i + s \cdot (n - n_i))$ resources. This problem can be described by a N dimensional Markovian process with state space $(x_1, x_2, ..., x_N)$ where $x_i = c_i \cdot d_i$ and c_i represent the number of connections for operator i. The

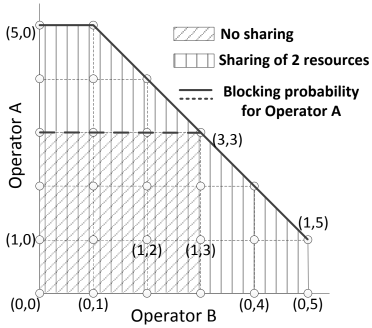


Figure 6.5: State transition diagram for sharing radio resources at single BS between two operators

restrictions to the state space can be defined as:

$$0 \leq x_i \leq (n_i^s), \quad \sum_{i=1}^{N} x_i \leq n, \quad \sum_{i=1}^{N} n_i^s \geq n \tag{6.1}$$

Fig. 6.5 presents the state transition diagram for resource sharing of a single BS with $n = 6$ RR between 2 operators. When there is no sharing, the resources are divided equally to each operator. In this case the state probabilities $p_i(x_i)$ for each operator are independent of the others, and can be calculated by a classical one-dimensional BPP loss system. In case of sharing, where $s = 2/3$, as in the figure, each operator can use at most 5 resources. In this case a diagonal truncation is present, and the blocking probability of one operator depends on the number of connections for both operators. The full line in the figure represents the states that define the time congestion for Operator A in case of sharing resources.

As the system is reversible and has product form, the convolution algorithm [181] can be applied to derive the individual performance metrics of each operator. The global state probabilities, $p(x)$, where $x = x_1 + x_2 + ... + x_N$, can be obtained by successive convolution of one flow at a time, where normalization need to be performed due to the diagonal restriction. Then the system can be described with one dimensional vector (* denotes the convolution operation):

$$p(x) = p_1(x_1) * p_2(x_2) * ... * p_N(x_N) \tag{6.2}$$

By convolving one operator's flow at a time, all flows except i can be aggregated into global state probability vector $p_{N/i}$. Then the performance metrics, such as call, time, and traffic congestion (carried traffic), for operator i can be determined. The full line in Fig. 6.5 indicates the states that determine the time congestion for Operator A, which can be calculated through the following equation:

$$E_i = \frac{1}{Q} \cdot \sum_{x \in S_i}^{max} p_{N/i}(x - x_i) \cdot p_i(x_i) \tag{6.3}$$

where Q is the normalization constant. The set S_i is a sum over all states such that a connection for operator i is blocked and can be defined as:

$$S_i = (x_i, x) | x_i \leq x \leq n \wedge (x_i > n_i^s - d_i) \vee (x > n - d_i) \tag{6.4}$$

The system described above is insensitive to the service time distribution, and each operator can have individual arrival rate, service time and bandwidth.

6.3.2 Numerical Analysis

This section provides analysis on the sharing degree among independent operators. Without loss of generality, the case study is restricted to two mobile operators that have flows of Poisson traffic ($Z_i = 1$), and require $d_i = 1$ resources for each connection. As the stationary Poisson arrival process does not depend on the system state, the probability that an arbitrary call is blocked is equal to the proportion of time when all resources are occupied. Due to this property, call, time and traffic congestions are equal and the time congestion as defined in eq. (6.3) is used as performance metric.

Two operators A, and B, own $n_A = 20$ and $n_B = 30$ RRs respectively, while the total resources are $n = 50$. The offered traffic is dimensioned using the improvement function principle. The improvement function is defined as difference in the carried traffic when the number of resources is increased by one. By choosing the same improvement value for all operators the ratio $\frac{\Delta A}{\Delta n}$ is the same for all operators. This principle, compared to the principle of fixed blocking probability, will allocate more resources to higher traffic load, but less resources to low traffic load. The principle reduces the overload sensitivity and it is more robust to traffic increment. In this study a fixed improvement value of 0.2 is considered, which defines traffic flow with mean value of $A_A = 14.83$ Erl for Operator A and $A_B = 23.27$ Erl for Operator B.

Fig. 6.6 and Fig. 6.7 show the benefit of sharing resources in terms of reduction of the individual blocking probabilities. The x-axis define the percentage of resources that are shared between the operators. For example, 20% means that Operator A has 16, while Operator B has 24 resources reserved, and the remaining 10 resources are shared and used on a first come first served basis. At 40% sharing both operators reduce the blocking probability to almost 1%, and larger sharing does not further decrease the congestion.

Fig. 6.6 and Fig. 6.7 also indicate that sharing resources means that the variation of the traffic flow at one operator influences the congestion of the other

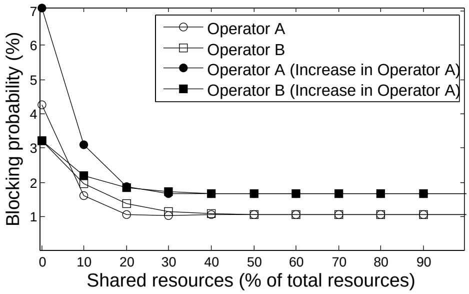


Figure 6.6: Impact on blocking probability due to 10 % increase in offered traffic for Operator A.

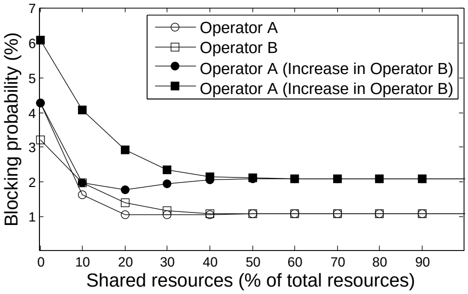


Figure 6.7: Impact on blocking probability due to 10 % increase in offered traffic for Operator B.

operator. By only 10% increase of the mean value for Operator A (Fig. 6.6) or Operator B (Fig. 6.7) the congestion for both operators is increased by 1% or more in case of traffic variation of Operator B. For that reason, it is important to investigate the possible strategies for sharing resources at Operator A, in case of traffic increase at Operator B. Fig 6.8 depicts the blocking probability for Operator A, as a function of the mean value of the Operator's B traffic flow and different degrees for sharing resources. Fig. 6.9 shows the blocking probability for Operator B under the same conditions.

The following can be observed from these figures. As long as $A_B \leq 20.63$, both operators benefit most by 100% sharing. For $A_B \in (20.63, 23.64]$, Operator A benefits most by 30% sharing. If $A_B \in (23.64, 26.65]$, operator A benefits most by 20% sharing. Sharing of 10% gives the lowest congestion if $A_B \in (26.65, 37.65]$. For any value $A_B > 37.65$, "no sharing" would be the preferred strategy for Operator A. Fig. 6.9 show the opposite trend for Operator B as full sharing strategy gives the lowest congestion. Any other sharing strategy would give less optimal congestion,

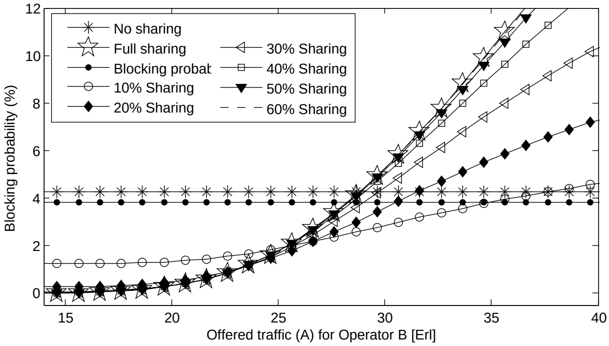


Figure 6.8: Sharing strategies for Operator A under traffic variations of Operator B.

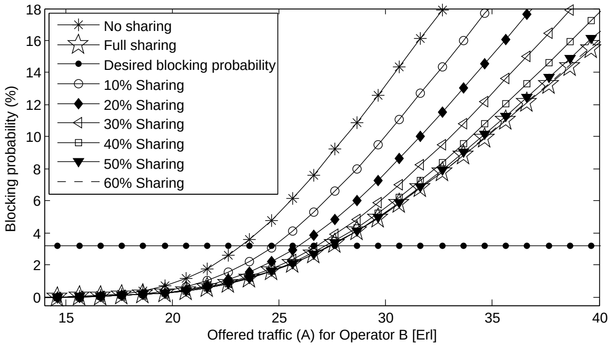


Figure 6.9: Sharing strategies for Operator B under traffic variations of Operator B.

but still better then "no sharing" strategy. Hence, if Operator A sets a threshold of blocking probability (Pb_{th}^{A}) that should not be exceeded, then suboptimal sharing strategies for Operator A can be chosen in order to improve the performance of Operator B. For example, if $Pb_{th}^{A} = 3.8\%$, then for $A_B \in (20.63, 28.27]$ any sharing strategy, except "no sharing" would improve the performances at Operator B, while still ensuring congestion lower than Pb_{th}^{A} for Operator A.

6.4 Resource sharing at a network level

The virtualization of the resources within one base station allows for creation of virtual base stations per operator. The abstraction of the control plane related to resource management and its centralization provides global knowledge of resource allocation within the individual elements. In this line, the decision on the sharing degree can be made globally, at an aggregation point versus locally within a limited area. Naturally the first approach allows for better resource utilization. However, the benefits for each operator might not be equal, and it is depended on the traffic load per area. Such an assessment is provided in this section.

6.4.1 Analytical Model

A heterogeneous network consisting of M small cells of n_k resources, covered by a macro cell with n resources is considered. In a cell k, operator i has a flow of a BPP traffic with mean value A_i^k, peakedness Z_i^k, and requires d_i^k resources for each connection. A call from a small cell can use resources from the macro cell, in case all resources at the small cell are occupied. Rearrangement of these calls is considered, such that when sufficient resources from the small cell become available, the call releases the resources at the macro cell and continues to use the resources at the small cell. The portion of resources that an operator i can use at a cell k is defined as n_i^k, when no sharing is present. If the operators decide to share a certain degree of the resources $s \in [0, 1]$, then an operator i has $((1 - s) \cdot n_i^k)$ number of resources reserved, but it can use at most $(n_i^k + s \cdot (n^k - n_i^k))$ RRs. In order to define the restrictions at each cell, the system can be described by a two dimensional matrix, where the columns define the individual traffic flows, while the cell restrictions are specified through the rows. Each element in the matrix, defines the number d_i^k of RRs that a connection from flow i requires within cell k. Then at a small cell k, the following restriction can be defined: $\sum_{i=1}^{N} x_i^k \leq (n_k + n)$. A restriction for operator i, within small cell k, can be defined as: $x_i^k \leq n_i^k$ in case resources are not shared, or as $x_i^k \leq (n_i^k + s \cdot (n_k - n_i^k))$ in case of resource sharing. The decision of sharing the resources can be made on a local level, per each small cell, or on a global level. Independently on where the decision is made, the total number of resources yields the following restriction:

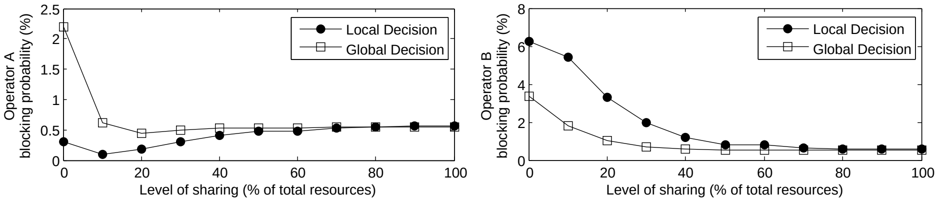
$$(\sum_{k=1}^{M}\sum_{i=1}^{N} x_i^k + \sum_{i=1}^{N} x_i^{macro}) \leq (\sum_{k}^{M} n_k + n) \tag{6.5}$$

In order to find the performance metrics for each operator, again the convolution method can be used on the flows for each operator. The aggregation of state probabilities for each operator into a global one, is done, one at a time, such that the rows define the restriction (diagonal truncation) to the state space. The restrictions define if the number of connections, or busy resources in this case, for each operator need to be tracked.

6.4.2 Numerical Analysis

In the numerical analysis a heterogeneous network composed of 3 small cells with $n_k = 15$ RRs and an overlay macro cell with $n = 5$ RRs is considered. Again two operators are considered, and each has a traffic flow to the small cells, such that $A_A^1 = 2$, $A_A^2 = 4$, $A_A^3 = 8$, $A_B^1 = 8$, $A_B^2 = 8$, and $A_B^3 = 8$. No traffic is offered at macro cell level, so the resources are used by the small cells in case of congestion. It is assumed that Operator A owns 40%, while Operator B owns 60% of the resources. The decision for sharing can be done locally (*local decision*) at each base station as in the previous case study. In case of "no sharing" strategy this would mean that at each small cell, Operator A has 6 RRs reserved, Operator B has 9 RRs reserved. The resources from the macro cell (2 RRs for Operator A and 3 RRs for Operator B) are shared among the small cells.

The decision of sharing can also be done at a central (aggregation) point such that a *global decision* is made. The global decision limits the sum of the resources from all cells that can be allocated to an operator. For the considered case above, Operator A has 20 RRs reserved, while Operator B has 30 RRs reserved within all cells in case of "no sharing" strategy. The obtained numerical results are presented in Fig. 6.10, Fig. 6.11 and Fig. 6.12. These figures show the blocking probabilities in case of local and global decision and for different degree of sharing per cell and per operator. The overall system performance in terms of total carrier traffic is presented in Fig. 6.13. The following can be observed from these figures:

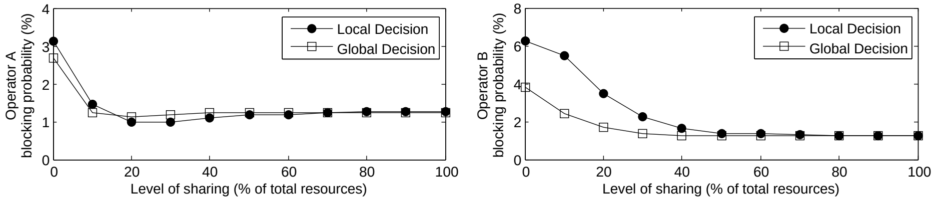


(a) Operator A

(b) Operator B

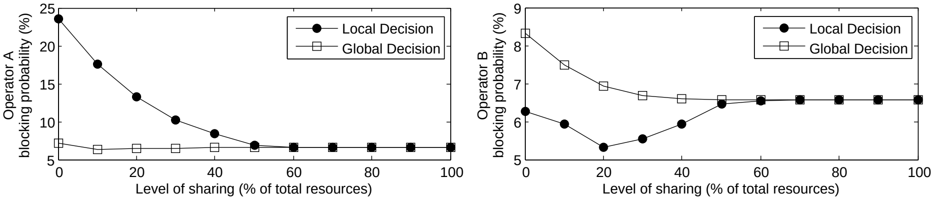
Figure 6.10: Blocking probabilities in Small Cell 1.

1.) "No sharing" strategy. It is important to underline the impact of the decision level on the performance for the individual operators in case of no resource sharing. The difference in the *local* and *global* decision is most evident in cell 3, where the local decision is more beneficial to Operator B, while Operator A benefits more from global decision. In case of local decision, Operator A achieves resource utilization ($\frac{carried\_traffic}{nr\_of\_resource}$) of approximately 60%, or only 25% in cell 1. In cell 3, both operators have the same offered traffic, but as Operator A has only 40% of the resources reserved, it will experience higher congestion. In case of global decision, there is no limitation per cell, hence the amount of resources that are not used by Operator A at cell 1, can be used by the operator at cell 3.



(a) Operator A (b) Operator B

Figure 6.11: Blocking probabilities in Small Cell 2.



(a) Operator A (b) Operator B

Figure 6.12: Blocking probabilities in Small Cell 3.

Thus for Operator A, congestion is slightly increased in cell 1, but significantly decreased in cell 3, from 24% to 6.5%. Fig. 6.13 shows the difference of the sum of the carried traffic per operator with respect to type of decision. It can be seen that the carried traffic is increased for Operator A in case of global decision, which increases the overall resource utilization.

Both operators benefit from global decision, as the individual blocking probability is decreased. This results from the fact that in case of global decision, the resources are distributed to the cell that has the highest load, and there is a degree of sharing of the resources from the macro cell among the operators. In this case Operator A benefits the most as the congestion is significantly reduced for cell 2 (from 11 % to 3.6 %) and cell 3 (28.5 % to 3.6 %). Furthermore, the global decision, ensures the similar level of congestions for all cells per operator.

2.) "Sharing" strategy. In case of local decision, Operator A benefits most at 30-40% sharing which decreases the congestion at cell 3. It can be also noticed that the performance for Operator B in small cell 3 is degraded in case of global

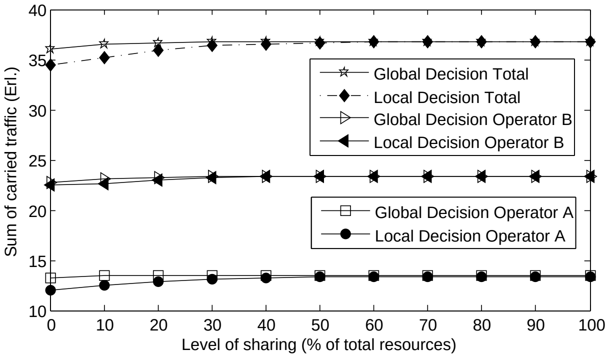


Figure 6.13: Comparison of carried traffic

decision. Hence, if Operator B needs to ensure certain grade of service, some local decisions on resource sharing need to be implemented. From Fig. 6.13 can be noticed that in case of global decision, both operator do not have significant increase of the carried traffic after 10-20% sharing. Thus, global decision is more robust as it requires less resource sharing. As such the variation of the operator's traffic will influence less to the performance of the other operators. As the sharing degree is increased, the two approaches converge to the same performance, which is expected, as all resources are shared in both cases.

Another way of analyzing the results is to look at the individual performances per cell. For both cell 1 and cell 2, it is more beneficial for Operator B to enter a sharing agreement, with higher degree of sharing. Thus for operators that have higher load and own large portion of resources is beneficial to enter a sharing agreement with global decision. However, for Operator A for the two cells, the benefits are highest only after 10-20 % sharing. If we look at the performances at cell 3, it is evident that local decisions must be enforced as well, in order to ensure that the performance of the operators is not compromised.

6.5 Conclusion

This chapter has elaborated on radio access networks with shared resources, where mobile operators are encouraged to enter a sharing agreement in order to relax the expenditures. The research in this chapter has been strongly motivated by the potentials of the network sharing related to improved spectrum and infrastructure utilization, as well as by the advances resulting from the emerging technologies such as C-RAN, virtualization, SDR and SDN. This chapter provides a quantitative study of dynamic resource sharing among operators and confirms the benefits seen from overall network utilization. More importantly, the numerical analysis provided gives insights on the performances of the individual operators, which reveals that the benefits are unequal for operators and dependent on several parameters.

In the analysis two operators and a shared pool of resources is considered. The operators contribute to the shared pool of resources by deciding on a sharing percent from its own resource pool. The sharing decision strategies have been analyzed at a local level at a single cell, as well as on a global level that relates to an aggregation point in a heterogeneous network. Although the numerical analysis is given for two operators, the analytical method proposed is general and it can be applied to larger number of operators and various scenarios such as in case of multi-RAT deployments.

The results from the study clearly show that by increasing the level of sharing, the two operators can benefit by having a decreased blocking probability, both for local and global decision. The novelty in this work is the analysis of the influence of traffic variation for Operator A, on the performance to Operator B for different level of sharing. This way a sensitivity study is performed, in order to quantify the derivative of the congestion for operator A with respect to the variation of the offered traffic of Operator B. The results can be used by the RAN provider to dimension the level of sharing in a way to maximize the overall performance without compromising the individual performances of each operator. Additionally, the study implies limitation of the allowed traffic per operator in order to optimize the congestion per operator.

Sharing of the resources in the RAN requires clear bilateral agreements among operators that will be supported and approved by the regulators, and strict orchestration that will preserve and enforce the agreements. In this work, the agreement has been considered as a sharing percent, and two cases of enforcement have been considered: locally within a cell area and globally at an aggregation point in a network composed of macro and small cells. It was demonstrated that the later one provides better resource utilization in the network, and by only 20 % sharing, the total blocking probability is reduced to the case of a system with full accessibility (100% share). This means that such system is less sensitive to traffic variations. However, a careful analysis per cell reveals that local decisions need to be made in order to ensure fair benefits to both operators.

CHAPTER 7

Thesis summary

This PhD thesis elaborates on the essentials for contemporary successful resource management and various performance improvements in heterogeneous networks. The presented novelties provide a comprehensive overviews and discussions of the main issues that need to be considered for better resource utilization, customer satisfaction and fair share for different providers within a heterogeneous network. Analytical approach and used models allows for generalization such that various case studies can be studied. The performances of the schemes have been assessed with system level simulations and/or analytical methods and comparisons with the previous case studies have been evaluated. The obtained numerical results show expected improvements which lead to valuable recommendation to the actors in the planning activities.

Chapter 2 presented a novel universal approach for joint resource management that facilitates the emerging technologies for spectrum, energy and cost efficiency. The proposed framework is generic and can be applied for various scenarios and network deployments. The different functionalities related to resource management in a multi-RAT network, (dense) small cell deployments and BS architectures have been addressed as well. The proposed hierarchical approach provides balance between centralized management characterized with global knowledge but excessive signaling and distributed management with local control of resources. Most importantly the individual modules are enhanced with cognitive behavior and learning abilities in order to assist the decision process and timely adapt to the changes in the network. The cognitive concept and the SON principles augment the proposed framework allowing for more efficient and automated resource management.

The third chapter dealt with the dual connectivity – a very important new form of multi-cell cooperation for further densification and spectrum extension through exploration of higher spectrum frequencies for the small cells. Bearer split as a part of dual connectivity, extends carrier aggregation for small cells, allowing for resource aggregation across carries without the need of ideal backhaul connection. The work presented considered the problem of user power allocation scheme for dual connectivity with bearer split. It proposed that two separate configurations per component carrier are maintained: one for single transmission and the other in case of dual transmissions within a TTI. The design on the power control has been simplified as to ensure minimal coordination among base stations as well as minimal air interface signaling. An important outcome of this study is that with bearer split high gains can be achieved for the median and cell edge user throughput. For cell edge user throughput almost threefold gain is accomplished for low load, while approximately 30% gain is achieved at high load (compared to the case when DC is not configured). The results have indicated that the distribution of the power is important in order to reduce the possible interference increase in the aggregation layer. However, a deeper study through investigations on the implications on higher layers and other mechanism is also required.

The Chapter 4 has discussed the problem of adaptive traffic allocation in cooperative and heterogeneous network. A novel optimization problem has been proposed with the objective of improving the traffic transmission quality under QoS constraints for different services. Compared to the previous research, the work presented considered three levels of heterogeneity among the available access networks: capacity, delay variations and QoS provisioning. Furthermore, services with different quality requirements have been considered. The numerical results for different case studies confirmed the benefits of multi-stream aggregation over heterogeneous networks. The performance and superiority of the proposed allocation is demonstrated by assessment with traditional allocation methods known in the literature. The performance is evaluated with respect to the achieved throughput, maximum transmission delay among the available interfaces as well as the buffering requirements needed to reduce the packet reordering at the receiver. The proposed solution provides flexibility and can be applied for different deployments.

The Chapter 5 has focused on C-RAN as an important architecture that offers cost benefits and improved radio resource utilization. As energy and cost savings are related to multiplexing gain, this chapter evaluated such gains under different network deployments and traffic profiles. The work concluded the optimal conditions for dense cell deployments under which the multiplexing gain is maximized. The numerical analysis indicated that not only cost, but also sensitivity to traffic variations need to be investigated when dimensioning the pool of baseband units as the pool size is increased. Highest multiplexing gains on computational resources can be achieved in case of full baseband centralization. However, such architecture requires the highest fronthaul capacity and may only be viable for operators that can have access to such network at a reasonable cost. Deployments with heterogeneous functional splits have been evaluated in order to give indications on how much of the traffic can be carried on specific interface. The numerical results suggested that heterogeneous deployments should be considered in order to optimally balance the multiplexing gains, spectral efficiency and cost of the fronthaul infrastructure.

The last but not least Chapter 6 of this work has paid attention on radio access networks with shared resources where mobile operators are encouraged to enter a sharing agreement in order to relax the expenditures. This chapter also investigates strategies for active sharing of the radio access among multiple operators and has analyzed the individual benefits depending on the sharing degree. The results from the study clearly showed that by increasing the level of sharing, the two operators can benefit by having a decreased blocking probability, both for local and global decision. The novelty in this work is the analysis of the influence of traffic variation for Operator A on the performance to Operator B for different level of sharing. This way a sensitivity study has been performed in order to quantify the derivative of the congestion for Operator A with respect to the variation of the offered traffic of Operator B. The results can be used by the RAN provider to dimension the level of sharing in a way to maximize the overall performance without compromising the individual performances of each operator. Additionally, the study implies limitation of the allowed traffic per operator in order to optimize the congestion per operator.

7.1 Future Research Directions

The work performed during the Ph.D. study considered different challenges arising from heterogeneity in the mobile networks. The findings in this thesis provide means for a converged wireless communication that will enhance the seamless user experience of mobile broadband communications. This is of high importance for the future Fifth Generation (5G) of mobile communication.

In the future, it is expected even more diversity in mobile networks and diversity in mobile devices and service requirements resulting from new business opportunities. Enhancements of the existing mobile communication system will be developed in order to address the need for higher network capacity and enrich the mobile broadband experience. In this context, there is a need for further exploration in the millimeter wave bands to augment the communication capacity. There is enormous potential in utilizing mmWave for small cell access and for the wireless fronthaul. There are still relevant open issues that need to be addressed in order to facilitate deployments of mmWave in future networks [182].

A novel proposal for a new mobile access medium includes the visible light bands. The Visible Light Communications (VLC) also called LiFi, enables data transmission on the intensity of lightning source. This new technology creates an additional tier in HetNets formed by the LiFi-enabled lights that will be used to offload the traffic from the other tiers and WiFi hotspots. In this context, [183] indicated the importance of coexistence of WiFi and LiFi, and demonstrated a framework for close integration of these technologies. Their results show that both technologies together can do more than triple the throughput for individual users.

The network society will be further expanded through new use cases being created by various vertical industries, such as public safety, e-health etc. In this context, an important area of future study is the Machine Type Communication

(MTC). The number of machine devices that are expected to be network connected will outline the number of human centric communication devices. But handling massive number of devices is not the only challenge - the specific applications that these devices have will result in diverse amount of data that need to be transmitted, as well as can have different requirements for cost, power consumption and coverage.

Bibliography

[1] A. Checko^{1st}, A. P. Avramova^{1st}, M. Berger, and H. L. Christiansen. “Evaluating C-RAN fronthaul functional splits in terms of network level energy and cost savings”. In: *Journal of Communications and Networks* (2016). accepted for publication (cit. on pp. xi, 73).

[2] A. P. Avramova, H. Wang, L. Dittmann, and K. I. Pedersen. “ Optimal Uplink Power Control for Dual Connected Users in LTE Heterogeneous Networks”. In: *IEEE 83rd Vehicular Technology Conference: VTC2016-Spring*. accepted for publication. May 2016 (cit. on pp. xi, 33).

[3] A. P. Avramova and V. Iversen. “Optimal Traffic Allocation for Multi-Stream Aggregation in Heterogeneous Networks”. In: *IEEE 7th Int. Workshop on Heterogeneous and Small Cell Networks (HetSNets), IEEE GLOBECOM 2015*. Dec. 2015 (cit. on pp. xi, 57).

[4] A. P. Avramova, H. L. Christiansen, and V. Iversen. “Cell Deployment Optimization for Cloud Radio Access Networks using Teletraffic Theory”. In: *The 11th Advanced International Conference on Telecommunications, AICT 2015*. July 2015 (cit. on pp. xi, 73, 83).

[5] A. P. Avramova and V. Iversen. “Radio Access Sharing Strategies for Multiple Operators in Cellular Networks”. In: *The First IEEE International Workshop on 5G & Beyond - Enabling Technologies and Application, IEEE ICC 2015*. June 2015, pp. 1113–1118 (cit. on pp. xi, 101).

[6] A. Zakrzewska, A. P. Avramova, H. Christiansen, Y. Yan, A. Checko, A. Dogadaev, S. Ruepp, M. Berger, and L. Dittmann. “A Framework for Joint Optical-Wireless Resource Management in Multi-RAT, Heterogeneous Mobile Networks”. In: *Workshop on Optical-Wireless Integrated Technology for Systems and Networks (OWITSN) 2013, IEEE ICC 2013*. June 2013, pp. 895–899 (cit. on pp. xii, 17).

[7] A. Zakrzewska, A. P. Avramova, S. Ruepp, M. Berger, and L. Dittmann. “CSMA-based SON Mechanism for Greening Heterogeneous Networks”. In: *INFOCOM Student Session, INFOCOM 2013*. Apr. 2013 (cit. on pp. xii, 17).

[8] E. Kisielius, A. Zakrzewska, A. P. Avramova, and S. Ruepp. “Energy Efficiency in Self Organising Networks”. In: *OPNETWORK 2013*. Aug. 2012 (cit. on pp. xii, 17).

[9] A. P. Avramova, Y. Yan, and L. Dittmann. “Evaluation of a Cross Layer Scheduling Algorithm for LTE Downlink.” In: *Telfor Journal* 5.1 (2013), pp. 26–31 (cit. on p. xii).

[10] A. P. Avramova, Y. Yan, S. Ruepp, and L. Dittmann. “Resource Allocation: Current Issues and Future Directions”. In: *4th Nordic Workshop On System And Network Optimization For Wireless (SNOW 2013)*. Apr. 2013 (cit. on p. xii).

[11] A. Dogadaev, A. Checko, A. P. Avramova, A. Zakrzewska, Y. Yan, S. Ruepp, M. Berger, L. Dittmann, and H. Christiansen. “Trafic Steering Framework for Mobile-Assisted Resource Management in Heterogeneous Networks”. In: *Ninth International Conference on Wireless and Mobile Communications, ICWMC*. July 2013 (cit. on p. xii).

[12] A. P. Avramova, Y. Yan, and L. Dittmann. “Cross layer scheduling algorithm for LTE downlink”. In: *20th Telecommunications Forum (TELFOR), 2012*. Nov. 2012, pp. 307–310 (cit. on p. xii).

[13] A. P. Avramova, Y. Yan, and L. Dittmann. “Modeling and Simulation of Downlink Subcarrier Allocation Schemes in LTE”. In: *OPNETWORK 2012*. Aug. 2012 (cit. on p. xii).

[14] A. P. Avramova, Y. Yan, and L. Dittmann. “Modeling of Bandwidth Aggregation over Heterogeneous Wireless Access Networks”. In: *OPNETWORK 2012*. Aug. 2012 (cit. on pp. xii, 17).

[15] B. Stojkoska and A. P. Avramova. “Wireless Sensor Networks Framework for Indoor Temperature Regulation”. In: *8th Annual South-East European Doctoral Student Conference*. Sept. 2013, pp. 354–364 (cit. on p. xiii).

[16] B. Stojkoska, A. P. Avramova, and P. Chatzimisios. “Application of Wireless Sensor Networks for Indoor Temperature Regulation”. In: *International Journal of Distributed Sensor Networks* (2014). Article ID 502419 (cit. on p. xiii).

[17] S. Michail, A. P. Avramova, and L. Dittmann. “MIH Based Mobility for TETRA-LTE Network”. In: *ICT Innovations 2013*. Sept. 2013 (cit. on p. xiii).

[18] A. P. Avramova, S. Ruepp, and L. Dittmann. “Towards Future Broadband Public Safety Systems: Current Issues and Future Directions”. In: *International Conference on Information and Communication Technology Convergence (ICTC)*. Oct. 2015, pp. 74–79 (cit. on p. xiii).

[19] *Global Mobile Data Traffic Forecast Update, 2015–2020.* Tech. rep. Cisco Visual Networking Index, 2016 (cit. on pp. 1, 74).

[20] 3GPP Technical Report (TR) 22.912. *Technical Specification Group Services and System Aspects; Study into network selection requirements for non-3GPP Access (Release 10).* Tech. rep. 3^{rd} Generation Partnership Project (3GPP), March 2011, available at www.3gpp.org (cit. on p. 3).

[21] 3GPP Technical Report (TR) 22.934. *Technical Specification Group Services and System Aspects; 3GPP system to Wireless Local Area Network (WLAN) interworking; System description (Release 10).* Tech. rep. 3^{rd} Generation Partnership Project (3GPP), March 2011, available at www.3gpp.org (cit. on p. 3).

[22] R. Ferrus, O. Sallent, and R. Agusti. "Interworking in heterogeneous wireless networks: Comprehensive framework and future trends". In: *IEEE Wireless Communications* 17.2 (Apr. 2010), pp. 22–31 (cit. on p. 3).

[23] S. Velentzas and T. Dagiuklas. "4G Cellular/WLAN Interworking". In: *Third International Working Conference HET-NETs 2005.* July 2005 (cit. on p. 3).

[24] C. Makaya and S. Pierre. "An Interworking Architecture for Heterogeneous IP Wireless Networks". In: *Wireless and Mobile Communications, 2007. ICWMC '07. Third International Conference on.* Mar. 2007, pp. 16–16 (cit. on p. 3).

[25] R. Beaubrun. "Heterogeneous Wireless Access Networks: Architectures and Protocols". In: ed. by E. Hossain. Boston, MA: Springer US, 2009. Chap. Integration of Heterogeneous Wireless Access Networks, pp. 1–18 (cit. on p. 3).

[26] E. W. C. Perkins. *IP Mobility Support for IPv4.* RFC 5944. Nov. 2010 (cit. on p. 3).

[27] 3GPP Technical Specification (TS) 23.234. *3GPP system to Wireless Local Area Network (WLAN) interworking (Release 8).* Tech. rep. 3^{rd} Generation Partnership Project (3GPP), Dec. 2008, available at www.3gpp.org (cit. on p. 3).

[28] 3GPP Technical Specification (TS) 43.318. *Generic Access Network (GAN) (Release 8).* Tech. rep. 3^{rd} Generation Partnership Project (3GPP), Dec. 2008, available at www.3gpp.org (cit. on p. 3).

[29] 3GPP. *Enhanced Generic Access Networks (EGAN) study.* TR 43.902 version 11.0.0 Release 11. 3^{rd} Generation Partnership Project (3GPP), Oct. 2012 (cit. on p. 3).

[30] P. TalebiFard, T. Wong, and V. C. Leung. “Heterogeneous Wireless Access Networks: Architectures and Protocols”. In: ed. by E. Hossain. Boston, MA: Springer US, 2009. Chap. Integration of Heterogeneous Wireless Access Networks with IP-based Core Networks: The Path to Telco 2.0, pp. 1–36 (cit. on p. 4).

[31] L. M. Gavrilovska and V. M. Atanasovski. “Resource management in wireless heterogeneous networks (WHNs)”. In: *Telecommunication in Modern Satellite, Cable, and Broadcasting Services, 2009. TELSIKS ’09. 9th Int. Conf. on.* 2009, pp. 97–106 (cit. on pp. 4, 18).

[32] *NATO SfP-982469 RIWCoS (Reconfigurable Interoperability of Wireless Communications Systems) project, 2007 - 2010.* `http://riwcos.comm.pub.ro` (cit. on pp. 4, 18).

[33] J. Perez-Romero, O. Sallent, R. Agusti, P. Karlsson, A. Barbaresi, L. Wang, F. Casadevall, M. Dohler, H. Gonzalez, and F. Cabral-Pinto. “Common radio resource management: functional models and implementation requirements”. In: *Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005. IEEE 16th Int. Symposium on.* Vol. 3. 2005, 2067–2071 Vol. 3 (cit. on pp. 4, 18).

[34] L. Giupponi, R. Agusti, J. Perez-Romero, and O. Sallent. “Joint radio resource management algorithm for multi-RAT networks”. In: *Global Telecommunications Conf., 2005. GLOBECOM ’05. IEEE.* Vol. 6. 2005 (cit. on pp. 4, 18).

[35] F. Zhiyong, Y. Kai, J. Yang, Z. Ping, V. O. K. Li, and Z. Yongjing. “Multiaccess radio resource management using multi-agent system”. In: *Wireless Communications and Networking Conf., 2006. WCNC 2006. IEEE.* Vol. 1. 2006, pp. 63–68 (cit. on pp. 4, 19).

[36] *A cognitive radio project (ARAGORN).* `http://www.ict-aragorn.eu` (cit. on pp. 4, 19).

[37] Y. Jian, C.-F. Shih, B. Krishnaswamy, and R. Sivakumar. “Coexistence of Wi-Fi and LAA-LTE: Experimental evaluation, analysis and insights”. In: *Communication Workshop (ICCW), 2015 IEEE International Conference on.* June 2015, pp. 2325–2331 (cit. on p. 5).

[38] S. Dama, A. Kumar, and K. Kuchi. “Performance Evaluation of LAA-LBT Based LTE and WLAN’s Co-Existence in Unlicensed Spectrum”. In: *2015 IEEE Globecom Workshops (GC Wkshps).* Dec. 2015, pp. 1–6 (cit. on p. 5).

[39] A. Khalid, P. Lassila, and S. Aalto. “Load Balancing of Elastic Data Traffic in Heterogeneous Wireless Networks”. In: *Teletraffic Congress (ITC), 2013 25th International.* Sept. 2013, pp. 1–9 (cit. on pp. 6, 59).

[40] H. Lian, et al. “Efficient Traffic Allocation Scheme for Multi-flow Distribution in Heterogeneous Networks”. In: *IEEE Globecom Workshops.* Dec. 2013, pp. 923–928 (cit. on pp. 6, 59, 60).

[41] A. Dilawari and M. Tahir. “Optimal Flow Splitting for Multi-Path Multi-Interface Wireless Data Streaming Networks”. In: *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on.* Sept. 2013, pp. 1878–1882 (cit. on pp. 6, 59).

[42] S. Gubner and C. Lindemann. “Aggregation-aware routing in wireless multi-hop networks with frame aggregation”. In: *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013 IEEE 14th International Symposium and Workshops on a.* June 2013, pp. 1–6 (cit. on pp. 6, 59).

[43] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi. “A survey on 3GPP heterogeneous networks”. In: *Wireless Communications, IEEE* 18.3 (June 2011), pp. 10–21 (cit. on p. 6).

[44] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel, and S. Geirhofer. “Network densification: the dominant theme for wireless evolution into 5G”. In: *Communications Magazine, IEEE* 52.2 (Feb. 2014), pp. 82–89 (cit. on p. 6).

[45] X. Chu, Y. Y. David Lopez-Perez, and F. Gunnarsson. *Heterogeneous Cellular Networks: Theory, Simulation and Deployment.* University Cambridge Press, 2013 (cit. on p. 6).

[46] J. Andrews. “Seven ways that HetNets are a cellular paradigm shift”. In: *Communications Magazine, IEEE* 51.3 (Mar. 2013), pp. 136–144 (cit. on p. 6).

[47] D. Lopez-Perez, I. Guvenc, and X. Chu. “Mobility management challenges in 3GPP heterogeneous networks”. In: *Communications Magazine, IEEE* 50.12 (Dec. 2012), pp. 70–78 (cit. on p. 7).

[48] K. Smiljkovikj, L. Gavrilovska, and P. Popovski. “Efficiency analysis of downlink and uplink decoupling in heterogeneous networks”. In: *Communication Workshop (ICCW), 2015 IEEE International Conference on.* June 2015, pp. 125–130 (cit. on p. 7).

[49] 3GPP Technical Report (TR) 36.842. *Study on Small Cell enhancements for E-UTRA and E-UTRAN; Higher layer aspects.* Tech. rep. July 2014, available at www.3gpp.org (cit. on pp. 7, 34).

[50] 3GPP Technical Report (TR) 36.932. *Scenarios and Requirements for Small Cell Enhancements.* Tech. rep. 2013, available at www.3gpp.org (cit. on pp. 7, 34).

[51] Y. Jin, L. Qiu, and X. Liang. “Small cells on/off control and load balancing for green dense heterogeneous networks”. In: *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*. Mar. 2015, pp. 1530–1535 (cit. on p. 8).

[52] Y. S. Soh, T. Quek, M. Kountouris, and H. Shin. “Energy Efficient Heterogeneous Cellular Networks”. In: *Selected Areas in Communications, IEEE Journal on* 31.5 (May 2013), pp. 840–850 (cit. on p. 8).

[53] J. Peng, P. Hong, and K. Xue. “Energy-Aware Cellular Deployment Strategy Under Coverage Performance Constraints”. In: *Wireless Communications, IEEE Transactions on* 14.1 (Jan. 2015), pp. 69–80 (cit. on p. 8).

[54] M. Peng, C. Wang, J. Li, H. Xiang, and V. Lau. “Recent Advances in Underlay Heterogeneous Networks: Interference Control, Resource Allocation, and Self-Organization”. In: *Communications Surveys Tutorials, IEEE* 17.2 (Secondquarter 2015), pp. 700–729 (cit. on p. 9).

[55] H. Wang, C. Rosa, and K. Pedersen. “Uplink Inter-Site Carrier Aggregation between Macro and Small Cells in Heterogeneous Networks”. In: *Vehicular Technology Conference (VTC Fall), 2014 IEEE 80th*. Sept. 2014, pp. 1–5 (cit. on pp. 9, 39, 45, 52, 53).

[56] G. Kardaras and C. Lanzani. “Advanced multimode radio for wireless and mobile broadband communication”. In: *Wireless Technology Conference, 2009. EuWIT 2009. European*. Sept. 2009, pp. 132–135 (cit. on p. 11).

[57] Common Public Radio Interface (CPRI). *Interface Specification*. August 2013, available at http://www.cpri.info (cit. on p. 11).

[58] ETSI GS ORI 002-2 V1.1.1. *Open Radio equipment Interface (ORI); ORI Interface Specification; Part 2: Control and Management (Release 1)*. August 2012 (cit. on p. 11).

[59] Open Base Station Architecture Initiative (OBSAI). *BTS System Reference Document, Version 2.0*. Tech. rep. 2006, available at http://www.obsai.org (cit. on p. 11).

[60] R. Wang, H. Hu, and X. Yang. “Potentials and Challenges of C-RAN Supporting Multi-RATs Toward 5G Mobile Networks”. In: *Access, IEEE* 2 (2014), pp. 1187–1195 (cit. on pp. 11, 104, 105).

[61] J. Liu, S. Zhou, J. Gong, Z. Niu, and S. Xu. “On the statistical multiplexing gain of virtual base station pools”. In: *Global Communications Conference (GLOBECOM), 2014 IEEE*. Dec. 2014, pp. 2283–2288 (cit. on pp. 11, 76, 83).

[62] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann. “Cloud RAN for Mobile Networks - a Technology Overview”. In: *IEEE Communications Surveys & Tutorials, IEEE* 17.1 (Firstquarter 2015), pp. 405–426 (cit. on pp. 12, 74).

[63] A. Davydov, G. Morozov, I. Bolotin, and A. Papathanassiou. “Evaluation of Joint Transmission CoMP in C-RAN based LTE-A HetNets with large coordination areas”. In: *Globecom Workshops (GC Wkshps), 2013 IEEE*. Dec. 2013, pp. 801–806 (cit. on p. 12).

[64] C.-L. I, J. Huang, R. Duan, C. Cui, J. Jiang, and L. Li. “Recent Progress on C-RAN Centralization and Cloudification”. In: *Access, IEEE* 2 (2014), pp. 1030–1039 (cit. on pp. 12, 74).

[65] S. Ferreira, et al. “An Architecture to offer Cloud-Based Radio Access Network as a Service”. In: *European Conference on Networks and Communications*. June 2014, pp. 1–5 (cit. on pp. 12, 105, 106).

[66] S. Nanba and A. Agata. “A new IQ data compression scheme for front-haul link in Centralized RAN”. In: *Personal, Indoor and Mobile Radio Communications (PIMRC Workshops), 2013 IEEE 24th International Symposium on*. Sept. 2013, pp. 210–214 (cit. on p. 12).

[67] *Small Cell Virtualization Functional Splits and Use Cases*. Tech. rep. Small Cell Forum, June 2015 (cit. on pp. 13, 90, 91).

[68] P. Rost, C. Bernardos, A. Domenico, M. Girolamo, M. Lalam, A. Maeder, D. Sabella, and D. Wubben. “Cloud technologies for flexible 5G radio access networks”. In: *Communications Magazine, IEEE* 52.5 (May 2014), pp. 68–76 (cit. on pp. 13, 90).

[69] *Next Generation Fronthaul Interface*. Tech. rep. China Mobile Research Institute, June 2015 (cit. on pp. 13, 90).

[70] U. Dotsch, M. Doll, H.-P. Mayer, F. Schaich, J. Segel, and P. Sehier. “Quantitative analysis of split base station processing and determination of advantageous architectures for LTE”. In: vol. 18. 1. June 2013, pp. 105–128 (cit. on pp. 13, 75, 90).

[71] C. Chen, J. Huang, W. Jueping, Y. Wu, and G. Li. *Suggestions on Potential Solutions to C-RAN*. Tech. rep. 2013, NGMN Alliance (cit. on p. 13).

[72] C. Dehos, J. González, A. De Domenico, D. Kténas, and L. Dussopt. “Millimeter-wave access and backhauling: the solution to the exponential data traffic increase in 5G mobile communications systems?” In: *Communications Magazine, IEEE* 52.9 (Sept. 2014), pp. 88–95 (cit. on p. 13).

[73] M. Artuso, A. Marcano, and H. Christiansen. “Cloudification of mmwave-based and packet-based fronthaul for future heterogeneous mobile networks”. In: *Wireless Communications, IEEE* 22.5 (Oct. 2015), pp. 76–82 (cit. on p. 13).

[74] Y.-N. R. Li, H. Xiao, J. Li, and H. Wu. “Wireless backhaul of dense small cell networks with high dimension MIMO”. In: *Globecom Workshops (GC Wkshps), 2014*. Dec. 2014, pp. 1254–1259 (cit. on p. 13).

[75] *Evolutionary Strategies for Radio Resource Management in Cellular Heterogeneous Networks (EVEREST)*. `http://cordis.europa.eu/projects/rcn/74605_en.html` (cit. on p. 18).

[76] S. Hamalainen and H. Sanneck and C. Sartori. *LTE Self-Organizing Networks (SON)*. John & Wiley Sons, Ltd., 2012 (cit. on pp. 22, 26).

[77] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming. *3G Evolution. HSPA and LTE for Mobile Broadband*. Elsevier, 2007 (cit. on p. 22).

[78] J. P. Vasseur, M. Pickavet and P. Demeester. *Network Recovery, Protection and Restoration of Optical, SONET-SDH, IP, and MPLS*. Morgan-Kaufmann Publishers, Elsevier, 2004 (cit. on p. 22).

[79] K. Dimou, M. Wang, Y. Yang, M. Kazmi, A. Larmo, J. Pettersson, W. Muller, and Y. Timner. “Handover within 3GPP LTE: Design Principles and Performance”. In: *Vehicular Technology Conf. Fall (VTC 2009-Fall), 2009 IEEE 70th*. 2009 (cit. on p. 22).

[80] H. Holma, K. Hooli, P. Kinnunen, T. Kolding, P. Marsch, and X. Wang. “Coordinated Multipoint Transmission and Reception”. In: *LTE-Advanced*. Ed. by H. Holma and A. Toskala. John Wiley & Sons, 2012, pp. 184–205 (cit. on p. 22).

[81] J. Sachs and J. Rune. “Access Network Selection in Multi-Access Network Environment”. 20110110300. May 2011 (cit. on p. 22).

[82] H. Wang, L. Ding, P. Wu, Z. Pan, N. Liu, and X. You. “Dynamic load balancing and throughput optimization in 3GPP LTE networks”. In: *IWCMC ’10 Proceedings of the 6th Int. Wireless Communications and Mobile Computing*. 2010, pp. 939–943 (cit. on p. 22).

[83] *5G systems, ENABLING INDUSTRY AND SOCIETY TRANSFORMATION*. Tech. rep. Ericsson, January 2015 (cit. on p. 26).

[84] 3GPP. *Telecommunication management; Self-Organizing Networks (SON); Concepts and requirements*. TR 32.500. 3^{rd} Generation Partnership Project (3GPP), Feb. 2009 (cit. on p. 26).

[85] C. Fortuna and M. Mohorcic. “Trends in the development of communication networks: Cognitive networks”. In: *Computer Networks* 53.9 (2009), pp. 1354–1376 (cit. on p. 27).

[86] J. Mitola III and J. Maguire G.Q. “Cognitive radio: making software radios more personal”. In: *Personal Communications, IEEE* 6.4 (Aug. 1999), pp. 13–18 (cit. on p. 27).

[87] 3GPP. *Telecommunication management; Self-Organizing Networks (SON) Policy Network Resource Model (NRM) Integration Reference Point (IRP); Information Service (IS).* TR 32.522. 3^{rd} Generation Partnership Project (3GPP), Dec. 2012 (cit. on p. 29).

[88] X. Chu, D. Lopez-Perez, Y. Yang, and G. F. *Heterogeneous Cellular Networks: Theory, Simulation and Deployment.* Cambridge University Press, July 2013 (cit. on p. 34).

[89] D. Astely, E. Dahlman, G. Fodor, S. Parkvall, and J. Sachs. “LTE release 12 and beyond [Accepted From Open Call]”. In: *Communications Magazine, IEEE* 51.7 (July 2013), pp. 154–160 (cit. on p. 34).

[90] Wireless World Research Forum. *LTE Small Cell Enhancements by Dual Connectivity.* White Paper. November 2014, Version 1.1 (cit. on pp. 34, 38).

[91] Y. Kishiyama, A. Benjebbour, T. Nakamura, and H. ISHII. “Future steps of LTE-A: evolution toward integration of local area and wide area systems”. In: *Wireless Communications, IEEE* 20.1 (Feb. 2013), pp. 12–18 (cit. on p. 35).

[92] Parkvall, S. and Dahlman, E. and Jongren, G. and Landstrom, S. and Lindbom, L. *Heterogeneous network deployments in LTE: The Soft-cell approach.* Tech. rep. Ericsson Review, 2011 (cit. on p. 35).

[93] H. ISHII, Y. Kishiyama, and H. Takahashi. “A novel architecture for LTE-B :C-plane/U-plane split and Phantom Cell concept”. In: *Globecom Workshops (GC Wkshps), 2012 IEEE.* Dec. 2012, pp. 624–630 (cit. on p. 35).

[94] A. Zakrzewska, et al. “Dual connectivity in LTE HetNets with split control- and user-plane”. In: *Globecom Workshops (GC Wkshps), 2013 IEEE.* Dec. 2013, pp. 391–396 (cit. on p. 35).

[95] G. Pocovi, S. Barcos, H. Wang, K. Pedersen, and C. Rosa. “Analysis of Heterogeneous Networks with Dual Connectivity in a Realistic Urban Deployment”. In: *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st.* May 2015, pp. 1–5 (cit. on p. 36).

[96] A. Mukherjee. “Macro-small cell grouping in dual connectivity LTE-B networks with non-ideal backhaul”. In: *Communications (ICC), 2014 IEEE International Conference on.* June 2014, pp. 2520–2525 (cit. on p. 38).

[97] M. Kim, S. Y. Jung, and S.-L. Kim. “Sum-rate maximizing cell association via dual-connectivity”. In: *Computer, Information and Telecommunication Systems (CITS), 2015 International Conference on.* July 2015, pp. 1–5 (cit. on p. 38).

[98] H. Wang, C. Rosa, and K. Pedersen. “Inter-eNB Flow Control for Heterogeneous Networks with Dual Connectivity”. In: *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st.* May 2015, pp. 1–5 (cit. on p. 38).

[99] A. Mukherjee. “Optimal flow bifurcation in networks with dual base station connectivity and non-ideal backhaul”. In: *Signals, Systems and Computers, 2014 48th Asilomar Conference on.* Nov. 2014, pp. 521–524 (cit. on p. 38).

[100] S. Jha, K. Sivanesan, R. Vannithamby, and A. Koc. “Dual Connectivity in LTE small cell networks”. In: *Globecom Workshops (GC Wkshps), 2014.* Dec. 2014, pp. 1205–1210 (cit. on p. 39).

[101] H. Holma, A. Toskala, and J. Reunanen. John Wiley and Sons Ltd, 2015. ISBN: 9781118912560 (cit. on pp. 40, 41).

[102] L. Zhao, X. Huang, Y. Li, L. Liu, H. Jiang, K. Takeda, and X. Ji. “Uplink Power Control for Dual Connectivity”. In: *Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on.* Dec. 2014, pp. 1412–1416 (cit. on p. 40).

[103] J. Liu, J. Liu, and H. Sun. “An Enhanced Power Control Scheme for Dual Connectivity”. In: *Vehicular Technology Conference (VTC Fall), 2014 IEEE 80th.* Sept. 2014, pp. 1–5 (cit. on p. 40).

[104] S. Ahmad and D. Datla. “Distributed Power Allocations in Heterogeneous Networks With Dual Connectivity Using Backhaul State Information”. In: *Wireless Communications, IEEE Transactions on* 14.8 (Aug. 2015), pp. 4574–4581 (cit. on pp. 40, 46).

[105] X. Xu, G. He, S. Zhang, Y. Chen, and S. Xu. “On functionality separation for green mobile networks: concept study over LTE”. In: *Communications Magazine, IEEE* 51.5 (May 2013), pp. 82–90 (cit. on p. 42).

[106] E. Ternon, P. Agyapong, L. Hu, and A. Dekorsy. “Database-aided energy savings in next generation dual connectivity heterogeneous networks”. In: *Wireless Communications and Networking Conference (WCNC), 2014 IEEE.* Apr. 2014, pp. 2811–2816 (cit. on p. 42).

[107] A. Prasad and A. Maeder. “Energy Saving Enhancement for LTE-Advanced Heterogeneous Networks with Dual Connectivity”. In: *Vehicular Technology Conference (VTC Fall), 2014 IEEE 80th.* Sept. 2014, pp. 1–6 (cit. on p. 42).

[108] S. Barbera, L. Gimenez, L. Luque Sanchez, K. Pedersen, and P. Michaelsen. "Mobility Sensitivity Analysis for LTE-Advanced HetNet Deployments with Dual Connectivity". In: *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*. May 2015, pp. 1–5 (cit. on p. 42).

[109] M.-S. Woo, S.-M. Kim, S.-G. Min, and S.-E. Hong. "Micro mobility management for dual connectivity in LTE HetNets". In: *Communication Software and Networks (ICCSN), 2015 IEEE International Conference on*. June 2015, pp. 395–398 (cit. on p. 42).

[110] K. Pedersen, P. Michaelsen, C. Rosa, and S. Barbera. "Mobility enhancements for LTE-advanced multilayer networks with inter-site carrier aggregation". In: *Communications Magazine, IEEE* 51.5 (May 2013), pp. 64–71 (cit. on p. 42).

[111] 3GPP TS 36.213. *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures*. Tech. rep. June 2013, available at www.3gpp.org (cit. on p. 42).

[112] C. Castellanos, D. Villa, C. Rosa, K. Pedersen, F. Calabrese, P.-H. Michaelsen, and J. Michel. "Performance of Uplink Fractional Power Control in UTRAN LTE". In: *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*. May 2008, pp. 2517–2521 (cit. on p. 43).

[113] M. Boussif, et al. "Load adaptive power control in LTE Uplink". In: *Wireless Conference (EW), 2010 European*. Apr. 2010, pp. 288–293 (cit. on pp. 43, 50).

[114] H. Wang, C. Rosa, and K. Pedersen. "Performance of Uplink Carrier Aggregation in LTE-Advanced Systems". In: *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*. Sept. 2010, pp. 1–5 (cit. on p. 48).

[115] 3GPP TS 36.101. *Further advancements for E-UTRA physical layer aspects (Release 11)*. Tech. rep. Jan 2015, available at www.3gpp.org (cit. on p. 48).

[116] H. Wang, C. Rosa, and K. I. Pedersen. "Radio resource management for uplink carrier aggregation in LTE-Advanced". In: *EURASIP Journal on Wireless Communications and Networking* 2015.1 (2015), pp. 1–15 (cit. on pp. 48, 50, 51).

[117] 3GPP Technical Report (TR) 36.814. *Further advancements for E-UTRA physical layer aspects (Release 9)*. Tech. rep. March 2010, available at www.3gpp.org (cit. on p. 49).

[118] F. D. Calabrese, et al. "Adaptive Transmission Bandwidth Based Packet Scheduling for LTE Uplink". In: *Vehicular Technology Conference, IEEE 68th*. Sept. 2008, pp. 1–5 (cit. on pp. 50, 51).

[119] C. Rosa, et al. “Performance of Fast AMC in E-UTRAN Uplink”. In: *Communications, IEEE International Conference on.* May 2008, pp. 4973–4977 (cit. on p. 51).

[120] A. S. et al. “LTE and WiFi: Experiences with Quality and Consumption”. In: *Procedia Computer Science* 34 (2014), pp. 418–425 (cit. on p. 58).

[121] R. Alkhansa, H. Artail, and D. M. Gutierrez-Estevez. “LTE-WiFi Carrier Aggregation for Future 5G Systems: A Feasibility Study and Research Challenges”. In: *Procedia Computer Science* 34 (2014), pp. 133–140 (cit. on p. 59).

[122] Qualcomm Research. *Improving LTE and Wi-Fi integration with PDCP aggregation.* `https://www.qualcomm.com/invention/research/projects/lte-advanced/lte-wi-fi-interworking`. [Online; accessed 01-07-2015]. 2015 (cit. on p. 59).

[123] Alcatel-Lucent. *Wireless Unified Networks.* `https://www.alcatel-lucent.com/press/2015/alcatel-lucent-combines-best-wi-fir-and-lte-enhance-mobile-performance-and-offer-consistent-high`. [Online; accessed 01-07-2015]. 2015 (cit. on p. 59).

[124] *KT’s demonstrations of LTE-H and LTE-U.* Tech. rep. Korea Communication Review, April 2015 (cit. on p. 59).

[125] K. Evensen, et al. “A Network-Layer Proxy for Bandwidth Aggregation and Reduction of IP Packet Reordering”. In: *IEEE 34th Conference on Local Computer Networks.* Oct. 2009, pp. 585–592 (cit. on pp. 59, 63).

[126] S. A. Latif, M. H. Masud, F. Anwar, and M. K. Alam. “An Investigation of Scheduling and Packet Reordering Algorithms for Bandwidth Aggregation in Heterogeneous Wireless Networks”. In: *Middle-East Journal of Scientific Research* 16 (2013), pp. 1613–1623 (cit. on pp. 59, 64).

[127] W. Fu, X. Wang, and D. Agrawal. “Multi-subnets Selection and Rate Allocation in a Heterogeneous Wireless Network”. In: *Mobile Adhoc and Sensor Systems (MASS), 2010 IEEE 7th International Conference on.* Nov. 2010, pp. 715–720 (cit. on p. 59).

[128] L. Sun, et al. “Traffic Allocation Scheme with Cooperation of WWAN and WPAN”. In: *Communications Letters, IEEE* 14.6 (June 2010), pp. 551–553 (cit. on p. 60).

[129] C. Wang, H. Tian, and J. Miao. “Dynamic Traffic Allocation Scheme for Optimum Distribution in Heterogeneous Networks”. In: *Vehicular Technology Conference (VTC Fall), 2011 IEEE.* Sept. 2011, pp. 1–5 (cit. on p. 60).

[130] J. Wu, et al. “Distortion-Aware Concurrent Multipath Transfer for Mobile Video Streaming in Heterogeneous Wireless Networks”. In: *Mobile Computing, IEEE Transactions on* 14.4 (Apr. 2015), pp. 688–701 (cit. on p. 60).

[131] X. Zhu, et al. “Rate Allocation for Multi-user Video Streaming over Heterogenous Access Networks”. In: *Proceedings of the 15th International Conference on Multimedia*. ACM, 2007, pp. 37–46 (cit. on p. 60).

[132] V. B. Iversen. *Teletraffic Engineering. Chapter 10: Applied Queueing Theory.* Technical University of Denmark, 2013 (cit. on pp. 61, 62).

[133] D. Kaspar, K. Evensen, A. F. Hansen, P. Engelstad, P. Halvorsen, and C. Griwodz. “An analysis of the heterogeneity and IP packet reordering over multiple wireless networks”. In: *Computers and Communications, 2009. ISCC 2009. IEEE Symposium on.* July 2009, pp. 637–642 (cit. on p. 63).

[134] T. K. H. Guan and P. Merz. “Discovery of Cloud-RAN”. In: *Proc. of NSN Cloud-RAN Workshop.* Apr. 2010 (cit. on p. 74).

[135] *C-RAN The Road Towards Green RAN.* Tech. rep. China Mobile Research Institute, 2011 (cit. on pp. 74, 84).

[136] J. Wu, Z. Zhang, Y. Hong, and Y. Wen. “Cloud radio access network (C-RAN): A primer”. In: *Network, IEEE* 29.1 (Jan. 2015), pp. 35–41 (cit. on p. 74).

[137] D. Wubben, P. Rost, J. Bartelt, M. Lalam, V. Savin, M. Gorgoglione, A. Dekorsy, and G. Fettweis. “Benefits and Impact of Cloud Computing on 5G Signal Processing: Flexible centralization through cloud-RAN”. In: *Signal Processing Magazine, IEEE* 31.6 (Nov. 2014), pp. 35–44 (cit. on p. 74).

[138] T. Werthmann, H. Grob-Lipski, and M. Proebster. “Multiplexing gains achieved in pools of baseband computation units in 4G cellular networks”. In: *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on.* Sept. 2013, pp. 3328–3333 (cit. on pp. 74, 75, 77).

[139] S. Bhaumik, S. P. Chandrabose, M. K. Jataprolu, A. Muralidhar, V. Srinivasan, G. Kumar, P. Polakos, and T. Woo. “CloudIQ: A framework for processing base stations in a data center”. In: *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM* (2012), pp. 125–136 (cit. on pp. 74, 76).

[140] S. Namba, T. Matsunaka, T. Warabino, S. Kaneko, and Y. Kishi. “Colony-RAN architecture for future cellular network”. In: *Future Network Mobile Summit (FutureNetw), 2012.* July 2012, pp. 1–8 (cit. on pp. 74, 76).

[141] M. Madhavan, P. Gupta, and M. Chetlur. “Quantifying multiplexing gains in a Wireless Network Cloud”. In: *Communications (ICC), 2012 IEEE International Conference on*. 2012, pp. 3212–3216 (cit. on pp. 74, 76, 77).

[142] A. Checko, A. Checko, H. Holm, and H. Christiansen. “Optimizing Small Cell Deployment by the Use of C-RANs”. In: *Proceedings of 20th European Wireless Conference*. May 2014, pp. 1–6 (cit. on pp. 74, 76, 85, 87).

[143] C. Liu, K. Sundaresan, M. Jiang, S. Rangarajan, and G.-K. Chang. “The case for re-configurable backhaul in cloud-RAN based small cell networks”. In: *IEEE INFOCOM*. Apr. 2013, pp. 1124–1132 (cit. on pp. 74, 76).

[144] S. Namba, T. Warabino, and S. Kaneko. “BBU-RRH Switching Schemes for Centralized RAN”. In: *7th International ICST Conference on Communications and Networking in China*. Aug. 2012, pp. 762–766 (cit. on pp. 74, 76).

[145] M. Peng, Y. Li, J. Jiang, J. Li, and C. Wang. “Heterogeneous cloud radio access networks: a new perspective for enhancing spectral and energy efficiencies”. In: *Wireless Communications, IEEE* 21.6 (Dec. 2014), pp. 126–135 (cit. on p. 75).

[146] M. Gerasimenko, D. Moltchanov, R. Florea, S. Andreev, Y. Koucheryavy, N. Himayat, S.-P. Yeh, and S. Talwar. “Cooperative Radio Resource Management in Heterogeneous Cloud Radio Access Networks”. In: *Access, IEEE* 3 (2015), pp. 397–406 (cit. on p. 75).

[147] H. Dahrouj, A. Douik, and O. Dhifallah. “Resource allocation in heterogeneous cloud radio access networks: advances and challenges”. In: *Wireless Communications, IEEE* 22.3 (June 2015), pp. 66–73 (cit. on p. 75).

[148] V. B. Iversen. *Teletraffic Engineering. Chapter 7: Multi-dimensional loss systems*. Technical University of Denmark, 2013 (cit. on p. 76).

[149] M. Stasiak, M. Głąbowski, A. Wiśniewski, and P. Zwierzykowski. *Modeling and dimensioning of mobile networks : from GSM to LTE*. John Wiley & Sons Ltd., 2011 (cit. on pp. 76, 81).

[150] V. B. Iversen, V. Benetis, and P. D. Hansen. *Wireless Systems and Mobility in Next Generation Internet: First International Workshop of the EURO-NGI Network of Excellence, Dagstuhl Castle, Germany, June 7-9, 2004. Revised Selected Papers*. Ed. by G. Kotsis and O. Spaniol. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. Chap. Performance of Hierarchical Cellular Networks with Overlapping Cells, pp. 7–19 (cit. on pp. 76, 78).

[151] V. B. Iversen. “The Exact Evaluation of Multi-Service Loss Systems with Access Control”. In: *Teleteknik, English ed.* 31.1 (Firstquarter 1987), pp. 56–61 (cit. on p. 78).

[152] M. Y. Arslan, K. Sundaresan, and S. Rangarajan. “Software-defined networking in cellular radio access networks: potential and challenges”. In: *IEEE Communications Magazine* 53.1 (Jan. 2015), pp. 150–156. ISSN: 0163-6804. DOI: `10.1109/MCOM.2015.7010528` (cit. on p. 87).

[153] I. Da Silva, G. Mildh, J. Rune, P. Wallentin, J. Vikberg, P. Schliwa-Bertling, and R. Fan. “Tight Integration of New 5G Air Interface and LTE to Fulfill 5G Requirements”. In: *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*. May 2015, pp. 1–5 (cit. on p. 91).

[154] B. Haberland. *Cloud RAN architecture evolution from 4G to 5G Mobile Systems*. 2014 (cit. on p. 92).

[155] B. Mondal, E. Visotsky, T. Thomas, X. Wang, and A. Ghosh. “Performance of downlink comp in LTE under practical constraints”. In: *Personal Indoor and Mobile Radio Communications (PIMRC), 2012 IEEE 23rd International Symposium on*. Sept. 2012, pp. 2049–2054 (cit. on p. 95).

[156] D.-E. Meddour, T. Rasheed, and Y. Gourhant. “On the Role of Infrastructure sharing for Mobile Network Operators in Emerging Markets”. In: *Computer Networks* 55.7 (2011), pp. 1576–1591 (cit. on p. 102).

[157] M. Marotta, N. Kaminski, I. Gomez-Miguelez, L. Zambenedetti Granville, J. Rochol, L. DaSilva, and C. Both. “Resource sharing in heterogeneous cloud radio access networks”. In: *Wireless Communications, IEEE* 22.3 (June 2015), pp. 74–82 (cit. on p. 103).

[158] *RAN Sharing: NEC’s Approach towards Active RAdio Access Network Sharing*. Tech. rep. NEC Corporation, 2013 (cit. on p. 103).

[159] N. Goods, B. Burtshcer, and C. Carnabuci. “Mobile Matters: A Network Shared”. In: (Aug. 2010). Freshfields Bruckhaus Deringer (cit. on p. 103).

[160] D. Leza. “Mobile Infrastructure Sharing: Trends in Latin America”. In: *ITU Regional Economic and Financial Forum of Telecommunications/ICTs for Latin America and the Caribbean* (Mar. 2014) (cit. on p. 103).

[161] M. Song, C. Xin, Y. Zhao, and X. Cheng. “Dynamic spectrum access: from cognitive radio to network radio”. In: *Wireless Communications, IEEE* 19.1 (Feb. 2012), pp. 23–29 (cit. on p. 103).

[162] H. Zamat and C. R. Nassar. “Introducing software defined radio to 4G wireless: Necessity, advantage, and impediment”. In: *Communications and Networks, Journal of* 4.4 (Dec. 2002), pp. 1–7 (cit. on p. 103).

[163] J. Liu, T. Zhao, S. Zhou, Y. Cheng, and Z. Niu. “CONCERT: a cloud-based architecture for next-generation cellular systems”. In: *IEEE Wireless Communications* 21.6 (Dec. 2014), pp. 14–22. ISSN: 1536-1284. DOI: `10.1109/MWC.2014.7000967` (cit. on p. 105).

[164] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan. “Radio access network virtualization for future mobile carrier networks”. In: *Communications Magazine, IEEE* 51.7 (July 2013), pp. 27–35 (cit. on p. 105).

[165] M. Yang, Y. Li, D. Jin, L. Su, S. Ma, and L. Zeng. “OpenRAN: A Software-defined RAN Architecture via Virtualization”. In: *Proceedings of the ACM 2013 Conference on SIGCOMM*. 2013, pp. 549–550 (cit. on p. 105).

[166] A. Gudipati, D. Perry, L. E. Li, and S. Katti. “SoftRAN: Software Defined Radio Access Network”. In: *Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking*. HotSDN ’13. Hong Kong, China: ACM, 2013, pp. 25–30. ISBN: 9781450321785 (cit. on pp. 105, 106).

[167] M. Artuso, C. Caba, H. L. Christiansen, and J. Soler. “Towards Flexbile SDN-based Management for Cloud-Based Mobile Networks”. In: *IEEE/IFIP Network Operations and Management Symposium*. accepted for publication. 2015 (cit. on p. 106).

[168] *The Benefits of Cloud-RAN Architecture in Mobile Network Expansion*. Tech. rep. Fujitsu, 2014 (cit. on p. 106).

[169] W. Kiess, P. Weitkemper, and A. Khan. “Base station virtualization for OFDM air interfaces with strict isolation”. In: *IEEE International Conference on Communications Workshops*. June 2013, pp. 756–760 (cit. on p. 106).

[170] A. Gudipati, L. E. Li, and S. Katti. “RadioVisor: A Slicing Plane for Radio Access Networks”. In: *Proceedings of the Third Workshop on Hot Topics in Software Defined Networking*. Chicago, Illinois, USA, 2014, pp. 237–238 (cit. on p. 106).

[171] A. Zakrzewska and V. Iversen. “Resource Sharing in Heterogeneous and Cloud Radio Access Networks”. In: *4th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops*. Oct. 2012, pp. 34–39 (cit. on p. 108).

[172] M. Bennis and J. Lilleberg. “Inter Base Station Resource Sharing and Improving the Overall Efficiency of B3G Systems”. In: *IEEE 66th Vehicular Technology Conference*. Sept. 2007, pp. 1494–1498 (cit. on p. 108).

[173] H. Kamal, M. Coupechoux, and P. Godlewski. “Inter-Operator Spectrum Sharing for Cellular Networks using Game Theory”. In: *IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*. Sept. 2009, pp. 425–429 (cit. on p. 108).

[174] Y.-T. Lin, H. Tembine, and K.-C. Chen. "Inter-Operator Spectrum Sharing in Future Cellular Systems". In: *IEEE Global Communications Conference (GLOBECOM)*. Dec. 2012, pp. 2597–2602 (cit. on p. 108).

[175] B. Leng, P. Mansourifard, and B. Krishnamachari. "Microeconomic Analysis of Base-Station Sharing in Green Cellular Networks". In: *Proceedings IEEE INFOCOM*. Apr. 2014, pp. 1132–1140 (cit. on p. 108).

[176] I. Malanchini, S. Valentin, and O. Aydin. "Generalized Resource Sharing for Multiple Operators in Cellular Wireless Networks". In: *International Wireless Communications and Mobile Computing Conference*. Aug. 2014, pp. 803–808 (cit. on p. 108).

[177] J. Panchal, R. Yates, and M. Buddhikot. "Mobile Network Resource Sharing Options: Performance Comparisons". In: *IEEE Transactions on Wireless Communications* 12.9 (Sept. 2013), pp. 4470–4482 (cit. on p. 108).

[178] R. Kokku et al. "CellSlice: Cellular Wireless Resource Slicing for Active RAN Sharing". In: *Fifth International Conference on Communication Systems and Networks*. Jan. 2013, pp. 1–10 (cit. on p. 109).

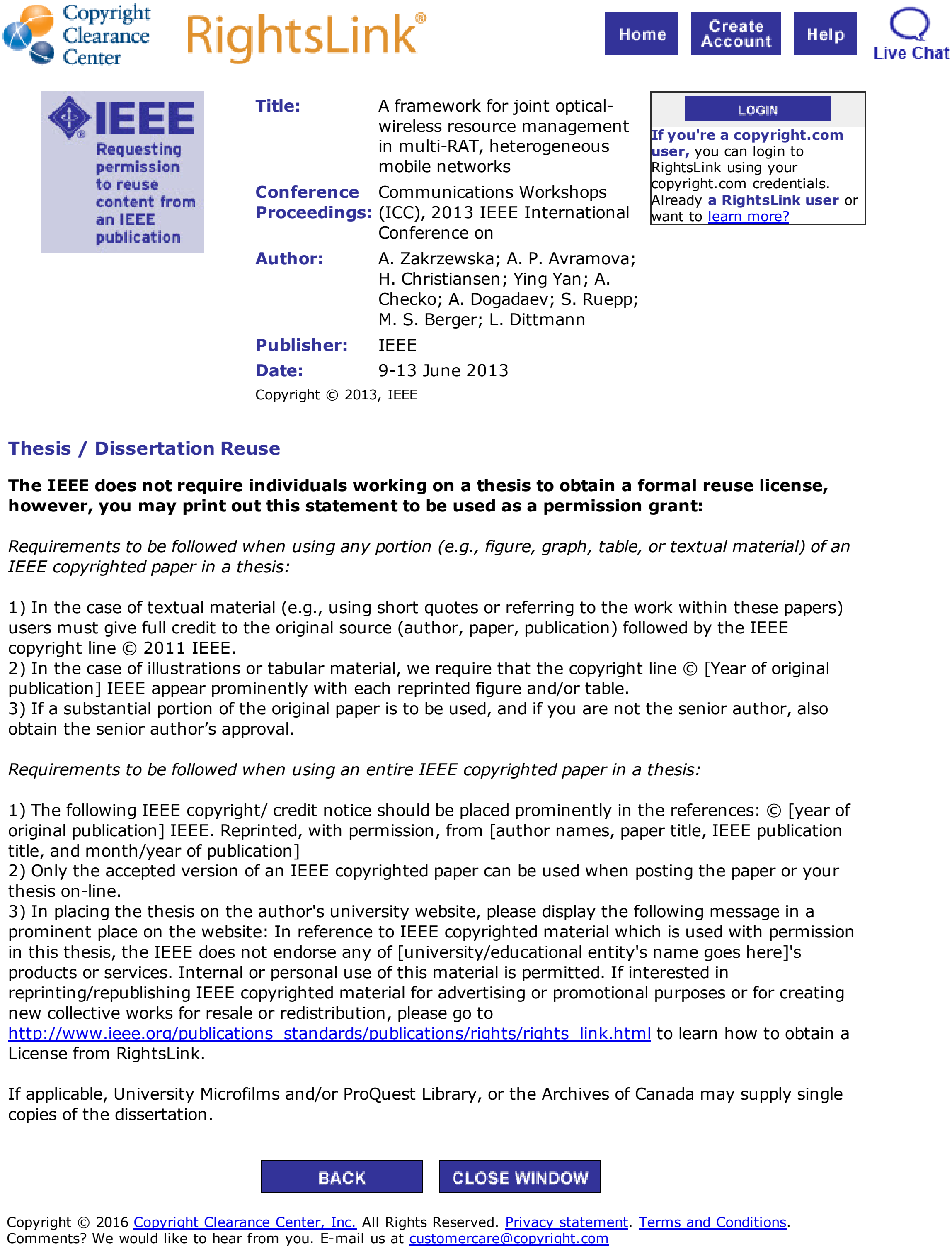
[179] T. Guo and R. Arnott. "Active LTE RAN Sharing with Partial Resource Reservation". In: *IEEE 78th Vehicular Technology Conference*. Sept. 2013, pp. 1–5 (cit. on p. 109).

[180] R. Mahindra et al. "Radio Access Network sharing in Cellular Networks". In: *21st IEEE International Conference on Network Protocols*. Oct. 2013, pp. 1–10 (cit. on p. 109).

[181] V. B. Iversen. "The Exact Evaluation of Multi-Service Loss Systems with Access Control". In: *Teleteknik, English ed.* 31 (1987), pp. 56–61 (cit. on p. 110).

[182] Y. Niu, D. Li Yong and Jin, L. Su, and A. V. Vasilakos. "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges". In: *Wireless Networks* 21.8 (2015), pp. 2657–2676 (cit. on p. 121).

[183] M. Ayyash, H. Elgala, A. Khreishah, V. Jungnickel, T. Little, S. Shao, M. Rahaim, D. Schulz, J. Hilt, and R. Freund. "Coexistence of WiFi and LiFi toward 5G: concepts, opportunities, and challenges". In: *IEEE Communications Magazine* 54.2 (Feb. 2016), pp. 64–71 (cit. on p. 121).



Title: A framework for joint optical-wireless resource management in multi-RAT, heterogeneous mobile networks

Conference Proceedings: Communications Workshops (ICC), 2013 IEEE International Conference on

Author: A. Zakrzewska; A. P. Avramova; H. Christiansen; Ying Yan; A. Checko; A. Dogadaev; S. Ruepp; M. S. Berger; L. Dittmann

Publisher: IEEE

Date: 9-13 June 2013

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already **a RightsLink user** or want to learn more?

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

IEEE
Requesting permission to reuse content from an IEEE publication

Title:	Optimal Traffic Allocation for Multi-Stream Aggregation in Heterogeneous Networks
Conference Proceedings:	2015 IEEE Globecom Workshops (GC Wkshps)
Author:	A. P. Avramova; V. B. Iversen
Publisher:	IEEE
Date:	6-10 Dec. 2015

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

IEEE Requesting permission to reuse content from an IEEE publication

Title:	Radio access sharing strategies for multiple operators in cellular networks
Conference Proceedings:	Communication Workshop (ICCW), 2015 IEEE International Conference on
Author:	A. P. Avramova; V. B. Iversen
Publisher:	IEEE
Date:	8-12 June 2015

LOGIN

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already **a RightsLink user** or want to learn more?

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK | CLOSE WINDOW

Copyright Release

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission to reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.