Load-Based Traffic Steering in heterogeneous LTE Networks

A Journey from Release 8 to Release 12

Fotiadis, Panagiotis

Publication date:
2014

Document Version
Accepted author manuscript, peer reviewed version

Load-Based Traffic Steering in Heterogeneous LTE Networks

– A Journey From Release 8 to Release 12 –

PhD Thesis

by

Panagiotis Fotiadis

A dissertation submitted to
Department of Electronic Systems
the Faculty of Engineering and Science, Aalborg University
in partial fulfillment for the degree of
PhD Degree,
Aalborg, Denmark
August 2014

Supervisors:
Preben E. Mogensen, PhD,
Professor, Aalborg University, Denmark
Principal Engineer, Nokia Networks, Aalborg, Denmark
Klaus I. Pedersen, PhD,
Professor, Aalborg University, Denmark
Senior Wireless Network Specialist, Nokia Networks, Aalborg, Denmark
Opponents:
Tatiana K. Madsen, PhD,
Associate Professor, Aalborg University, Denmark
Juan Ramiro Moreno, PhD,
Director of Technology Strategy, Ericsson, Madrid, Spain
Raquel Barco, PhD,
Associate Professor, University of Malaga, Spain

List of published papers:

- P. Fotiadis, Michele Polignano, D. Laselva, B. Vejlgaard, P. Mogensen, R. Irmer and N. Scully, "Multi-Layer Mobility Load Balancing in a Heterogeneous LTE Network," *Vehicular Technology Conference (VTC Fall), 2012 IEEE*, pp. 1-5, September 2012.
- P. Fotiadis, M. Polignano, L. Chavarria, I. Viering, C. Sartori, A. Lobinger and K. Pedersen, "Multi-Layer Traffic Steering: RRC Idle Absolute Priorities & Potential Enhancements," *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, pp. 1-5, June 2013.
- P. Fotiadis, Michele Polignano, K. I. Pedersen and P. Mogensen, "Load-Based Traffic Steering in Multi-Layer Scenarios: Case with & without Carrier Aggregation," *Wireless Communications and Networking, 2014 IEEE*, pp.1-5, April 2014.
- P. Fotiadis, I. Viering, K. I. Pedersen and P. Zanier, "Abstract Radio Resource Management Framework for System Level Simulations in LTE-A Systems," *Vehicular Technology Conference, 2014 IEEE*, pp.1-5, May 2014.
- P. Fotiadis, Michele Polignano, I. Viering and P. Zanier, "On the Potentials of Traffic Steering in HetNet Deployments with Carrier Aggregation," *Vehicular Technology Conference, 2014 IEEE*, pp.1-5, May 2014.

Abstract

The explosion of mobile broadband compels operators to migrate towards multilayer deployments, denoted as Heterogeneous Network (HetNet), consisting of different cell types and carrier frequencies. In this context, functionalities such as load balancing and mobility management have to be properly engineered so that the HetNet benefits are fully exploited. The dissertation aims at developing automated solutions for distributing the traffic across the deployed network layers. This function is referred to as load-based Traffic Steering (TS) and utilizes mobility procedures for steering devices to the least loaded cell. The alignment of TS decisions in both idle and connected mode is of key importance in order to achieve dynamic load balancing with a reasonable signaling cost. To further analyze the impact of different LTE releases on load balancing, cases with Carrier Aggregation (CA) are considered as well. Release 10 intra-eNB CA allows users to maintain connectivity to multiple macro carriers, while Dual Connectivity (DC) expands the concept to have it working between macrocells and picocells deployed at different frequencies. Among the serving cells, the Primary Cell (PCell) is responsible for all higher layer processes such as mobility support and connection maintenance, while the remaining serving cells are referred to as Secondary Cells (SCell).

For each investigated topic, appropriate algorithms are designed and their performance is evaluated by means of extensive system level simulations. These are conducted in 3GPP-defined scenarios, including widely accepted stochastic radio propagation models, explicit representation of mobility procedures in both idle and connected mode, CA functionalities and abstract radio resource management models. For deployments where macrocells and picocells share the same frequency, load balancing is performed by means of Range Extension (RE) that virtually expands the picocell service area. The developed solution adjusts the picocell coverage subject to cell load and mobility performance observations. Compared to a fixed RE configuration, the same user satisfaction is achieved while Radio Link Failures (RLFs) can be reduced up to ∼30-50%. On the contrary, the challenge for dedicated carrier deployments is to discover Inter-Frequency (IF) cells without excessive physical layer measurement rates as they cost both in terminal power and perceived throughput. To ensure their

efficient utilization for load balancing purposes, inter-frequency measurements are explicitly requested whenever cell overload is detected. Based on the associated measurement reports, users are steered to less loaded cells by means of forced handovers. Moreover, idle mode parameters are adjusted according to the cell load conditions, so that the TS decisions in idle and connected mode are aligned. The designed framework does not compromise offloading to picocells as well as it guarantees low handover/cell reselection rates.

With CA, the simultaneous connectivity to multiple carriers offers opportunities to further perform load balancing via collaborative packet scheduling schemes. To provide a load–aware PCell management, the load metric used by TS is neatly modified so as to consider multi-carrier connectivity. By means of that, the TS algorithms developed for Release 8/9 LTE can be reused for balancing the load of multi-layer HetNet deployments supporting intra-eNB CA together with the scheduler. If DC is further enabled, load-based TS is only relevant for users with single-carrier connectivity capabilities. Nevertheless, it is of key importance to apply proper cell management policies for DC capable users. Particularly, it is proposed to relax the requirement of always maintaining the PCell of all DC users on the macro overlay by configuring nomadic slowly moving hotspot users with a small cell PCell. The associated simulation results have shown that the proposed method ensures a high utilization of the picocell layer while significant signaling gains can be achieved – without any capacity loss – if a suitable SCell management policy is applied that configures macro SCells only for the cell edge hotpot users.

Dansk Resumé[1]

Eksplosionen af mobilt bredbånd tvinger operatørerne til at migrere til flerlags implementeringer, betegnet Heterogene Netværk (HetNet), som består af forskellige celletyper og bærefrekvenser. I denne sammenhæng er funktioner såsom load balancing og mobility management vigtige, således at HetNet fordelene udnyttes fuldt ud. Afhandlingen sigter mod at udvikle automatiserede løsninger til fordeling af trafikken på tværs af de tilgængelige netværks lag. Denne funktion kaldes load-baserede Traffic Steering (TS) og udnytter mobilitets mekanismerne. Tilpasningen af TS beslutninger i både tomgang og forbundet tilstand er af afgørende betydning for at opnå dynamisk load balancing med en rimelig signalering omkostninger. For yderligere at analysere virkningen af forskellige LTE udgivelser på load balancing, er tilfælde med Carrier Aggregation (CA) også analysert. Release 10 intra-eNB CA giver brugerne mulighed for at bevare forbindelsen til flere makro bærefrekvenser, mens Dual Connectivity (DC) udvider begrebet til at få det til at virke mellem makrocelle- og picoceller som bruger forskellige frekvenser. Blandt de betjener celler, den Primary Cell (PCell) er ansvarlig for alle højere lag processer såsom mobilitet support og vedligeholdelse-forbindelse, mens de resterende betjener cellerne betegnes som Secondary Cells (SCell).

For hver undersøgt emne, er egnede algoritmer designet og deres fordele undersøgt ved hjælp af omfattende systemniveau simuleringer. Disse er udført i 3GPP-definerede scenarier, baseret på bredt accepterede stokastisk radioudbredelses modeller, eksplicit repræsentation af mobilitet procedurer i både tomgang og tilsluttet tilstand, og abstrakt radio ressource management-modeller. Til installationer hvor makrocelle- og picoceller deler samme frekvens, udføres load balancing ved hjælp af Range Extension (RE) for at udvide picocellens serviceområde. Den udviklede løsning justerer picocellen dækning, baseret på celle belastning og mobilitets observationer. Sammenlignet med en fast RE konfiguration kan samme brugertilfredshed opnås samtidig med at Radio Link Failures (RLFs) kan reduceres med op til ∼30-50%. Udfordringen for dedikerede carrier implementeringer er at opdage Inter-Frekvens (IF) celler uden overdreven fysisk

[1]The author would like to express his gratitude to Klaus I. Pedersen of Nokia, Aalborg, Denmark for translating and proofreading this section.

lag målinger, da disse koster både i terminal strømforbrug og data hastighed. For at sikre en effektiv udnyttelse af load balancing, er inter-frekvensmålinger udtrykkeligt anmodet om, når der registreres celle overbelastning. På baggrund af de tilknyttede måling rapporter, styres brugerne til mindre belastede celler ved hjælp af tvungne overdragelser. Desuden justeres inaktive tilstands parametre i henhold til celle belastningsforhold, således at trafikstyringsmæssige afgørelser i tomgang og tilsluttet tilstand er i overensstemmelse.

CA giver mulighed for yderligere at udføre load balancing via kollaborative schedulering af brugerne. Ved at bruge en belastnings afhængig PCell algoritme, hvor load metrikken til dette er modificeret til CA tilfældet, opnås der en simpel form for TS. Ved hjælp af dette, kan de trafikstyringsmæssige algoritmer udviklet til Udgivelse 8/9 LTE genanvendes til balancering af belastningen af multi-layer HetNet implementeringer understøtter intra-eNB CA. Hvis DC er yderligere aktiveret, så er load baseret trafik styring kun relevant for brugere der er serviceret på en enkelt bærefrekvens. Ikke desto mindre, er det af afgørende betydning at anvende passende celle tildelings algoritmer for brugere der understøtter DC. Især foreslås det at lempe kravet om altid at opretholde PCell af alle DC brugere på makro overlay ved at konfigurere nomadiske langsomt bevægende hotspot-brugere med picocellen som PCell. De tilhørende resultater har vist, at den foreslåede metode sikrer en høj udnyttelse af picocelle laget mens betydelige signalerings gevinster kan opnås - uden nogen kapacitet tab - hvis en passende SCell management politik anvendes der konfigurerer makro SCells kun for cellekant hotpot brugere.

Preface & Acknowledgments

This PhD dissertation is the result of a three-year research project carried out at the Radio Access Technology (RATE) Section, Department of Electronic Systems, Aalborg University (AAU), Denmark. The work was conducted alongside with the obligatory courses and teaching obligations required for attaining the PhD degree. The research was supervised by Professor Preben E. Mogensen (Aalborg University, Nokia, Aalborg, Denmark) and Senior Wireless Network Specialist Dr. Klaus I. Pedersen (Aalborg University, Nokia, Aalborg, Denmark). The project was co-funded by the Faculty of Engineering and Science, Aalborg University and Nokia, Aalborg, Denmark.

The high level target of this project is to provide solid guidelines for load balancing in mature LTE deployments, where multiple frequency layers are deployed and carrier aggregation might be supported as well. Particularly, the thesis investigates different load balancing solutions for HetNet LTE deployments from Release 8 to Release 12. These include the usage of mobility procedures for steering traffic from one cell to another together with collaborative packet scheduling schemes being able to achieve a more efficient utilization of the deployed spectrum. The latter is enabled by the emergence of carrier aggregation which allows users to concurrently receive data from more than a single cell. The reader is expected to have fundamental knowledge of the mobility and radio resource management framework governing LTE/LTE Advanced systems.

First and foremost, I would like to express my sincere gratefulness to my supervisors for their continuous support and the countless discussions we had throughout the study. Their guidance together with the trust they showed in my skills are highly appreciated. I extend this gratitude also to Wireless Networks Specialist Daniela Laselva for her guidance during the first year of this research project.

I really feel fortunate for being part of Nokia and RATE section. Both groups constitute an ideal working environment characterized by passionate researchers and joyful people. Specifically, I would like to thank Michele Polignano, Simone Barbera and Lucas Chavarria Jimenez from AAU for embracing the simulator work. Furthermore, I want to show my gratitude to Mads Brix and Per Henrik

Michaelsen from Nokia for their support with regards to code development. Dr. Ingo Viering, Cinzia Sartori and Dr. Paolo Zanier from Nokia, Munich, Germany deserve my gratefulness for their valuable feedback throughout my studies. In addition to this, I really appreciate the assistance of Lisbeth, Jytte, Eva and Dorthe for handling all bureaucratic issues; a fact that essentially allowed me to stay 100% focused on my research. Ultimately, I would like to thank Davide Catania with whom I have closely worked since 2008 and was the first person that I met in Denmark.

I am truly thankful to the Greek and Serbian community that lives in Aalborg and helped me feel like home whenever I had the need of such feeling. And lately to Stefania, whose affection made me strong during the difficult periods. Last but not least, I am deeply grateful to my parents and my beloved sister. Your love has been a mandatory fuel for running this work and the thesis is fully dedicated to you.

Panagiotis Fotiadis,

Aalborg, July 2014

Contents

Chapter 1

Introduction

1.1 Preliminaries

As the number of mobile subscriptions approaches the global population, the end-user demand for higher bandwidth is constantly increasing [1, 2]. It is an undeniable fact that mobile broadband has deeply penetrated into our daily life due to the "always-on" experience that it offers. In contrast to the Second Generation (2G) Global System for Mobile Communications (GSM) [3], which was designed for delivering voice services, modern cellular networks provide significantly larger transmission bandwidths along with a broad range of attractive data applications fueled by the usage of tablets and smartphones. Indoubtfully, the massive deployment of Third Generation (3G) cellular technologies, such as High Speed Packet Access (HSPA) and HSPA+ [4, 5], is one of the main drivers for migrating towards networks where the data traffic volume overwhelms voice service consumption.

Nevertheless, the global success of the aforementioned network deployments did not decelerate the need for designing evolved systems, capable of maintaining the mobile broadband evolution sustainable in the future. In fact, a novel cellular technology, denoted as Long Term Evolution (LTE), came as part of the Third Generation Partnership Project (3GPP) Release 8 standardization [6, 7]. Compared to its 3GPP 3G predecessors, LTE introduces an innovative radio interface based on Orthogonal Frequency Division Multiplexing (OFDM) [8], whilst supporting larger transmission bandwidths up to 20 MHz. However, the peak LTE data rate of 300 Mbps is still far away from the International Mobile Telecommunications Advanced (IMT-A) requirement of 1 Gbit/s for candidate Fourth Generation (4G) systems [9]. Building on the existing LTE standardization, Release 10 specifications introduce LTE-Advanced including several enhancements in order to fulfill the IMT-A targets.

Except for the connection speed amelioration, the emergence of more capable

devices is foreseen as a critical contributor to the mobile data increase. Tablets, laptops and smartphones exploit high rate applications such as high definition video content, gaming and video conference services multiplying the end-user generated traffic by tens of times compared to conventional telephony terminals [2, 10, 11]. Given that the penetration of advanced mobile devices is expected to rapidly increase in the following years, the imposed pressure on the network infrastructure to satisfy the capacity demand will become even more stressful. Fig. 1.1 illustrates the impact of the aforementioned factors on the mobile data consumption, as it is envisaged by Cisco. Particularly, mobile broadband is growing with an Compound Annual Growth Rate (CAGR) of 66%, resulting in a 13-fold increase by 2017, as compared to 2012.

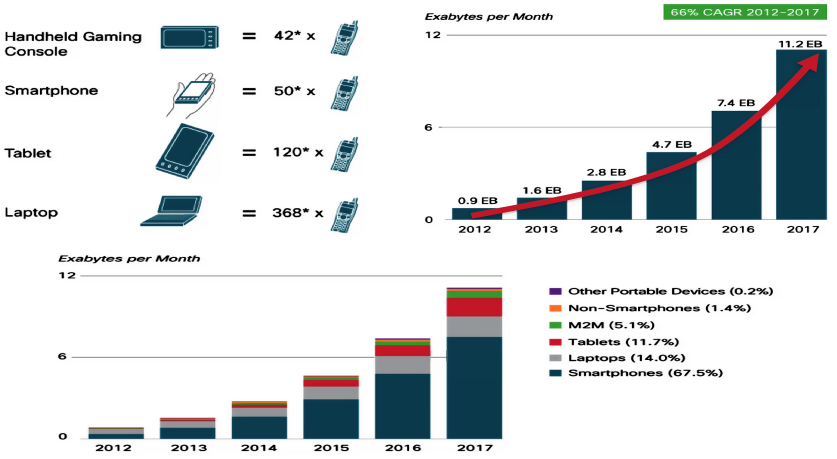


Fig. 1.1: The mobile broadband traffic explosion and the contribution of advanced mobile devices, as foreseen by Cisco in [2].

To cope with the exponential growth of mobile broadband, operators have to upgrade their networks in order to meet the forthcoming capacity demand. As shown in Fig. 1.2, this calls for the migration to hierarchical deployments, also denoted as Heterogeneous Network (HetNet), containing overlapping networks with divergent characteristics in terms of frequency bands, Radio Access Technology (RAT), cell sizes and backhaul support. However, the deployment cost along with the system optimization effort of such diverse and complex infrastructure might not be reimbursed by supplementary revenue. Flat rate pricing policies caused by the relentless market competition have decreased the average income per subscriber. To reduce operation costs by introducing autonomous system optimization functions, the concept of Self Organizing Networks (SON) has been singled out – by both academia and industry collaborations – as a key system design requirement. A set of different SON use cases can be

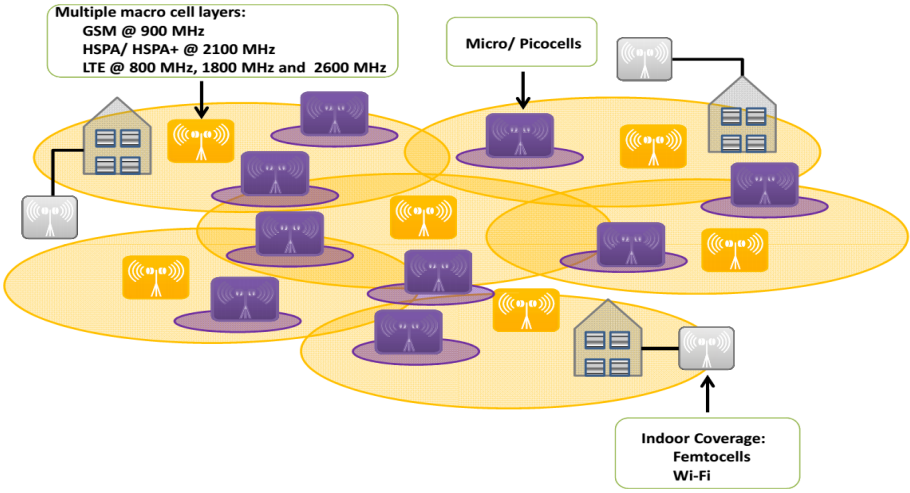


Fig. 1.2: Example of heterogeneous deployment. Macro cells are supplemented with outdoor low-power small cells for enhancing both outdoor coverage and capacity, while indoor traffic is primarily served by femtocells and Wi-Fi [12].

found in [13–15], published by 3GPP and the Next Generation Mobile Networks (NGMN) Alliance, respectively. Among others, traffic steering is defined as the ability to control and direct users to the best suitable cell and distribute traffic among the different layers [16]. To achieve this goal, factors such as User Equipment (UE) capabilities, power consumption, cell load, terminal velocity and backhaul capacity can be utilized for performing adaptive traffic steering subject to the desired network operator policy.

This thesis focuses primarily on the development of load-based traffic steering solutions for HetNet LTE deployments that find an attractive trade-off between gain and complexity. The analysis is conducted on the bases of different use cases, including the explicit modeling of mobility management procedures, analytical Radio Resource Management (RRM) models and advanced features such as Carrier Aggregation [17, 18]. All in all, the main scope is to provide solid recommendations for an autonomous load balancing framework subject to realistic network constraints like device capabilities, physical layer measurement availabilities, mobility performance, signaling overhead and UE power consumption.

The remainder of the chapter is structured as follows. Section 1.2 outlines the HetNet evolution addressing several deployment aspects for handling the foreseen increase of mobile broadband traffic. The key use cases of SON are discussed in Section 1.4, while Section 1.5 describes the major objectives of this thesis dissertation. The main research methodology is outlined in Section 1.6, followed by a list of contributions in Section 1.7. Finally, Section 1.8

concludes the chapter by presenting the overall thesis structure.

1.2 The Heterogeneous Network Evolution

To meet the ever growing traffic demand, operators can rely on a wide range of access technologies and base station types, jointly operating for ubiquitous communication and QoS experience. LTE networks – initially single carrier and later on multi-carrier at different frequency bands – will overlay the legacy 2G/3G infrastructure enhancing system capacity and mobile broadband coverage. Complementary upgrades such as spectrum aggregation, higher order sectorization and spatial multiplexing techniques will also be necessary for further boosting performance at the macro layer.

Nonetheless, once the gains of the aforementioned enhancements saturate, the deployment of low-power small cells is envisaged to be the most appropriate solution for improving the spectral efficiency per area unit. Targeting on offloading a significant amount of traffic towards them, small cells will be widespread adopted, bringing the network closer to the end-user both in outdoor and indoor areas.

The co-existence of several RATs along with the large scale small cell deployment will result in a divergent cellular environment. Within the context of HetNets, this section is dedicated to the aforementioned deployment aspects.

1.2.1 The Multi-RAT Relevance

Albeit the emergence of LTE and its evolved releases, immediate transition to LTE-only networks is not expected to occur in the mid term for various practical reasons. Terminal penetrations and existing investments on 2G/3G systems will prolong the lifespan of the RATs prior to LTE.

Fig. 1.3 depicts the evolution of mobile subscriptions by technology, as these are foresighted in [19]. Although 2G subscriptions have been declining after 2012, GSM EDGE Radio Access Network (GERAN) will continue operating for service suitability purposes. Having gained significant momentum in recent years, Machine-to-Machine (M2M) type applications can utilize 2G as the mobile interface for the low bit rate communication between electronic devices. Additionally, such networks are still offering good voice coverage.

At the meantime, HSPA connections will hold almost 45% of the global market share by 2017 [19]. As the LTE penetration is foreseen to represent ∼10% of the worldwide subscriber base and not exceed ∼30% in mature regions such as North America and West Europe by the same year, LTE should be gradually

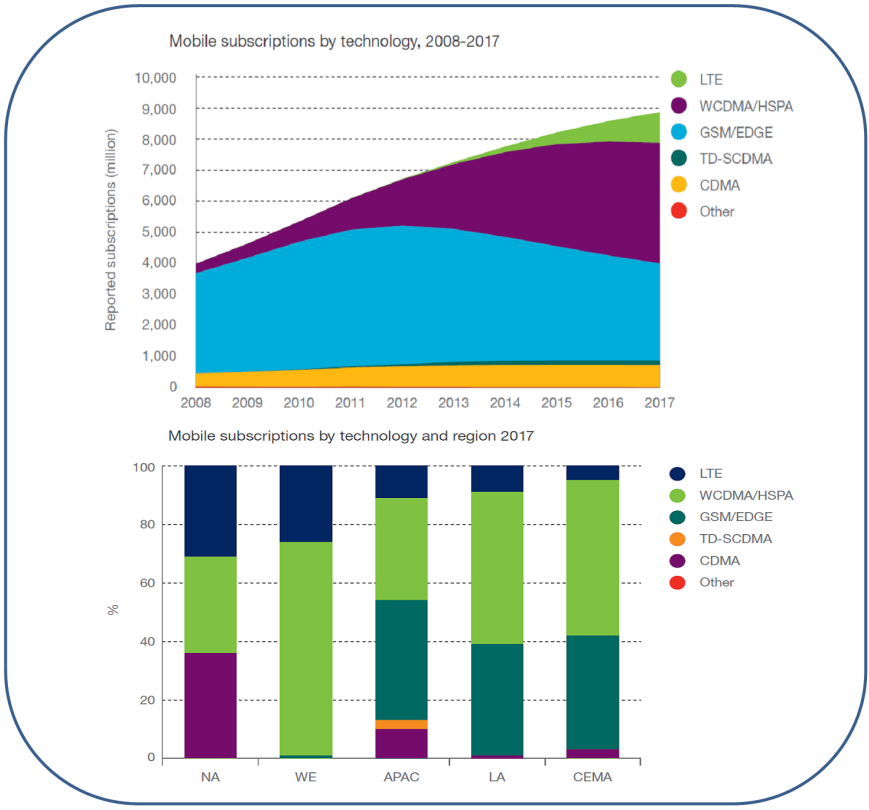


Fig. 1.3: Forecast of the mobile subscription evolution by technology, both regionally and globally. Subscriptions are defined by the most advanced RAT that the terminal and the network support. Source: [19].

deployed in order to be cost-effective. Early phase roll-outs comprise LTE as an overlay to the existing 2G/3G infrastructure providing broadband coverage in dense urban zones and rural areas with poor fixed-line connection by utilizing the digital dividend (800 MHz band) [20]. If spectrum available and terminal penetration allows it, additional LTE carriers at higher frequency bands should be added for further increasing capacity in traffic-critical network locations.

Upgrading the macro layout by means of reusing the existing site locations is a viable approach for enhancing system performance. Adding more carriers to existing base stations, cell-splitting via higher order sectorization [21] and antenna tilting optimizations [22] are cost-efficient solutions that do not require any site acquisition costs.

1.2.2 Small Cells Deployment Aspects

To further densify the network layout – e.g below 200-300 m – by solely deploying new macro sites would be rather impractical; especially in capacity-limited dense urban areas, where the aforementioned enhancements may not suffice and site acquisition can get prohibitively high [23]. To increase the base station density in a cost-effective manner, low-power small cells should be deployed. Being significantly less costly than the macro base stations, their large-scale adoption is envisioned to offload traffic from the macro overlay to the small cell layer. Typical deployment scenarios involve small cells to be installed on the street level in dense urban areas, enhance capacity in indoor regions with high user density (i.e. airports, shopping malls, etc.) and eliminate coverage holes in network locations with poor macro coverage [24].

1.2.2.1 Small Cell Classification

Depending on the base station technology, low-power nodes are usually classified into four categories, also denoted as microcells, picocells, femtocells and relays. Their characteristics are summarized in Table 1.1.

Table 1.1: Small Cell Classification

Base Station	Power	Cell Radius	Deployment	Backhaul
Micro	<36 dBm	100-300 m	Outdoor	Wired
Relay	30 dBm	<200 m	Outdoor	Wireless
Pico	>24 dBm	<200 m	Outdoor/indoor	Wired
Femto	<24 dBm	10-25 m	Indoor	Wired

Transmission power commonly differs subject to the use case, varying from 36 dBm to 30 dBm for outdoor deployments and not more than 23 dBm for indoor usage. In more detail, outdoor small cells are deployed by the operator in order to maximize their offloading potentials and mitigate interference. Unlike to micro/picocells, which employ ordinary backhauling via wireline solutions (i.e. leased lines, fiber, etc.), relays operate with a wireless backhaul link utilizing the radio interface resources. Comparative studies between picocells and relay performance are available in the literature, both in 3GPP defined scenarios [25] and site-specific irregular layouts [26]. The bottom line is that although relays can enhance network coverage by extending the macrocell umbrella, the wireless backhauling can be a significant capacity bottleneck, making picocells a more viable solution in dense high traffic areas. Finally, femtocells are user-deployed nodes designed for residential and enterprise areas, reusing the customer's wired broadband connection for providing the attachment with the core network. Being the primary competitor of Wi-Fi [12] for the dominant local

area solution, femtocells support 3GPP technologies, utilize licensed spectrum as well as they can maintain adequate interoperability with the outdoor cellular environment and support speech service.

1.2.2.2 Associated Challenges

Although HetNets will naturally increase the spectral efficiency, such a paradigm shift introduces a handful of challenges in terms of load balancing, interference mitigation and mobility management. A high level description of the related requirements when migrating towards HetNet scenarios is illustrated in Table 1.2.

Table 1.2: HetNet Paradigm Shift

Challenge	Macro-only scenarios	HetNet
Cell Association	Strongest cell	Cell providing the highest data rate
Interference	Few interfering nodes	Many interfering nodes
Mobility	Stable macrocell connection	Different cell sizes make mobility challenging

Associating users with the strongest cell typically results in a poor utilization of small cells. The reason is that the larger macrocell transmission power unavoidably shrinks the coverage footprint of low power nodes, leaving users in their vicinity to be served by the macro overlay. Although traditional cellular networks associate users to the strongest base station, several studies have shown that user association to the cell able to provide the highest data rate is the proper way forward for HetNets [27–29]. To achieve this goal, the service area of small cells is virtually expanded by means of biasing techniques that make handovers executions towards them seem more attractive.

Interference management tops the list as well, as HetNets include a larger number of interference sources. In addition to this, the fact that users no longer connect to the cell providing the best Signal-to-Interference plus Noise Ratio (SINR) makes interference mitigation even more important. Such functionality is rather crucial in co-channel deployments where macrocells and small cells share the same carrier frequency. For that purpose, both academia and industry have put significant effort in investigating solutions for resolving this problem, with enhanced Inter-Cell Interference Coordination (eICIC) [30, 31] being a well-known outcome of this joint work.

Finally, mobility for ensuring reliable support of mobile connections is of key importance for cellular networks. Commonly, mobility performance is evaluated

by various such as handover rates, ping-pong events and the probability of Radio Link Failure (RLF) and Handover Failure (HOF), respectively [32]. Nowadays, field results from commercial macrocell LTE networks illustrate an exceptional mobility performance with a low RLF probability [33]. However, this is not the case for small cells, where recent 3GPP studies [32] have shown that mobility performance degrades when migrating towards HetNet deployments. This mainly involves users moving at a medium-to-high velocity and one of the main reasons is that the receiving signal from small cells appears and disappears more frequently owing to their propagation properties. By means of that, time-accurate small cell related handovers – especially when the UE device leaves the low-power node – become a challenging issue.

1.3 Carrier Aggregation in Heterogeneous Networks

Given that the multi-carrier upgrades have taken place for capacity reasons, peak data rates can further improve by utilizing spectrum aggregation schemes introduced in Release 8 HSPA+ (Dual Cell HSPA [34]), and Release 10 LTE-Advanced specifications. Focusing on the LTE-Advanced use case, Carrier Aggregation (CA) allows for a transmission bandwidth larger than the 20 MHz bound of Release 8 LTE. In particular, resources from either contiguous or non-contiguous spectrum chunks – also denoted as Component Carrier (CC) – are aggregated, resulting up to a maximum bandwidth of 100 MHz. Being backward compatible with legacy devices, each CC reuses the typical Release 8 LTE numerology.

A CC, denoted as the Primary Cell (PCell), is responsible for all higher layer processes such as mobility support, RLF monitoring, connection maintenance, security, etc [17]. Handovers are solely executed at the PCell, while the remaining serving cells – each of them is also referred to as Secondary Cell (SCell) – are dynamically added, changed or removed subject to the SCell management policy. Figure 1.4 shows the fundamental CA scenarios specified in [7]. Intra-eNB CA allows the concurrent connectivity to multiple macrocell CCs, while inter-eNB CA expands the concept to have it working between small cells and the macro overlay. Being initially introduced for Remote Radio Head (RRH) (scenario 4) and frequency selective repeaters (scenario 5), Release 11 specifications support inter-eNB CA between macrocells and picocells as well. Such concept is generalized by Release 12, where the simultaneous connection to at least two network layers interconnected via the X2 interface is denoted as Dual Connectivity (DC) [35]. With DC, CA UE devices can maintain a stable PCell connection at the macro overlay, while SCells are configured on the small cell layer.

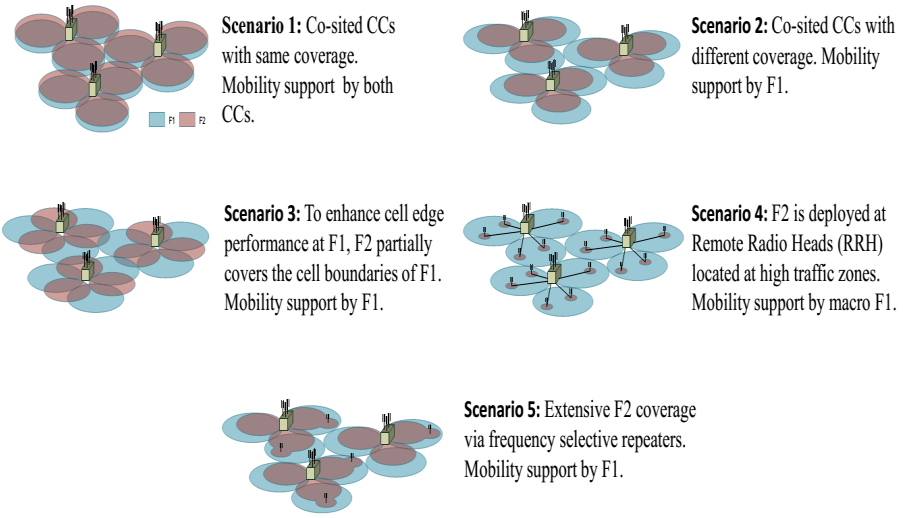


Fig. 1.4: Reference CA deployments as defined by 3GPP in [7].

1.4 Self Organizing Networks

Cost reduction and zero-touch network management are the main rationale for introducing the SON concept. Owing to the HetNet complexity, engineering functions need to be automatized so that the manual intervention is reduced to the minimum possible. By minimizing the human factor impact, procedures will become more robust to manual errors, accelerate deployment roll-out and essentially boost the overall performance by allowing fast re-configuration of the network parameters. The following subsections elucidate the SON framework by outlining the specified uses cases, discussing architectural aspects as well as explaining its distinction from RRM.

1.4.1 SON Use Cases

3GPP has worked on the standardization of several SON functions within the fields of self-configuration, self-optimization and self-healing (see Fig. 1.5). These can be shortly described as follows:

- *Self-Configuration*: It is the ability to bring a new network element into functional state with the minimum manual involvement [13]. I.e. once a new base station is deployed, the concerned SON mechanism should be responsible for transport link detection, connection setup with the core network and software upgrades [36]. Parameterization of the initial

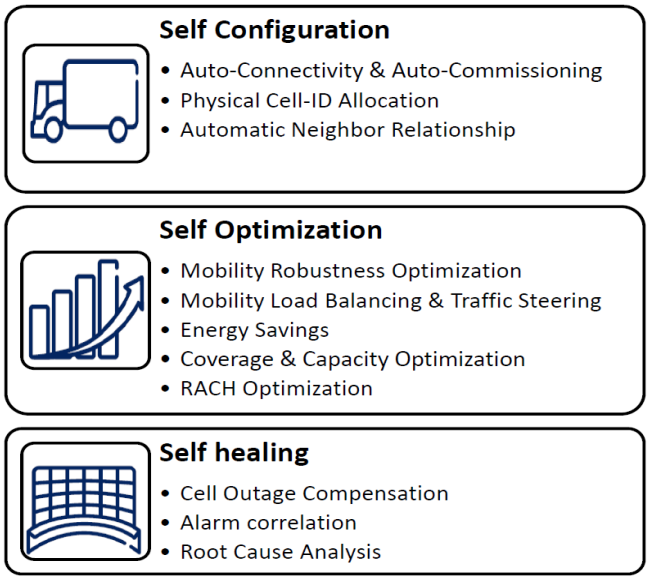


Fig. 1.5: SON use cases classification and examples of automated engineering functions.

transmission settings, dynamic Physical Cell Identity (PCI) selection and autonomous construction of the neighbor relationship list [37,38] are further included in the self-configuration process.

- *Self-Healing*: It aims at alleviating network failures by automatically activating the proper cell outage compensation algorithms. However, fault diagnosis might not be straightforward, since the detected symptom is usually associated with several failure causes. To complement the diagnosis process with proper correlation between symptoms and failure causes, artificial intelligence schemes can be utilized [39]. Given the reader's interest, more information can be found in [40,41].

- *Self-Optimization*: It refers to the group of functions that operate while the network is commercially active and adjust system parameters subject to the current network environment. The associated SON functions aim at enhancing network coverage [42], Random Access Channel (RACH) performance [43], provide network energy savings [44] as well as optimizing mobility performance and load balancing.

Focusing on traffic steering and Mobility Load Balancing (MLB), the goal is to optimally distribute traffic among neighboring cells. Adaptive tuning of mobility configurations and explicit handover executions based on cell load information are some of the basic tools for exploiting fairly utilized cells [45–49]. However, to efficiently exploit the common pool of network resources, factors such

as UE speed, requested service, backhaul capacity, UE power consumption and terminal capabilities can be considered as well.

Another important SON feature is Mobility Robustness Optimization (MRO), being responsible for ensuring autonomous mobility management. As HetNets involve a remarkable number of cells, the manual configuration of mobility parameters for each individual cell boundary will be rather impractical. To overcome configuration complexity while maintaining good mobility performance, MRO periodically adjusts mobility parameters on the basis of mobility performance indicators that are collected online [50, 51].

1.4.2 SON Architecture & Challenges

SON functions can be implemented either in a distributed or centralized fashion. However, employing the same architecture for all automation mechanisms would be suboptimal. The reason is that the design requirements and the operational time scale of each SON function may distinctively differ from one to another. This leads to a hybrid SON architecture that enables the simultaneous operation of both distributed and centralized schemes, as shown in Fig. 1.6.

In centralized architectures, the decision points are located at higher network management elements, being responsible for managing a large number of cells. SON servers collect the associated network measurements as well as informing the concerned cells about the parameters to be adjusted. In principle, centralized implementations can potentially achieve better performance than distributed solutions as they allow for global optimization. Nevertheless, sending information to central servers (*and vice versa*) can be rather costly, consuming network resources for signaling purposes.

On the other hand, distributed implementations limit the related signaling inbetween neighboring cells as the decision point is the base station itself. By means of that, SON decisions are taken based on information exchange that occurs over the X2 interface used for interconnecting adjacent cells. Apart from scaling better to a larger number of cells, such an approach further allows for a faster reaction to any network changes.

Albeit a hybrid architecture can better exploit the potentials of SON, it does not resolve the problem of functionality conflicts between different SON mechanisms. Such operational collisions might occur whenever two SON functions simultaneously attempt to adjust the same parameter in different directions or they mutually cancel out the performance gains made by each other. Although a simple full instances serialization could be easily realized, its static nature essentially limits the SON potentials [52]. To avoid such an undesirable effect, SON functions should be properly coordinated.

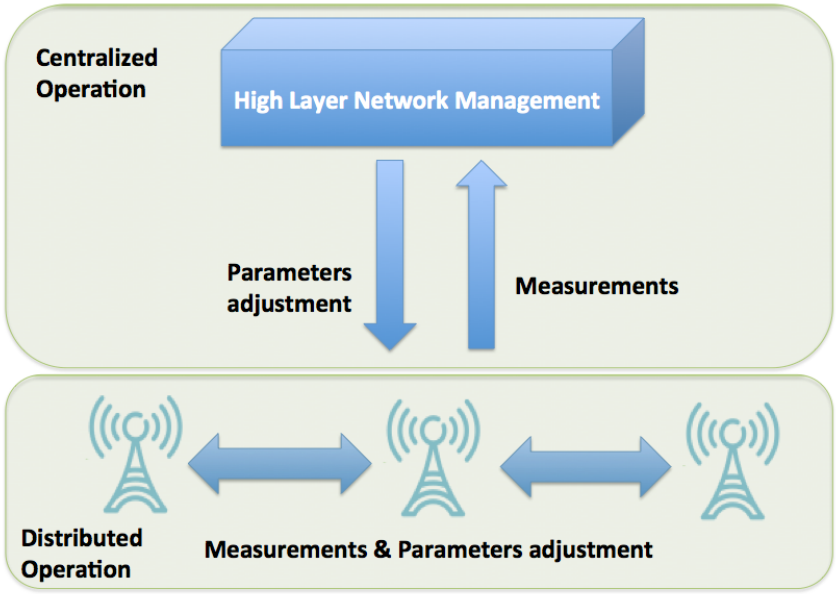


Fig. 1.6: Hybrid SON architecture. Simultaneous operation of centralized and distributed SON functions within the same network.

1.4.3 SON versus RRM

As both SON and RRM functions aim at optimizing system performance, a lot of confusion may occur whenever trying to classify an algorithm within the context of SON or RRM. To clarify this misunderstanding, this short subsection focuses on shedding some light on the distinctive difference between RRM and SON.

RRM includes all these procedures that are responsible for the adaptive allocation and the sharing of the radio resources. Among others, functionalities such as admission control, packet scheduling and mobility management are typical RRM paradigms. In particular, they are governed by specific parameters, according to which, dynamic decisions are taken in order to improve network performance.

On the other hand, SON algorithms operate in time periods longer than the millisecond basis and monitor network statistics. Based on these observations, they access RRM parameters and fine-tune them so that the related RRM functionalities adapt to the network environment. A typical example of such interaction between RRM and SON is whenever MRO adjusts mobility parameters for the sake of mobility robustness. Finally, SON may even reuse RRM functionalities without necessarily adjusting any related parameter. The explicit triggering of

a forced handover for load balancing purposes is another representative case of this interaction.

1.5 Thesis Scope

The general scope of this dissertation is to propose distributed traffic steering solutions that achieve dynamic load balancing in multi-layer HetNet deployments. To further investigate the impact of different LTE Releases on load balancing performance, intra/inter-eNB CA is enabled whenever applicable. Focusing solely on downlink, the developed self-optimizing schemes should boost network capacity by reacting autonomously to the cell load variations. Enablers such as terminal measurements, handoff procedures, information exchange between neighboring cells and packet scheduling functionalities could facilitate this purpose.

For scenarios with UE devices without multi-carrier connectivity capabilities, load balancing can be solely performed by means of mobility procedures. Thus, handovers in connected mode and cell reselections in idle mode must be employed so that traffic is pushed towards less loaded cells. The high level targets for studies conducted in network environments prior to Release 10 are outlined below:

- Develop schemes for co-channel and inter-frequency load balancing in HetNet deployments.
- Evaluate the potentials of idle and connected mode load balancing.
- Align the traffic steering decisions for idle UE devices with the ones taken in connected mode.
- Ensure that load balancing does not disrupt mobility robustness.

With CA, the concurrent connectivity to multiple carriers allows for the development of collaborative packet scheduling schemes that can provide a more efficient utilization of the system resources. For that purpose, the focus is put on the following research aspects whenever CA is considered:

- Understand how to perform load balancing by means of mobility procedures when this is further assisted by the scheduler.
- Identify suitable SCell management policies subject to the deployment type.
- Evaluate the viability of load-based traffic steering from Release 10 and hereinafter.

1.6 Research Methodology

Performing load balancing involves various network operations that mutually interact with each other. As mobility procedures in both idle and connected mode top this list, the derivation of a analytical load balancing solution – which is also capable of capturing mobility effects – becomes a rather challenging task. Thus, a heuristic approach is adopted in this dissertation, where appropriate schemes are designed for each studied topic and their performance is evaluated by means of extensive system level simulations.

Relying on simulations demands for a simulation tool that can generate reliable results. For that purpose, a significant amount of time was spent on calibrating and validating results based on others simulators used within Nokia/Aalborg University together with advising similar open literature material. Notice that the developed simulation tool is capable of reproducing the results from the HetNet mobility 3GPP studies presented in [32, 35]. This necessarily strengthens the reliability of our simulation tool.

Particularly, the simulation environment is aligned with the generic 3GPP guidelines for conducting system level simulations with small cells. The adopted network layout is the one defined in [53], being the most common baseline HetNet scenario used in open literature. Radio propagation lies on widely accepted stochastic models including the effect of distance-dependent pathloss, shadow fading, fast-fading, etc. Furthermore, physical layer abstraction models provide the necessary link-to-system level mapping so as to represent physical layer procedures in time periods longer than the millisecond basis with an attractive trade-off between computational complexity and accuracy. Nevertheless, it is worth mentioning that the reader should focus on the relative trends of the results depicted in this PhD dissertation and not on the absolute values, as the latter could naturally differ from those obtained in a real LTE system. More information about the underlying modeling framework can be found in Appendix A.

Finally, the conducted investigations involve long simulation times so as to ensure that convergence is achieved and the obtained results are statistically reliable. The performance evaluation relies on statistics collected after the convergence period. In more detail, these include several performance indicators that essentially reflect not only the capacity gains of the investigated algorithms but also the associated costs for achieving a particular performance.

1.7 List of Contributions

The following publications have been authored during the PhD study:

- P. Fotiadis, Michele Polignano, D. Laselva, B. Vejlgaard, P. Mogensen, R. Irmer and N. Scully, "Multi-Layer Mobility Load Balancing in a Heterogeneous LTE Network," *Vehicular Technology Conference (VTC Fall), 2012 IEEE*, pp. 1-5, September 2012.

- P. Fotiadis, M. Polignano, L. Chavarria, I. Viering, C. Sartori, A. Lobinger and K. Pedersen, "Multi-Layer Traffic Steering: RRC Idle Absolute Priorities & Potential Enhancements," *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, pp. 1-5, June 2013.

- P. Fotiadis, Michele Polignano, K. I. Pedersen and P. Mogensen, "Load-Based Traffic Steering in Multi-Layer Scenarios: Case with & without Carrier Aggregation," *Wireless Communications and Networking, 2014 IEEE*, pp.1-5, April 2014.

- P. Fotiadis, I. Viering, K. I. Pedersen and P. Zanier, "Abstract Radio Resource Management Framework for System Level Simulations in LTE-A Systems," *Vehicular Technology Conference, 2014 IEEE*, pp.1-5, May 2014.

- P. Fotiadis, Michele Polignano, I. Viering and P. Zanier, "On the Potentials of Traffic Steering in HetNet Deployments with Carrier Aggregation," *Vehicular Technology Conference, 2014 IEEE*, pp.1-5, May 2014.

In addition, one patent application has been submitted via the Nokia patent department. Several deliverables has been written and internally presented, while the study has also provided input on a research project jointly ran by Nokia and Vodafone. Moreover, some of the main project results have been disseminated into the following Nokia Network Solutions white paper:

- Nokia Solutions Networks, "Load balancing mobile broadband traffic in LTE HetNets," *White Paper*, October 2013.

Significant simulator development effort has been contributed as well. The dissertation in based on a single carrier LTE Release 8 simulator from Nokia, which has been neatly evolved so that it supports simulations in multi-layer HetNet deployments. This involves the addition of several features such as user mobility models, inter-frequency mobility support, CA modeling, particular packet scheduling schemes, etc. Furthermore, simulator maintenance in terms of regression testing has been conducted during the duration of this PhD study. Last but not least, it is worth mentioning that the current state of the simulator is widely used by Nokia colleagues for providing input on 3GPP work items as well as on studies within the context of SON and European Union research collaborative projects.

1.8 Thesis Outline

The thesis is divided into 7 chapters and 5 appendices. A brief overview of the following chapters is provided below:

- Chapter 2: *Setting the Scene* – This chapter sets the framework for the conducted investigations. Initially, the standardized mechanisms and interfaces for facilitating automated load balancing are presented. The design prerequisites of the developed solutions are thoroughly discussed and the main performance indicators are defined. Finally, an overview of the considered scenarios is provided.
- Chapter 3: *Co-channel Load Balancing in HetNet Deployments* – This chapter is dedicated to load balancing in co-channel HetNet deployments. A joint MLB/MRO framework is developed that adjusts handover offsets based on cell load and mobility observations. Its performance is compared against a range extension scheme that statically biases measurements in favor of the small cells.
- Chapter 4: *Inter-Frequency HetNet Load Balancing* – The problem of inter-frequency load balancing in HetNet deployments is addressed in this chapter. A traffic steering solution is developed that moves users to less loaded inter-frequency neighbors via handovers and cell reselections in connected and idle mode, respectively. The obtained results illustrate that the proposed scheme enhances network capacity while maintaining the associated signaling and UE power consumption costs relatively low.
- Chapter 5: *Load Balancing in HetNets with Intra-eNB CA* – The study expands to Release 10 LTE HetNets with intra-eNB CA support. The joint interaction of load-based traffic steering with CA is investigated for different CA UE penetrations and deployment scenarios.
- Chapter 6: *HetNet Load Balancing with Dual Connectivity* – Unlike to Chapter 5, DC in the form of inter-eNB CA between picocells and the macro overlay is now enabled. In particular, a cell management framework is proposed for managing the PCells and the SCells of DC users.
- Chapter 7: *Conclusions and future work* – Based on the overall study findings, recommendations for autonomous load balancing are given subject to the different LTE Releases. Ultimately, some topics for future work are discussed.

As for the attached appendices, these are outlined below:

- Appendix A: *Modeling Framework* – This appendix describes the modeling framework supporting the conducted studies. It covers several aspects

such as propagation modeling, SINR calculation, user scheduling assumptions, throughput estimation, etc.

- Appendix B: *MRO Reliability Analysis* – This appendix provides an approximation of the statistical sample required for achieving a reliable estimation of the RLF probability used for realizing MRO decisions.
- Appendix C: *Complementary Results for Chapter 4* – The study of inter-frequency load balancing conducted in Chapter 4 is complemented with some additional results. Emphasis is put on further enhancing the network capacity as well as estimating the UE power consumption gains achieved by the developed solution.
- Appendix D: *Emulating Different Scheduling Policies with CA* – In this appendix, an abstract RRM framework is presented for emulating packet scheduling in LTE-Advanced systems. The focus is put on the targeted fairness to be maintained between non-CA and CA users when allocating transmission resources. It is shown that the proposed model captures the main properties of different state-of-art schedulers without the need of detailed modeling at a subframe basis.
- Appendix E: *Complementary Results for Chapter 5* – It includes a paper reprint so as to give further insight on the benefits of the proposed framework for performing load-based traffic steering in HetNet deployments with intra-eNB CA.

Chapter 2

Setting the Scene

2.1 Introduction

The design requirements of traffic steering strategies depend on the network maturity in terms of terminal penetrations, spectrum availability, cell densification, etc. Following the trends of network evolution, traffic management policies have to migrate from simple static approaches to more intelligent schemes, fully functional within a multi-layer HetNet environment.

Typically, early-stage LTE roll-outs provide mobile broadband coverage only in strategically selected areas with high traffic density or poor 3G support. To take full advantage of the deployed LTE capacity, traffic steering mechanisms should ensure that LTE-capable devices are pushed to LTE whenever feasible. Switching back to the 3G overlay should solely occur either due to the lack of adequate LTE coverage or for service suitability purposes. Simple schemes based on UE/RAT capabilities [54] together with static configurations of mobility management functionalities can be employed for facilitating that purpose.

Nevertheless, the ever-growing demand for mobile broadband will force operators to acquire more spectrum so that additional carriers are deployed and the increasing capacity requirements are met. An illustrative spectrum allocation for a European operator being capable of investing in more LTE spectrum is shown in Table 2.1, including the option of refarming part of the GSM spectrum to HSPA and LTE [20,55]. Mobile broadband coverage is ensured by macrocells deployed at lower frequency bands, while more remarkable data rates are experienced at higher carrier frequencies, fully dedicated to small cells or shared with the macro overlay.

As the coverage area of these layers is commonly overlapped, there is a greater degree of freedom for operators to manage UE distributions so that they achieve an optimized utilization of the network resources. This involves traffic steering

Table 2.1: European spectrum availability and associated RAT

Carrier Frequency (MHz)	RAT	Service
800	LTE	Coverage
900	GSM ⇒ HSPA	Coverage
1800	GSM ⇒ LTE	Capacity
2100	HSPA	Capacity
2600	LTE	Capacity

decisions based on either a single goal or contain the combination of several targets subject to the desired operator policy. However, the envisioned deployment heterogeneity in terms of deployed RATs, cell sizes and propagation properties of different carrier frequencies makes such functionality a challenging task. Furthermore, it is of paramount importance that the same algorithm covers multiple scenario cases. As network upgrades take place – e.g. addition of a new network layer –, traffic steering should adapt to the deployment changes and eventually optimize system performance with the minimum manual intervention. In this context, the derivation of simplified traffic rules no longer suffices, postulating evolved solutions to be developed. Focusing on dynamic load balancing, this is essentially performed via distributed schemes that exploit mobility management functionalities for steering users from one

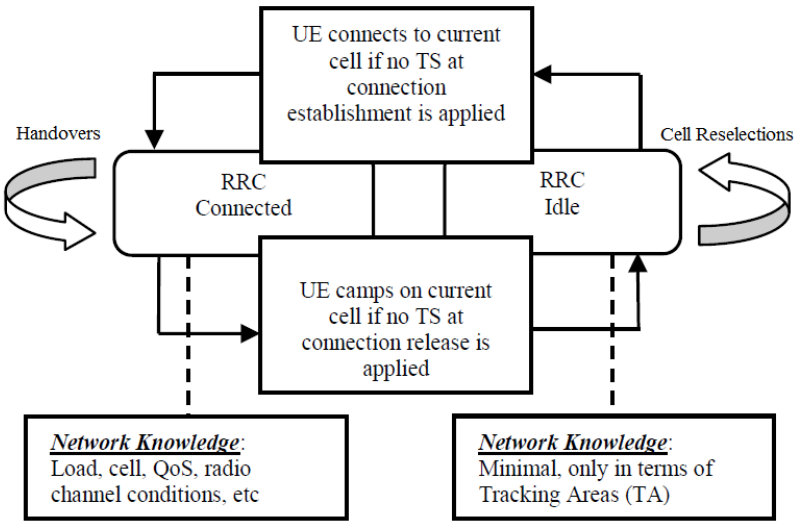


Fig. 2.1: Interworking between load balancing and the associated RRC UE state machine instances.

cell to another. Information exchange – in the form of load reporting – between adjacent eNBs assists the process, allowing overloaded cells to identify under-utilized neighbors and shift traffic to them.

This chapter aims at giving insight into the problem of dynamic load balancing in HetNet LTE deployments, pointing out the main rationale for the later designed solutions. The problem itself is delineated in Section 2.2 depicting a generic framework for performing traffic steering. The potentials of idle and connected mode load balancing together with the associated challenges are discussed in Section 2.3 and Section 2.4, respectively. Section 2.5 describes the developed model associated with the 3GPP-defined exchange of load information between adjacent cells. The simulation scenarios together with the related assumptions are presented in Section 2.6, followed by the precise definition of the considered performance indicators in Section 2.7.

2.2 Traffic Steering Framework

As illustrated in Fig. 2.1, traffic steering can be employed at any instance of the Radio resource Control (RRC) UE state machine. This includes steering the UE device while being in idle mode, in connected mode or whenever switching from one RRC state to another. The main problem addressed in this dissertation is how to develop a traffic steering framework, capable of achieving dynamic load balancing in multi-layer HetNet deployments.

Load balancing in idle mode is performed by means of cell reselections and involves users that do not claim any network resources, since they do not have any radio bearer established. The main motivation for steering idle devices is that such an approach does not cause any signaling overhead to the network. Given that idle mode UE distributions are balanced, devices are more likely to establish their RRC connection at a non-congested layer, saving signaling overhead from potential load-driven handover executions. Nonetheless, cell reselections should be economically utilized as they cost in terms of UE power consumption [56, 57]; hence, they may jeopardize the battery life of the UE device.

Idle mode load balancing is not a trivial task due to the UE-controlled nature of the procedures governing mobility management in this RRC state. Terminals autonomously perform reselections from one cell to another according to parameters that are broadcast on the system information (i.e. hysteresis values, camping priorities, etc). In this context, there is no other means of modifying camping decisions at a user-specific basis once devices have switched to idle mode. Moreover, the limited network knowledge regarding user location poses an additional challenge. Traditionally, several cells are grouped into a Tracking Area (TA) and reselections between cells belonging to the same TA are transparent to the network [58, 59]. This higher level of abstraction in terms

of location management does not allow spatial UE distributions to be accurately monitored, a fact that may endanger the creation of high concentrations of devices camping on a specific cell, even in the presence of a load balancing mechanism.

Whenever a idle device switches to the RRC connected, the radio bearer is commonly established at the latest camping cell. Nevertheless, it might still occur that the cell does not have adequate resources to serve the newly arrived user. To resolve overload conditions and provide fast load balancing, the victim UE can be directed to a different carrier via a network-controlled operation referred to as redirection. Redirection-based schemes mainly tackle inter-frequency load balancing and are triggered once the radio bearer is established so that excessive delays in the connection establishment are avoided. The decision is realized by means of a forced handover towards the redirected carrier.

Given that redirections do not suffice for balancing the load, additional actions should be taken during the connection lifetime. Connected mode load balancing is undeniably the most effective solution, since the network owns full control over users. In this context, it can quickly react to inter-layer load variations and take the proper traffic steering decisions either in the form of modifying mobility parameters or executing forced handovers. Irrespective of the adopted method, the challenge here is to maintain handovers at a reasonable level, as they cost in terms of signaling overhead together with causing service interruption whenever executed.

To enhance idle mode performance, user-specific information can be exploited before the device releases its connection and switches to idle mode. In fact, the RRC protocol [60] specifies the possibility of explicitly providing the UE with a dedicated mobility configuration via the RRC CONNECTION RELEASE message. No additional signaling overhead is required, as the associated information field is part of the standardized messaging format. In such a manner, the device may be forced to camp at a different cell for load balancing purposes. The rationale for exploiting this feature is that the network improves its degree of control over idle users, allowing for the development of more advanced idle mode load balancing solutions.

As handovers and cell reselections come at the expense of signaling overhead and UE power consumption respectively, the question to answer is how they should be efficiently exploited so that load balancing finds a good trade-off for the overall network performance and the aforesaid factors. To achieve this goal, the alignment of load balancing schemes in the different RRC states is of key importance in order to avoid conflicting situations, where users switching to connected are immediately handed over to a different cell either due to radio conditions or load balancing purposes. Such an undesirable event is also denoted as idle-to-connected ping-pong [16].

2.3 On Idle Mode Functionalities

Conventionally, cell reselections are performed on the bases of the R_s and R_n cell ranking criteria, that evaluate the camping cell s and the neighboring cell n as follows:

$$\begin{aligned} R_s &= Q_{meas,s} + Q_{hyst} \\ R_n &= Q_{meas,n} + Q_{offset}, \end{aligned} \tag{2.1}$$

where $Q_{meas,s}$ and $Q_{meas,n}$ are the corresponding measurements in terms of Reference Signal Received Power (RSRP) or Reference Signal Received Quality (RSRQ) [61]. While the former constitutes the typical signal strength measurement in 3GPP terminology, the RSRQ is defined the RSRP over the total wideband received power, also referred to as Received Signal Strength Indicator (RSSI). Lastly, Q_{hyst} specifies the serving cell hysteresis and Q_{offset} is utilized for compensating propagation properties between different carriers. Both Q_{hyst} and Q_{offset} are broadcast on the system information. Measurement availability is managed by absolute thresholds relative to the camping cell so that UE battery is economized. More specifically, the UE device initiates intra-frequency measurements whenever the camping cell drops below the $S_{intraFreqSearch}$ threshold, while $S_{interFreqSearch}$ controls the inter-frequency neighbor cell discovery. Note that the measurement periodicity is determined by the idle mode Discontinuous Reception (DRX) configuration[1]. Among the set of discovered cells, the terminal reselects to cell k that achieves the highest ranking for a predefined $T_{reselection}$ time period:

$$k = \arg\max_{k \in n} \{R_n,\ R_s\} \tag{2.2}$$

Load balancing by means of ranking-based criteria is essentially performed by auto-tuning Q_{hyst} and Q_{offset} so that reselections to under-utilized cells seem more attractive. This concept is also known as Basic Biasing (BB) [16]. Although BB is widely adopted in single-carrier scenarios, the divergent propagation properties of different carrier frequencies makes it difficult to control inter-frequency/RAT user distributions via BB-based schemes.

To overcome this challenge, Release 8 LTE introduces a new reselection mechanism – denoted as Absolute Priorities (AP) [62] – for managing inter-frequency and inter-RAT mobility. The AP framework allows for the prioritization of specific carriers in idle mode. Carrier priorities are broadcast on the system information and devices reselect to a higher priority carrier whenever the target signal strength or quality $Q_{meas,n}$ is above the absolute threshold $Thresh^{AP}_{2High}$, i.e.:

[1] During each DRX cycle, idle UE devices have to wake up so as to listen to the paging channel, perform physical layer measurements and potentially reselect to a neighboring cell.

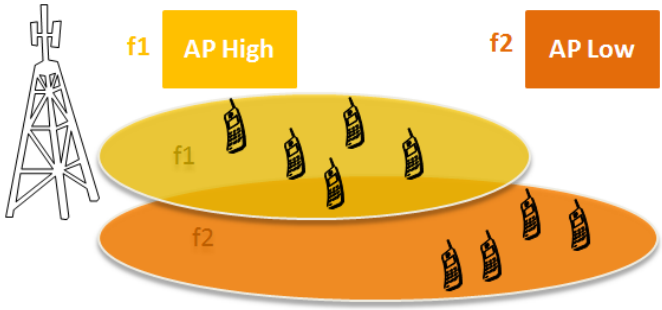


Fig. 2.2: AP applicability in a dual-band macrocell deployment. High priority is assigned to the high carrier frequency in order to be mainly exploited by cell center users.

$$Q_{meas,n} > Thresh^{AP}_{2High}, \tag{2.3}$$

By contrast, cell reselections towards a lower priority carrier are triggered whenever the camping cell drops below $Thresh^{AP}_{sLow}$ and a neighbor is better than $Thresh^{AP}_{High2Low}$:

$$Q_{meas,s} < Thresh^{AP}_{sLow} \quad \wedge \quad Q_{meas,n} > Thresh^{AP}_{High2Low} \tag{2.4}$$

Notice that (2.3), (2.4) are solely eligible for cell reselections between carriers with different priorities. For carrier being assigned with the same camping priority, cell reselections are still performed based the R_s and R_n criteria in (2.2).

Fig. 2.2 illustrates the applicability of the AP framework in a multi-layer macrocell deployment consisting of 2 carrier frequencies with $f_1 > f_2$. By prioritizing f_1 over f_2, cell center UE devices camp on the high carrier frequency as its coverage is mainly controlled by the $Thresh^{AP}_{2High}$ threshold, overruling the propagation disparities of the two layers. On the other hand, cell edge UEs camp on the low carrier frequency, which is the desirable behavior. AP can be adopted in HetNet scenarios as well. In this case, the highest priority should be assigned to the small cell carrier in order to maximize the offloading capabilities of the low-power nodes.

Undoubtedly, AP-based cell reselections provide operators with a greater degree of flexibility for controlling idle mode UE distributions, since the priority assignment can be done according to their specified objectives. However, there are still some limitations related to the AP operation:

- Inter-frequency measurements are inefficiently utilized when UEs are configured to perform AP-based reselections. Albeit measurements towards

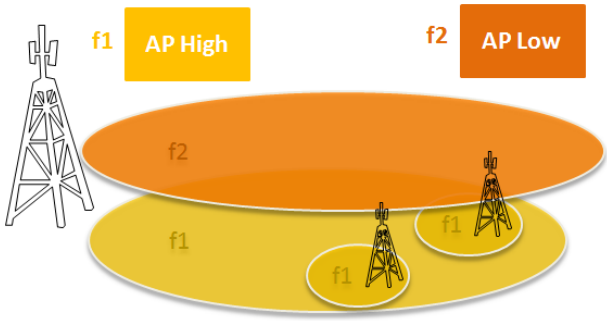


Fig. 2.3: Example of a co-channel deployment at $f1$, supplemented by an additional macro-cell carrier at $f2$ with $f2 < f1$.

a lower priority carrier are typically performed whenever the camping cell fall below the $Thresh^{AP}_{sLow}$ threshold, terminals camping at a lower priority carrier always search for prioritized carriers. Apparently, this may severely impact the power consumption of devices located in areas not covered by the higher priority carrier. A typical situation involves idle users away from the small cell vicinity which will repetitively perform inter-frequency measurements since the small cell carrier will most likely be prioritized.

- AP may overload the high priority carrier due to the limited location management information that is available in the RRC idle mode. I.e. a large concentration of idle users switches to RRC connected congesting the serving cell. To avoid such undesired effect, SON-based schemes could dynamically adjust the coverage of the high priority carrier. However, this may come at the expense of increasing the UE power consumption of devices camping at a lower priority carrier [16], as aforedescribed.

- Different layers deployed at the same carrier must be assigned with the same priority. I.e. such a limitation is relevant in the scenario depicted in Fig. 2.3, where the co-channel macro-pico deployment at f_1 is supplemented by an additional macro carrier at f_2. In this case, the dynamic adjustment of the AP thresholds controls inter-frequency load balancing only at carrier basis. This necessarily implicates that layer-specific load balancing (macro $f1 \leftrightarrow$ macro $f2$ or pico $f1 \leftrightarrow$ macro $f2$) cannot be performed by means of AP.

These observations indicate that there are still potentials for improving RRC idle operations, so that they better facilitate load balancing in multi-layer HetNet deployments. All in all, this involves enhanced control of user distributions in the RRC idle state along with energy-efficient inter-frequency cell neighbor mechanisms.

2.4 On Connected Mode Functionalities

Mobility management for RRC connected users is based on a network controlled and UE assisted paradigm. This essentially means that handover decisions are taken by the network according to measurement reports that are provided by the UE device in the uplink. Albeit a periodical reporting configuration is possible, event-triggered is usually employed in order to save messaging overhead. The standardized measurement event [60] for both intra-LTE and inter-RAT mobility are outlined in Table 2.2.

Table 2.2: Measurement Reporting Events

Event	Description
A1	Serving cell becomes better than threshold
A2	Serving cell becomes worse than threshold
A3	Neighbor cell becomes offset better than serving
A4	Neighbor cell becomes better than threshold
A5	Serving cell becomes worse than threshold_1 and neighbor cell better than threshold_2
A6	Neighbor cell becomes offset better than Secondary cell
B1	Inter-RAT neighbor cell becomes better than threshold
B2	Serving cell becomes worse than threshold_1 and inter-RAT neighbor cell becomes better than threshold_2

Intra-frequency handovers are typically triggered by the A3 event, as shown in Fig. 2.4. The UE device is configured to periodically measure intra-frequency cell neighbors in order to identify potential handover candidates. Given that a target cell meets the A3 condition for a specific time duration, referred to as Time-To-Trigger (TTT) window, the event is reported to the network and the handover procedure is initiated.

Unlike intra-frequency measurements, which are performed rather often for the sake of mobility robustness on the serving carrier, this is not the case for inter-frequency cell discovery. Particularly, inter-frequency measurements are typically configured for UE devices that are about to experience coverage problems on their serving carrier (see Fig. 2.5). The reason is that the terminal has to change its frequency oscillators to the measured carrier frequency; a fact that costs energy and time in which the terminal cannot receive data [63]. To limit such measurements owing to the aforedescribed costs, inter-frequency measurement availability is controlled by the A2 event and is commonly triggered when the signal of the serving cell gets weak. Once the A2 condition is met, inter-frequency mobility is enabled by configuring the related measurement event (either A3 or A5).

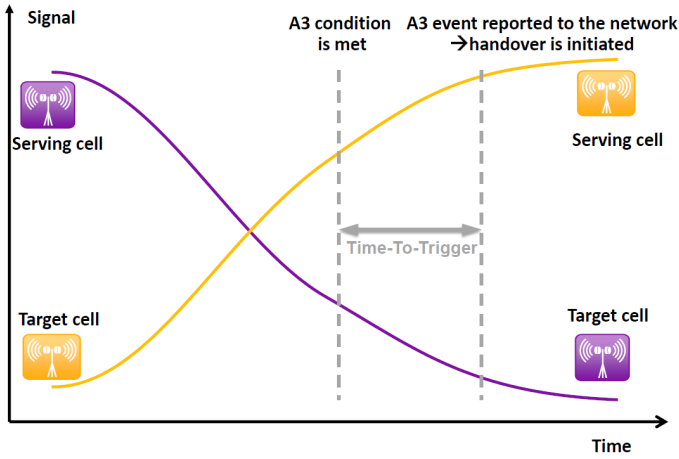


Fig. 2.4: Handover execution based on A3 event.

Adjusting handover parameters or performing user-explicit forced handover executions are the main features to employ for connected mode load balancing. The former modify the service area of adjacent cells so that the coverage of an overloaded cell is shrunk, while enlarging underutilized neighbors. This concept is widely adopted in co-channel deployments, where interference limits load balancing in the cell boundaries of neighboring cells. In this context, such a scheme constitutes the simplest traffic steering approach, as no user-specific information is required and dynamic load balancing is achieved by favoring handover decisions towards fairly utilized cells. The challenge in this case is to maintain mobility robustness, since load balancing is performed at the expense of radio conditions; hence, the environment enmity in terms of inter-cell interference may degrade mobility performance.

On the other hand, dedicated carrier deployments do not naturally suffer from strong co-channel interference. Consequently, steering users against radio conditions for load balancing purposes does not result in noticeable mobility problems in most of the cases. However, the major problem for inter-frequency load balancing is how to provide sufficient measurement availability in order to trigger handover events without impacting the UE power consumption and the perceived data rates. In such deployments, forced handovers could be an alternative solution. Formally, these overrule the configured mobility events in a sense that handovers are triggered – for the sake of load balancing – even if the associated event is not yet met. The drawback of this approach is that traffic steering requires additional user-specific information for explicitly triggering inter-frequency measurements and selecting the appropriate UE devices to be offloaded to a neighboring cell. Thus, the algorithm complexity may increase.

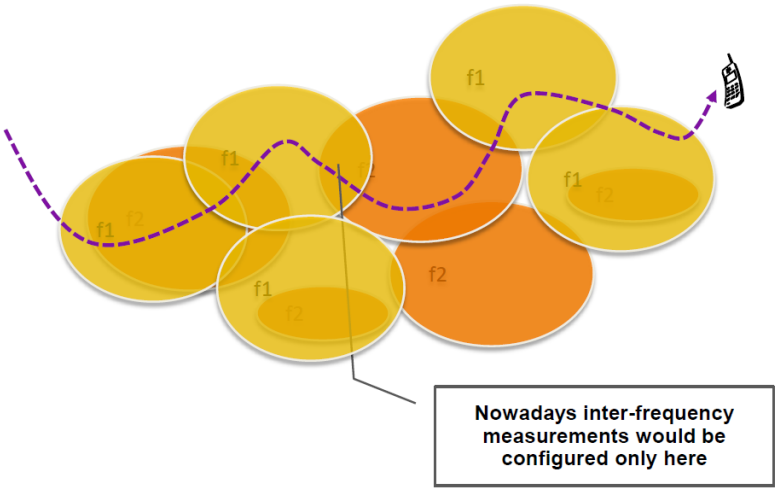


Fig. 2.5: Typical configuration of inter-frequency measurements in nowadays deployments.

2.5 Load & Composite Available Capacity

To apply any of the aforedescribed traffic steering actions, cells should be aware of their own load conditions as well as the load of their neighbors. For that purpose, they periodically monitor their own load conditions and exchange load information in the form of Composite Available Capacity (CAC) over the X2 interface [64]. Formally, CAC is implementation specific and expresses the amount of load that a particular cell is willing to accept subject to several factors such include resource utilization, QoS requirements, backhaul capacity and the load of control channels. The following subsections describe how these mechanisms are modeled in this PhD dissertation.

2.5.1 Cell Load Definition

The most intuitive cell load expression is evidently the utilization of transmission resources. Given that S_u is the set of users scheduled in cell c, the corresponding resource utilization is defined as follows:

$$\rho_c = \frac{1}{B_c} \sum_{u \in S_u} f_u \in [0, 1], \tag{2.5}$$

where B_c is the cell bandwidth in terms of Physical Resource Block (PRB) and f_u corresponds to the amount of PRBs allocated to user u. However, the

resource usage is not always a good indication of the experienced QoS within the cell [45, 63, 65]. Typically, packet scheduling with different QoS UE classes prioritizes real-time traffic when assigning resources. If all real-time users are satisfied and some resources are still available, they are distributed among bearers with elastic traffic. Under such scheduling assumptions, neighboring cells with active best effort users are less likely to perform any load balancing actions during the duration of large file downloads, if PRB-based traffic steering is applied. Cells will always look fully loaded, while no information regarding the UE experience within the cell is provided. Consequently, this may limit the potentials of performing fast load balancing at a finer time granularity, i.e. in the time scale of minute or below.

To overcome the aforementioned limitation, a more useful load definition is provided in this thesis. Since the PRB usage does not reflect the UE experience within the cell, a notion of user satisfaction is introduced in (2.5) by modifying f_u as below:

$$\widehat{f}_u = min\{\frac{f_u \widehat{R}_u}{r_u}, \rho_{max}\}, \tag{2.6}$$

where $\widehat{R}_u$ is the targeted data rate to be met by user u; and r_u is the data rate that is actually achieved for a given resource share f_u. In case of real-time traffic flows, $\widehat{R}_u$ should be set according to the associated QoS requirements, while an equivalent acceptable data rate may be defined for best effort users as well. Note that (2.6) bounds $\widehat{f}_u$ to some maximum ρ_{max} value in order to avoid situations where a single UE in poor channel conditions could declare the cell in overload by requesting a lot of transmission resources. As $\widehat{f}_u$ provides an estimation of the resources which are required to satisfy user u, a virtual cell load metric, $\widehat{\rho}_c$, can be defined as follows:

$$\widehat{\rho}_c = \frac{1}{B_c} \sum_{u \in S_u} \widehat{f}_u \tag{2.7}$$

A direct comparison between the cell load definitions in (2.5) and (2.7) can be found in Fig. 2.6. The illustrated example assumes 4 best effort users with equal resource shares, occupying the whole transmission bandwidth. Unlike ρ_c that equals to 1, $\widehat{\rho}_c$ shapes the system load according to the configured objectives, expressed in the form of $\widehat{R}$ in this specific example. By means of that, cell may accept additional traffic – for load balancing purposes – by claiming resources from the "*satisfied*" users. To avoid any misconceptions from the reader, the later designed schemes will be engineered on the bases of balancing $\widehat{\rho}_c$ across the deployed network layers. Hence, whenever referring to cell load hereinafter, the definition in (2.7) is considered.

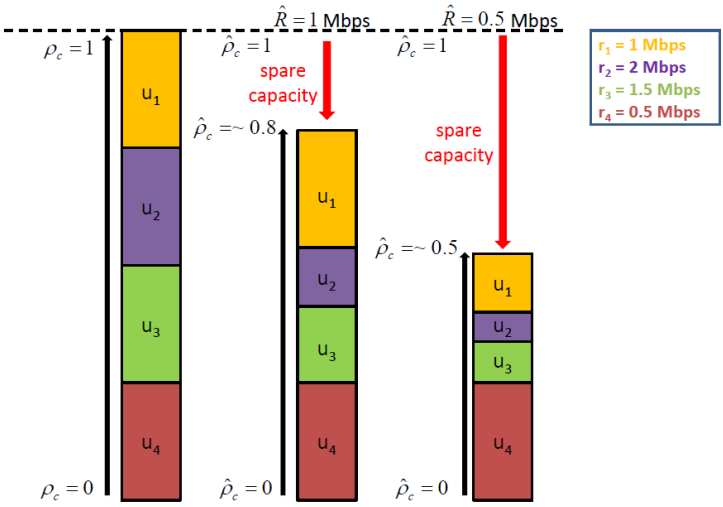


Fig. 2.6: Properties of virtual load metric. Resource usage is shaped according to the predetermined target bit rate requirements.

2.5.2 CAC Definition

To model the periodic cell load monitoring, a first order auto-regressive filter is employed as follows:

$$\widetilde{\rho}_c^{sm}(t_k) = (1 - \alpha_l) \cdot \widetilde{\rho}_c^{sm}(t_{k-1}) + \alpha_l \cdot M_c(t_k), \tag{2.8}$$

where $M_c(t_k)$ is the instantaneous load measurement at the time instance t_k; and $\widetilde{\rho}_c^{sm}(t_k)$ is the corresponding smoothed value subject to the α_l parameter that determines the filter memory. Ultimately, CAC is defined as the proportional difference between $\widetilde{\rho}_c^{sm}$ and the target operational cell load, ρ_{target}:

$$CAC = 1 - \frac{\widetilde{\rho}_c^{sm}}{\rho_{target}}, \tag{2.9}$$

It is worthy to mention that ρ_{target} is a operator-configurable parameter and it actually indicates the desired load conditions within the cell. Unless CAC $>$ 0, cells do not accept more traffic, protecting the users that they are currently serving.

2.6 Scenarios & Assumptions

In order to capture the impact of load balancing in different network environments, a wide range of scenarios have been considered throughout this PhD study. These are summarized in Table 2.3 starting from a single-carrier study case and then moving forward to more mature LTE deployments. The latter consider multiple frequency layers as well as the usage of intra/inter-eNB CA whenever this is applicable. Once these are described, the fundamental simulation assumptions are outlined.

2.6.1 Considered Deployments

Scenario A is utilized for investigating the potentials of co-channel HetNet load balancing. Picocells are deployed in full reuse with the macro overlay at 2.6 GHz with 10 MHz of transmission bandwidth. As the experienced interference between the two layers is a challenging issue, load balancing should be accompanied by suitable MRO functionalities so that capacity enhancements are achieved without a noticeable mobility degradation.

Although *Scenario A* is preferred for operators with moderate spectrum [20, 66], a handful of deployment alternatives can be realized given that the spectrum availability is larger. Inter-frequency load balancing prior to Release 10 is conducted in *Scenario B*. Particularly, the bandwidth of the co-channel 2.6 GHz HetNet deployment is expanded to 20 MHz, being overlaid by an additional macrocell 10 MHz carrier at the 800 MHz band. The main rationale for considering such a case study instead of a dedicated carrier deployment (e.g. by removing the 2.6 GHz macrocell layer) is that the former is more challenging. Apart from efficiently discovering cells deployed at a neighbor carrier, traffic steering must manage both load balancing between the 2 macrocell carriers as well as between the 800 MHz carrier and the 2.6 GHz picocell layer. Ultimately, steering UE devices from 800 MHz to the 2.6 GHz carrier should not jeopardize mobility robustness on the directed carrier given the strong interference experienced at the 2.6 GHz band.

With CA, the packet scheduler can assist the load balancing process as CA UE devices can have data reception from more than a single cell. This essentially means that it is quite important to understand the viability of load-based traffic steering in such deployments. For that purpose, intra-eNB CA is enabled at the macro layers of *Scenario B* and its joint interaction with load-driven mobility procedures is evaluated for different penetrations of CA UE devices. Together with *Scenario B*, a complementary case study is considered as well. Specifically, this is denoted as *Scenario C* and involves the deployment of an additional 1.8 GHz macrocell carrier – with 10 MHz of bandwidth – over *Scenario B*. The only difference is that CA UEs do not own full access to the macro

Table 2.3: Considered HetNet Scenarios

Co-channel load balancing			
Network Deployment			**2.6 GHz (10 MHz)**
Scenario A			Macro Pico
Inter-frequency load balancing with \ without CA UEs			
Network Deployment	**800 MHz (10 MHz)**	**1.8 GHz (10 MHz)**	**2.6 GHz (20 MHz)**
Scenario B	Macro	——	Macro Pico
Scenario C	Macro	Macro	Macro Pico
Scenario D	Macro	Macro	Pico

resources in *Scenario C* due to RF constraints that commonly constraint inter-band CA to 2 serving CCs [67]. The direct comparison between the two case studies allows us to draw valuable conclusions regarding the relevance of traffic steering in deployments where the number of CCs is greater (*Scenario C*) or equal (*Scenario B*) to the number of serving cells that a CA UE supports.

Finally, the load balancing performance of Dual Connectivity (DC)[2] is evaluated in *Scenario D*. In particular, this scenario constitutes a dedicated carrier deployment where the 2.6 GHz band is fully dedicated to the picocells with two additional macrocell carriers at 800 MHz and 1.8 GHz, respectively. Albeit CA UEs away of the picocell vicinity still aggregate resource from the two macrocell CCs, users within the picocell coverage can operate in DC mode. The picocells are assumed to be interconnected with the macro overlay via a high-bandwidth low-latency interface. This necessarily enables the usage of collaborative multicell scheduling for offering load balancing between the two heterogeneous layers.

2.6.2 Modeling Assumptions

Regardless of the simulated scenario, the adopted network layout is the one defined in [53], consisting of a regular hexagonal grid of 7 macro sites with 3 sectors per site. The inter site distance between adjacent macro eNBs is 500 m. Picocell placement within each macrocell area is based on a spatial

[2]Adopting the Release 12 3GPP terminology, inter-eNB CA between small cells and macrocells is denoted as Dual Connectivity (DC) [35].

uniform distribution, fulfilling specific distance constraints subject to different base station types. In particular, the minimum macro-to-pico cell distance is set to 75 m, while the minimum distance between picocells is 40 m. Macro base stations are equipped with directional antennas and are deployed at a height of 30 m. Unlike macro eNBs, picocells with omni-directional antennas are placed below rooftop level at 5 m height. Propagation modeling comprises the effects of distance-dependent pathloss, shadowing, fast fading, 3-dimension antenna patterns and fixed penetration loss due to building obstacles [53, 68].

Picocells are assumed to be deployed in high traffic density locations. For that purpose, a non-uniform spatial user distribution is considered, where 2/3 of the users per macrocell area are confined within 40 m of picocells creating hotspot zones. The remaining 1/3 is uniformly distributed in the macrocell area, moving at straight line trajectories with constant velocity. Unlike hotspot UEs – they always move at 3 km/h –, the speed of free moving terminals may vary from one study to another, reaching significantly higher values. To avoid unwanted cell border effects, the wrap-around technique is applied and free moving UEs appear at the other network side whenever reaching the layout borders, as shown in Fig. 2.7.

Last but not least, both Constant Bit Rate (CBR) and best effort traffic is simulated depending on the study scope. Similarly to [69], data rates are estimated by means of rate functions that map the wideband user SINR to throughput values including the effect of Link Adaptation (LA), Hybrid Automatic Repeat Request (HARQ) management, spatial multiplexing and Frequency Domain Packet Scheduling (FDPS) diversity gain [70]. Users are scheduled only if their perceived SINR is greater than -8 dB; otherwise it is assumed that synchronization with the control channel is lost. The most generic simulation assumptions are summarized in Table 2.4, while a more detailed description of the modeling framework can be found in Appendix A.

2.7 Key Performance Indicators

Subject to the simulated traffic type, system performance is mainly measured by means of the following Key Performance Indicators (KPI):

- *Coverage throughput*: This is the 5^{th} percentile worst user throughput over all simulated users and transmission instances.
- *Average throughput*: This is the arithmetic mean of user throughput averaged over all simulated users and transmission instances.
- *User Satisfaction*: The percentage of users meeting their QoS requirement in terms of Constant Bit Rate (CBR). It is averaged over all CBR users and transmission instances.

Table 2.4: Main Simulation Parameters

Parameter	**Value**
Network Layout	Hexagonal grid with wrap-around (7 sites, 3 cells per site)
Inter site distance	500 m
Carrier frequencies	{800, 1800, 2600} MHz
Minimum macro-to-pico cell distance	75 m
Minimum pico-to-pico cell distance	40 m
Base station height	Macro: 30 m, Pico: 5 m
Transmit power	Macro: 46 dBm, Pico: 30 dBm
User distribution	Hotspot
Hotspot radius	40 m
Macro antenna pattern (horizontal)	$A_H(\phi) = -min\{12(\frac{\phi}{\phi_{3B}})^2, A_m\}$ $\phi_{3B} = 70^o$, $A_m = 25dB$
Macro antenna pattern (vertical)	$A_V(\theta) = -min\{12(\frac{\theta-\theta_{etilt}}{\theta_{3B}})^2, SL_V\}$ $\theta_{3B} = 10^o$, $SL_V = 20dB$
Antenna gain	Macro: 14 dBi, Pico: 5 dBi
Antenna configuration	1x2
Macro Pathloss	800MHz:$119.8 + 37.6 \cdot log10(R)$ 1.8GHz:$127.2 + 37.6 \cdot log10(R)$ 2.6GHz:$130.5 + 37.6 \cdot log10(R)$
Pico Pathloss	2.6GHz:$140.7 + 36.7log_{10}(R) - \Delta_{cor}$ $\Delta_{cor} = 3.5dB$
Shadowing standard deviation	Macro: 8 dB, Pico: 10 dB
Shadowing decorrelation distance	Macro: 50 m, Pico: 13 m
Penetration Loss	20 dB

Moreover, Jain's fairness index has been adopted in this PhD study for visualizing the achievable degree of load balancing. It is defined as the square root of the cosine of the angle between the data set x_i and the hypothetical equal allocation [71]:

$$J(x_1, x_2, \cdots, x_n) = \frac{(\sum_{i=1}^{n} x_i)^2}{n \cdot \sum_{i=1}^{n} x_i^2} \tag{2.10}$$

x_i corresponds to the average cell load of layer i in this case; and n is the number of deployed layers. If $J = 1$, then all cells are equally loaded, while greater load imbalances are indicated as J approaches the $1/n$ value.

The following KPIs are used for measuring mobility performance:

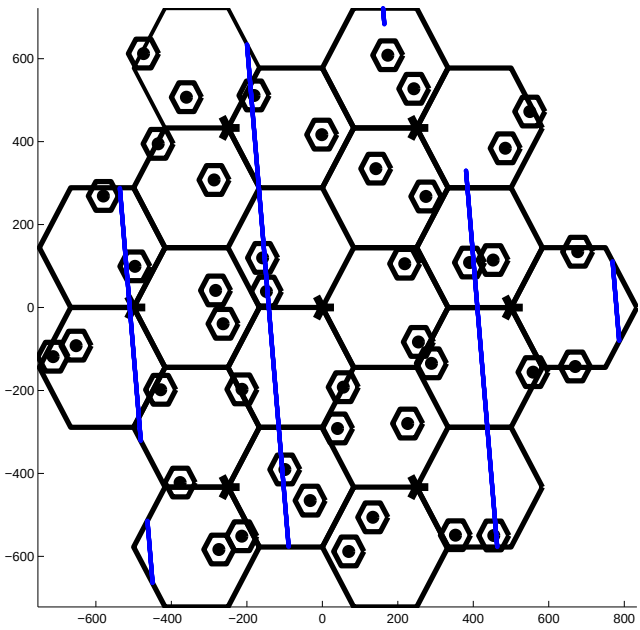


Fig. 2.7: Network layout assuming 2 picocells per macrocell area. The blue line refers to the trace of a free moving user.

- *Handover Rate*: It is defined as the total number of handovers averaged over the simulation time and the number of users. It further gives a notion of the required signaling overhead for realizing load balancing decisions in connected mode.
- *Cell Reselection Rate*: This is the total number of cell reselections averaged over the simulation time and the number of users. As cell reselections may impact UE power consumption, this particular KPI gives insight into the potentials UE battery savings in idle mode.
- *RLF Rate*: Similarly to the aforementioned metrics, the RLF rate is defined as the total number of RLFs averaged over the simulation time and the number of users. A low RLF occurrence is desirable so that service continuity is not jeopardized. Note that a RLF is declared whenever the downlink wideband SINR has been below -8 dB and does not exceed -6 dB within 1 second [32].

Lastly, the rate of SCell additions, removals and changes are utilized for measuring the overhead introduced by CA. This comes from the required RRC

signaling for realizing any SCell related action. Note that if any other additional metric is required for supporting the conducted PhD study, it will be introduced before its first appearance. This mainly involves statistics regarding the spatial user distribution, the rate with which physical layer measurements are performed, etc.

Most of the presented KPIs are tightly coupled. I.e. frequent inter-frequency handover executions would most likely result in high throughput values simply because users would always be steered to the least loaded carrier. Similarly, high intra-frequency handover rates could eliminate the risk of RLF declarations. However, the signaling overhead at the network side would be rather unacceptable in any of the aforedescribed situations. Therefore, it is of key importance to always measure the cost associated with a particular system performance before concluding on the suitability of any simulated scheme.

Chapter 3

Co-channel Load Balancing in HetNet Deployments

3.1 Introduction

Deploying small cells in full frequency reuse with the macro overlay poses a handful of challenges. As co-channel deployments naturally suffer from strong interference between the two layers, it is not a trivial task to maintain mobility robustness while performing inter-layer load balancing. To cope with the aforementioned issues, deployment automation plays a key role for achieving such a goal. Adopting the SON terminology, autonomous functions tackling dynamic load balancing and mobility optimization are commonly referred to as Mobility Load Balancing (MLB) and Mobility Robustness Optimization (MRO), respectively. As illustrated in Fig. 3.1, the development of such self-optimization features is facilitated by a handful of mechanisms that enable the tight coordination of adjacent cells. In this context, neighboring base stations can exchange information such as cell load conditions, handoff parameters, and eventually fine-tune system configurations for either load balancing or mobility robustness purposes.

The remainder of the chapter is organized as follows: Section 3.2 formulates the problem and elucidates the targets of the conducted study, while relevant state-of-art literature is reviewed in Section 3.3. The proposed framework is presented in Section 3.4 focusing on both architectural and feature implementation aspects. Section 3.5 summarizes the associated simulation assumptions followed by the system performance results in Section 3.6. Finally, Section 3.7 concludes the chapter.

3.2 Problem Formulation

As mentioned in Chapter 2, intra-frequency handover procedures are triggered based on the A3 event, hence, whenever the following inequality is met:

$$Q_j + bias_j > Q_i + CIO_{i \to j}, \tag{3.1}$$

where Q_i, Q_j are the signal strength measurements from the serving cell i and the target cell j, respectively; $bias_j$ is the measurement biasing of cell j; and $\text{CIO}_{i \to j}$ is the handover offset for the $i \to j$ cell pair, also referred to as Cell Individual Offset (CIO). Given that cells i and j are a macro and a pico node, respectively, there are two means of pushing excess traffic to the small cell layer. These are either by increasing $bias_j$ or decreasing $CIO_{i \to j}$. The former is denoted as Range Extension (RE) and corresponds to an effective ΔCIO adjustment of $(CIO_{i \to j} - bias_j)$ dB. In fact, both methods enlarge the coverage footprint of the small cell so that handovers towards that particular cell seem more attractive. Notice that a similar positive adjustment should be employed for the $\text{CIO}_{j \to i}$ configuration (or negative modification of the $bias_i$) so that handed over UE devices are not pushed back to the macro overlay due to radio conditions. Assuming fewer users on the small cell layer, the SINR deterioration after the handover completion is compensated by the larger resource availability; a fact that essentially improves the spectral efficiency. However, this naturally comes at the risk of mobility problems, as the increased interference

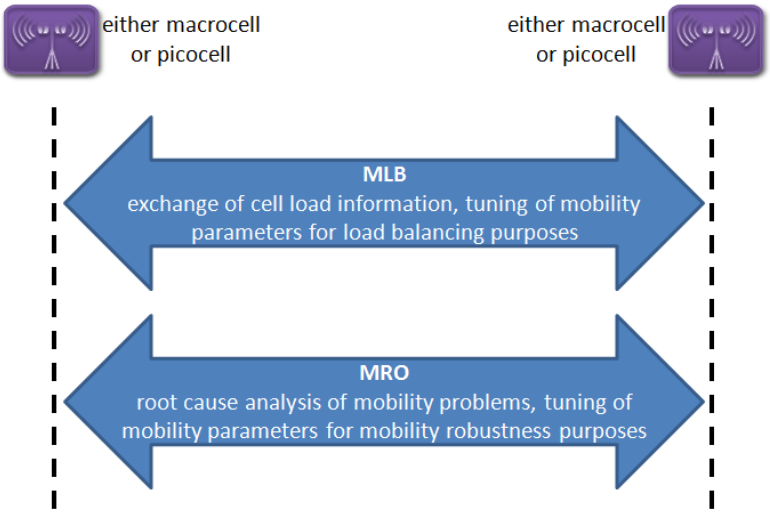


Fig. 3.1: Multi-cell coordination over the X2 interface for MLB and MRO functionalities

may make the UE device lose connection with the network due to RLFs and HOFs.

The problem at hand is to find for each macro-pico cell pair the appropriate CIO configuration that balances the load between the two layers, while guaranteeing an acceptable mobility performance over the cell boundary. This task is evidently more challenging in HetNet scenarios rather than in macro-only networks owing to the stronger interference experienced at the small cell layer. Moreover, as both MLB and MRO functions can modify the CIO parameterization, it is of paramount importance to coordinate their joint operation and eventually find an attractive trade-off between user experience and mobility robustness. It is worth mentioning that although the aforementioned SON functions can operate between any type of cells (e.g. macro-to-macro, macro-to-pico, pico-to-pico), we limit the problem only between macrocells and picocells for the sake of simplicity.

3.3 State-of-Art

This section presents the main state-of-art with regards to co-channel load balancing and mobility robustness. Initially, material from prior studies being available in open literature is reviewed and the major findings are summarized. After that, the emphasis is put on presenting how the 3GPP specifications facilitate the MRO employment.

3.3.1 Literature Review

For co-channel HetNet deployments, state-of-art literature mainly faces the problem of load balancing separately from mobility management. The former is often investigated together with eICIC, allowing resource partitioning between the two network layers. Macro base stations are muted on specific subframes, during which, small cells can schedule their cell edge users that otherwise would suffer from severe interference. In fact, eICIC can significantly improve offloading towards small cells since larger RE values can be employed for attracting UE devices to low-power nodes [72, 73]. The impact of eICIC on mobility robustness is investigated in [32, 81, 82], The bottom line is that the synchronized eICIC employment essentially decreases the RLF rates; however, it does not fully resolve the problem as the mobility performance is still dominated by problematic outbound small cell handovers.

An insightful analysis regarding HetNet load balancing can be found in [74], proposing different biasing configurations subject to the small cell location. The reason is that small cells positioned closer to the macro eNB have a significantly smaller coverage footprint compared to these located at more distant

positions. Hence, a common RE value for all small cells – favoring handover executions towards them – will essentially aid cell center low-power nodes to increase their resource utilization; however, it may create undesired overload problems at cell edge ones. Undeniably, these findings motivate the adoption of MLB-based schemes that dynamically adjust CIO configurations on a cell boundary granularity.

Notice that theoretical load balancing studies – e.g. [28, 29] – typically expand the cell-pair RE adjustment even to a UE dedicated basis. In this context, the users with the maximum offloading gain are selected and appropriate RE values are applied so that these particular UE devices are pushed towards the small cells. Nevertheless, albeit their valuable contribution, such results are often obtained under semi-ideal conditions, neglecting the impact of several effects in order to derive closed-form expressions. E.g. mobility is typically not modeled or ideal estimation of the achievable throughput per layer is considered. In addition, the signaling overhead for explicitly configuring UE devices with suitable RE values is commonly ignored.

With regards to HetNet mobility management, state-of-art studies have shown that the most challenging handoff type is the outbound small cell handover (pico-to-macrocell) [76, 77] for medium to high speed UE devices. For that purpose, a small Layer-3 measurement coefficient together with shortening the TTT window is recommended in an attempt to increase the probability of timely-accurate handover executions. In addition, the adoption of velocity-dependent handover settings are also suggested in open literature in combination with Mobility State Estimation (MSE). Specifically, MSE allows UEs to calculate their handoff rate and classify themselves into either a low, medium or high mobility state according to thresholds provided by the network. Users in medium/high MSE state can be excluded from connecting to a low-power node or scale down their TTT configuration even more so that the outbound small cell handover is triggered fast enough. Notwithstanding its satisfactory performance in macro-only scenarios [78, 79], the disparities in terms of cell sizes make conventional MSE fall short of adequately estimating user velocity in HetNet deployments. To enhance MSE operation, the work in [80] proposes that handover counts associated to small cells should be weighted less in order to achieve proper correlation between the MSE state and the terminal velocity.

Last but not least, solutions that jointly tackle load balancing and mobility robustness can be found [83, 84]; nevertheless, they solely refer to macro-only deployments. These are realized on the basis of a Fuzzy Logic Controllers (FLC) that translates human knowledge into a set of rules for adjusting CIOs based on cell load, call drop ratio and mobility performance indicators. In fact, FLC are widely used in open literature for fine-tuning mobility parameters (e.g. [83–86]); however, they require significant configuration effort. For that purpose, the work in [84, 87] has proposed the adoption of reinforcement learning techniques so as to train the FLC decision making and essentially minimize the human

intervention.

3.3.2 3GPP Support for MRO

3GPP specifications [88] provide a solid framework for identifying potential mobility problems and solving them. Based on tight multi-cell cooperation and online monitoring of mobility KPIs on a cell boundary basis, the following root causes of mobility problems can be identified (see Fig. 3.2):

- *Too Late Handover*: The UE device is connected to cell D and is about to perform a handover to Cell B. However, it experiences a RLF and re-establishes its connection at cell B. At that point, cell B requests from the terminal a RLF report including the last measurements and the serving cell before the RLF declaration. By means of that, cell B is informed that the "guilty" neighbor is cell D which is notified that its mobility settings create too late handovers on this specific cell boundary. Hence, the $\mathrm{CIO}_{D\rightarrow B}$ should decrement.
- *Too Early Handover*: The UE device is connected to cell A, moving towards cell D and the handover to that particular cell is successfully executed. However, shortly after the handover completion, the connection is lost and the UE re-connects to cell A via the re-establishment procedure. The associated UE report will indicate that the RLF was declared while the terminal was connected to cell D; hence, the potentially problematic cell will be informed via the X2 interface. Upon receiving the concerned notification, cell D has to identify whether or not the RLF was caused due to its own mobility configuration (e.g. too late handover). As former serving cells keep UE history information stored for some time after a handover/RLF occurrence, cell D can check the UE time-of-stay and hereby conclude that the RLF was caused by a too early handover, notifying cell A about the root cause of the observed mobility problem. Apparently, the handover execution is delayed if $\mathrm{CIO}_{A\rightarrow D}$ increments.
- *Wrong Handover*: The UE device is connected to cell D, moving towards cell B but wrongly hands over to cell C where it experiences a RLF. After re-establishing its RRC connection to cell B, the new serving eNB informs cell C about the associated RLF. Cell C checks the UE history and discovers that there was a handover shortly beforehand, being initiated by cell D. It then signals to cell D that its mobility configuration causes wrong handovers. To increase the probability of avoiding such erroneous handoff decision between cell D and C, the $\mathrm{CIO}_{D\rightarrow C}$ has to increment.

The applicability of the aforedescribed framework has been mainly investigated in macrocell deployments both in intra-LTE and inter-RAT mobility studies

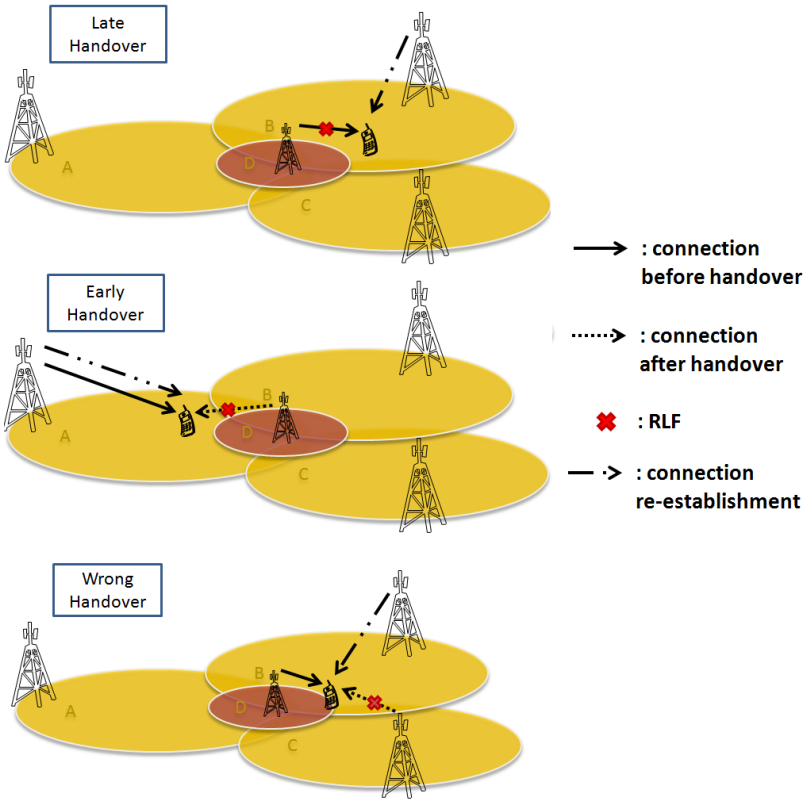


Fig. 3.2: Root cause identification of mobility problems

[50, 51, 89, 90]. Specifically, the associated results have shown satisfactory performance in terms of mobility robustness. This motivates for its adoption also in HetNet deployments, where the remarkable increased number of cell boundaries makes the manual optimal CIO configuration rather impractical.

3.4 Joint MLB & MRO Solution

The problem of coordinating the interaction between MLB and MRO functionalities is also denoted as a double responsibility conflict [16]. This is due to the fact that both SON functions are allowed to access and modify the mobility parameters. MLB commonly attempts to increase the time-of-stay in underutilized cells endangering the RLF/HOF occurrence due to too late outbound handovers. At the same time, MRO will basically try to limit the

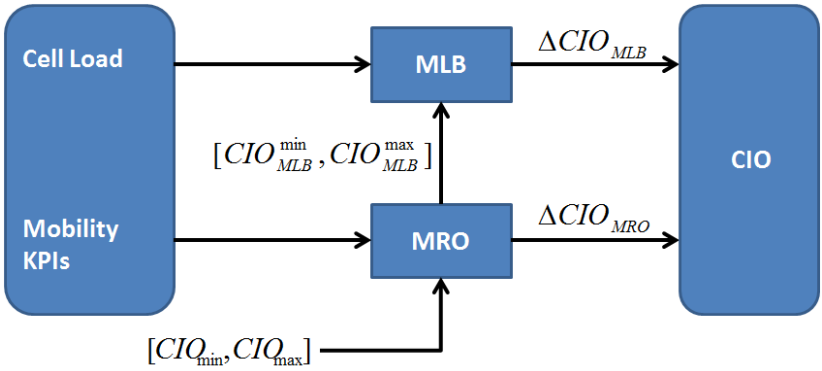


Fig. 3.3: Proposed architecture for joint MLB and MRO operation.

load balancing effect by decreasing the MLB-applied CIO adjustments so that mobility robustness improves. As this behavior can create conflicting actions where MRO neutralizes MLB decisions and vice versa, it is of key importance to define an architectural framework that avoids MLB/MRO conflicts.

The proposed architecture assigns higher responsibility to the MRO function for controlling mobility parameters. In more details, MRO modifies the CIO parameterization based on mobility success/failure observations as well as it further controls the MLB aggressive in terms of load balancing. A schematic representation of the designed solutions is illustrated in Fig. 3.3. The upper and lower bounds of the eligible CIO configurations are specified [CIO$_{min}$, CIO$_{max}$]. This range is operator-configurable and is fed as an input to the MRO block. At the meanwhile, MLB is allowed to modify the CIO within the interval signaled by MRO. This is expressed by [CIO$_{MLB}^{min}$, CIO$_{MLB}^{max}$] and its initialization should typically set to [0 0] dB. By means of that, MRO is let to adaptively relax or constrain MLB operations subject to their impact on mobility performance. Thus:

$$CIO \in [CIO_{MLB}^{min},\ CIO_{MLB}^{max}] \in [CIO_{min},\ CIO_{max}] \tag{3.2}$$

The aforedescribed framework has as its major target to maintain connection stability. The rationale for such decision is that connectivity problems could make subscribers much more unsatisfied compared to cases where the user may perceive suboptimal throughput but at least its connection is stable. The designed MLB and MRO algorithms are described in the following subsections.

3.4.1 MLB Algorithm

The MLB algorithm relies on the principles described in Section 2.5. Hence, cells periodically perform load measurements and estimate their own CAC sub-

ject to the targeted operational load status ρ_{target}. As seen by Table 3.1, cells are classified into three categories given their own load conditions. Passive cells correspond to nodes whose load is below the ρ_{low} threshold and are willing to accept more traffic. On the contrary, those above ρ_{high} are denoted as active and constitute cells declared in overload. Ultimately, cells operating within the $[\rho_{low}, \rho_{high}]$ range (i.e. hysteresis region) are characterized as neutral ones as they do not participate in the MLB procedures. The reason is for avoiding system instability due to load fluctuations around ρ_{target}, which in turn may trigger repetitive load balancing handovers.

Table 3.1: MLB Cell State Characterization

Cell Status	Own load Condition
Passive	$\widetilde{\rho}_c^{sm}(t_k) < \rho_{low}$
Neutral	$\rho_{low} \leq \widetilde{\rho}_c^{sm}(t_k) \leq \rho_{high}$
Active	$\widetilde{\rho}_c^{sm}(t_k) > \rho_{high}$

Formally, all MLB procedures are triggered by active cells. Therefore, upon overload declaration, an active cell has to estimate a CIO decrement so that excess traffic is offloaded to adjacent cells. In addition, it informs its passive neighbors over the X2 interface to report back a proposed ΔCIO modification subject to their own cell load conditions. Both active and passive cells need to estimate the required load shift, l_s, so that they achieve the targeted cell load status. Since CAC expresses the proportional load difference from ρ_{target}, l_s can be defined as:

$$l_s = \frac{1}{1 - CAC} \tag{3.3}$$

Once l_s is estimated, it is basically mapped to a CIO modification, ΔCIO_p, as follows:

$$\Delta CIO_p = max\{min\{10 \cdot log_{10}(l_s), \Delta CIO_{p,bound}\}, -\Delta CIO_{p,bound}\}, \tag{3.4}$$

where $\Delta\text{CIO}_{p,bound}$ specifies the bounds of the ΔCIO_p suggestion. Albeit larger values of $\Delta\text{CIO}_{p,bound}$ can be employed in macro-only networks, more conservative adjustments are recommended for HetNet deployment owing to the associated mobility challenges. For that purpose, ΔCIO_p is considered to be bounded within the [-1,1] dB interval.

To illustrate the properties of (3.4), ΔCIO_p is plotted versus the own cell load in Fig. 3.4 for different ρ_{target} configurations. It can be observed that $\Delta CIO_p < 0$ and $\Delta CIO_p > 0$ for active and passive cells respectively. Apparently, this is

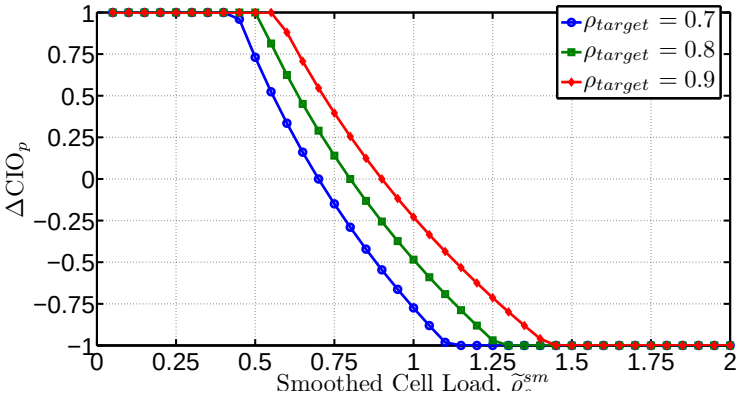


Fig. 3.4: Mapping l_s to ΔCIO_p for different ρ_{target} configurations.

a desired behavior, i.e. active cells make handovers towards passive neighbors more likely to happen, while passive cells decrease the probability of triggering handoff procedures towards active neighbors. Moreover, $|\Delta CIO_p|$ increases with l_s. Therefore, cells with $\widetilde{\rho}_c^{sm} \ll \rho_{target}$ or $\widetilde{\rho}_c^{sm} \gg \rho_{target}$ propose $\Delta CIO_p =$ -/+1 dB, while smaller ΔCIO_p values correspond to cell load conditions with $\widetilde{\rho}_c^{sm} < \rho_{target}$ or $\widetilde{\rho}_c^{sm} > \rho_{target}$, respectively.

Notice that the estimated ΔCIO_p might differ between peer neighbors due to the fact that each cell calculates the proposed coverage adjustment based on its own load conditions. To avoid asymmetrical cell range shrinking/extension, the minimum value is selected for each negotiated cell pair as follows:

$$\Delta CIO = min\{\left|\Delta CIO_p^{active}\right|, \Delta CIO_p^{passive}\} \tag{3.5}$$

By means of that, the configured CIO values for each direction are updated as follows:

$$\begin{aligned} CIO_{i\to j} &\leftarrow max\{CIO_{i\to j} - \Delta CIO, CIO_{MLB}^{min}\} \\ CIO_{j\to i} &\leftarrow min\{CIO_{j\to i} + \Delta CIO, CIO_{MLB}^{max}\}, \end{aligned} \tag{3.6}$$

where i is the active cell and j is its passive peer neighbor.

As such procedures require resources over the X2 interface, their event-based triggering – i.e. whenever a cell is declared active – tries to maintain the related overhead at a reasonable level. This essentially means that the scheme tolerates slight load imbalances as long as users are satisfied within the cell. By means of that, the signaling overhead over the X2 interface is limited. Notice

that communication between peer neighbors freezes for some specific time – denoted as T_{Wait} – after a CIO reconfiguration. The reason is for letting cell load stabilize after the associated handover executions. If T_{Wait} expires and the cell is still tagged as active, then the process is triggered again. T_{Wait} is a parameter that can be configured by the operator.

3.4.2 MRO Algorithm

The MRO solution builds on the work in [63]. Unlike MLB, MRO decisions are conducted periodically based on mobility statistics which are collected during observation periods, also referred to as MRO cycles. In particular, the following handoff KPIs are recorded at every MRO cycle on a cell pair basis:

- N_{sHO} is the total number of successful handovers
- N_{RLF} is the total number of declared RLFs
- N_{WHO} is the total number of wrong handovers
- N_{TLHO} is the total number of too late handovers
- N_{TEHO} is the total number of too early handovers
- N_{PP} is the total number of ping pong handovers

The associated MRO decisions are realized according to the flow chart depicted in Fig. 3.5. Notice that MRO actions are triggered only if a minimum amount of samples, N_{min}, is collected during the MRO cycle. If not, then MRO continues to gather samples and waits for the next cycle to determine whether N_{min} is achieved. This is for ensuring MRO decisions based on a statistically reliable KPI collection. More information about the number of samples required for obtaining a reliable estimation of the measured mobility performance can be found in Appendix B.

Given that the amount of mobility observations suffices, MRO checks whether the estimated RLF probability, defined as $\pi_{RLF} = N_{RLF} \,/\, (\, N_{sHO} + N_{RLF})$, is below the targeted T_{RLF} threshold. If so, no major mobility problems are detected allowing the extension of the operational MLB range by Δ dB. Otherwise, the algorithm has to determine which type of handoff problems dominates the cell boundary.

To achieve this goal, MRO groups mobility problems into 2 generic handover types, denoted as "too fast handovers (N_{TFHO})" and "too slow handovers (N_{TSHO})" respectively. Albeit $N_{TSHO} = N_{TLHO}$, N_{TFHO} comprises the impact of too early handovers, wrong handovers and ping-pong events. Hence, N_{TFHO} is defined as follows:

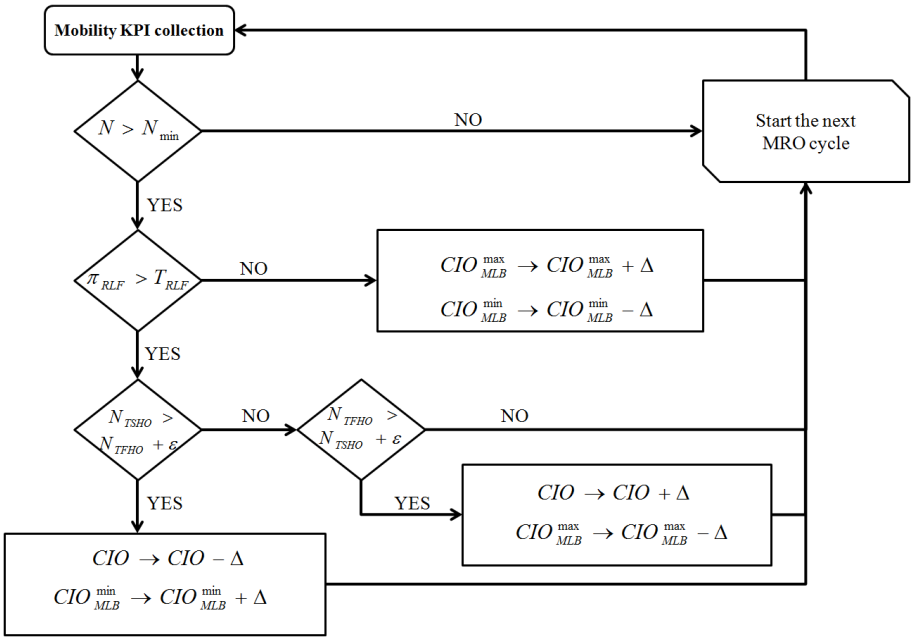


Fig. 3.5: Flowchart of the joint MLB and MRO flowchart.

$$N_{TFHO} = N_{TEHO} + N_{WHO} + w_{PP} \cdot N_{PP}, \tag{3.7}$$

where $w_{PP} \in [0,1]$ and weights the associated ping pong counts. The reason is that ping pong handovers waste network resources but at the same time they might save the connection from too early or wrong handovers. By means of that, w_{PP} is operator configurable and essentially determines the ping-pong severity. E.g. when w_{PP} equals to 1, ping-pong handovers are fully counted as mobility problems while they are simply ignored if w_{PP} is set to 0. Once N_{TFHO} and N_{TSHO} are calculated, the following inequalities are checked:

$$\pi_{TSHO} > \pi_{TFHO} + \varepsilon \tag{3.8}$$

$$\pi_{TFHO} > \pi_{TSHO} + \varepsilon, \tag{3.9}$$

where ε is the parameter utilized for determining which problematic handover group dominates the cell boundary; while π_{TSHO} and π_{TSHO} are associated probabilities of too fast and too slow handovers, respectively. Apparently, solely one or even none of (3.8), (3.9) can be met at each MRO cycle. In the second case, no actions are taken since the algorithm cannot figure out the dominant root cause of mobility problems and prefers to maintain the same handoff configuration for the sake of system stability.

On the contrary, fulfilling (3.8) indicates that mobility performance is primarily degraded by too late handovers. To increase the probability of faster handover executions, MRO has to decrement the latest CIO configuration by Δ dB, while incrementing – by the same Δ value – the MLB associated CIO^{min}_{MLB} limit. The latter restricts MLB in a sense that any increase of the CIO^{min}_{relax} parameter inevitably shortens the time-of-stay in underutilized cells for mobility robustness purposes. Similarly, given that (3.9) is met, the root cause of the mobility deterioration comes from too fast handovers. To resolve this undesired effect, parameters should be adjusted in a manner that handovers executions are slightly delayed. Thus, MRO increases the CIO configuration while limiting the upper MLB bound in an effort to make mobility setting less aggressive.

3.5 Simulation Assumptions

System level simulations are conducted in the co-channel deployment described in Section 2.5, where both macrocells and picocells are deployed at the 2.6 GHz sharing 10 MHz of bandwidth. Two high traffic areas (hotspots) are generated per macrocell area and picocells are concentrically deployed. CBR traffic is simulated for a fixed number of $N_u = 30$ users is considered per macrocell area and the associated offered traffic demand is defined as the product of N_u with the requested CBR rate[1]. Lastly, low mobility at 3 km/h is assumed for hotspot users, while a mixture of different speeds – with equal probability – is considered for free moving UEs at 3 km/h and 50 km/h, respectively. The handover procedure is completed 100 msec after the event reception at the network side, taking into account the time required for preparing and executing the handover.

To evaluate the trade-off between the capacity enhancements provided by load balancing and the potential mobility performance degradation, the following cases are simulated:

- *No RE*: No CIO adjustments are performed. A fixed handover offset of 3 dB is assumed for all cell pairs; hence, the inbound and outbound picocell CIO configuration equals to $\text{CIO}_{m\to p} = \text{CIO}_{p\to m} = 3$ dB.
- *Fixed RE*: Picocell measurements are virtually biased by 3 dB so that they attract more users. As discussed in Section 3.2, this essentially means that $\text{CIO}_{m\to p}$ and $\text{CIO}_{p\to m}$ effectively equal to 0 dB and 6 dB, respectively. Notice that the applied RE value is chosen according to the recommendations in [72], for cases when no eICIC is employed.
- *MLB*: In this case, $\text{CIO}_{m\to p}$ and $\text{CIO}_{p\to m}$ are dynamically adjusted within the range of $[\text{CIO}^{min}_{MLB}, \text{CIO}^{max}_{MLB}] = [0\ 6]$ dB. Cell load obser-

[1]For more information regarding the scheduling assumptions, please refer to Appendix A

Table 3.2: Simulation Parameters

Parameter	Value
Picocells/Macrocell area	2
Number of Users	30
Traffic Type	CBR @{512:64:832} kbps
$\text{TTT}_{m\to m}$	480 msec
$\{\text{TTT}_{m\to p}, \text{TTT}_{p\to m}\}$	256 msec
Layer-3 Filtering Coefficient	1
MLB	
$[\text{CIO}_{MLB}^{min}, \text{CIO}_{MLB}^{max}]$	[0 6] dB
Cell Load smoothing coefficient, α_l	0.2
Cell Load Measurement Interval	0.5 sec
$\{\rho_{low},\ \rho_{target},\ \rho_{high}\}$	{0.7, 0.8, 0.9}
T_{Wait}	4 sec
MRO	
$[\text{CIO}_{min}, \text{CIO}_{max}]$	[0 6] dB
MRO cycle	90 sec
Δ	0.5 dB
$\{\varepsilon, \text{w}_{PP}\}$	{0.03, 0.25}
Fixed RE	
$\{\text{CIO}_{m\to p}, \text{CIO}_{p\to m}\}$	{0, 6} dB

vations are performed periodically every 500 msec and the associated measurements are smoothed with a forgetting factor of $a_l = 0.2$. Lastly, ρ_{target} is set to 0.8, while ρ_{low} and ρ_{high} equal to 0.7 and 0.9, respectively.

- *MLB+MRO*: The MLB range is controlled by MRO. $[\text{CIO}_{MLB}^{min}, \text{CIO}_{MLB}^{max}]$ is set to [0 0] dB at the beginning of the simulation and is dynamically modified at each MRO cycle subject to the measured mobility performance. The duration of a MRO cycle is 90 sec and N_{min} is set to 20 handover attempts[2]. Notice that the targeted RLF probability is set to 2%, while all MRO-applied decisions are realized by steps of Δ = -/+ 0.5 dB. Ultimately, the maximum allowable CIO range is configured to $[\text{CIO}_{min}, \text{CIO}_{max}]$ = [0,6] dB.

The simulation length is constituted by 15 MRO cycles and system performance is evaluated by taking into account only the last 5 observation periods. Note that other mobility parameters such as TTT settings and the Layer-3 filter-

[2] As seen by Appendix B, a significantly larger amount of samples is required for ensuring a small error in the estimation of the measured RLF probability. However, this would made the simulation time practically unfeasible.

ing co-efficient of UE measurements are fixed for all simulated cases in line with the recommendations in [76].The most relevant simulation parameters are summarized in Table 3.2.

3.6 Performance Results

Fig. 3.6 illustrates the configured CIO values for all recorded successful inbound and outbound picocell handovers. The depicted statistics correspond to a high offered traffic simulation and are presented in the form of empirical Cumulative Distribution Functions (CDF). Note that the fixed RE case is not included as $\text{CIO}_{m\to p}$ and $\text{CIO}_{p\to m}$ are always equal to 0 dB and 6 dB respectively. Particularly, it can be observed that even in the standalone MLB operation, CIO configurations do not always reach the maximum relaxation values as they depend on the cell-pair load conditions at the time of the handover. Moreover, the larger variability of the outbound CIO parameterization further indicates that there are time instances when even picocells are overloaded, requesting $\text{CIO}_{p\to m}$ to be decreased. When MRO is enabled, its effect is mainly visible at the outbound picocell boundary in order to resolve the mobility problems cause by too late outbound picocell handovers. Hence, the

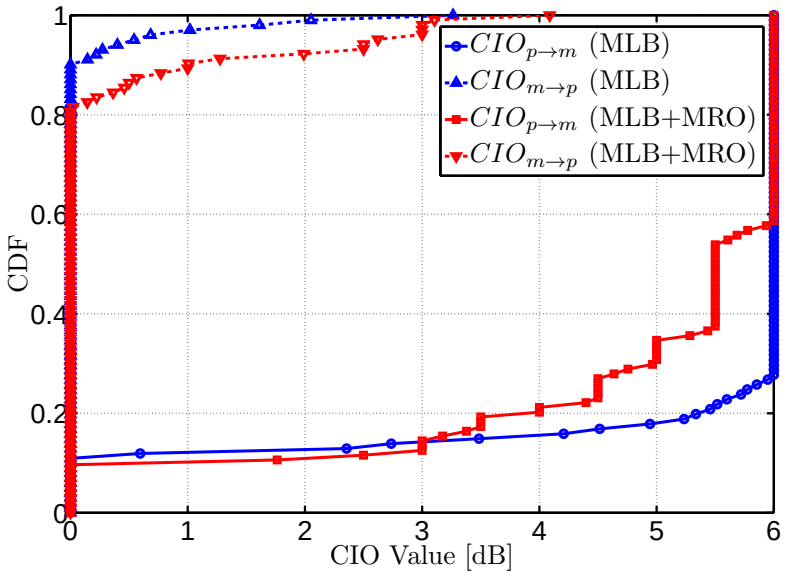


Fig. 3.6: Configured CIO values for all recoded successful handovers in both cell type directions. High traffic conditions are assumed.

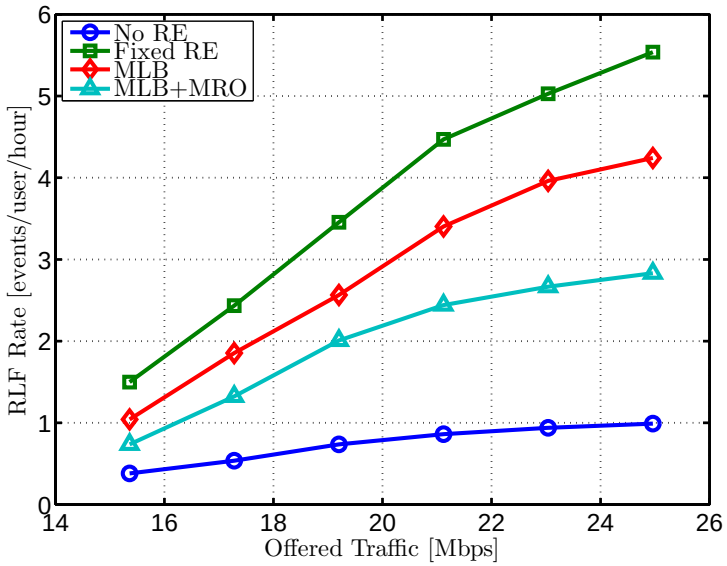
MLB applied CIO adjustments are constrained in an attempt to make handover executions faster. On the contrary, the MRO impact on $\text{CIO}_{m \to p}$ is much less noticeable since inbound picocell handovers do not typically suffer from severe mobility problems.

The estimated RLF rates versus the offered load for all simulated schemes are shown in Fig. 3.7a. It can be seen that the RLF occurrence increases monotonically with the offered traffic demand. This is due to the co-channel interference that naturally becomes stronger; making the outbound picocell handover even more challenging. Albeit no major problems are observed when none load balancing mechanism is activated, both RE and MLB degrade handoff performance significantly. Notice that the RLF rates for the standalone MLB operation are lower than the fixed RE case simply because MLB does not always configure the maximum allowable CIO adjustment of 3 dB. On the other hand, MRO tries to bring RLF performance within a more acceptable range. E.g. at the high load conditions of 25 Mbps, the MLB+MRO configuration decreases the estimated RLF rates by ∼33% relative to the standalone MLB operation, while higher reduction gains of ∼50% are achieved over the fixed RE configuration. Nonetheless, even when MRO is enabled, the RLF performance is still almost three times worse than the case without load balancing.

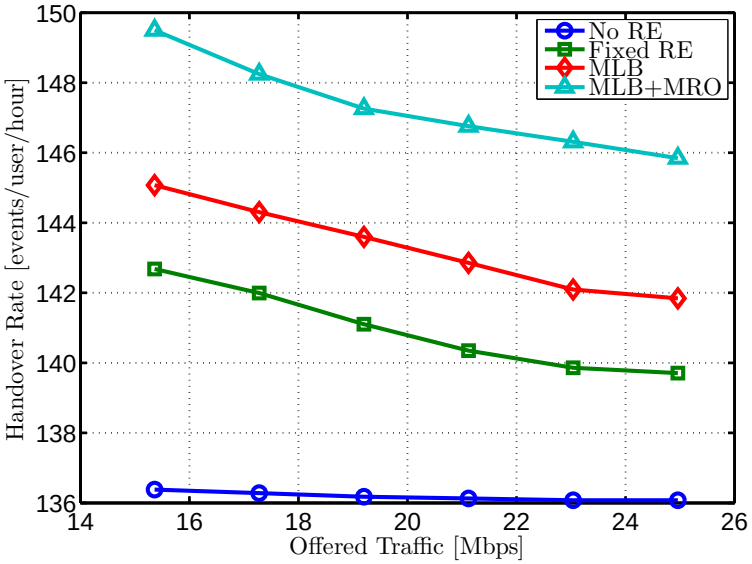
Fig. 3.7b depicts the corresponding handover rates. As expected, all employed load balancing schemes result in increased handoff rates compared to the case without load balancing. The reason that more handovers are triggered in an attempt to push excess traffic to the deployed picocells. Among the simulated cases, MRO+MLB is apparently the most costly approach due to the faster outbound picocell handover executions induced by MRO. However, the associated cost is rather affordable since the estimated handover increase over the static RE configuration is only ∼5%. Ultimately, it is worth investigating the RLF probability achieved by MLB+MRO. This can be extracted by dividing the RLF rate over the rate of the overall handover attempts (i.e. successful handovers plus RLFs). In fact, the proposed solution meets the 2% target for all offered load conditions providing a convenient means of autonomously controlling mobility robustness.

The ratio of users offloaded to the picocell layer together with the associated Jain's fairness index[3] are shown in Fig. 3.8a and Fig. 3.8b, respectively. It can been observed that the fixed RE configuration maximizes the picocell offloading since the small cell measurements are always biased by 3 dB. Particularly, it offloads ∼9% more users to the picocell layer compared to when no load balancing mechanism is used. For MLB+MRO, the related picocell offloading ratio is ∼1.5% less than the one achieved by the static RE configuration. The reason is that MRO essentially reduces the time-of-stay in the picocell layer for the sake of mobility robustness. As for the Jain's index, similar trends are seen. The

[3]The definition of Jain's fairness index is provided in (2.13) and illustrates the degree of achievable load balancing.

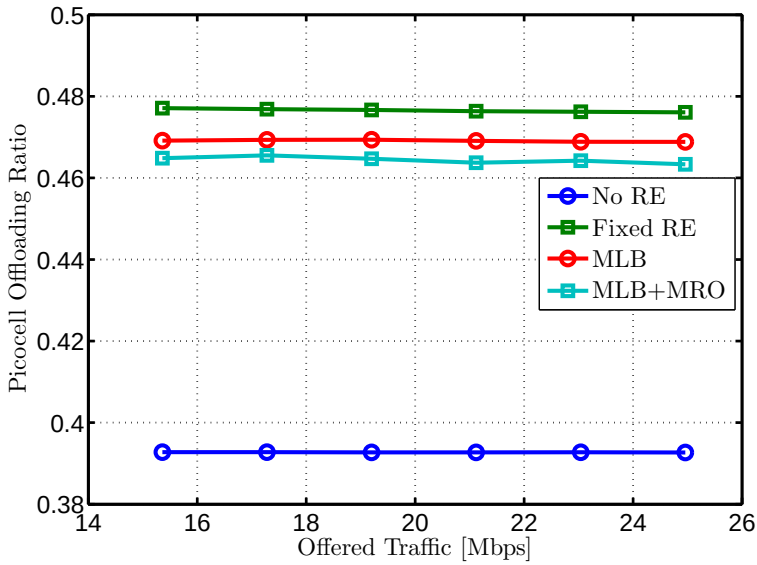


(a) RLF rates

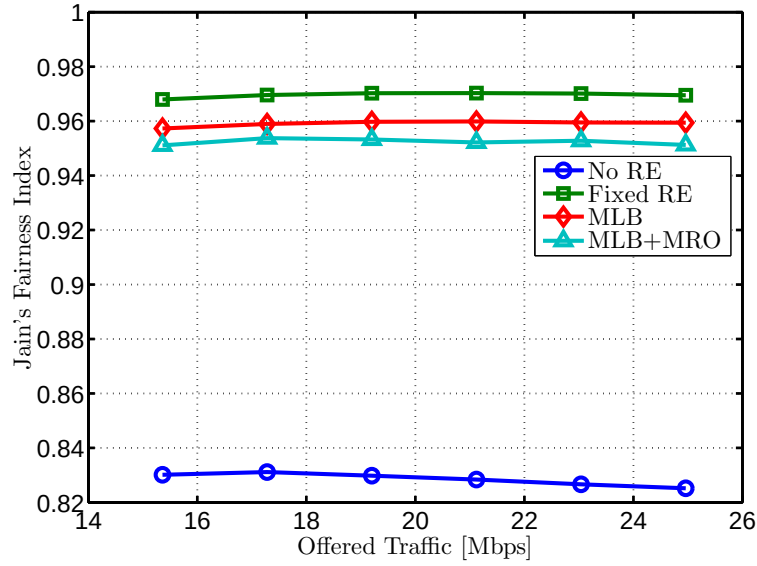


(b) Handover rates

Fig. 3.7: Estimated mobility performance for different CIO configuration policies and offered traffic demands.



(a) Ratio of users offloaded to picocells



(b) Jain's fairness index

Fig. 3.8: Picocell user offloading and degree of achievable load balancing for different CIO configuration policies and offered load conditions.

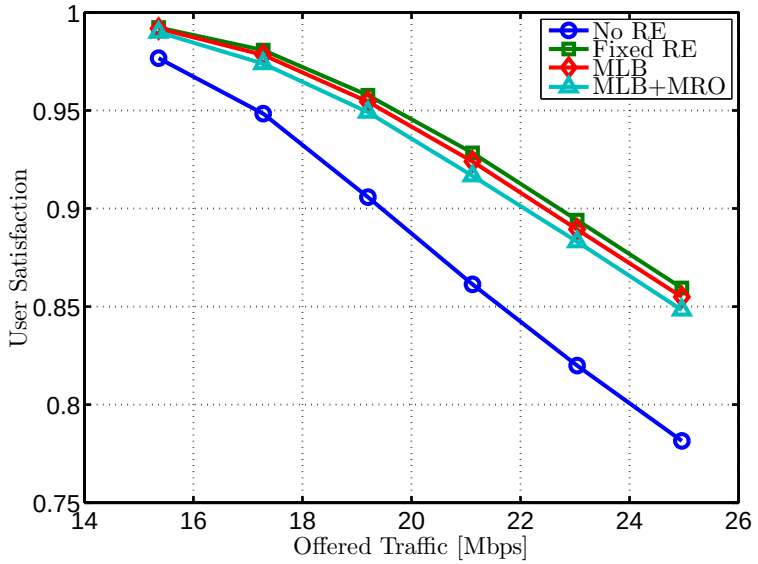


Fig. 3.9: User satisfaction for different CIO configurations and offered load conditions. The user satisfaction also includes the impact of RLFs, during which, users are essentially declared unsatisfied.

fixed RE configuration achieves the load balancing performance as this methods maximizes the user offloading to the picocell layer. It is worth mentioning that the aforedescribed KPIs are primarily impacted by the CIO adjustment range. Consequently, they cannot further improve unless larger fixed/dynamic RE values are configured. However, this is feasible only if interference coordination – i.e. eICIC – is applied on top.

Lastly, Fig. 3.9 illustrates the user satisfaction for all simulated configurations and offered traffic demands. Recall from Section 2.7 that this is defined as the ratio of users that on average meets the QoS requirement. Despite the significant mobility enhancements achieved by MRO, MLB+MRO results in a marginally worse performance than the fixed RE scheme. The reason is that the associated MRO improvements in terms of RLF reduction refer to time instances when the users experience very low SINR conditions (i.e. below -8 dB). As the number of these time instances constitute less than 2% of the overall pool of samples used for the post-processing of statistics, user satisfaction is mainly affected by the picocell user offloading ratio. By means of that, the better offloading performance of the static RE scheme is also reflected in the UE experience metric. Assuming a target outage of 10%, the capacity gains over the case without load balancing are in the order of ∼15%, as this potion

of additional traffic can be carried by the network with the same outage.

3.7 Conclusions

This chapter has addressed the problem of co-channel HetNet load balancing and mobility robustness. A joint MLB/MRO framework has been proposed that modifies the handover CIO based on both cell load and mobility observations. By means of that, the load balancing aggressiveness is autonomously controlled given its impact on mobility robustness. Performance evaluation has been conducted by means of system level simulations that explicitly model the LTE handoff procedures as well as 3GPP proprietary functionalities that allow the tight coordination between adjacent cells.

Simulation results have shown that although a static RE value maximizes the user offloading to the picocell layer, it could result in unacceptable RLF rates. The reason is that the outbound picocell handover becomes even more challenging when RE is employed since users are forced to stay longer connected to low-power nodes for load balancing purposes. On the other hand, the dynamic adaptation of the handoff parameterization based on load and the estimated mobility performance seems a more attractive solution. Specifically, the coordinated MLB/MRO operation halves the RLF rates – when these are compared against the fixed RE configuration – maintaining adequate user satisfaction with a marginal increase in terms of handover events.

All in all, the proposed MLB/MRO solution finds a better trade-off between the capacity gains and the associated mobility robustness. However, it does not fully resolve the mobility problems. For that purpose, it is recommended to be complemented with additional features for fulfilling this goal. These may involve state-of-art solutions that explicitly manage high velocity UE devices or the usage of interference coordination schemes.

Chapter 4

Inter-Frequency HetNet Load Balancing

4.1 Introduction

As mobile broadband traffic continues to grow, operators are compelled to deploy more carriers. Macrocell upgrades by means of adding more spectrum in existing site locations will constitute the initial network evolution phase, followed by the strategical placement of small cells in traffic-critical urban areas. Apparently, balancing the load across the deployed frequency layers is a key prerequisite for boosting network performance; however, it has to be performed in a efficient manner. This calls for the development of solutions that not only achieve high system performance, but also maintain the associated load balancing costs at a reasonable level. These include the signaling burden for triggering load-driven handovers together with the impact of inter-frequency cell discovery mechanisms on user throughput and terminal power consumption.

The remainder of the chapter is organized as follows: The problem of inter-frequency load balancing is discussed in Section 3.2, followed by the literature review in Section 3.3. Section 3.4 describes the proposed inter-frequency load balancing framework covering traffic steering schemes for both idle and connected UE devices. The simulation assumptions are summarized in Section 3.5, while the corresponding performance results are presented in Section 3.6. Lastly, the chapter concludes with Section 3.7.

4.2 Problem Delineation

The dilemma for inter-frequency load balancing is that while the associated measurements increase the offloading opportunities towards neighbor carriers, they come at the expense of terminal power and transmission gaps. The latter may impact significantly the perceived data rates since the UE device cannot be scheduled during these measurement periods. For that purpose, inter-frequency measurements in connected mode are essentially controlled by the A2 event[1], meaning that these are triggered whenever the serving radio conditions drop below a network configurable threshold. We denote the related parameter as $A2_{Thresh}$.

Typically, $A2_{Thresh}$ is set to a relatively low value so that inter-frequency measurements are configured in areas where the UE device is about to experience coverage problems at its serving carrier. Performing load balancing by means of adjusting handover parameters may fall short of moving cell center users from carrier to another or pushing excess traffic to small cells deployed at a different carrier. The reason is that inter-frequency handovers are seldom triggered in network locations with good macrocell coverage due to the lack of UE physical layer measurements.

In a attempt to fully exploit the large overlapping coverage of inter-frequency deployments, a rather simple approach would be to configure all handover procedures on RSRQ measurements, while increasing $A2_{Thresh}$. As it can be seen by Fig. 4.1, the RSRQ measure partially captures cell load information due to the RSSI contribution[2]. Such property is more visible at cell center locations, where the measured RSRQ may vary several dB subject to the load conditions. On the contrary, the observed dependency diminishes while moving towards the cell edge. Given that the $A2_{Thresh}$ configuration provides sufficient measurement availability to discover cells deployed at a different carrier, RSRQ-based mobility can operate as a passive traffic steering mechanism, where inter-frequency handovers/cell reselections are triggered due to the load imbalance of neighbor carriers. However, such an approach does not ensure that cell discovery mechanisms are efficiently utilized. By means of that, the resulting cost in terms of handovers might be relatively high, while the unnecessary physical measurements could jeopardize UE power consumption and perceived throughput.

This calls for the development of load balancing solutions that elegantly control standardized mechanisms such as measurement rules, mobility configurations

[1]For more details refer to Table 2.2, where all 3GPP defined measurement events are outlined.

[2]Recall that RSRQ is defined as the RSRP over the RSSI, where RSSI comprises the linear average of the received wideband power including co-channel serving and non-serving cells, adjacent channel interference, thermal noise, etc [61]. For more information refer to Appendix A.

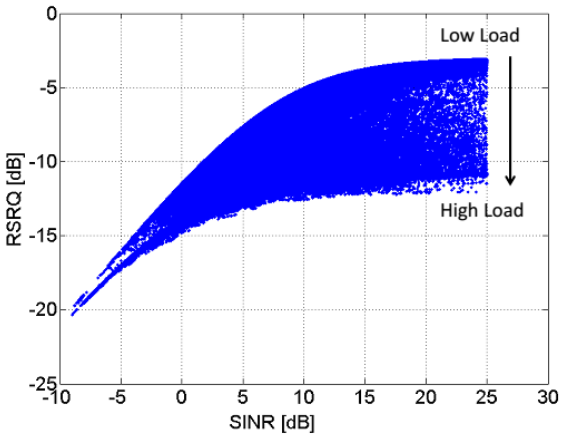


Fig. 4.1: Load dependent relation between SINR and RSRQ

and information exchange between neighboring cells or between the network and the UE device. It is worth mentioning that the problem is not only limited to connected mode. Specifically, further emphasis is put on the idle mode AP framework used for triggering priority-based cell reselections. As pointed out in Section 2.3, the target is to enhance the degree of network control over idle UE devices in combination with providing more energy efficient mechanisms for discovering inter-frequency cell neighbors in idle mode. The latter mainly involves users camping on a lower priority carrier which typically scan for prioritized carriers even if they are out of their coverage.

All in all, the derived traffic steering framework will concurrently perform load balancing in both RRC modes. However, as idle and connected mode procedures distinctively differ, the proposed methods have to avoid decisions that result in conflicting policies between the different RRC states. This is for minimizing the occurence of idle-to-connected (and vice versa) ping pong events[3] which naturally increase the network RRC signaling as well as they impact UE power consumption. In this context, aligning idle and connected mode is of crucial importance so as to achieve dynamic load balancing with a reasonable cost in terms of handovers and cell reselection.

4.3 State-of-Art

State-of-art literature covers a handful of research topics in deployments with multiple frequency layers. Among others, these include the derivation of rules

[3]Recall that an idle-to-connected ping pong refers to the situation when a user switching to connected mode is immediately handed over to a different cell.

for assigning users onto different carriers, evaluation of inter-frequency mobility performance, UE power savings via measurement optimizations and dynamic load balancing by means of adjusting handover parameters.

Simplified carrier load balancing studies can be found in [49,91–93]. User assignment is typically conducted based on factors such as the achievable throughput per carrier, number of served users, UE capabilities, coverage, etc. Albeit their valuable contributions in terms of deriving the optimal user distribution across the deployed carriers, they commonly neglect mobility effects as well as the associated RRC signaling for achieving a particular performance.

Concerning relevant mobility investigations, the work in [94] compares the suitability of RSRQ and RSRP measurements for triggering inter-frequency handovers. The bottom line is that RSRQ measurements result in higher handover rates and ping pong events due to their sensitivity to load fluctuations. In an attempt to improve inter-frequency mobility, a joint handoff measure based on both measurement entities is proposed in [95]. The suggested scheme achieves better performance than the case when a single measurement entity – either RSRQ or RSRP – is used for triggering inter-frequency handovers.

Although the aforementioned mobility studies refer solely to UE devices in connected mode, some work is done for idle mode as well. In particular, the impact of idle mode mobility on UE power consumption is analyzed in [57], employing inter-frequency ranking-based cell reselections. Results indicate that power savings up to 30% can be achieved if cell reselections are kept at a reasonable level by optimizing measurement thresholds and hysteresis values. However, the study is performed in the form of a sensitivity analysis for different mobility settings, without proposing any method for their autonomous configuration as well as investigating the impact on user throughput. Similarly, the work in [96] gives a deep insight into the suitability of AP-based cell reselections as a means of idle mode traffic steering. The limitations regarding the assignment of camping priorities are further discussed but no concrete solution is recommended.

Connected mode load balancing by means of modifying mobility parameters is thoroughly investigated in macro-only deployments, focusing mainly on inter-RAT scenarios. The reason is that inter-system handovers are usually triggered by traffic steering mechanisms aiming at realizing a particular operator policy. In fact, the methods described in [97, 98] illustrate a promising load balancing performance. Moreover, their application to inter-frequency deployments within the same RAT should be rather straightforward, requiring minimal modifications. Nevertheless, they hardly provide any information about the associated cost in physical layer measurements.

With the emergence of HetNets, the topic of efficient cell neighbor discovery has gained significant momentum. To discover small cells deployed at a different carrier without performing unnecessary measurements, several state-of-art solutions are already proposed. Specifically, the work in [99] proposes the re-

laxation of the measurement periodicity for low mobility users so that they measure other carrier frequencies less often. Some more advanced schemes are proposed in [100,101], where users are explicitly configured with inter-frequency measurements subject to geo-location information. In other words, users within the small cell vicinity are signaled to start measuring neighbor carriers. Given the reader's interest, a set of candidate solutions can be found in [32,102]. Nevertheless, they solely focus on connected mode, ignoring how idle mode functionalities can assist small cell discovery.

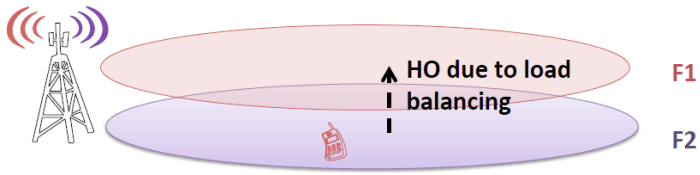
4.4 Inter-Frequency Load balancing Framework

The proposed load balancing framework assumes that the $A2_{Thresh}$ is configured to a relatively low value, while all mobility functionalities are based on RSRP measurements. To avoid unnecessary UE measurements, inter-frequency cell discovery – for load balancing purposes – is explicitly triggered whenever cell overload is detected. In other words, the network can request measurement reports from particular users so that potential offloading target cells are discovered. Such an approach decouples mobility management from load balancing, allowing the classification of inter-frequency events into load-driven and mobility-driven handovers/cell reselections.

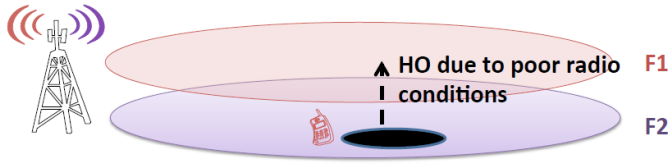
To comprehend the main principles of the suggested solution, Fig. 4.2 illustrates a representative example for connected mode devices. A load-driven handover is described in Fig. 4.2a. F2 is overloaded and the concerned UE is configured to measure neighbor carriers even if the A2 condition is not yet met. Based on the associated measurement reports and the inter-cell information exchange in terms of CAC, the cell at F1 is discovered. By means of that, the user can be forced to connect to F1 for the sake of load balancing.

On the other hand, mobility-driven inter-frequency handovers are solely triggered due to poor radio conditions. This case is illustrated in Fig. 4.2b, where the serving radio conditions drop dramatically such that the A2 condition is fulfilled and the terminal is configured to measure F1. Given that the handover mechanisms are timely executed, the UE device is handed over to F1 before experiencing a RLF at F2. Lastly, no action is taken in Fig. 4.2c. F2 is not overloaded as well as it provides full coverage. Therefore, the terminal stays connected to the same carrier.

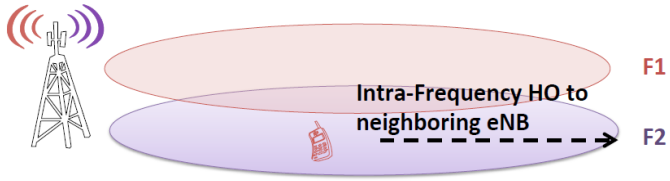
Notice that albeit the aforedescribed examples refer to connected mode, the same principles can also be applied for idle UE devices. Nevertheless, as the network does not own full control over idle mode mobility, the traffic steering decisions are taken before the connection release. By means of that, the concerned UE can be provided with suitable idle mode parameters so that it is forced to camp on the desired layer once switching to idle mode.



(a) F2 is overloaded and a load-driven handover to F1 is initiated



(b) Coverage hole on F2. A mobility-driven handover is triggered towards F1 regardless of the inter-layer load conditions so that the RLF is avoided.



(c) F2 is not overloaded as well as provides full coverage. No action is taken.

Fig. 4.2: Decoupling inter-frequency mobility from load balancing

In a nutshell, the developed framework exploits the switching instances of UE RRC state machine for realizing load balancing decisions. This essentially means that it steers UE devices that have just established a radio bearer or they are about to switch to idle mode. It is worth mentioning that the load thresholds introduced in Section 3.4 are still applicable; hence, cells whose load is above ρ_{high} are declared in overload, while those below ρ_{low} are only willing to accept excess traffic.

4.4.1 Forced Handovers Upon Connection Setup

Terminals switching to connected mode are explicitly requested to initiate inter-frequency measurements, whenever the serving load exceeds the overload detection threshold ρ_{high}. Once the associated measurements are reported to the network, a single cell per carrier is selected as follows:

$$c^* = \arg\max_k \{Q_{k,m}^{RSRP}\}, \tag{4.1}$$

Algorithm 1 Forced Handovers upon Connection Setup (FHO@CS)

$\{\mathbf{C}\} = \emptyset$
for $m = 1$ to *number of carriers* **do**
  $c^* = \arg\max_k \{Q_{k,m}^{RSRP}\}$,
  $\{\mathbf{C}\} \leftarrow \{\mathbf{C}\} \cup c^*$
end for
select target cell $c = \arg\max_{c^*} \{CAC_{c^*}\}$
if $\widetilde{\rho_c^{sm}} < \rho_{low}$ **then**
  if $Q_c^{RSRP} \geq A_{thresh}^{RSRP} \wedge Q_c^{RSRQ} \geq A_{thresh}^{RSRQ}$ **then**
    initiate load balancing handover to cell c
  end if
end if

where $Q_{k,m}^{RSRP}$ is the measurement report for cell k on carrier m in terms of RSRP; and c^* is the strongest RSRP measured cell per carrier. We denote as C the constructed set of candidate target cells; hence $c^* \in \mathbf{C}$. The cells in $\mathbf{C}$ are sorted in a descending CAC order and the neighbor c with the highest capacity is chosen as the final handover target. The load state of cell c is extracted by the CAC value and if it is below ρ_{low}, an inter-frequency handover to that particular cell is initiated, given that the following constraints are satisfied:

$$Q_c^{RSRP} \geq A_{thresh}^{RSRP}, \tag{4.2}$$

$$Q_c^{RSRQ} \geq A_{thresh}^{RSRQ}, \tag{4.3}$$

where Q_c^{RSRP}, Q_c^{RSRQ} are the related RSRP and RSRQ measurements reports for cell c; and A_{thresh}^{RSRP}, A_{thresh}^{RSRQ} correspond to the minimum RSRP and RSRQ requirements that should be met by the handover target. It is worth mentioning that the A_{thresh}^{RSRP} parameter in (4.2) is configured Δ dB higher than $A2_{Thresh}$ value so that the handed over UE device does not initiate inter-frequency measurements after the handover completion. In such a manner, the probability of an inter-frequency ping-pong handover is significantly reduced, while interference information about cell c is extracted by (4.3).

The aforedescribed scheme is denoted as Forced Handovers upon Connection Setup (FHO@CS) and steers users to cells that provide sufficient coverage and capacity. As it can be seen by Fig. 4.3, this does not necessarily mean that forced handovers are always initiated towards the least loaded neighbor. It is of vital importance to ensure that the handed over UE device is well-covered by the target cell. Apart from avoiding inter-frequency ping-pong events, such an approach guarantees that mobility robustness on the redirected carrier is not jeopardized by extra handovers or RLF declarations.

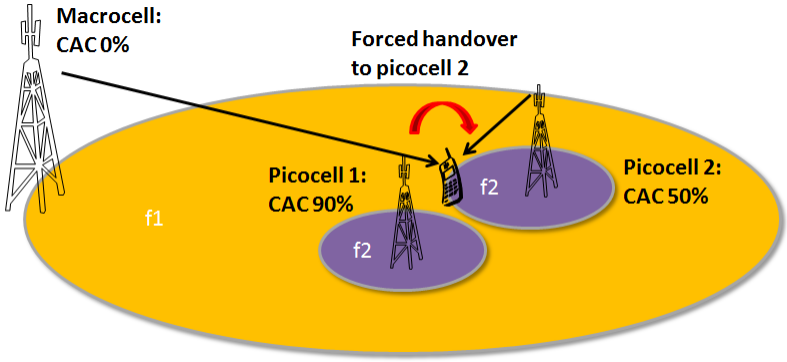


Fig. 4.3: Example operation of Algorithm 1 (Forced handover upon connection establishment). Albeit the picocell 1 is the least loaded neighbor, the UE device is steered to picocell 2, since it is the best cell at the redirected carrier. By means of that, extra handovers or RLFs on f2 are avoided.

4.4.2 Dedicated Priorities at Connection Release

To improve the assignment flexibility of camping priorities, a new scheme is proposed – denoted as Dedicated Absolute Priorities (DAP) – that expands the AP framework on a UE specific resolution. As discussed in Section 3.2, this can be realized by exploiting the RRC Connection Release message for providing the UE device with dedicated mobility parameters. By means of that, different camping priorities can be signaled to users subject to their physical location and the load per carrier.

The priority adjustment mechanism is described in Algorithm 2 and it generally follows the same principles as Algorithm 1. In particular, the highest priority is assigned to the least loaded carrier that well-covers the concerned UE. For that purpose, both (4.2) and (4.3) are used for the construction of the **C** set. The measurements for checking the coverage of a particular carrier might be available due to previously performed connected mode measurements; otherwise, an explicit measurement trigger could be used for requesting such measurements before the RRC connection release.

Moreover, A^{RSRP}_{Thresh} should be configured to a value that is greater or equal to $Thresh^{AP}_{sLow}$. The reason is that $Thresh^{AP}_{sLow}$ is the parameters that determines the coverage of a prioritized carrier. Given that (4.2) is satisfied, then DAP ensures that the concerned UE device will camp on the selected carrier as it is assigned with the highest priority. Nevertheless, it is recommended to set $Thresh^{AP}_{sLow}$ equal to $A2_{Thresh}$ in order to align the mobility-driven events in connected and idle mode respectively. Inevitably, carriers failing to meet (4.2) and (4.3) are given the lowest priority.

Algorithm 2 Dedicated Camping Priorities at Connection Release (DAP)

$\{\mathbf{C}\} = \emptyset$
for $m = 1$ to *number of carriers* **do**
  $c^* = \arg\max_k \{Q_{k,m}^{RSRP}\}$ *subject to* $Q_{k,m}^{RSRP} \geq A_{thresh}^{RSRP} \wedge Q_{k,m}^{RSRQ} \geq A_{thresh}^{RSRQ}$
  $\{C\} \leftarrow \{C\} \cup c^*$
end for
select target cell $c = \arg\max_{c^*} \{CAC_{c^*}\}$
if $\widetilde{\rho_c^{sm}} < \rho_{low}$ **then**
  sort cells in **C** in descending CAC order and derive the associated carrier list $\{f\}$.
  assign the highest priority to the first carrier of $\{f\}$ and continue in a descending order.
  if $|\{f\}| <$ *number of carriers* **then**
    find the unavailable carriers and move them to the end of the list by assigning them the remaining lower priorities based on CAC.
  end if
end if

4.5 Simulation Assumptions

DAP+FHO@CS is evaluated by means of system level simulations in scenario B, as presented in Section 2.6.1. Thus, both macrocells and picocells share a 20 MHz carrier at 2.6 GHz, being supplemented by an additional macrocell carrier at 800 MHz with 10 MHz of bandwidth. The standardized RSRQ-based mobility framework is used as a reference, considering three different $A2_{Thresh}$ setups. For these configurations, inter-frequency handover and cell reselections are triggered by exploiting the in-built load information that is available in the RSRQ measure. Broadcast AP are used in idle mode and the 2.6 GHz is prioritized over the 800 MHz band owing to its larger transmission bandwidth and the picocell deployment. No other means of traffic steering is employed and there is no discrimination between load-driven and mobility-driven events.

Finite buffer best-effort traffic is simulated for different offered traffic demands. A Poisson packet arrival is assumed and the size per burst is negatively exponentially distributed with a mean value of 3 Mbits. As the number of users is fixed for all simulated cases, the offered traffic demand per macrocell area is essentially controlled by the packet interarrival time. Shortly after the downlink buffers are emptied, users switch to idle mode. The connection release timer is set to 1 sec, while UE devices are assumed to switch to connected 0.1 sec after the packet arrival. Since the cell load conditions will differ significantly from one simulation to another, two different ρ_{target} configurations are considered for DAP+FHO@CS. The first assumes that $\rho_{target} = 0.8$ is fixed for all of-

Table 4.1: Generic Simulation Parameters

Parameter	**Value**
Number of picocells per macrocell area	2
Number of Users	30
User Distribution	Hotspot
Offered Traffic	{20:5:40} Mbps
Traffic Type	Finite-Buffer Best Effort
Mean Packet Size	3 Mbps
User Velocity	3 km/h
Measurement Filtering Coefficient	4
Intra Frequency Handover	RSRP-based A3 event
Intra A3 Offset	3 dB
Inter Frequency Handover	RSRP or RSRQ-based A3 event (depending on configuration)
TTT	456 msec (for any handover type)
Intra Frequency Cell Reselection	R criterion-based
Inter Frequency Cell Reselection	priority-based
Broadcast AP	{2600 MHz, 800 MHz} = {HIGH, LOW}
$T_{reselection}$	1.5 sec
RSRQ-based Configuration	
$A2_{Thresh}$	{-12,-14,-16} dB
$Thresh^{AP}_{sLow}$	$A2_{Thresh}$
$\{Thresh^{AP}_{High2Low}, Thresh^{AP}_{2High}\}$	$\{Thresh^{AP}_{sLow}$-2, $Thresh^{AP}_{2High}$-1$\}$ dB
DAP+FHO@CS	
$A2_{Thresh}$	-110 dBm
$Thresh^{AP}_{sLow}$	-110 dBm
$\{Thresh^{AP}_{High2Low}, Thresh^{AP}_{2High}\}$	{-106, -108} dBm
$A2^{RSRP}_{Thresh}$	-105 dBm
$A2^{RSRQ}_{Thresh}$	-14 dB
ρ_{target}	{0.2, 0.3, 0.5, 0.6, 0.8}
$\{\rho_{low}, \rho_{high}\}$	$\{\rho_{target}$-0.1, ρ_{target}+0.1$\}$

fered traffic conditions (*config. 1*), while ρ_{target} is properly adjusted according to the offered load in the second case (*config. 2*). $\widehat{R}$ is set to 3 Mbps. Note that this configuration is rather important for determining the load balancing

aggressiveness. Small $\widehat{R}$ values naturally decrease the virtual cell load, while significantly high values may declare all cells in overload.

Mobility management procedures are explicitly modeled for both idle and connected UE devices. The periodicity of the DRX cycle in idle mode for performing measurements and cell reselections is 1.28 sec, while measurements in connected mode are performed every 40 msec. Since the study focuses on offloading solutions for low velocity users, mobility at 3 km/h is assumed for all users[4]. Therefore, not so aggressive handover parameters are utilized. Further optimizations by means of MRO mechanisms are possible; however, these are out of the study scope. The main simulation parameters are summarized in Table 4.1.

4.6 Simulation Results

4.6.1 Interworking of Proposed Methods

Before presenting the performance comparison between DAP+FHO@CS and the RSRQ-based mobility configuration, this section is dedicated to the mutual interaction of the developed algorithms. For that purpose, their stand-alone performance is compared against the case where both schemes are concurrently operating. If DAP is not employed for idle mode UE devices, then broadcast AP is applied with the RSRP-based parameterization of Table 4.1. Note that the simulations are conducted for a fixed offered load of 40 Mbps per macrocell area.

Fig. 4.4 illustrates both the average and the coverage (5%-ile) UE throughput for all load balancing methods. In fact, DAP performs the worst due to the time instance that traffic steering decisions are applied. Specifically, camping priorities are adjusted during the connection release; hence, there is no guarantee that the selected camping carrier will still be in similar load conditions when the UE devices switches to connected mode again. Notice that for this particular simulation campaign, users typically stay in idle mode for ∼5-7 sec before establishing a new RRC connection. Owing to the up-to-date load information, FHO@CS further enhances the perceived data rates. Nevertheless, this case is a representative example where idle and connected mode are severely misaligned. In more detail, broadcast AP pushes idle users to cells deployed at the 2.6 GHz, whereas FHO@CS steers UE devices to the 800 MHz carrier. This essentially means that – after that the data session completion – 800 MHz connected mode users will most likely to the 2.6 GHz band, regardless of the offered load conditions. This undesired effect is avoided when both DAP and FHO@CS

[4]Medium/ high velocity UE devices should be pushed to the 800 MHz carrier in order not to endanger their handoff performance due to the deployed picocells at the 2.6 GHz carrier.

are jointly operating; a fact that makes the DAP+FHO@CS configuration the best amongst the simulated ones.

The related cost in terms of handovers and cell reselections is shown in Fig. 4.5. DAP results in a cell reselection rate that is 2.6 times higher than the one estimated for handovers. The reason is that load balancing is solely performed by means of idle mode functionalities. Nonetheless, the standalone FHO@CS operation constitutes the most costly approach due to the conflicting policies applied in idle and connected mode, respectively. Idle-to-connected ping pong events (and vice versa) dominate mobility performance, increasing the associated handover and cell reselection rates by ∼240% and ∼60% over the DAP configuration. The importance of aligning the traffic steering decisions between the two different RRC UE states is highlighted by the estimated DAP+FHO@CS mobility performance. In this case, the aforedescribed FHO@CS behavior is eliminated and both handover and cell reselection rates decrease significantly. By means of that, the associated cost in mobility events is kept at an affordable level. Notice that the cell reselections are constantly more than the estimated handover events. This is due to the non-optimal configuration of the R criterion used for handling idle mode mobility between macrocells and picocells at

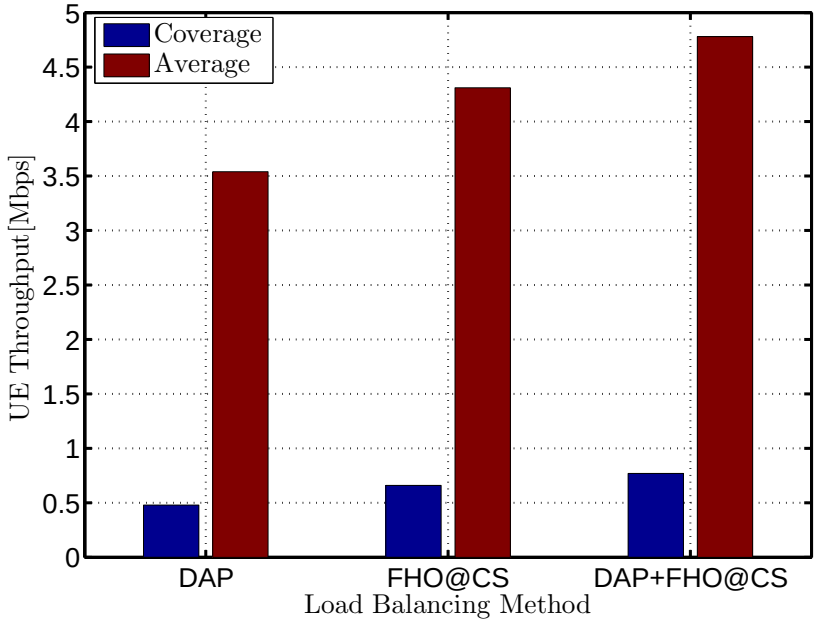


Fig. 4.4: Throughput performance for the developed load-based traffic steering policies. 40 Mbps per macrocell area are considered.

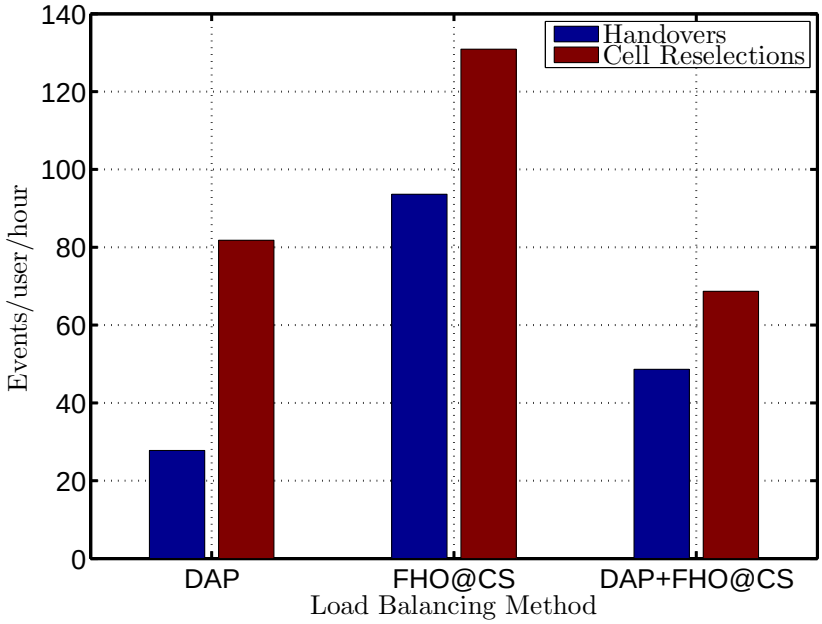


Fig. 4.5: Mobility performance for the developed load-based traffic steering policies. 40 Mbps per macrocell area are considered.

the 2.6 GHz carrier. Applying a hysteresis value to the serving picocells would essentially resolve the problem by increasing the camping time on the small cell layer.

4.6.2 Performance Comparison between DAP+FHO@CS and RSRQ-based Mobility

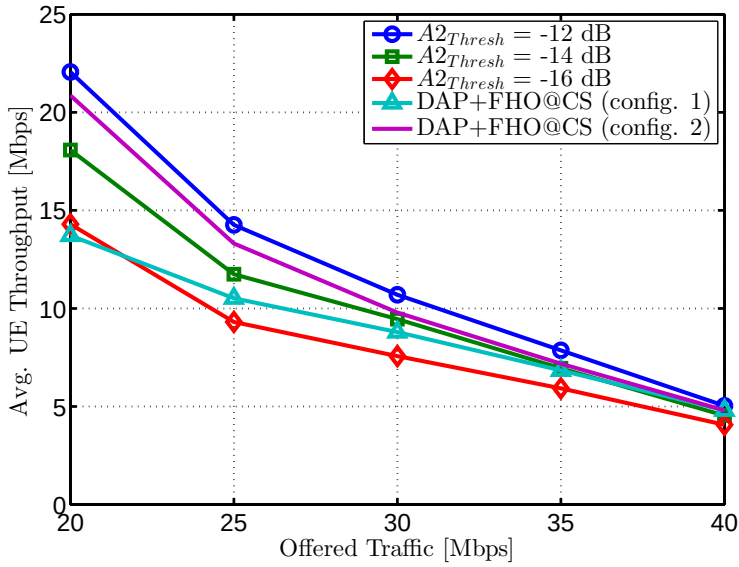
Fig. 4.6a and Fig. 4.6b depict the average and coverage UE throughput for the simulated load balancing configurations. All in all, the same trends are observed in both figures. The $A2_{Thresh} = -16dB$ setup results in the worst performance regardless of the offered load conditions, while the $A2_{Thresh} = -12dB$ configuration achieves the highest UE throughput since it offers sufficient inter-frequency measurement availability to exploit the load information carried by the RSRQ measure. Note that the UE experience gap between the $A2_{Thresh} = -14dB$ and the $A2_{Thresh} = -12dB$ setting broadens as the offered load decreases. The reason is that the $Thresh^{AP}_{sLow} = A2_{Thresh} = -14dB$ configuration can not trigger enough cell reselections to the lower priority 800 MHz carrier at lower offered traffic demands. By means of that, the utilization

of the 800 MHz carrier gradually decreases. On the other hand, the proposed DAP+TS@CS framework results in similar performance to one achieved by the $A2_{Thresh} = -12dB$ configuration only if ρ_{target} is adjusted according to the offered traffic (*config. 2*). Unless ρ_{target} is properly configured, the number of load-driven events essentially decreases at lower offered load conditions since the traffic steering framework is less frequently triggered.

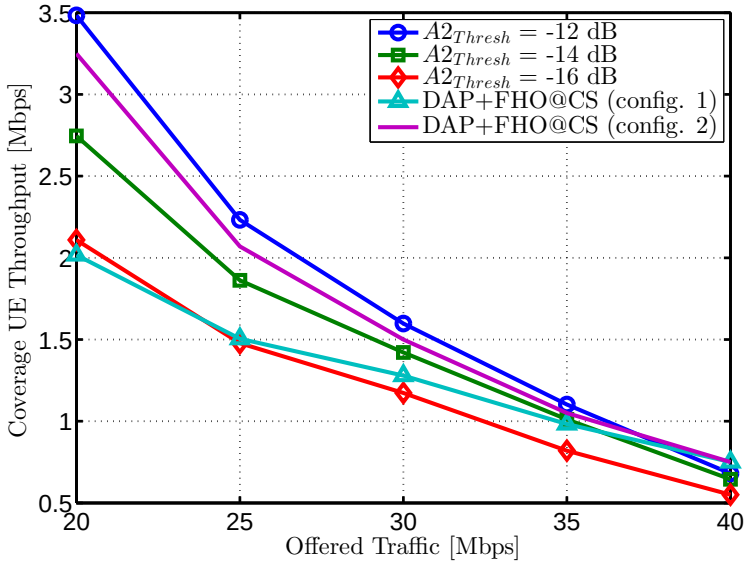
The corresponding mobility performance in terms of handovers and cell reselections is shown in Fig. 5.9c and Fig. 5.9d, respectively. With regards to the RSRQ-based mobility configurations, it is alleged that the capacity gain comes at the expense of RRC signaling. Thus, the $A2_{Thresh} = -12dB$ setup constitutes the most costly approach due to the inefficient use of inter-frequency measurements and the sensitivity of the RSRQ measure to the load variations. On the contrary, the proposed framework reduces the RRC signaling in terms of handovers. Between the two different DAP+FHO@CS configurations, the capacity-driven one (*config. 2*) essentially results in higher RRC signaling simply because more load-driven handovers are executed. Nevertheless, the associated overhead is well below the one estimated for the $A2_{Thresh} = -12dB$ case. In more detail, the reduction gains are in the order of ∼20% at high offered load conditions, while higher savings are achieved at lower traffic demands. It is worth mentioning that the inter-frequency handovers correspond to ∼50-60% of the total handovers for the $A2_{Thresh} = -12$ setting, whereas for DAP+FHO@CS the corresponding percentages are in the order of ∼20-40% depending on the offered load conditions.

The trends are similar for idle mode as well. Compared to DAP+FHO@CS, the RSRQ-based setup of broadcast AP results in noticeably higher cell reselection rates. This is due to the measurement configuration of users camping on the lower priority 800 MHz carrier that always scan for prioritized 2.6 GHz cells if broadcast AP are employed. Hence, even if a cell reselection to the 800 MHz carrier is performed, the concerned user can reselect back to the 2.6 GHz band at any time, given that the associated condition is met. This behavior is more evident at low offered load conditions where the load is rather unstable and a single transmission can impact RSRQ measurements by several dB. On the other hand, the usage of DAP in idle mode ensures that users almost always camp on the frequency layer with the highest priority; hence, inter-frequency measurements towards lower priority carriers are only triggered whenever the UE is about to experience coverage problems on its camping carrier. Note that a high reselection rate increases UE power consumption in idle mode as it extends the DRX cycle of idle users. An estimation of the achievable UE battery savings achieved by DAP+FHO@CS relative to the other simulated RSRQ-based configurations is available in Appendix C. In particular, it is shown that the proposed inter-frequency load balancing framework can reduce the idle mode UE power consumption up to ∼45%.

Fig. 4.8 illustrates the user distribution across the deployed layers together

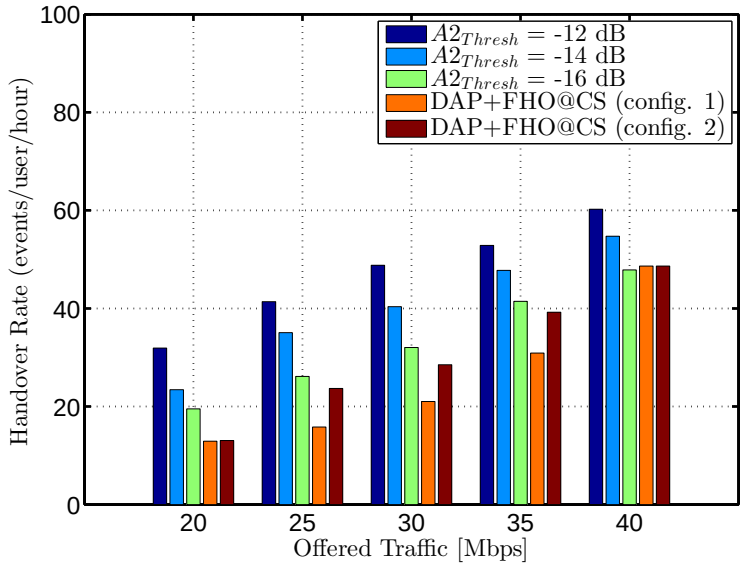


(a) Avg. UE throughput

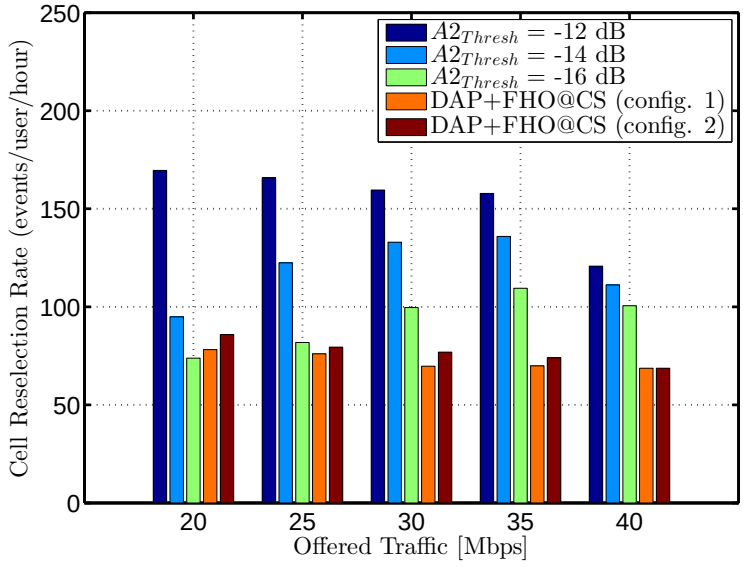


(b) Coverage UE throughput

Fig. 4.6: Throughput performance for different load balancing configurations and offered traffic demands.



(a) Handover rates



(b) Cell reselection rates

Fig. 4.7: Mobility performance performance for different load balancing configurations and offered traffic demands.

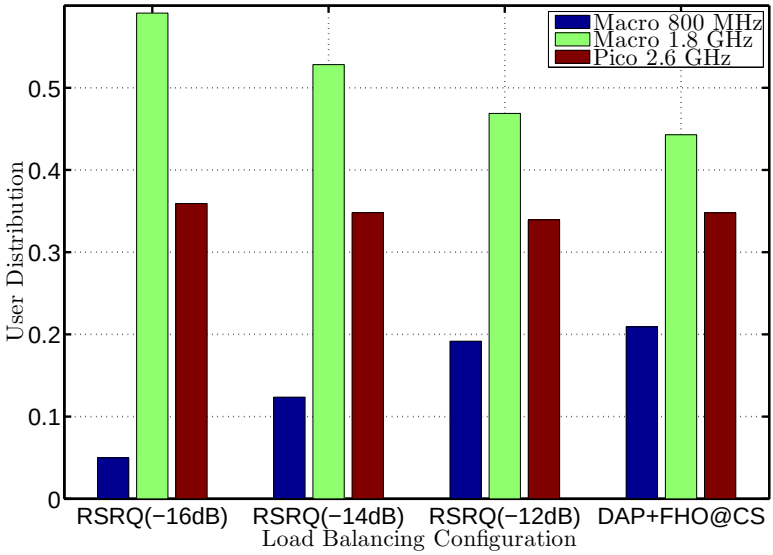
with the rate that connected mode users perform inter-frequency measurements. The illustrated KPIs correspond to high offered load conditions of 40 Mbps per macrocell area, where measurement gaps may have a critical impact on the user throughput. Regardless of the load balancing configuration, it is observed that most of the users are offloaded to the 2.6 GHz carrier as it owns double bandwidth than the 800 MHz band. With regards to the RSRQ-based configurations, the $A2_{Thresh} = -12$ dB case provides the best utilization of the 800 MHz carrier; a fact that results in the throughput gains depicted in Fig. 4.6a and 4.6b, respectively. Nevertheless, DAP+FHO@CS achieves almost the same user distribution with a noticeably lower cost in physical layer measurements and RRC signaling, as seen by Fig. 4.8b. In more detail, the related measurement rate is ∼22 times lower than the one estimated for the $A2_{Thresh} = -12dB$ configuration proving that inter-frequency cell discovery mechanisms are utilized much more efficiently when DAP+FHO@CS is employed.

Lastly, it is worth mentioning that the user distribution illustrated in Fig. 4.8 is not the optimal one. To support this argument, Fig. 4.9 depicts the Jain's fairness index for both DAP+FHO@CS and the $A2_{Thresh} = -12dB$ configuration. Jain's index is calculated twice with and without the impact of the picocell layer. Although the investigated methods provide exceptional macro-only load balancing performance, the fairness index deteriorates when the picocell load is also taken into account. The reason is that no traffic steering mechanism is utilized for balancing the load of the two layers sharing the 2.6 GHz carrier. Given the reader's interest, Appendix C shows that the supplementary usage of co-channel MLB at the 2.6 GHz further improves the picocell user offloading and eventually results in better load balancing performance, compared to when DAP+FHO@CS is solely employed.

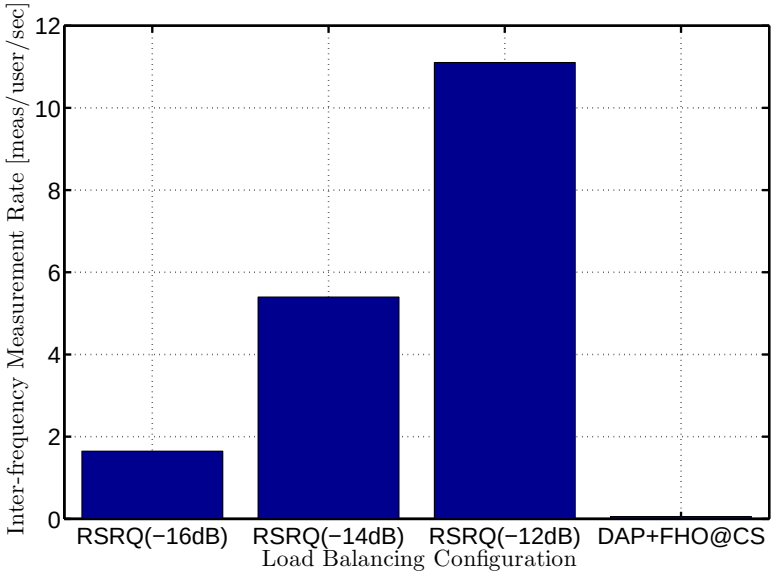
4.7 Conclusions

This chapter has addressed the problem of inter-frequency load balancing in multi-layer HetNet deployments. Specifically, a traffic steering framework has been designed, denoted as DAP+FHO@CS, that explicitly configures UE devices with inter-frequency measurements whenever cell overload is detected. In such a manner, less loaded neighbors at different carriers are discovered and UE devices are steered towards them via idle and connected mode functionalities. The proposed framework is compared against different RSRQ-based configurations of the standardized mechanisms used for managing inter-frequency mobility.

Simulation results have shown that RSRQ-based mobility requires adequate measurement availability so as to perform as a passive traffic steering mechanism. However, such an approach does not avoid UE devices from perform-



(a) User distribution averaged over all users and the simulation time. Both connected and idle mode samples are included.



(b) Connected mode inter-frequency measurement rate.

Fig. 4.8: User distribution and corresponding inter-frequency measurement rate for different load balancing configurations. 40 Mbps per macrocell area are considered.

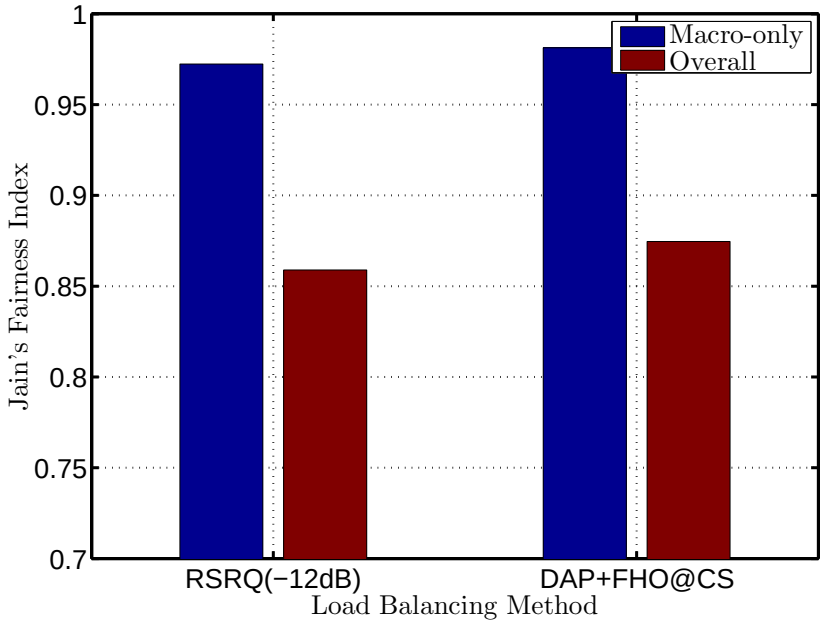


Fig. 4.9: Jain's index for different load balancing configurations. 40 Mbps per macrocell area are considered.

ing unnecessary inter-frequency measurements, while it further results in handover/cell reselection rates that are relatively high. On the contrary, the developed framework uses more efficient cell discovery mechanisms. In such manner, DAP+FHO@CS achieves a better trade-off between the derived capacity gains and the cost paid in handovers, cell reselections and physical layer measurements. Therefore, it could be considered as an attractive traffic steering solution for performing inter-frequency load balancing in multi-layer HetNet deployments.

Note that the complementary UE throughput benefits of DAP+FHO@CS cannot be seen with the adopted best-effort traffic model since RSRQ already performs well for such type of traffic. To make a more fair comparison between the investigated schemes, QoS should also be introduced into the traffic model, which is left for future work.

Chapter 5

Load Balancing in HetNets with Intra-eNB CA

5.1 Introduction

Carrier Aggregation (CA) enables UE devices to concurrently connect to multiple carriers, also denoted as Component Carriers (CC). Albeit the main rationale for introducing CA has been the amelioration of the peak data rates, the multi-CC connectivity gives operators the opportunity to better exploit their fragmented spectrum. This can be achieved by advanced packet scheduling schemes that enhance user fairness and perform fast load balancing across the aggregated CCs. In this context, load-based forced handovers might become less relevant in CA network environments.

This chapter aims at identifying of viability of load-based traffic steering in HetNet deployments with intra-eNB enabled at the macro overlay. Unlike previous studies, (e.g. [103, 104]), we adopt a more realistic approach where mobility procedures and the associated latencies for configuring SCells are explicitly modeled as well as both idle and connected mode UE devices are considered. Furthermore, more challenging CA configurations are simulated where the experienced interference and the bandwidth per CC differs. To understand the mutual interaction between CA and load-based handovers/cell reselections, the DAP+FHO@CS framework is adopted for providing load-aware PCell management, while the scheduler assists the load balancing process. To smoothly integrate inter-frequency traffic steering with CA, the load metric defined in Section 2.5 is modified so as to consider multi-CC connectivity as well. Ultimately, system level simulations are conducted in scenarios B and C, as described in Section 2.6. Recall that the reason for simulating both scenarios is for analyzing the relevance of load-based traffic steering in deployments where the number of deployed CCs is equal or greater than the number of CCs that

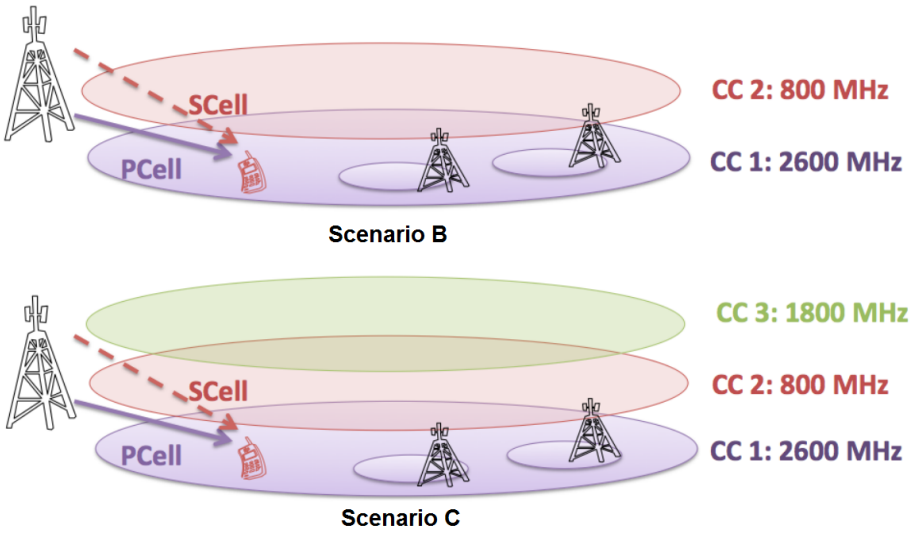


Fig. 5.1: Considered scenarios with intra-eNB enabled at the macro overlay. CA UE devices can be configured with up to 2 serving cells, while conventional handovers/cell reselections are required for accessing the picocell resources.

a CA UE supports. Since inter-band CA is commonly constrained to only 2 serving cells, CA UE devices do not own access to the overall macrocell spectrum in scenario C (see Fig. 5.1). To our knowledge, such case has not been yet analyzed.

The remainder of the chapter is structured as follows: Section 5.2 gives insight into the main state-of-art with regards to cell management and packet scheduling in CA network. The proposed load metric modification for providing load-aware PCell management by means of traffic steering is described in Section 5.3, while the simulation parameters are outlined in Section 5.4. The associated performance results are shown in Section 5.5, whereas the chapter is concluded with Section 5.6.

5.2 State-of-Art CA: Mobility & Scheduling

5.2.1 PCell & SCell Management

As discussed in Section 1.3, all higher layer procedures are handled by one of the serving cells, denoted as the PCell. Among others, mobility procedures are managed by the PCell. This essentially means that handovers and RLF declarations can only take place on the PCell complying to the same proce-

dures as without CA. Since connected mode UE devices can solely operate in CA mode, the PCell is commonly assigned on the cell that the UE devices establishes its RRC connection once switching to connected mode.

The remaining serving cells – denoted as SCells – are dynamically configured on a UE dedicated basis. Specifically, such actions are triggered by measurements reports sent by the UE device in the uplink. Therefore, whenever a SCell event is fulfilled, the network informs – via downlink RRC signaling – the concerned user about the action to be performed (i.e. adding, removing or changing a SCell). It is worth mentioning that the cost of inter-frequency measurements in terms of transmission gaps is less relevant for CA terminals, since they may concurrently receive data on one carrier, while performing measurements on a different carrier [105]. By means of that, inter-frequency measurements for CA UE devices are not controlled by the $A2_{Thresh}$ configurations, but they are typically performed with a certain periodicity. Nevertheless, the A2 event can be still utilized for activating inter-frequency PCell mobility; thus, for changing the PCell via an inter-frequency handover.

Albeit the tremendous peak data rate enhancements, CA inevitably increases the overhead burden due to the associated RRC signaling for configuring SCells. In fact, the impact of various PCell and SCell management policies on network signaling has been investigated in [106] for the LTE-Advanced Scenario 1 and Scenario 3[1], respectively. The corresponding simulations have shown that a dynamic SCell policy is more relevant in deployments where the coverage of the deployed CCs differs significantly (i.e. Scenario 3). In such, both the PCell and SCell SINR distribution improve; however, at the expense of a 3-fold increase in RRC reconfigurations.

5.2.2 Packet Scheduling

Layer-2 packet scheduling is responsible for allocating physical resources to the multiple users. It can be performed independently per CC or jointly across several CCs. The first approach is similar to packet scheduling in a conventional single carrier system, meaning that CA users are assigned resources in each CC without considering the scheduling decisions in other CCs. On the contrary, the latter approach takes into account the CA UE statistics of all configured CCs.

For single carrier systems, the Proportional Fair (PF) scheduler is a famous packet scheduling paradigm that ensures user fairness by exploiting multi-user diversity [107]. By means of that, UE devices are allocated an equal amount of transmission resources in the long term irrespective of fading characteristics and radio conditions [108]. The degree of fairness is typically by a utility

[1] The reader should refer to Fig. 1.4 in Section 1.4. Both F1 and F2 have the same coverage in Scenario 1. On the contrary, different coverage is provided by F1 and F2, while F2 partially overlaps F1 so that to improve cell edge throughput at F1.

function that is defined as the sum of the logarithmic user throughput. However, guaranteeing user fairness within each CC does not essentially means that global fairness over all CCs is also achieved. This is of paramount importance in scenarios where both non-CA and CA UE devices co-exist in the system, as the former naturally own limited access to the overall spectrum resources (they can solely connect to a single CC).

The problem of resource allocation in LTE-Advanced systems with a mixture of non-CA and CA users has been investigated in [103, 109], showing that the independent PF scheduler results in resource starvation for non-CA UE devices. To enhance user fairness between the different UE categories, a joint PF metric is proposed as follows:

$$u_{i,j} = \underset{u}{\text{argmax}}\{\frac{R_{u,i,j}}{\sum_{i=1}^{n} \widetilde{R}_{u,i}}\}, \tag{5.1}$$

where $u_{i,j}$ is the user assigned the j^{th} PRB on the i^{th} CC; $R_{u,i,j}$ is the instantaneous user throughput on this particular PRB; and $\widetilde{R}_{u,i}$ is the past average user throughput on the i^{th} CC. Compared to the traditional PF metric, (5.1) considers the aggregated past throughput over all CCs so as to decrease the scheduling priority of CA users in particular CCs. In other words, CA UE devices are primarily scheduled on the CC that can serve them with the highest data rate. By means of that, non-CA UE devices are assigned more transmission resources, which in turn enhances the global user fairness. In fact, the metric in (5.1) maximizes the sum of logarithmic throughput in scenarios with a mixture of non-CA and CA devices [104], which essentially means that it is the optimal scheduler given the aforementioned utility function.

A representative example of how the independent PF and the cross-CC scheduler allocate transmission resources is depicted in Fig. 5.2. Specifically, 2 CCs with similar coverage and bandwidth are considered. A single legacy UE is assigned onto each of them, while a CA user aggregates resources from both CC1 and CC2. As the independent PF metric guarantees fairness within each CC, the CA UE device will be allocated half of the bandwidth in each CC; thus, it will be assigned with double transmission resources compared to the non-CA users. Evidently, this is avoided if the cross-CC scheduler is used for achieving fair resource allocation. In this case, all UEs are assigned 1/3 of the total system resources.

For scenarios where all CCs provide similar coverage, both schedulers perform the same when the CA UE penetration reaches to 100%. Nonetheless this is not the case for deployments with CCs of different coverage. For such scenarios, the conducted investigations in [93, 110] have shown that cross-CC scheduling is essential even if all users are CA capable. The reason that this type of scheduler can compensate for potential SINR disparities between the aggregated CCs and

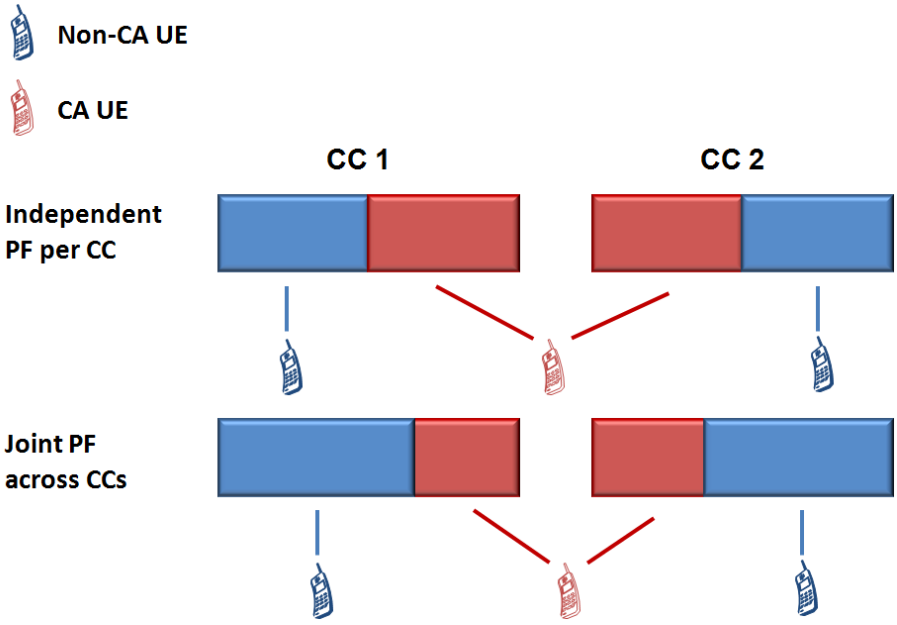


Fig. 5.2: Example of resource allocation for different packet scheduling schemes and a mixture of non-CA and CA users.

eventually provide fast inter-layer load balancing.

5.3 Integrating Traffic Steering with CA

The prior state-of-art studies indicate that CA relaxes the requirements of balancing the load by means of mobility procedures, as the scheduler can serve this purpose as well. However, these investigations are typically conducted without simulating any mobility mechanism in idle and connected mode respectively. Particularly, newly arrived users – generated by dynamic birth-death process – are assigned onto one or more CCs based on suitable Layer-3 carrier load balancing mechanisms and disappear from the system when their data session ends. In a real network, such Layer-3 carrier assignment procedures are realized by the associated PCell and SCell management policies. A forced PCell handover towards a less loaded layer might still be relevant for a CA UE that establishes its connection at an overloaded layer and the scheduler is not able to compensate the load imbalance. In addition, idle mode functionalities should ensure that users start service on the best cell without having to perform any PCell handover whenever switching to connected mode.

To provide traffic steering support in such network environments, the load

metric defined in (2.6) has to be neatly modified so that it takes CA into account. The reason is that (2.6) does not consider the throughput that CA UE devices perceive on the remaining carriers, which in turn may result in cell load misconceptions and triggering of unnecessary traffic steering events. These include PCell handovers triggered by FHO@CS or inter-frequency reselections based on the DAP operation. To avoid such undesired effect, the virtual load contribution of user u to one of its serving cells c is now defined as follows:

$$\widehat{f}_{u,c} = min\{\frac{f_{u,c}\widehat{R}_u}{\sum_{j\in S} r_{u,j}}, \rho_{max}\}, \tag{5.2}$$

where $f_{u,c}$ is the resource share of user u in cell c; and $\sum_{j\in S} r_{u,j}$ represents the aggregated user throughput over the set $\mathbf{S}$ of all configured cells. Similar to cross-CC scheduling, the metric in (5.2) assumes that the individual schedulers exchange information on the CC-specific CA UE throughput. On the other hand, the load contribution of non-CA users does not alter compared to when (2.6) is used, since such users solely receive data from a single cell. Note that the cell load of cell c is eventually derived by summing $\widehat{f}_{u,c}$ over all users being served by that particular cell, as discussed in Section 5.2. Except for the modification in the load metric used for realizing any traffic steering action, the developed DAP+FHO@CS framework is directly applicable in a CA environment. This necessarily means that both algorithms operate as described in Section 4.4, aiming at load-aware PCell management irrespective of the UE category since the load metric in (5.2) now considers multi-cell connectivity.

At the meanwhile, SCell management can be decoupled from the PCell traffic steering framework. Specifically, SCell additions, removals and changes could be governed by the 3GPP-defined measurement events outlined in Table 2.2. A RSRQ-based configuration of the associated SCell events is generally suggested so that the load information carried by the RSRQ measure is rapidly exploited. By means of that, a load-aware SCell association is guaranteed as well. All in all, the suggested integration method results in independent PCell and SCell operations. The advantage of this approach is its simplicity as the only required modification lays on the definition of a proper load metric for managing PCells. Furthermore, it is quite generic in a sense that it can be straightforwardly applied in any deployment scenario, even if CA is not supported by all network layers.

Scenario B and Scenario C essentially belong to this deployment category as it assumed that CA is only enabled at the macro overlay. This necessarily means that CA terminals can access the picocell spectrum only by means of handovers or cell reselections. Similarly, whenever these users are connected to a small cell, they cannot add a macro SCell. Cross-CC scheduling together with DAP+DFHO@CS is responsible for balancing the load across the deployed macrocells, while load balancing between the small cell layer and the

overlaid inter-frequency macrocells is conducted via traffic steering functionalities. Hence, it is of paramount importance to ensure that macro CA users are offloaded to the picocell layer – they leave CA mode – only if they will truly benefit from such action. This is even more crucial in Scenario C, where hotspot CA users should either aggregate resources between the macrocell 800 MHz and 1.8 GHz carriers or be connected to small cells depending on the network load and the perceived interference at the 2.6 GHz band. Using a load metric – such as the one defined in (5.2) – that comprises the effect of multi-cell connectivity essentially increases the probability of taking the proper traffic steering decisions.

5.4 Simulation Assumptions

The applied simulation setup is in line with the one outlined in Table 4.1 in terms of considered users, traffic model, mobility, DAP+FHO@CS configuration, etc. To emulate the cross-CC scheduler behavior, an abstract RRM management model is developed capable of capturing the main properties of collaborative PCell/SCell scheduling. Given the reader's interest, the adopted modeling framework is described in Appendix A. The main simulation parameters are outlined in Table 5.1. Notice that the duration of a SCell execution refers to the overall time required for starting the data reception on the SCell. Thus, it includes the time required for setting up, configuring and activating the SCell.

Table 5.1: Simulation Parameters

Parameter	Value
CA UE Penetration	{0, 20, 40, 50, 60, 70, 80, 100} %
Traffic Type	Finite Buffer
Offered Load	20:5:50 Mbps
Macrocell Scheduling	Cross-CC
Picocell Scheduling	Proportional Fair
SCell Support	1 SCell
SCell Execution Time	50 msec

Particularly, the following cases are simulated for different penetrations of CA UE devices:

- AP+CA: This configuration refers to the case when RSRP-based broadcast AP are applied for idle UE devices prioritizing the 2.6 GHz carrier. No means of connected mode traffic steering is employed and the scheduler solely offers load balancing at the macro overlay. With this

setup, all PCells are assigned on the 2.6 GHz carrier regardless of the CA penetration and the offered load conditions. This is due to the load insensitive RSRP configuration and the conservative absolute thresholds used for triggering cell reselections towards lower priority carriers, which makes all users establish their RRC connection at the 2.6 GHz carrier. By means of that, CA UEs are solely configured with SCells on the remaining lower macrocell carriers. Albeit being a rather extreme case, it is an interesting configuration in order to investigate whether the CA scheduler is capable of achieving inter-layer load balancing regardless of the great PCell imbalance.

- TS+CA: Load-aware PCell management is performed via the employment of DAP+FHO@CS.

Table 5.2: SCell Events Definition

SCell Action	Event	Value	TTT
Addition	target becomes better than threshold	-16 dB	128 msec
Removal	SCell becomes worse than threshold	-18 dB	256 msec
Change	target becomes better than SCell	3 db	256 msec

The configured SCell measurement events for adding, changing or removing a SCell are outlined in Table 5.2. Specifically, a target cell whose RSRQ is above -16 dB can be added as SCell. The motivation for configuring such a low threshold is for exploiting CA as much as possible. Apparently, if more than a single cell fulfill the addition condition, the best among them is added. The threshold for realizing a SCell removal is set 2 dB lower than the associated addition threshold in order to decrease the probability of triggering repetitive additions and removals of the same SCell owing to the RSRQ sensitivity to load variations. In fact, SCell are also removed whenever a CA user performs a PCell handover or switches to idle mode. Lastly, the SCell can be changed if a 3 dB stronger inter-frequency neighbor is discovered. Such an event is only feasible in Scenario C, where CA terminals cannot connect to all macrocell carriers. Notice that a short TTT window is configured for SCell additions in order to increase the probability of exploiting the SCell before the transmission buffer empties. On the other hand, SCell removals and changes are configured with a longer TTT window; again for the sake of SCell stability[2].

[2]For SCell removals, the TTT is only applicable to the case when the SCell has to be removed due to radio conditions. Given that the user performs a PCell handover or switches to idle mode, then the SCell is instantly removed.

Table 5.3: PCell & SCell Distribution (AP+CA)

	Scenario B		
	Macro 800 MHz	**Macro** 2.6 GHz	**Pico** 2.6 GHz
PCell	0	0.6	0.4
SCell	1	0	0
	Scenario C		
	Macro 800 MHz	**Macro** 1.8 GHz	**Macro** 2.6 GHz	**Pico** 2.6 GHz
PCell	0	0	0.6	0.4
SCell	0.5	0.5	0	0

5.5 Simulation Results

Scenario B is initially simulated for different traffic demands with and without CA. The CA UE penetration is set to 50%. When CA is not supported, load balancing is solely achieved by means of DAP+FHO@CS; hence, the associated performance results obtained in Chapter 4 are reused for comparison purposes. After that, both scenarios B and C are simulated for fixed high load conditions and different CA UE penetrations. The focus is put on analyzing how load balancing is affected by the UE connectivity capabilities and the CA UE penetration. As the PCell/SCell distributions are insensitive to the offered load and the CA penetration for the AP+CA configuration, the estimated distributions – expressed in the form of per-layer ratio – are summarized in Table 5.3. Hence, the reader can refer to that particular table, whenever is needed for comparison purposes hereinafter.

5.5.1 Impact of Offered Traffic

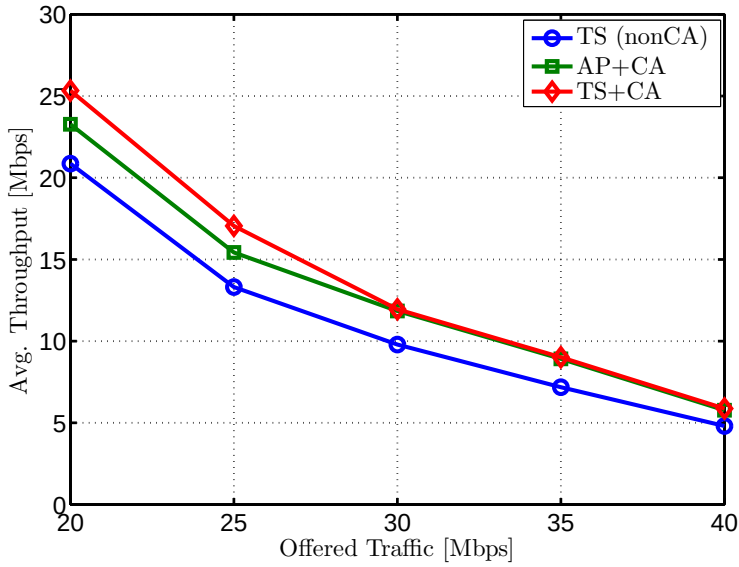
Fig. 5.3a and Fig. 5.3b depict the average and coverage UE throughput with and without CA for the simulated load balancing configurations in scenario B. Compared to the non-CA case, data rates essentially improve when CA is introduced. Nevertheless, the associated non-CA and CA curves come closer as the offered load increases. This is a typical CA behavior, where at low load conditions the extended transmission bandwidth results in great capacity enhancements, while at high load conditions the system saturates and lower gains are achieved mainly by exploiting multi-user diversity. Focusing on the CA cases, it can be observed that TS+CA outperforms AP+CA only at low load conditions. At such traffic demands, the time in connected mode is not long enough to exploit the CA scheduler for balancing the load. Hence, it is

important to ensure that all UE devices start service on the best cell via traffic steering. When the offered load increases, the gains of TS+CA over AP+CA diminish as the CA scheduler is now offering load balancing. Particularly, even if all UE devices have their PCells on the 2.6 GHz carrier when AP+CA is employed, CA UE devices are primarily scheduled on their 800 MHz SCells so that resource fairness between the different UE categories is maintained. By means of that, the load-aware PCell management provides no gain for CA penetrations above 50% and the scheduler performs fast inter-layer at the macro overlay. At the meanwhile, idle mode AP ensures the high utilization of the deployed picocells. Thus, global load balancing can be achieved.

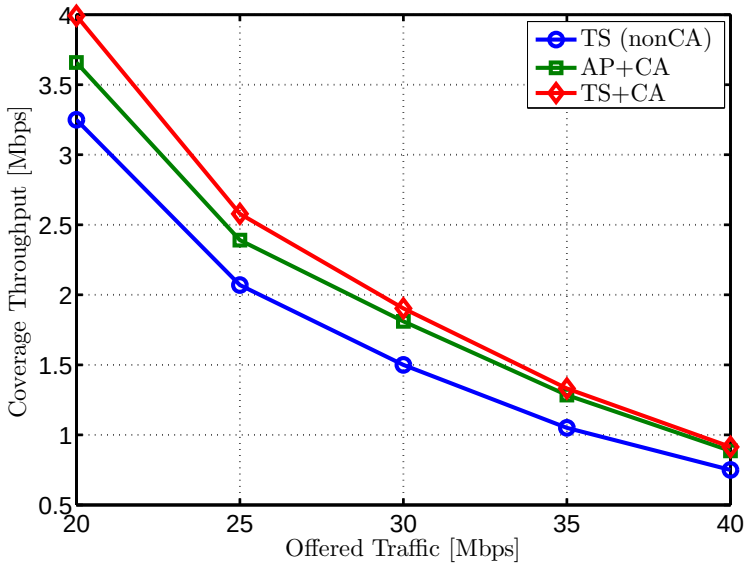
To measure the associated RRC signaling for all simulated cases, the estimated handover and SCell event rates are illustrated in Fig. 5.4a and Fig. 5.4b, respectively. In fact, handovers essentially decrease with CA simply because less inter-frequency handovers – triggered by FHO@CS – are required for balancing the load. Note that almost no inter-frequency handovers are triggered in the case of AP+CA since the bandwidth of the 800 MHz carriers is accessed only by means of SCell additions. Nevertheless, SCell actions dominate RRC signaling when the SCell event rates are compared with the associated handover ones. Furthermore, the applied PCell policy does not impact the SCell events. This is due to the rather aggressive configuration of the SCell addition threshold, which makes macrocell CA users almost always add a SCell whenever switching to connected mode. In other words, the generated signaling induced by SCell additions – and removals as there is an 1:1 relationship between them – heavily depends on the rate that user switch to RRC connected. One way to decrease the RRC signaling would be to increase the RSRQ-based threshold for adding SCells. However, the throughput gains provided by CA would naturally decrease as such an action would limit the amount of UE devices in CA mode.

5.5.2 Impact of CA UE Penetration & multi-CC Connectivity Capabilities

Fig. 5.5 depicts the associated PCell and SCell mass probability function for the TS+CA configuration in both investigated scenarios and for different CA penetrations. In general, it can be seen that traffic steering distributes PCells across the deployed layers, as this is performed via forced handovers and cell reselections in connected and idle mode, respectively. By means of that, a portion of users is steered to the inter-frequency carriers, while the addition of the 1.8 GHz band further offloads the 2.6 GHz macrocell layer, when comparing the PCell distribution between the two scenarios. With regards to SCells, their assignment is impacted by the TS+CA configuration, in a sense that – unlike to AP+CA – CA UEs are now configured with 2.6 GHz SCells as well. Such actions are performed by CA terminals having their PCell on either the 800 MHz or 1.8 GHz carrier. Nevertheless, as the CA UE penetration increases, both

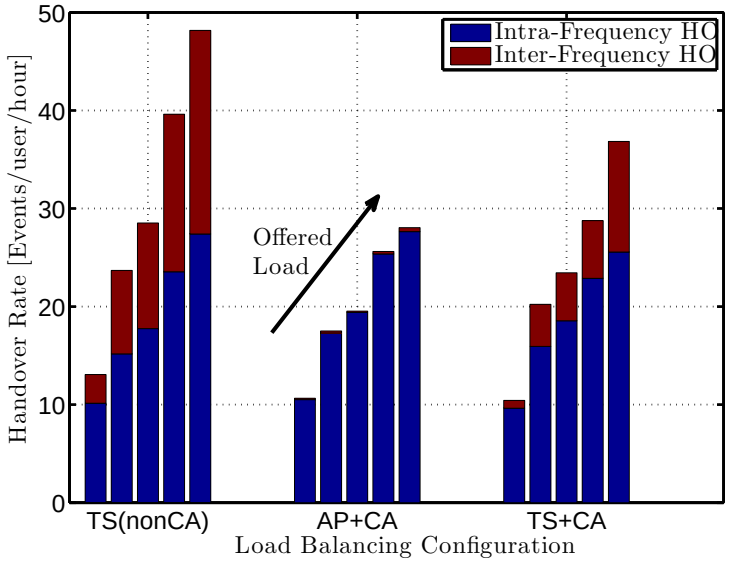


(a) Avg. throughput versus offered load

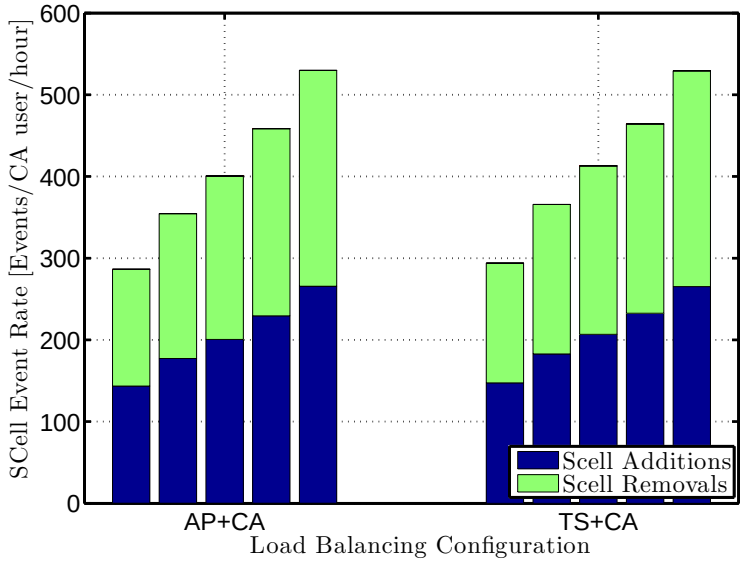


(b) Coverage throughput versus offered load

Fig. 5.3: Throughput performance with and without CA in scenario B for different offered traffic demands and load balancing configurations. 50% of the users are CA capable.

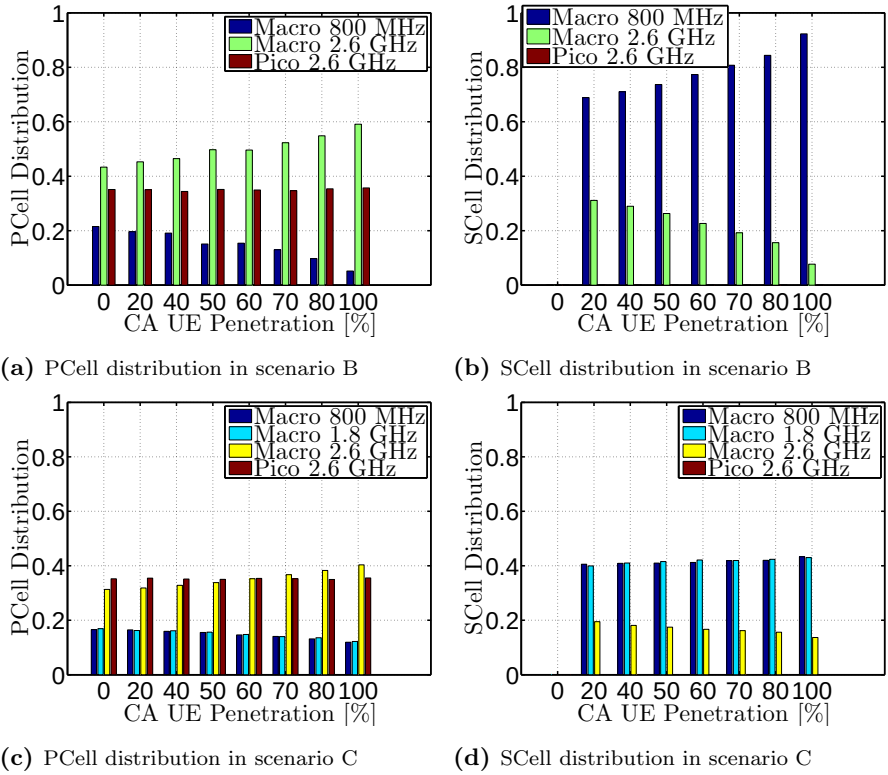


(a) Handover rates versus offered load. The offered load range is from 20 Mbps to 40 Mbps in steps of 5 Mbps.



(b) SCell event rates versus offered load. The offered load range is from 20 Mbps to 40 Mbps in steps of 5 Mbps.

Fig. 5.4: Mobility performance with and without CA in scenario B for different offered traffic demands and load balancing configurations. 50% of the users are CA capable.



(a) PCell distribution in scenario B

(b) SCell distribution in scenario B

(c) PCell distribution in scenario C

(d) SCell distribution in scenario C

Fig. 5.5: PCell and SCell distribution for both scenarios and simulated CA UE penettrations. Fixed offered load conditions are considered.

the PCell and the SCell distribution start converging to the ones estimated for the AP+CA configuration. The reason is that less traffic steering events are required for balancing the load. In this context, CA tries to compensate the load imbalances caused by suboptimal PCell assignments by means of cross-CC scheduling.

Fig. 5.6 shows the event rates in terms of forced handovers – triggered by FHO@CS – for both scenarios and all considered CA UE penetrations. It can be seen that both curves steadily decrease with the CA UE penetration. This indicates that CA truly relaxes the requirements of performing inter-frequency load balancing by means of mobility procedures. Specifically, the rate of load-based forced handovers is reduced to ∼82% when all users are CA capable in scenario B, while lower reduction gains of ∼55% are estimated for scenario C.

Jain's fairness index is presented in Fig. 5.7 in a attempt to capture the achievable degree of load balancing in both deployments. Regardless of the investigated scenario, the employment of DAP+FHO@CS improves the asso-

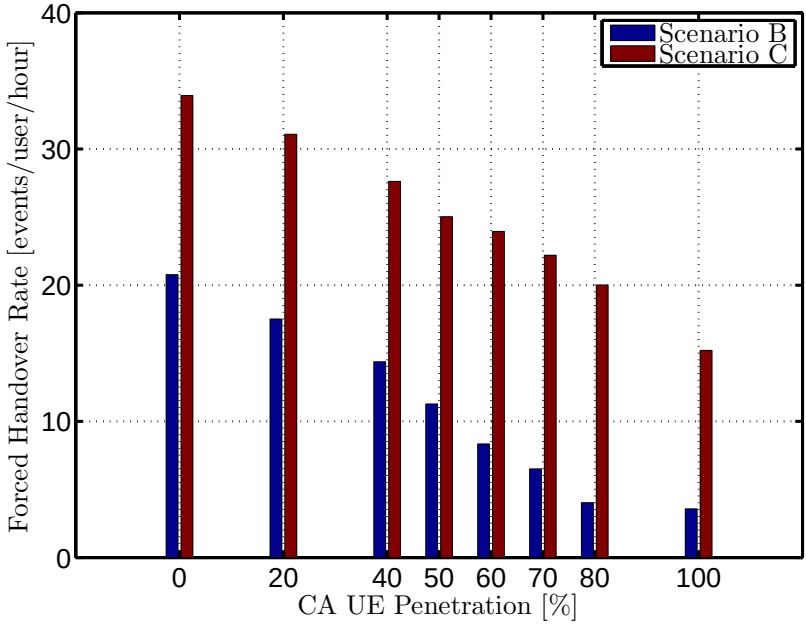
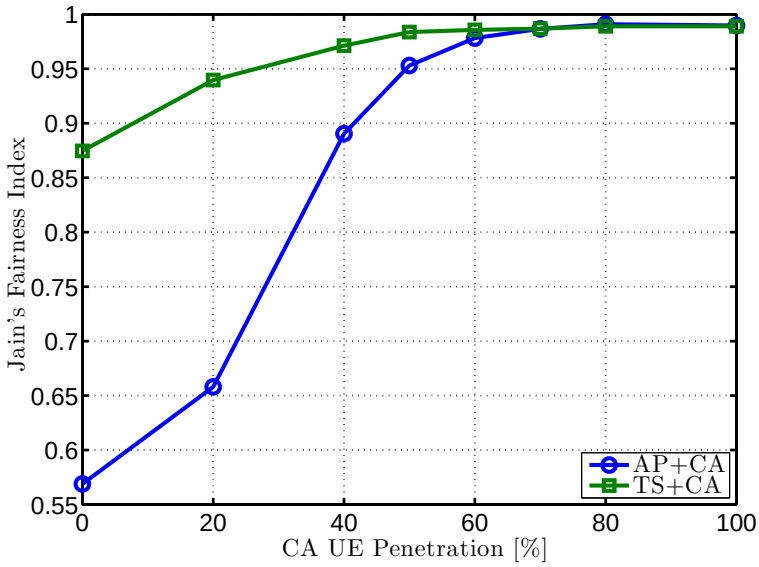


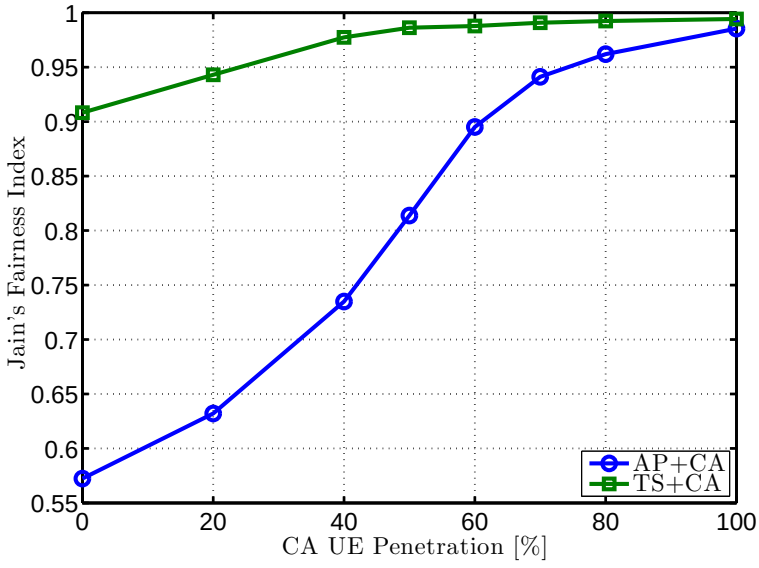
Fig. 5.6: Load-driven forced handover rates versus the CA UE penetration for both simulated deployments. Fixed offered load conditions are considered.

ciated fairness index significantly for CA penetrations below 50%. The reason is that the load imbalances induced by AP cannot be fully resolved by CA since the amount of CA users is not large enough to fully exploit cross-CC scheduling. Nevertheless, this is achieved for CA UE penetrations above 50% in scenario B, where TS+CA hardly provides any gain over the AP+CA configuration. Unlike to scenario B, TS+CA continues to enhance the load balancing performance for CA UE penetrations above 50% in scenario C. In this case, the fact that the number of deployed macrocell carriers is greater than the multi-CC connectivity capabilities of CA users poses a challenge on balancing the load solely via the CA scheduler. Undeniably, this observation proves that load-based traffic steering is still relevant in such scenarios, providing some gain even for a CA UE penetration of 100%.

Fig. 5.8a and 5.8b illustrate the estimated throughput gains of TS+CA over AP+CA for both scenarios B and C. Generally, the trends are similar to Jain's index performance shown in Fig. 5.7. Albeit the gains are noticeably higher for low CA penetrations – where the load-aware PCell management is of paramount importance – traffic steering does not any throughput benefit for CA UE ratios above 50% in scenario B. This is not the case for scenario C, where the stan-

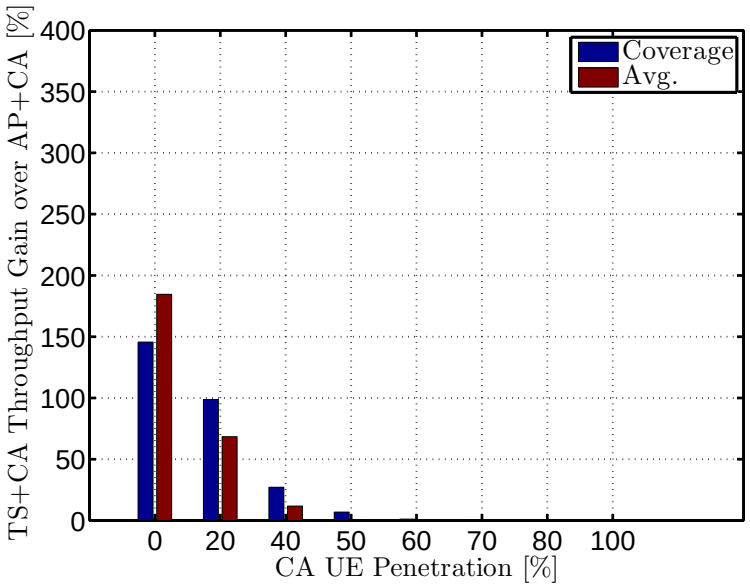


(a) Scenario B

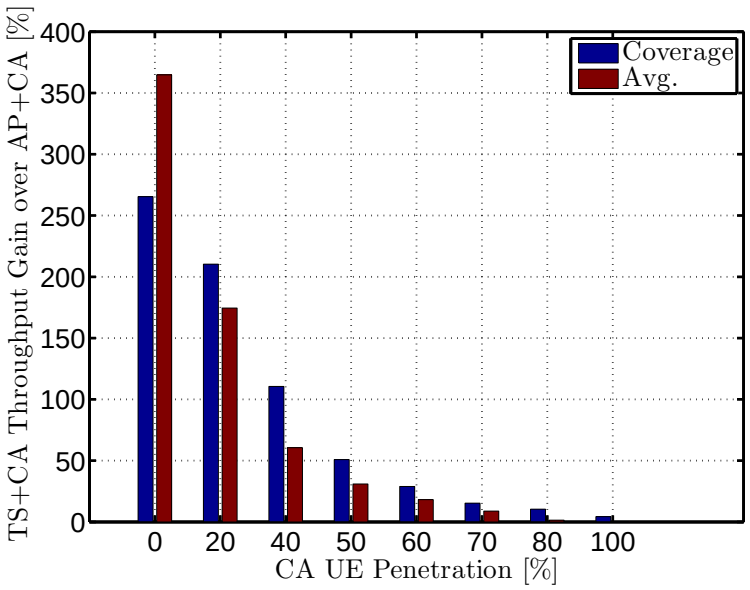


(b) Scenario C

Fig. 5.7: Jain's fairness index for all different load balancing configurations in both simulated scenarios and CA UE penetrations. Fixed offered load conditions are considered.

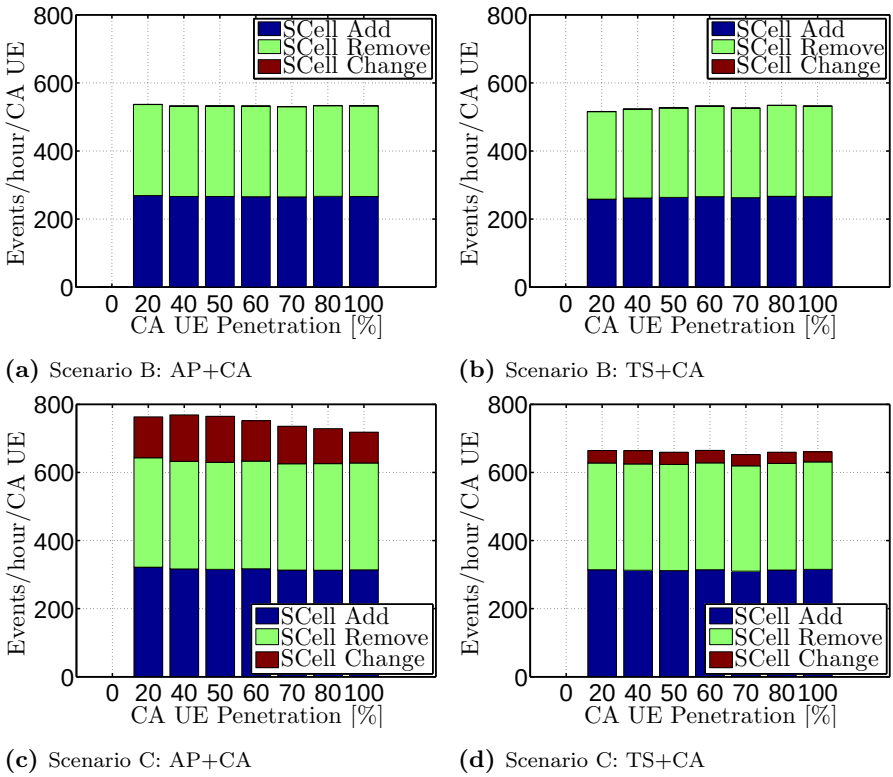


(a) Scenario B



(b) Scenario C

Fig. 5.8: Throughput gains of TS+CA over AP+CA for both simulated scenarios and CA UE penetrations. Fixed offered load conditions are considered.



(a) Scenario B: AP+CA

(b) Scenario B: TS+CA

(c) Scenario C: AP+CA

(d) Scenario C: TS+CA

Fig. 5.9: SCell events for both scenarios and load balancing configurations versus the CA penetration.

dalone operation of the CA scheduler is not sufficient for balancing the load. In such manner, the load-driven PCell management is still required and load balancing is eventually achieved jointly by the scheduler and the PCell traffic steering framework. Specifically, the gains of TS+CA over AP+CA in terms of coverage and average UE throughput are in the order of ∼50% and ∼28%, respectively, when half the considered users are CA capable. As the load balancing performance naturally increases with the CA UE penetration, the gains decrease for higher penetrations of CA users. E.g. the gain in coverage throughput is around 10% at 80% CA penetration, 5% at 100 % CA penetration, while the associated average throughput gains almost disappear for CA UE ratios above 70%. The reason is that in such high CA penetrations, traffic steering primarily enhances the user fairness between the different user types (i.e. non-CA and CA devices), while maintaining a similar average user throughput.

Lastly, the CA signaling overhead measured in terms of SCell additions, removals and changes is shown in Fig 5.9. Albeit no difference is observed for

the SCell additions and removals (for the reasons explained in subsection 5.5.1), traffic steering impacts the associated SCell changes estimated for scenario C. As already seen in Fig. 5.5 and Table 5.3, the RSRQ-based configuration of the associated SCell change event guarantees on average an equal split of SCells between the 800 MHz and the 1.8 GHz carrier for both the AP+CA and the TS+CA configuration. However, this requires many more SCell changes in the AP+CA case indicating that the macrocell load experiences more fluctuations. Hence, a larger amount of SCell changes is triggered owing to the load-sensitive RSRQ measure. On the contrary, the usage of the PCell traffic steering framework ensures a more stable SCell connection reducing the corresponding SCell change rate to ∼70% regardless of the CA UE ratio. The reason is that TS+CA balances better the macrocell load; a fact that equalizes the RSRQ distribution per carrier and decreases the probability of triggering SCell changes.

5.6 Conclusions

This chapter has evaluated the potentials of load-based traffic steering in multi-layer HetNet deployments supporting intra-eNB CA. This essentially means that CA users can aggregate resources from more than one co-located macrocells, while regular handovers/cell reselections are used for accessing the picocell bandwidth. To avoid cell load misconceptions, the associated metric defined in Chapter 2 has been neatly modified so that it considers multi-CC connectivity. By means of that, DAP+FHO@CS is capable of providing load-aware PCell management taking into the throughput that users perceive in all of the servings cells. On the other hand, SCell management is essentially decoupled from the PCell traffic steering framework as SCell additions, removals or changes are solely triggered based on RSRQ measurements. System level simulations are conducted in more challenging CA environments, where the aggregated carriers differ in terms of bandwidth, interference as well as CA UE devices may not be capable of aggregating resources from all of them.

The associated simulation results have shown that performing load balancing by means of mobility procedures is essential for any deployment and CA UE penetrations below 50%. For higher ratios of CA UE devices and medium/high load conditions, traffic steering hardly provides any throughput gains in scenarios where CA users can aggregate resources from all macrocell carriers. The reason is that joint PF scheduling across the macrocell carriers can balance the load, while broadcast AP in idle mode maximizes the picocell offloading potentials. By means of that, global load balancing is achieved. Nevertheless, at low offered load conditions traffic steering should still be employed as the time in connected mode is not sufficient for fully exploiting CA.

On the other hand, load-based handovers/cell reselections are still relevant

for CA penetrations above 50% in scenarios where the number of deployed macrocell carriers is greater than the multi-CC connectivity capabilities of the users. In such deployments, the load-aware PCell management can provide additional gains especially at the coverage throughput by improving the allocated resource shares between non-CA and CA users. Albeit the capacity enhancement gradually diminish with the CA penetration, traffic steering reduces SCell changes events irrespective of the CA UE ratio to 70% for the considered scenario. Nevertheless, the significant decrease of SCell changes is not reflected in the overall RRC signaling reduction achieved by traffic steering. The reason is that the absolute number of SCell additions/removals is noticeably higher than the SCell changes. This is due to the aggressive SCell addition policy applied in this study. By means of that, the RRC signaling reduction gains are reduced to ∼10-15% when all CA events are taken into account.

Chapter 6

HetNet Load Balancing with Dual Connectivity

6.1 Introduction

The concept of inter-eNB CA between macrocells and small cells deployed at different carrier frequencies is also known as Dual Connectivity (DC). By means of that, UE devices can experience higher data rates as they can be served by both layers. This further enables the usage of collaborative PCell/SCell scheduling for balancing the load not only between co-located macrocell carriers but also between small cells and the macro overlay. In addition, HetNet mobility becomes easier since UE devices in DC mode can be configured with a macrocell PCell, which naturally shows a better handoff performance.

This chapter aims at proposing methods for managing PCells and SCells in dedicated carrier deployments supporting DC. Unlike most of the prior art studies, a more realistic approach is adopted where user mobility in both RRC UE states is explicitly modeled and the related latencies for managing SCells are accounted as well. This allows to capture specific effects not seen otherwise, based on which, a cell management framework is proposed that ensures the high utilization of the small cell layer together with decreasing the RRC signaling induced by DC.

The remainder of the chapter is organized as follows: Section 6.2 provides an overview of DC focusing primarily on the main state-of-art in terms of packet scheduling and mobility management. The proposed PCell and SCell management framework is described in Section 6.3. Section 6.4 summarizes the main simulation assumptions followed by the associated performance evaluation in Section 6.5. Finally, the chapter concludes with Section 6.6.

6.2 DC Overview and State-of-Art

As shown in Fig. 6.1, UE devices in DC mode typically have their PCell on the macro layer, while small cells are dynamically configured as SCells within their vicinity. SCell additions, removals and changes are triggered based on UE RRM measurements, while RRC signaling is utilized for executing such decisions.

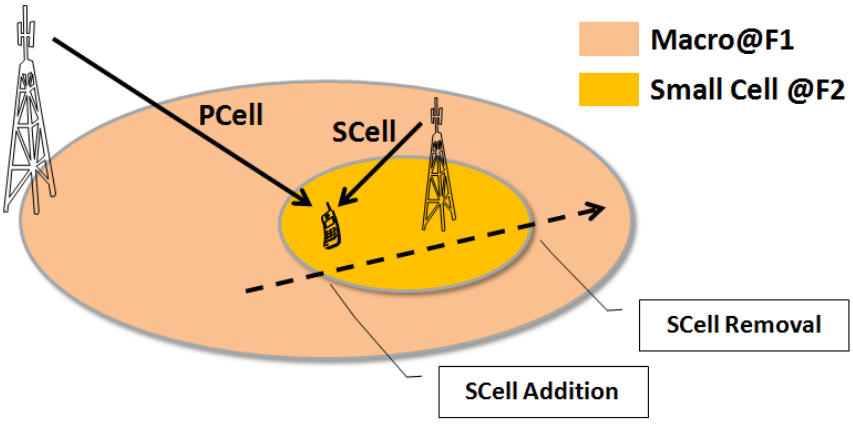


Fig. 6.1: UE device in DC mode. The PCell is assigned onto the macrocell while the small cell is typically configured as a SCell.

Mobility studies evaluating the handoff performance in dedicated carrier deployments with DC can be found in [106, 111, 112]. The bottom line is that DC naturally improves mobility robustness in HetNet scenarios. Challenging handovers between the two layers are avoided and users access the small cell bandwidth by means of SCell events. Therefore, even if the SCell connection on the small cell is lost, the UE still maintains a stable connection with the macro overlay; a fact that naturally reduces the likelihood of RLFs and HOFs.

Performance results evaluating the capacity gain of DC are available in [66, 113]. It is shown that cooperative multicell scheduling – e.g. cross-CC scheduler – is also essential for deployments supporting DC. To fully exploit the potentials of such scheduling scheme, a high bandwidth low latency interface is required for interconnecting the two layers. This enables their tight coordination so as to allocate the transmission resources in a fair manner; a fact that naturally results in fast load balancing between the two layers. An example of such cooperation is shown in Fig. 6.2. Specifically, *user C* is located at very close to small cell and it would perceive exceptional throughput even by being solely scheduled on the small cell layer. By means of that, the scheduler allocates transmission resources primarily on the small cell for that particular user, leaving more macrocell resources for UE devices away from the small cell vicinity (*user A*) and cell edge terminals operating in DC mode (*user B*).

Apart from enhanced mobility robustness and fast inter-layer load balancing, the work in [114] has shown that DC can effectively decrease the UE energy per delivered bit as long as the provided throughput benefits are above 20%. This is due to the faster transition to RRC idle that eventually compensates for the higher amount of power that the UE device consumes while operating in DC mode. Consequently, it is quite important to ensure that this requirement is met so that the power consumption of CA UE devices is not jeopardized by DC.

Unfortunately, the aforedescribed enhancements do not come without any cost. It is widely accepted that DC increases the network signaling. This is induced by the SCell management overhead. In fact, the aforementioned mobility studies estimate an overall RRC signaling increase of at least $\sim 30\%$ compared to when no DC is used. To reduce the overhead burden, the authors in [105] propose a novel method for enabling SCell management based on autonomous UE decisions assisted by some network control. By means of that, RRC signaling is essentially save as a portion of the standardized messaging exchange between the UE and the network can be bypassed, whenever a SCell action is realized.

6.3 Proposed PCell/SCell management for DC

6.3.1 Rationale

Albeit the valuable contributions of the prior state-of-art studies, they commonly neglect particular effects. E.g. having the PCell always on the macro overlay may limit the small cell offloading potentials, especially at low offered load conditions where the transmission time is short and it may not suffice for fully exploiting the small cell as a SCell. Such effects were not captured by the studies in [66, 113].

To make DC performance less dependent on the traffic type and the associated latencies whenever configuring SCells, it is suggested that the PCell of nomadic low mobility user within the small cell coverage is assigned on the small cell layer. By means of that, these users can instantly access the small cell bandwidth. Note that such an approach does not jeopardize mobility robustness as the low UE speed does not make the PCell management a challenging task. Furthermore, the co-channel interference is not strong enough to potentially degrade mobility performance as the small cells are deployed at a dedicated carrier. At the meantime, users that freely trespass several small cells with time-variant velocity should still be configured with macro PCells as their time-of-stay on a particular small cell is not long enough to worth a small cell PCell configuration.

The aforedescribed PCell management framework allows for a more flexible

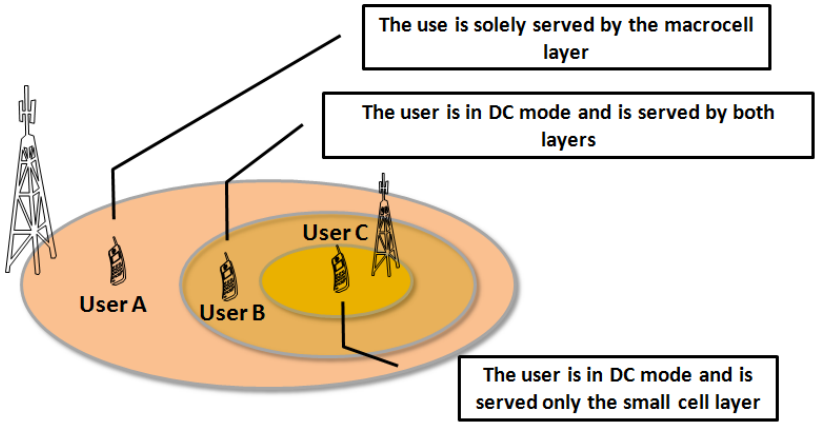


Fig. 6.2: Example of global resource allocation with DC and cross-CC scheduling. The two layers are assumed to be interconnected with a high bandwidth low latency interface so as to enable their tight coordination.

SCell policy as well. In principle, users with their PCell on the small cell layer should still be configured with macro SCells so that the benefits of the extended bandwidth are maintained. Nevertheless, as seen by Fig. 6.2, users experiencing exceptional radio conditions on the small cell are primarily scheduled on that particular layer for the sake global user fairness. Hence, defining a SCell policy that is capable of filtering out macro SCell configurations which bring small benefit to the user throughput could naturally decrease the DC RRC signaling without a noticeable cost in the UE experience.

The following subsections propose a cell management framework for realizing our reasoning. The major underlying assumption is that the network is capable of estimating user behavior. In practice, features such as UE History Information [115], Mobility State Estimation (MSE) and novel finger-printing techniques [101] could essentially provide an estimation about the speed and the location of UE devices. It is important to clarify that the suggested cell management framework solely refers to UE devices with DC capabilities. This essentially means that the cell management of legacy UE devices is still governed by load-based traffic steering, reusing the same algorithms as the ones developed for providing load balancing in Release 8/9 LTE.

6.3.2 Idle Mode Policy

Slowly moving hotspot UE devices are always provided with dedicated camping priorities – via the RRC connection release message – that prioritize the small cell layer over the macro overlay. In such a manner, these users are forced to camp on small cells and establish their PCell on that particular layer when

switching to connected mode again. Notice that they still can reselect to an overlaid macrocell if they are about to experience coverage problems on the small cell. Furthermore, intra-frequency cell reselections between neighboring small cells are supported via the R criterion.

On the other hand, free moving users are only configured with macrocell idle mode mobility. Hence, cell reselections to small cells can not occur. Similarly, the R criterion governs intra-frequency macro-to-macro cell reselections, while inter-frequency idle mode mobility and associated traffic steering schemes – e.g. DAP – can still be enabled if more than a single carrier are deployed at the macro overlay.

6.3.3 Connected Mode PCell Policy

For users having small cells configured as PCells, intra-frequency mobility support is provided via the A3 event for potential small cell-to-small cell handovers, while small cell-to-macro handovers are triggered whenever the connection on the small cell layer is about to be lost. Finally, users with a macrocell PCell are only allowed to perform handovers to neighboring macrocells. Apparently, if more than a single macrocell carrier is deployed, inter-frequency traffic steering policies such as FHO@CS can still realized at the macro overlay, as shown in the previous chapters of the dissertation.

6.3.4 Connected Mode SCell Policy

For users having their PCell on the small cell layer, it is desirable to configure macrocell SCells only for those that they will truly benefit from such an action. Particularly, a SCell addition event that relatively compares the signal quality between the two layers (i.e. A3 event) would serve that purpose. Hence:

$$Q_{SCell}^{RSRQ} > Q_{PCell}^{RSRQ} + CA_{Add}, \tag{6.1}$$

where Q_{PCell}^{RSRQ} and Q_{SCell}^{RSRQ} are the RSRQ measurements for the PCell and the target SCell, respectively; and CA_{Add} specifies the offset for adding a SCell. By setting $\mathrm{CA}_{Add} < 0$, then macro SCells will be configured only for those users whose signal quality on the small cell is worse or at least comparable with the one measured at the macro overlay. The degree of comparability between the experienced radio conditions in both layers is typically determined by the CA_{Add} parameterization. On the other hand, the macro SCell could be removed whenever the signal quality on the PCell becomes offset better than the SCell, i.e.

$$Q_{PCell}^{RSRQ} > Q_{SCell}^{RSRQ} + CA_{Rem}, \tag{6.2}$$

where CA_{Rem} is the associated offset of the SCell removal event and it is set to $CA_{Rem} = -CA_{Add}$ so that repetitive additions and removals of the same SCell are avoided.

With regards to free moving users with the macrocell configured as PCell, small cell SCell additions/removals are governed by the threshold-based events described in Table 5.2[1]. The reason is that a SCell addition on the small cell will almost always enhance their perceived data rates. Therefore, there is no need to relatively compare the signal quality of the two layers before adding a small cell SCell. If fact, even in the case of a too early small cell SCell addition, the scheduler would tackle this problem by assigning transmission resources on the macro overlay for the suffered UE device until the small cell radio conditions essentially improve.

Lastly, SCell changes can still be performed based on the A6 event defined in Table 5.2, regardless of the layer on which the PCell is configured.

6.4 Simulation Assumptions

The simulated deployment is Scenario D, as described in Section 2.6. Specifically, the 2.6 GHz band is now dedicated to the picocell layer with two additional macrocell carriers at 800 MHz and 1.8 GHz, respectively. Both intra-eNB and inter-eNB CA are supported in this scenario meaning that the scheduler offers fast load balancing without any inter-layer restriction.

The proposed PCell/SCell management policy is compared against the case when all CA UE devices have their PCell on the macro overlay, while SCell additions, changes and removals are performed according to the events of Table 5.2. Low mobility at 3 km/h is considered both for hotspot and free moving UE devices. It is assumed that the network provides the necessary means for identifying slowly moving CA users within a radius of 40 m from the picocell position. The applied CA_{Add} configurations are -4 dB, -2 dB and -1 dB, respectively. It is worth mentioning that picocell measurements are biased by 2 dB during the evaluation of a SCell condition. The considered CA UE ratios range from 0% to 100%.

Finite-buffer best effort traffic is simulated for different offered load conditions. Packet arrivals are modeled as a Poisson process and the packet size is negatively exponentially distributed with a mean value of 10 Mbits. Lastly, the

[1] A target cell is added as a SCell whenever its signal quality is above -16 dB (A4 event). On other hand, a SCell is removed whenever its signal quality becomes worse than -18 dB.

Table 6.1: Simulation Parameters for Scenario D

Parameter	Value
Macrocell Carriers	800 MHz & 1.8 GHz
Picocell Carrier	2.6 GHz
UEs per Macrocell Area	30
Picos per Macrocell Area	2
UE Distribution	Hotspot
Offered Load	20:10:70 Mbps
Traffic Type	Finite Buffer
mean packet Size	10 Mbits
Scheduling	Cross-CC PF
SCell Support	1 SCell
CA UE Penetration	{0, 20, 50, 80, 100} %
CA$_{Add}$	{-4, -2, -1} dB

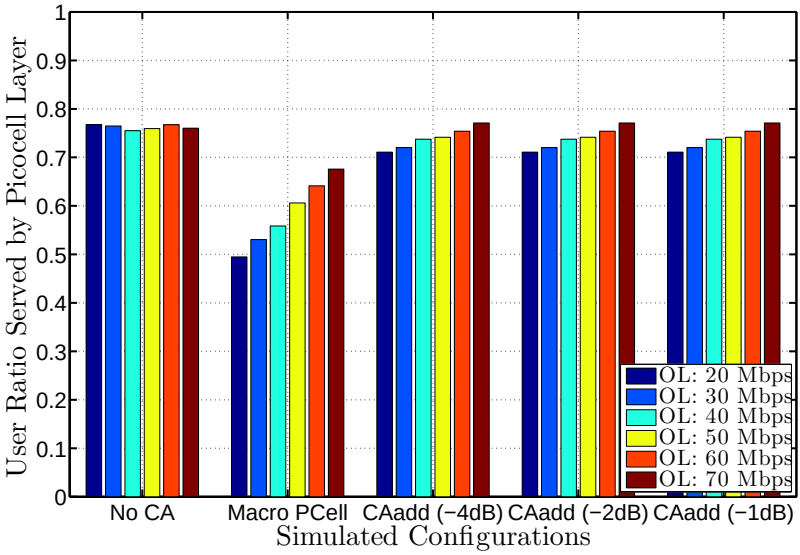
traffic steering (DAP+FHO@CS[2]) configuration as well as the mobility management parameterization are similar to the one applied in Chapter 4 and Chapter 5, respectively. The most relevant simulation assumptions are summarized in Table 6.1.

6.5 Impact of Offered Load and CA$_{Add}$

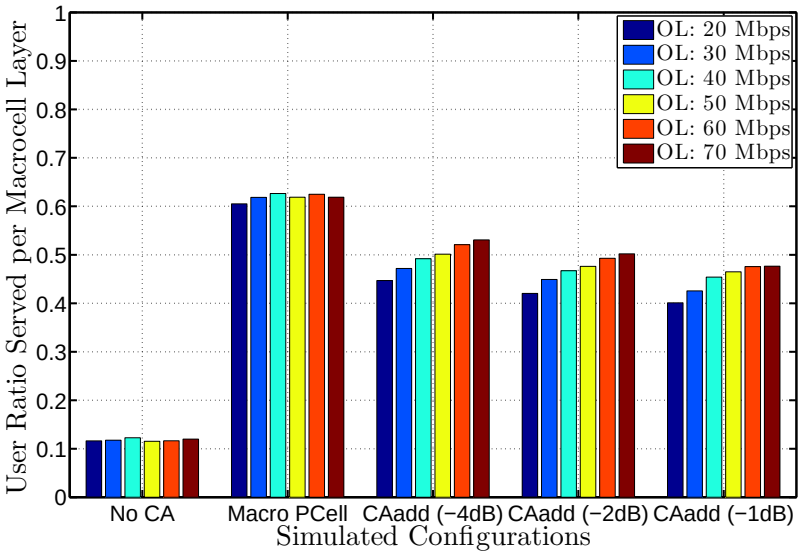
In this section, the two cell management policies are simulated for different offered load conditions and a CA UE ratio of 100%. Additionally, the case when all UE devices are legacy UE devices is simulated as well.

Fig. 6.3a illustrates the user ratio served by the picocell layer – averaged over the simulation time – for all simulated cases and offered traffic conditions. When no CA is used, DAP+FHO@CS pushes ∼76-78% of the total users to the small cell layer. Unlike the previously investigated scenarios where picocells were deployed in full reuse with the macro overlay, the picocell coverage is no longer limited by the macrocell interference. Consequently, a significant larger amount of UE devices can be offloaded to the small cell layer. However, this is not the case when DC is enabled with all PCells configured on the macro overlay. In principle, a smaller ratio of users is served by the deployed picocells, especially at low offered load conditions. The reason is that CA users may already have emptied their buffer before the SCell on the picocell layer is activated. If this

[2]DAP+FHO@CS is applied only for non-CA UE devices without inter-layer restrictions. On the other hand, traffic steering support for CA users is solely provided at the macro overlay, as discussed in Section 6.3



(a) Ratio of users served by the picocell layer.



(b) Ratio of users served per macrocell layer.

Fig. 6.3: User ratio served per layer in Scenario D for all simulated PCell/SCell management policies. All users are considered CA capable.

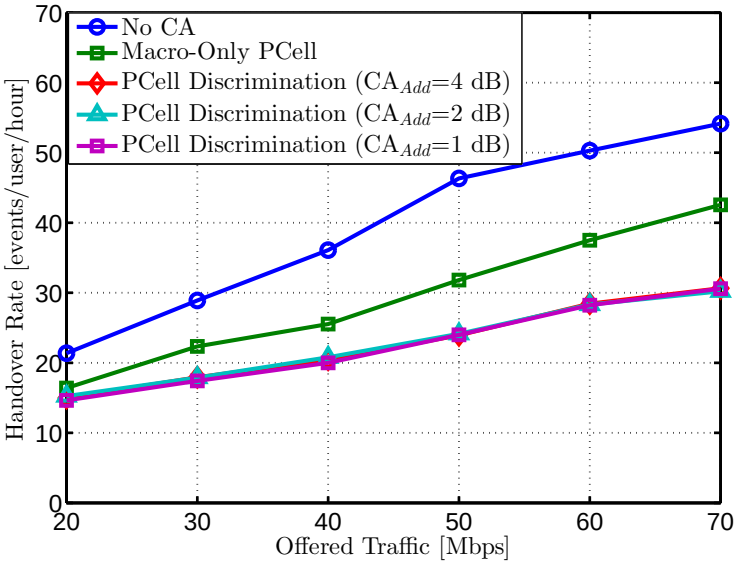
occurs, then the SCell is still configured but it can be exploited only if a new data session starts before the expiration of the connection release timer. As the offered load increases, sessions naturally last longer. Thus, there is a higher probability that users will experience throughput in their pico SCells. In such a manner, the number of served UE devices increases with the offered load when configuring the DC mode with macro-only PCell management.

This undesired behavior is avoided if hotspot users are configured with their PCells on the small cell layer. In this case, the picocell bandwidth is at least exploited by the hotspot UE devices which constitute 66% of the total users in the network. As explained above, the associated user ratio increases with the offered traffic simply because more free moving users access the picocell bandwidth by means of SCell additions at higher load conditions. Apparently, the CA_{add} parameterization does not affect the investigated metric since it solely governs the dynamic macro SCell management of hotspot UE devices.

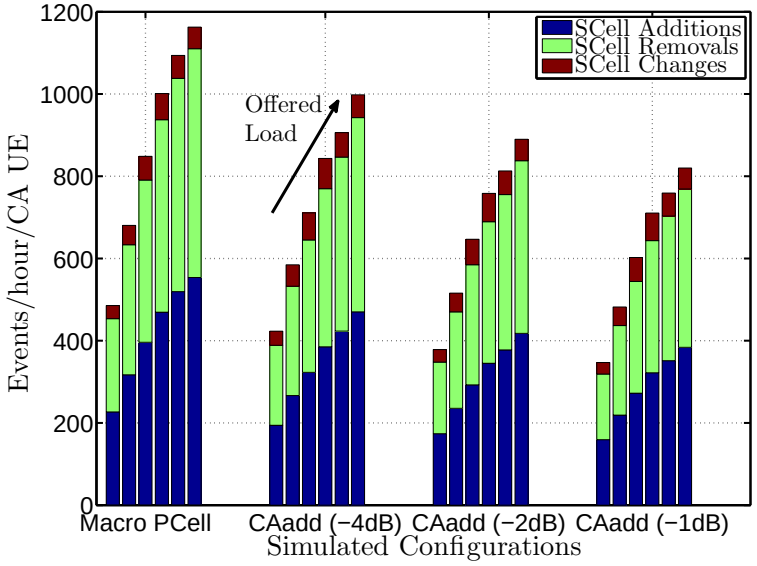
The effect of CA_{add} can be observed in Fig. 6.3b showing the average ratio of served users per macrocell layer. In fact, this particular KPI decreases for a larger CA_{add} value. The reason is that less users operate in DC mode, as hotspot UE devices located more distant from picocells are solely configured with a macro SCell. To avoid any misconception from the reader, summing the corresponding ratios over all layers essentially equals to 1 only in the non-CA case, where UE devices can solely connect to a single cell. On the other hand, samples from a CA UE device may count in more than a single layer depending on whether the user is scheduled on both of its serving cells.

The mobility performance in terms of handover rates and SCell events is depicted in Fig. 6.4. In line with previous results, handovers naturally decrease with CA. Nevertheless, the proposed PCell management method for hotspot users achieves a better performance compared to when all UE devices have their PCell on the macro overlay. The reason is that less PCells are assigned on the overlaid macrocell carriers which decreases the probability of triggering inter-frequency handovers between the macrocell carriers. Notice that hotspot UE devices can still perform picocell-to-picocell handovers. Nonetheless, the small cell deployment density of the considered scenario is not high enough to increase the RRC signaling caused by the intra-frequency hotspot mobility on the 2.6 GHz carrier.

As for SCell events, the proposed scheme reduces the DC RRC signaling. Compared to the macro-only PCell configuration, the associated event rates are reduced irrespective of the offered traffic conditions to $\sim$15%, $\sim$25% and $\sim$30% when CA_{Add} is set to -4 dB, -2 dB and -1 dB, respectively. Apparently, the gain comes by avoiding macro SCell additions for hotspot UE devices with adequate small cell coverage. As the proposed method solely tackles the RRC signaling induced by hotspot CA users, the concept of UE autonomous SCell management [105] discussed in Section 6.2 could be applied for managing the



(a) Handover Rates



(b) SCell Event Rates. The offered load ranges from 20 Mbps to 70 Mbps per macrocell area in steps of 10 Mbps.

Fig. 6.4: Mobility performance of Scenario D for all simulated PCell/SCell management policies. All users are considered CA capable.

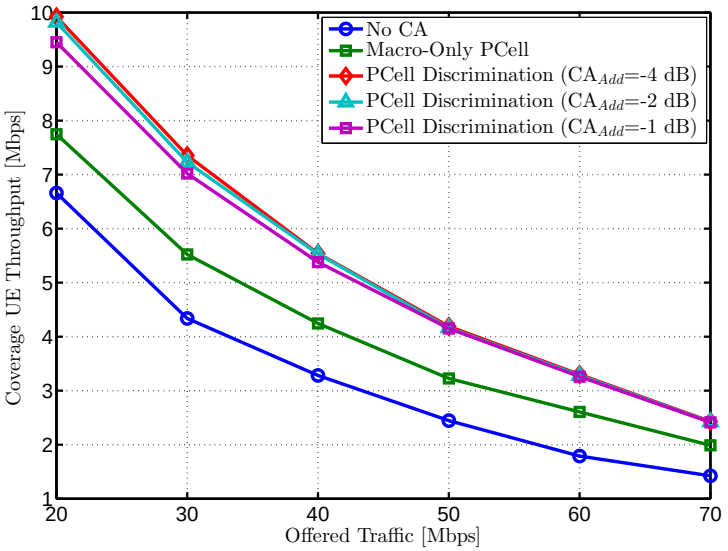
SCells of free moving users. By means of that, the signaling burden of each SCell action would decrease despite the fact that users would experience the same SCell event rates.

Fig. 6.5a and Fig. 6.5b show the coverage and average UE throughput for all simulated cases. When CA is enabled, data rates essentially improve as users benefit from the larger transmission bandwidth and the increased multi-user diversity. Nevertheless, the macro-only PCell configuration is constantly worse than the case when hotspot UE devices have their PCell on picocells. The reason is that the proposed method ensures a better utilization of the small cell layer. Ultimately, it can be observed that the CA_{Add} parameterization has almost no impact on the perceived throughput proving that the proposed SCell management policy is capable of filtering out unnecessary SCell configurations. Notice that there is some marginal loss when CA_{Add} increases at low load conditions. This is due to time instances when no active macrocell users exists in the network but hotspot UE devices are not configured with a macro SCell. Apparently, this situation is very unlikely to happen at higher load conditions where the optimal approach is to primarily schedule the cell center hotspot users on the picocell layer irrespective of whether a macro PCell is added or not.

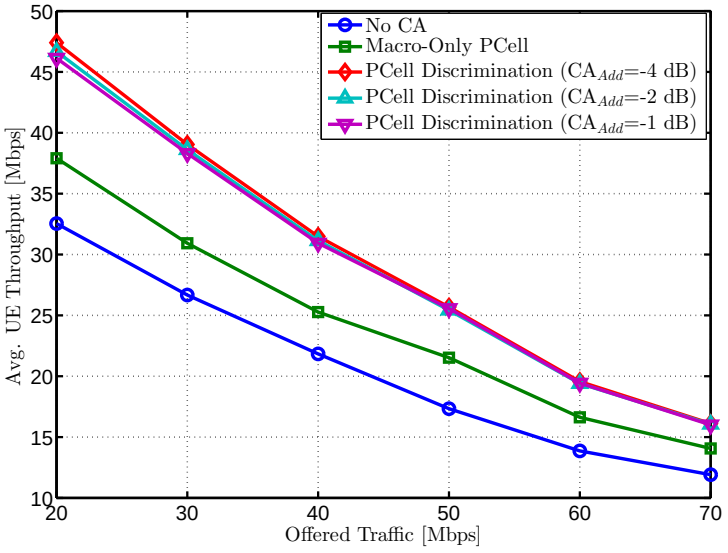
6.6 Impact of CA UE Penetration

In this subsection, the proposed PCell/SCell management policy is evaluated for different CA UE penetrations and fixed offered load conditions of 70 Mbps per macrocell area. Notice that the C_{Add} parameter is configured to -1 dB since it is the one providing the maximum RRC signaling reduction, as seen above.

Fig. 6.6 illustrates the coverage and average UE throughput versus the CA UE penetration for both PCell/SCell policies. It can be observed that relaxing the requirements of maintaining the PCell always on the macro overlay does not benefit the UE perceived data rates at low-to-medium CA UE penetrations. The reason is that DAP+FHO@CS still pushes non-CA users to the deployed picocells; hence, the small cell layer is well exploited given that the ratio of non-CA UE devices is large enough. Nevertheless, the performance gap between the two policies steadily broadens as the CA UE penetration increases simply because more PCells are assigned on the macro overlay; a fact that results in worse utilization of the picocell layer for the reasons explained in Section 6.5. It is worth mentioning that the gains of the proposed PCell discrimination policy over the conventional macro-only PCell configuration naturally depend

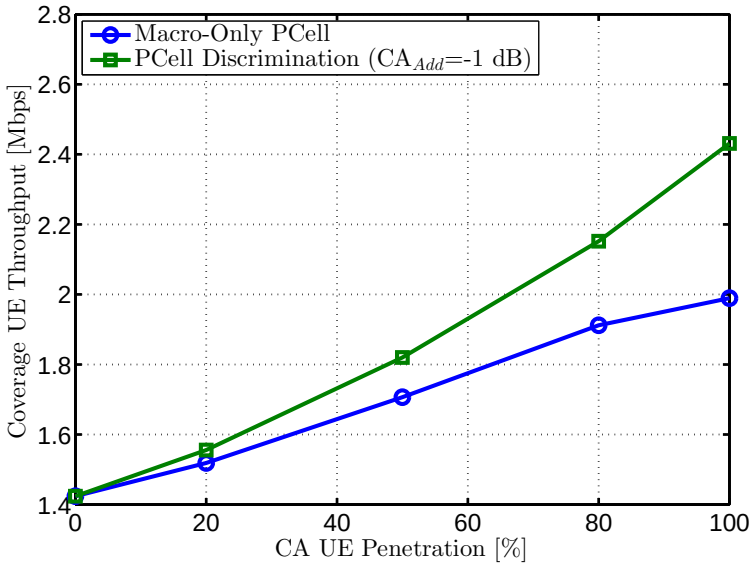


(a) Coverage UE Throughput

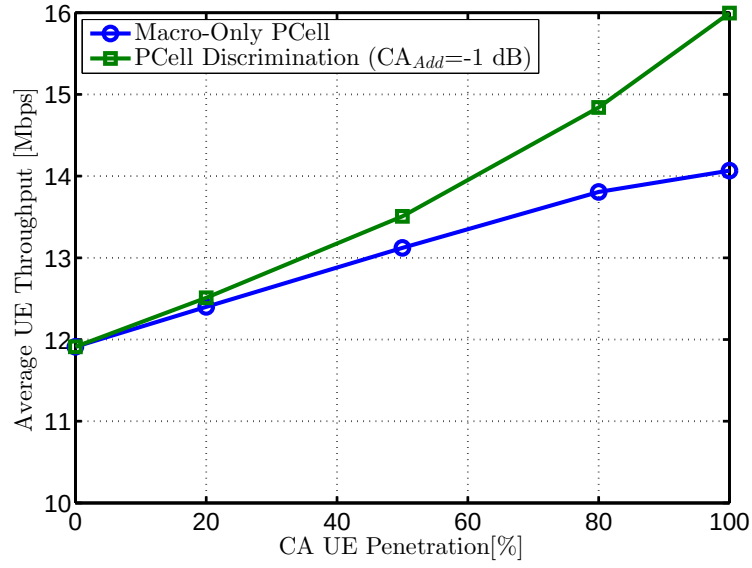


(b) Average UE Throughput

Fig. 6.5: UE throughput performance in Scenario D for all simulated PCell/SCell management policies. All users are considered CA capable.



(a) Coverage UE Throughput



(b) Average UE Throughput

Fig. 6.6: UE throughput performance for all simulated PCell/SCell management policies and different CA UE ratios. A fixed offered load of 70 Mbps per macrocell area is considered.

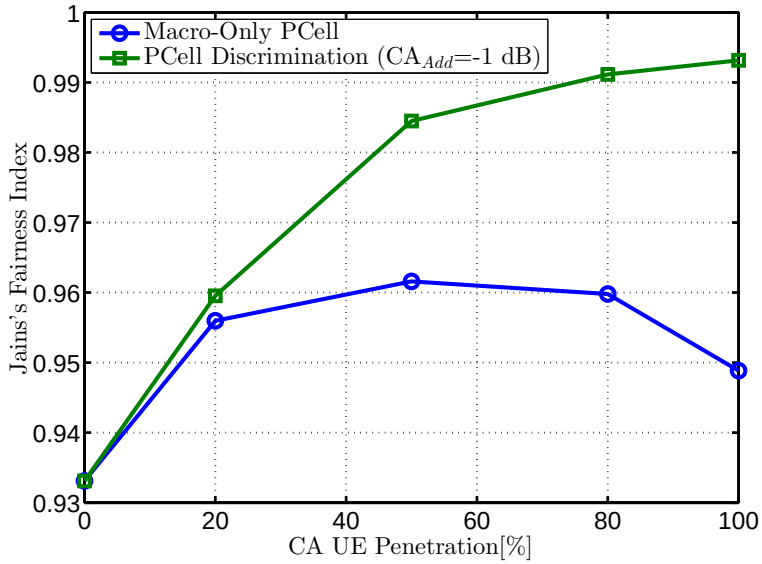


Fig. 6.7: Load balancing performance for all simulated PCell/SCell management policies and different CA UE penetration. A fixed offered load of 70 Mbps per macrocell area is considered.

on the offered load. Hence, higher gains would be achieved if lower offered load conditions were simulated.

The associated degree of achievable load balancing is depicted in Fig. 6.7, expressed in the form of Jain's fairness index. Albeit the load balancing performance achieved by the macro- only PCell configuration is generally high owing to the scheduler, it is still impacted by the buffer size and the associated latencies adding a SCell on the picocell layer. These two factors naturally reduce the time that is available for exploiting the picocell bandwidth by CA UE devices with a macro PCell. By means of that, the estimated Jain's index starts decreasing for CA UE ratios above 50%. The aforedescribed dependencies disappear with the proposed PCell policy since the slowly moving CA users can immediately start service on the picocell layer. This results in a better load balancing performance which is further reflected by the corresponding Jain's index that steadily increases with the CA UE penetration.

Ultimately, Fig. 6.8 presents the SCell event rates for all simulated cases. In particular, the relative SCell policy introduced for CA users with a picocell SCell decreases the CA RRC signaling by 30% regardless of the CA UE penetration. As these reduction gains are similar to the ones estimated in Section 6.5, it makes us concludes that the RRC signaling savings achieved by the pro-

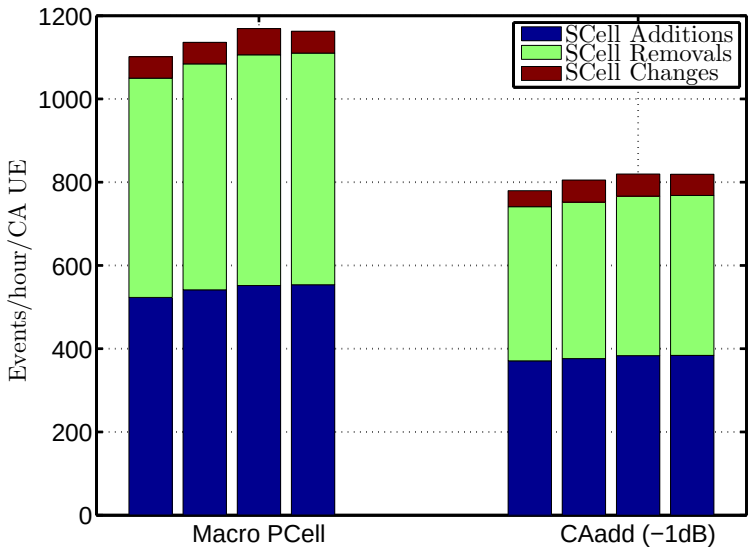


Fig. 6.8: SCell events rates for all simulated PCell/SCell management policies and different CA UE penetration. From left-to-right the considered CA UE penetrations are [20 50 80 100] . A fixed offered load of 70 Mbps per macrocell area is considered.

posed scheme mainly depend on the spatial UE distribution, irrespective of the offered load conditions and the CA UE ratio.

6.7 Conclusions

In this chapter, a cell management framework has been proposed for governing the PCells and SCells of users operating in DC mode. In a nutshell, the suggested scheme relaxes the requirement of maintaining the PCell solely on the macro overlay by allowing nomadic slowly moving hotspot UE devices to be configured with a picocell PCell. At the meantime, macrocell SCell configurations for these particular users are governed by events that relatively compare the signal quality of the two layers. The performance of the proposed cell management policy has been evaluated by means of system level simulations using as a reference the case when all CA UE devices have their PCell on the macro overlay and picocell SCell configurations are managed by threshold-based events.

Simulation results have shown that DC naturally diminishes the need of steering CA users to picocells by means of load-driven mobility events. The reason

is that these users can access them via SCell configurations, while the scheduler can also provide fast load balancing between the two layers. Therefore, load-based traffic steering remains relevant only for legacy UE devices, while proper cell management policies can be applied for users operating in DC mode. Compared to the macro PCell-only configuration, the proposed cell management framework improves the network capacity at high CA UE penetrations while it further reduces the DC RRC signaling up to ∼30% regardless of the CA UE ratio and the offered load conditions. These signaling savings are achieved without limiting the capacity gains of DC as macrocell SCells are only configured for the hotspot UE devices that truly benefit from such an action.

Chapter 7

Conclusions & Future Work

This dissertation has constituted a "*journey*" through the different LTE Releases analyzing their impact on load balancing performance. For that purpose, a handful of load balancing solutions has been evaluated in a set of different HetNet LTE deployments. These started from a single-carrier study and then moved forward to more mature scenarios with an increasing number of frequency layers allowing the usage of Release 10 intra-eNB Carrier Aggregation (CA) and Dual Connectivity (DC), respectively.

This final chapter summarizes the main findings from all conducted studies in the form of general recommendations. The given guidelines are organized per LTE release, while some ideas for future work are proposed as well.

7.1 Recommendations for Release 8/9 LTE

Exploiting mobility procedures is the only means of achieving dynamic load balancing in Release 8/9 LTE. However, this needs to be performed in a efficient manner so as to achieve a reasonable cost in handovers and cell reselections. It is generally suggested that:

- **Having both RRC modes push users to less loaded cells guarantees their mutual alignment.** This in turn reduces the amount of handovers and cell reselections required for balancing the load. At low load conditions, the capacity enhancements mainly come from idle mode traffic steering, while at higher offered traffic demands idle mode schemes primarily reduce the RRC signaling and connected mode traffic steering boosts network performance.

For co-channel load balancing, the modifications of the handover offset in connected mode should also applied in idle mode by modifying the camping cell

hysteresis correspondingly. For inter-frequency load balancing, load-adaptive UE dedicated camping priorities together with forced handover upon connection setup provide the necessary degree of alignment between the two RRC modes.

7.1.1 Co-channel Load Balancing

For co-channel HetNet load balancing, the main challenge is to maintain mobility robustness. To find an attractive trade-off between the capacity gain and the associated mobility performance, the following guidelines are given:

- **Dynamically modify cell-pair handover offsets based on Mobility Load Balancing (MLB) and Mobility Robustness Optimization (MRO) functions.** To solve the problem of double responsibility conflict between the two SON functions, MRO monitors the mobility performance and signals the range within which MLB is eligible to modify handover offsets for load balancing purposes.
- **Supplement MLB/MRO with interference mitigation schemes as well as provide differentiated mobility configurations based on UE speed.** The reason that the strong inter-cell interference makes MRO incapable of resolving all mobility problems.

The proposed framework coordinates the interaction between MLB and MRO, while it allows operators to control the load balancing aggressiveness based on the targeted mobility performance.

7.1.2 Inter-frequency Load Balancing

Load balancing across cells deployed at different carriers requires an efficient utilization of inter-frequency measurements. This is for reducing their cost in UE power consumption and transmission gaps together with avoiding unnecessary mobility events. To achieve this goal, it is highly recommended to:

- **Decouple inter-frequency mobility from load balancing.** This is achieved by configuring the associated idle and connected mode measurement events such that handovers/cell reselections are only triggered for mobility robustness purposes when the UE is about to experience coverage problems on its serving carrier.
- **Explicitly request measurements from load balancing UE candidates so as to push them to a less loaded carrier whenever overload is detected.** Measurements can be requested from users that are about to switch to idle or they just have established a radio bearer.

- **Use load-adaptive dedicated camping priorities together with forced handovers upon connection setup.** The former are provided via the Connection Release message and force users to camp on the desired carrier. On the other hand, forced handover upon the connection establishment can resolve outdated traffic steering decision caused by long periods in idle mode. If used do not frequently switch from one RRC mode to another, then forced handovers during the connection life-time should also be employed.

All in all, the proposed inter-frequency traffic steering framework ensures the efficient utilization of physical layer measurements while achieving a good load balancing performance with a reasonable cost in RRC signaling.

7.2 Recommendations for Release 10

The load balancing solutions developed for Release 8/9 LTE are also scalable with Release 10 and beyond. **The only required modification involves the re-definition of the load metric so as to comprise the effect of multi-carrier connectivity**. Based on this finding, the following guidelines are given in order to achieve dynamic load balancing in multi-layer HetNet deployment with intra-eNB CA:

- **Reuse the existing idle and connected mode schemes developed for Release 8/9 LTE so as to provide a load-aware PCell management.**
- **Exploit signal quality measurements for adding, removing and changing Secondary Cells (SCells).**
- **Apply collaborative PCell/SCell scheduling so as to assist load-based traffic steering in balancing the load across collocated cells.**

The proposed method for integrating load-based traffic steering with CA provides a good load balancing regardless of the considered scenario. Furthermore, it results in RRC signaling savings by reducing SCell changes in scenarios where the number of collocated carriers is greater than the multi-carrier connectivity capabilities of the CA users.

7.3 Recommendations for Release 12

With Release 12 DC, users can be offloaded to the small cell layer by means of SCell configurations, while the scheduler can further offer dynamic load

balancing between small cells and the macro overlay. In this context, **traffic steering becomes only relevant for legacy UE devices so as to ensure that these users are offloaded to the small cell layer by means of load-driven mobility procedures**. Nevertheless, suitable cell management policies should be applied for users operating in DC mode so as to optimize system performance. The associated guidelines are outlined below:

- **Relax the requirement of managing the mobility of all DC users by the macro overlay.** Nomadic slowly moving DC users should be provided with fixed dedicated priorities that prioritize the small cell layer. By means of that, their PCell is established to the small cell layer when switching to connected mode again. For free moving DC users, macro-only mobility support is provided.
- **Avoid macrocell SCell configurations that hardly provide any throughput benefit to users having their PCell on a small cell.** This is achieved by configuring SCell events that relatively compare the signal quality of the two layers. In such manner, macrocell SCells are solely configured for the cell edge hotspot UE devices, which typically are the ones that truly from benefit from DC.

The proposed cell management framework makes DC performance less dependent on the traffic type and the delays for configuring SCells. Albeit the resulting throughput gains are load-dependent, it reduces the SCell management overhead regardless of the offered load conditions and the DC user penetration.

7.4 Future Work

This section intends to propose some ideas for future work. Apparently, there are several directions for broadening the study scope; nevertheless, we limit the future endeavor to some interesting topics to be analyzed.

To start with, integrating MLB and MRO with eICIC is definitely a topic worth investigating. As seen in Chapter 3, MLB and MRO change mobility parameters on a cell-pair basis, while eICIC selects macrocell muting patterns on a cell basis. This essentially raises the issue of understanding how to coordinate these functionalities so as to take optimal decisions in an dynamic environment where many macrocells interact with many small cells.

Furthermore, a natural step forward is to remove the assumption of the regular-grid 3GPP network layout and evaluate the performance of the designed load balancing solutions in more realistic simulation environments. An attractive option is the explicit modeling of site-specific scenarios based on data obtained from operators. This involves the use of 3D topography maps for deriving both building and street information, site locations, small cell deployment in high

traffic hotspot zones, etc. Furthermore, ray-tracing tools should be utilized for providing a realistic representation of the propagation environment. Apparently, such scenarios illustrate a greater degree of irregularity and capture effects that can be seen in the regular grid 3GPP layout. Therefore, it is very interesting to analyze the behavior of load balancing when indoors users are considered and users realistically move along city streets. Complementing the 3GPP-based results with site-specific studies is the only means for making the drawn conclusions stronger.

Lastly, the seamless integration of Wi-Fi with LTE has gained significant momentum during the last years. Owing to the widespread adoption of Wi-Fi access points, operators desire enhanced features that adaptively control user offloading rather than blindly prioritize Wi-Fi over LTE. For that purpose, advanced traffic steering solutions need to be developed so as to enable the tight coordination between the two networks and further allow for dynamic offloading decisions based on load, user mobility, backhaul capacity, etc.

Appendices

Appendix A

Modeling Framework

This appendix aims at shedding some light on the modeling framework that supports this PhD study. For that purpose, the following sections give a deep insight into the underlying models used for representing radio propagation effects, physical layer aspects, radio resource management and higher layer mobility functionalities.

A.1 Propagation Environment

The macrocell pathloss is calculated based on the deterministic model in [68], being expressed in the logarithmic domain as follows:

$$L_{macro}^{dB} = 40(1 - 4 \cdot 10^{-3} h_c) log_{10}(d_{u,c}) - \\ - 18 log_{10}(h_c) - 21 log_{10}(f_c) + 80 \quad [dB], \quad \text{(A.1)}$$

where $d_{u,c}$ is the distance between user u and cell c; h_c is the base station height measured from above the average rooftop level (15 m); and f_c is the deployed carrier frequency. Similarly, the perceived pathloss between a small cell and the UE device is estimated using the model in [53], being necessarily adjusted to a carrier frequency of 2.6 GHz:

$$L_{pico}^{dB} = 140.7 + 36.7 log_{10}(d_{u,c}) - \Delta_{corr} \quad [dB], \quad \text{(A.2)}$$

where Δ_{corr} is the correction factor applied for carrier frequencies above 2 GHz and it is set to 3.5 dB [117, 118].

Large scale shadowing effects are explicitly modeled as a zero mean log-normal distribution with a different standard deviation σ_s subject to the base station technology. Specifically, σ_s equals to 8 dB for macrocells, while 10 dB are considered for the small cells. Notice that full correlation is assumed for cells served by the same eNB; otherwise, a correlation coefficient of 0.5 is accounted. In accordance with [119], the shadowing spatial variability is given by the decorrelation distance, d_s, which differs between macrocells and picocells. The considered values for macro base stations and small cells are 50 m and 13 m, respectively.

Antenna gains are calculated based on the elevation and azimuth angle, while a fixed loss of 20 dB is accounted for the penetration losses due to building obstacles [68]. Finally, velocity-dependent fast fading is only considered as a error component in the modeling of the physical layer measurements so as to emulate the fast fading smoothing that typically occurs at Layer-1 for mobility management purposes. Therefore, it is not taken into account whenever calculating wideband values.

A.2 Physical Layer

This section presents the assumptions related to the LTE physical layer modeling. First of all, the SINR formulation is provided followed by the abstraction model used for estimating the wideband user throughput. Finally, the last subsection focuses on the modeling of the physical layer measurements in terms of RSRP and RSRQ.

A.2.1 SINR Formulation

The wideband received power $P_{rx,c}$ from cell c is given by:

$$P_{rx,c} = P_{tx,c} L_{u,c} \; [dBm], \tag{A.3}$$

where $P_{tx,c}$ is the total transmission power of the base station; and $L_{u,c}$ is the link loss between user u and cell c comprising the impact of the propagation effects discussed in Section A.1 except for fast fading. Based on (A.3), the wideband SINR of user u with regards to cell c can be written as:

$$\gamma_{u,c} = \frac{P_{rx,c}}{\sum_{k \neq c} \rho_k P_{rx,k} + N}, \tag{A.4}$$

where N is the received noise power; and $\rho_k \in [0\ 1]$ is the resource utilization of each interfering node k. Particularly, ρ_k scales the interference power so as

to acquire a valid notion of the generated interference depending on the traffic conditions. For instance, cell k generates no interference when ρ_k equals to 0[1], while full interference is accounted whenever ρ_k is 1.

A.2.2 Estimation of User Throughput

The UE perceived data rates are estimated based on a set of rate functions that map the estimated wideband SINR γ_u to user throughput comprising the effect of link adaptation, HARQ management, spatial multiplexing, etc. The FDPS diversity gain is accounted by selecting the proper mapping curve subject to the number of active users U_c connected to cell c. Therefore, each rate function is typically denoted as $\eta(\gamma_u, U_c)$.

The impact of U_c on $\eta(\gamma_u, U_c)$ – expressed in the form of [kbits/s/PRB] – is depicted in Fig. A.1. It can be observed that the achievable throughput per PRB increases with U_c since more users exist in the cell and there is more multi-user diversity to be exploited by the scheduler. In addition to this, the model further captures the saturation of the FDPS gain as U_c increases. For that purpose, up to eight mapping curves are used in this work and the wideband data rate r_u can be essentially expressed as:

$$r_u = \eta(\gamma_u, min\{U_c, U_c^{max}\}) \cdot f_u, \tag{A.5}$$

where $U_c^{max} = 8$; and f_u is the amount of resources assigned to user u. Lastly, to avoid throughput estimations coming from unrealistically high SINR values that typically do not occur in real system due to RF impairments, the achievable throughput per PRB is bounded to 810 kbit/sec. This corresponds to a spectral efficiency of 4.5 bit/sec/Hz.

A.2.3 Physical Layer Measurements

RSRP is defined as the linear average of the received power measured on the resource elements that carry the reference signal [61]. Assuming that all resource elements are equally powered regardless of whether data or reference signal is carried, the measured RSRP from cell c is modeled as below:

$$RSRP_c = \frac{1}{N_{sbc}} P_{rx,c}^{PRB} \cdot e_f \cdot e_m \quad [Watt], \tag{A.6}$$

where $N_{sbc} = 12$ is the number of subcarriers over an OFDM symbol; P_{rx}^{PRB} is the wideband received power scaled down to a PRB basis; e_f is the error

[1] In case of $\rho_k = 0$, the generated interference due to the transmission of the reference signal is neglected by (A.4) for the sake of simplicity.

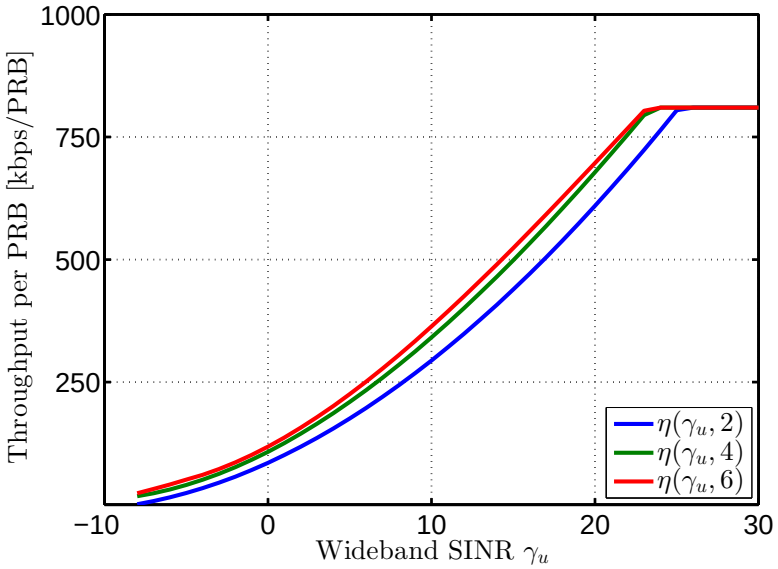


Fig. A.1: $\eta(\gamma_u, U_c)$ as a function of the wideband SINR γ_u and the number of scheduled users U_c

component that comprises the effect of Layer-1 fast fading averaging; and e_m is an additional error that emulates measurement imperfections. In particular, it is modeled as Gaussian random variable – truncated within the range of [-3, 3] dB – with a mean of 0 dB and a standard deviation of 1.21 dB, respectively [120].

As for RSRQ, it is defined as:

$$RSRQ = \frac{RSRP}{RSSI}, \tag{A.7}$$

where RSSI constitutes the total wideband received power on the carrier including serving cell, interference and noise. Thus, RSSI can be written down as:

$$RSSI = \sum_{k \in S_m} \rho_k \cdot P_{rx,k}^{PRB} + \sum_{k \in S_m} (1 - \rho_k) \cdot \lambda \cdot P_{rx,k}^{PRB} + N, \tag{A.8}$$

where S_m expresses the set of cells deployed at carrier m; and λ is the ratio of the received power when no data are transmitted meaning that only the reference signal power is measured in this case. Assuming that the reference

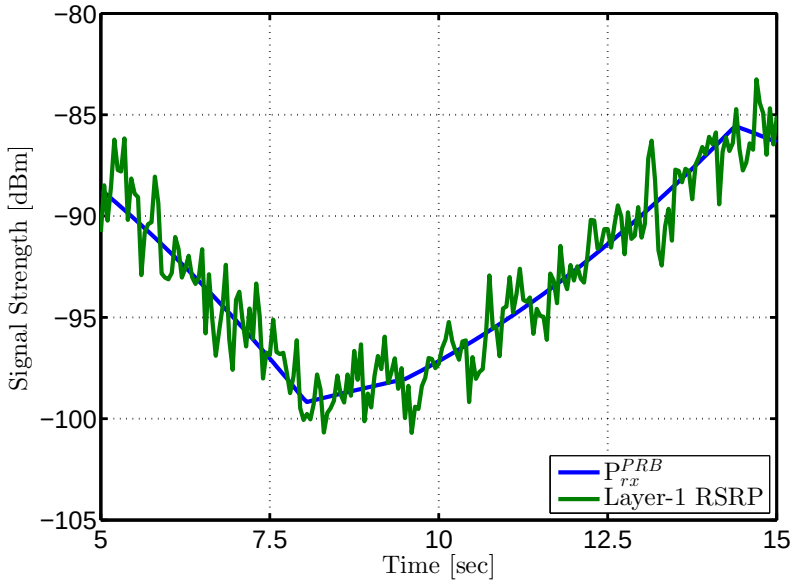


Fig. A.2: Time trace of PRB-scaled wideband received power together with the associated physical layer RSRP measurement. User velocity of 30 km/h is assumed.

symbols of the first antenna port are measured by the UE device, then λ equals to 2/12. By inserting (A.8),(A.6) in (A.7), RSRQ is expressed as:

$$RSRQ = \frac{\frac{1}{N_{sbc}} P_{rx}^{PRB} \cdot e_f \cdot e_{meas}}{\sum_{k \in S_m} \rho_k \cdot P_{rx,k}^{PRB} + \sum_{k \in S_m} (1 - \rho_k) \cdot \lambda \cdot P_{rx,k}^{PRB} + N} \tag{A.9}$$

A.3 Layer-3 Filtering

The physical layer measurements are subject to additional time-domain filtering at Layer-3, i.e:

$$\widetilde{M}_t = (1 - \alpha) \cdot \widetilde{M}_t + \alpha \cdot M_t, \tag{A.10}$$

where M_t is the latest physical layer measurement; $\widetilde{M}_t$ is the updated filtered value at time t; and α equals to $1/2^{(K/4)}$, where K is the Layer-3 coefficient and is responsible for further smoothing the physical layer measurements. As seen by Fig. A.3, large values of K limit the capability of following the signal

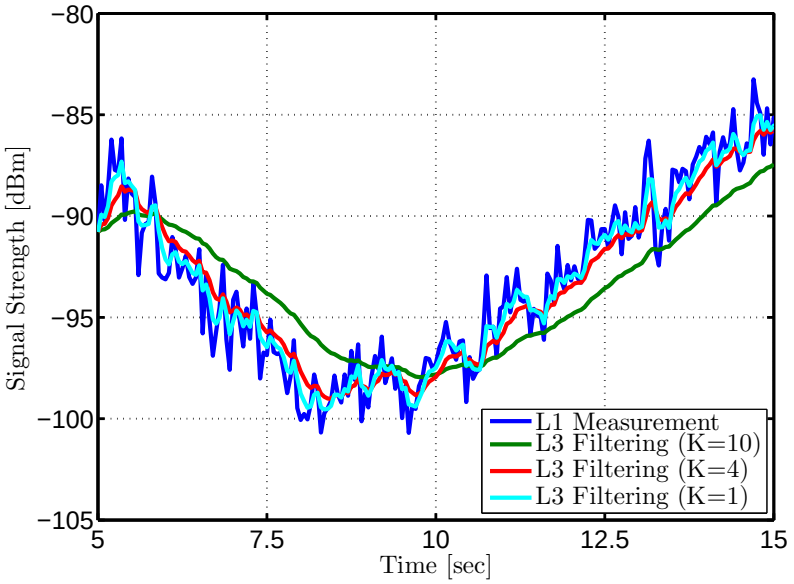


Fig. A.3: Impact of Layer-3 filter coefficient on physical layer measurements.

properly; a fact that can jeopardize mobility robustness. The Layer-3 coefficient is a network-configurable parameter with $K = 1$ and $K = 4$ to be the most typical configurations.

A.4 Radio Resource Management

This section describes the assumptions related to how users share transmission resources together with defining the associated cell resource utilization. Apparently, the derived formulations depend on the traffic QoS class and the user connectivity capabilities (i.e. non-CA or CA user). It is worth mentioning that the underlying models reuse and neatly extend the mathematical framework presented in [63, 116] so as to enable the explicit representation of such procedures over time periods longer than the millisecond basis with an attractive trade-off between accuracy and computational complexity. In other words, the reader should interpret the below calculations as average values for time periods in the order of tenths of milliseconds during which users move from one position to another.

A.4.1 Case without Carrier Aggregation

The following subsections refer to the case when only non-CA UE devices are considered. The associated definitions are given both for CBR and best effort users and are valid as long as a single user type exists in the cell.

A.4.1.1 CBR Users

CBR users are denoted as satisfied if the amount of allocated PRBs suffices for meeting their bit rate requirement D_u. Hence, f_u can be defined as follows:

$$f_u = \frac{B_c D_u}{r_u^{max}}, \tag{A.11}$$

where r_u^{max} is the wideband throughput that user u would perceived if it were allocated the whole transmission bandwidth[2] ; and B_c is the cell bandwidth in PRBs. By means of that, the PRB utilization of cell c can be written as:

$$\rho_c = \frac{1}{B_c} min\{\sum_{u=1}^{U_c} f_u, B_c\}, \tag{A.12}$$

All users are satisfied whenever $\rho_c < 1$. Nevertheless, (A.12) does not provide any information about which users to declare unsatisfied whenever $\rho_c = 1$ and B_c does not suffice for satisfying all users. For that purpose, it is assumed that the scheduler tries to maximize the user satisfaction by sorting users in descending SINR order so as to start the resource assignment from the highest SINR users. The reason is that such users naturally request fewer resources. This essentially means that the worse SINR users are always the ones declared unsatisfied.

A.4.1.2 Best Effort Users

A simple resource-fair policy is applied for best-effort users, i.e.

$$f_u = \frac{B_c}{U_c} \tag{A.13}$$

[2]the derived resource utilization depends on the user perceived throughput which in turn requires SINR information. To overcome this dependency, the fixed-point iteration solution proposed in [63, 116] is applied. Specifically, the resource utilization estimated for the time period t_{l-1} is used as an input for calculating the SINR for the current time period t_l.

As some users may empty their buffer and others may not, the average resource utilization for this particular time period is approximated by the following equation:

$$\rho_c = \frac{1}{B_c} \sum_{u=1}^{U_c} min\{f_u^b, f_u\}, \tag{A.14}$$

where f_u^b is the amount of resources required for emptying the buffer of user u. Evidently, $\rho_c = 1$ if full buffer traffic is considered, while $\rho_c \in [0\ 1]$ for finite buffer traffic.

A.4.2 Case with Carrier Aggregation

For cases with a mixture of non-CA and CA users, it is of key importance to model collaborative PCell/SCell scheduling. For that purpose, this subsection presents the developed modeling framework for capturing the main properties of such scheduling scheme without any need for an explicit implementation at a subframe granularity. Finally, please notice the model is only valid for users best-effort traffic.

To start with, for the case with CC-independent scheduling the resource share of each user – regardless of its connectivity capabilities – can still be approximated by using (A.13). However, this resource management policy results in poor coverage throughput owing to the resource starvation of non-CA users. To improve user fairness, a cross-CC scheduler would primarily assign transmission resources for CA users to the cells that can serve them with the highest throughput, as also discussed in Chapter 5. Clearly, the achievable data rates achieved with CC-independent scheduling – i.e. estimated from (A.14) – could provide an indication of where is preferable to allocate more resources for CA users. For that purpose, the concept of virtual user is introduced, according to which, the user presence in each cell is scaled as below:

$$w_{u,c} = \frac{r_{u,c}^{ind}}{\sum_{c \in X_c} r_{u,c}^{ind}}, \tag{A.15}$$

where $r_{u,c}^{ind}$ is the perceived user throughput on cell c assuming no PCell/SCell scheduling collaboration; and X_c constitutes the set of serving cells for user u. This essentially means that $w_{u,c}$ equals to 1 only for non-CA users, while $w_{u,c}$ ranges in-between 0 and 1 for CA users. In fact, if $w_{u,c} > w_{u,c'}$ then the CA user u should be assigned more resources in cell c rather than cell c', and vice versa.

To incorporate this effect into the resource allocation solution and eventually emulate the behavior of collaborative PCell/SCell scheduling, (A.13) can be modified as follows:

$$f_{u,c} = \frac{w_{u,c} B_c}{\sum_{u=1}^{U_c} w_{u,c}}, \tag{A.16}$$

where $\sum_{u=1}^{U_c} w_{u,c}$ is total number of virtual users connected to cell c; while the use of $w_{u,c}$ in the nominator is for ensuring that the sum of allocated resources does not exceed B_c.

To derive the average PRB utilization per cell, (A.14) can still be reused given that a virtual buffer is defined per cell whose size $b^v_{u,c}$ is proportional to $w_{u,c}$. Hence, $b^v_{u,c}$ can be expressed as the product of w_i with b_u, where b_u is the actual buffer size for user u. By means of that, ρ_c can be written as:

$$\rho_c = \frac{1}{B_c} \sum_{u=1}^{U_c} min\{f^{vb}_u, f_u\}, \tag{A.17}$$

where f^{vb}_u is the amount of resources required for emptying the virtual buffer of user c in cell c.

Appendix B

MRO Reliability Analysis

This appendix aims at calculating the amount of mobility observations required for a reliable MRO prediction of the RLF probability. To achieve this goal, each handover attempt is considered as a binary random variable, i.e. '1' whenever it is successfully completed and '0' in case of a RLF declaration. By means of that, the probability of having X successful handovers out of N handover attempts is calculated by the Binomial mass probability function as follows:

$$P(X) = \binom{X}{N} p^X (1-p)^{N-X}, \tag{B.1}$$

where p is the expected success probability of each independent handover attempt.

Let $\widehat{p}$ denote the estimation of p made by MRO during the length of a MRO cycle. Then the associated error $\mid p - \widehat{p} \mid$ can be written as:

$$\epsilon = \pm z \sqrt{\frac{1}{N} \widehat{p}(1-\widehat{p})}, \tag{B.2}$$

where z corresponds to the 1-0.5α percentile of the Gaussian distribution and α is the error percentile. The associated z values for different confidence intervals are shown in Table B.1.

Fig. B.1 illustrates the estimated error for $\widehat{p} = 0.98$ (i.e. $\widehat{p}_{rlf} = 0.02$) as a function of the statistical sample N for different confidence intervals. Generally, it can be observed that the error decreases with the number of collected samples, regardless of the considered confidence interval. Targeting on a reliable MRO estimation with $\epsilon = 0.01$, the number of required collected samples is $N_{min} = 756$ for a confidence interval of 95%. With this amount of samples, MRO is 95% confident that the RLF probability is between 0.01

Table B.1: Table of z values for confidence intervals

Confidence Level	z
70%	1.04
80%	1.28
90%	1.645
95%	1.96

and 0.03. Apparently, N_{min} decreases for less reliable confidence intervals. E.g. ∼550 samples are required for a confidence level of 90%, while N_{min} equals to 200 samples when the considered confidence level is 70%.

Notice that for the conducted simulation campaign of Chapter 3, N_{min} was set to 20 samples. The related error is 0.061, 0.052, 0.040 and 0.033 for a confidence interval of 95%, 90%, 80% and 70%, respectively. To improve the algorithm reliability, MRO should either wait for several cycles before taking any decision or adequately extend the duration of the KPI collection cycle. Unfortunately, none of the aforementioned modifications were finally applied as the associated simulation run-time was unfeasibly long.

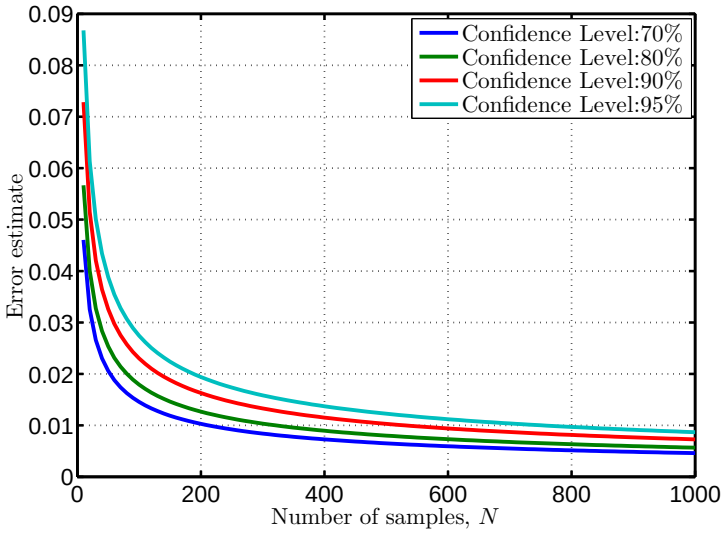


Fig. B.1: Estimation of error for the measured RLF probability as a function of the observed handover attempts for different confidence intervals. $\widehat{p}_{rlf}$ equals to 0.02.

Appendix C

Complementary Results for Chapter 4

This appendix aims at complementing Chapter 4 with some additional results. As already seen in Section 4.6, DAP+FHO@CS results in cell reselection rates which are significantly lower than the ones estimated for any RSRQ-based AP configuration. This effect clearly impacts the UE power consumption in idle mode. Hence, an estimation of the achievable UE battery savings is given in Section C.1. After that, the focus is put on analyzing the system performance when intra-frequency traffic is further enabled so as to balance the load between the two layers (macro/pico) deployed in full frequency reuse at the 2.6 GHz carrier.

C.1 Idle Mode UE Power Consumption Savings

Extending the duration of the DRX cycle for measuring neighbor cells as well as for performing cell reselections essentially increases the power consumption of idle UE devices. Specifically, the work in [56, 57] has shown that the relative increase/decrease of the UE power consumption in idle mode achieved by one mobility configuration with regards to another can be estimated by the following equation:

$$B = 100 \cdot \left(\frac{T^1_{tot} + T^1_{DRX}\left(\frac{I_{DRX}}{I_{idle}} - 1\right)}{T^2_{tot} + T^2_{DRX}\left(\frac{I_{DRX}}{I_{idle}} - 1\right)} - 1\right) \text{ [\%]}, \tag{C.1}$$

where T_{tot} is the total time that is on average spent by each UE device on idle mode; T_{DRX} is the associated time spent on DRX cycles; I_{idle} is the current drawn while the terminal is totally idle; and I_{DRX} is the current drawn

whenever the UE device wakes up so as to enter DRX. According to UE power measurement made on a LTE smartphone in [121], the I_{DRX}/I_{idle} ratio equals to ∼50. As for T_{DRX}, it can written as:

$$T_{DRX} = N_{DRX}(T_{page} + T_{meas} + T_{Res}), \tag{C.2}$$

where N_{DRX} is the average number of DRX cycles that each idle UE device entered; while T_{page}, T_{meas} and T_{res} correspond to the time spent at every DRX cycle on paging, measuring neighboring cells and performing cell reselections, respectively. For the sake of simplicity, let us consider that $T_{page} + T_{meas} = 30$ msec [56, 57]. Then, (C.1) can be written as:

$$B = 100 \cdot \left(\frac{T^1_{tot} + N_{DRX}(0.03 + N^1_{res}T^{Exec}_{res})(\frac{I_{DRX}}{I_{idle}} - 1)}{T^2_{tot} + N_{DRX}(0.03 + N^2_{res}T^{Exec}_{res})(\frac{I_{DRX}}{I_{idle}} - 1)} - 1\right) \text{ [\%]}, \tag{C.3}$$

where N_{res} is the average number of cell reselections per UE device; and T^{Exec}_{res} is the time required for completing a single cell reselection.

Using (C.3), the idle mode UE battery savings achieved by DAP+FHO@CS relative to the RSRQ-based mobility configurations are illustrated in Fig. C.1. In fact, DAP+FHO@CS provides significant idle mode UE power consumption savings regardless of the considered reference case. The highest reduction gains are in the order of ∼45% and are achieved over the $A2_{thresh} = Thresh^{AP}_{sLow} = -12$ dB case since this is the most costly configuration in terms of cell reselections, as already shown in Fig 4.7b of Chapter 4.

C.2 Joint MLB and DAP+FHO@CS Operation

The MLB algorithm – developed in Chapter 3 – is further enabled in Scenario B so as to better balance the load between the two layers deployed at the 2.6 GHz. Specifically, MLB is still configured as described in Table 3.2; however, both idle and connected mode UE devices are now considered. For that purpose, idle mode follows the MLB-applied ΔCIO adjustments by modifying the serving cell hysteresis Q_{hyst} accordingly. In such a manner, the probability of triggering idle-to-connected ping pong events between the two 2.6 GHz layers reduces significantly. The joint MLB and DAP+FHO@CS operation is evaluated for high offered traffic conditions of 40 Mbps per macrocell area, while the standalone DAP+FHO@CS performance is used as a reference.

Fig. C.2 illustrates the estimated user distribution with and without enabling co-channel MLB at the 2.6 GHz carrier. Owing to the exploitation of the cell

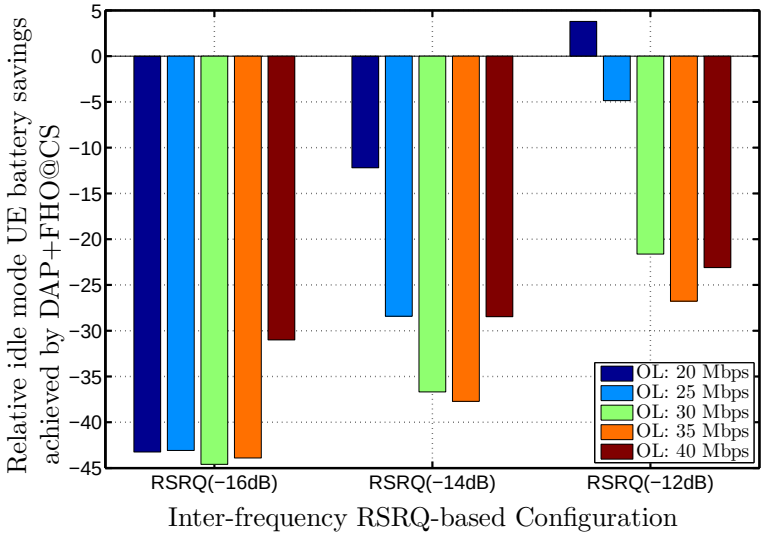


Fig. C.1: Esimated idle mode UE battery savings achieved by DAP+FHO@CS relative to the remaining simulated RSRQ-based configurations. α is set to 50 and $T_{res}^{exec} = 100$ msec.

load information, the associated handover and cell reselection parameters are modified in such a manner such that the picocell layer seems more attractive. By means of that, a user percentage of ∼5% is additionally offloaded to them, coming mainly from the 2.6 GHz macrocell layer.

The estimated Jain's fairness index together with the UE throughput performance are shown in Fig. C.3 for all load balancing configurations. It can be observed that the MLB usage improves Jain's index by a factor of ∼1.2 compared to when solely inter-frequency traffic steering is applied. Apparently, the reason is that the load of the underlaid co-channel deployment is more balanced when MLB is further enabled. However, this is not translated in significant benefits in terms of UE throughput since the strong interference between the two 2.6 GHz layers is a limiting factor. Compared to the standalone DAP+FHO@CS configuration, MLB enhances the coverage throughput by ∼15%, while the corresponding gain in average throughput is only ∼7%.

Fig. C.4a depicts the mobility performance in terms of handovers and cell reselections, respectively. MLB essentially increases the handoff rate by ∼10% since additional handovers are triggered for balancing the load of the 2.6 GHz carrier. Nevertheless, an interesting observation is that the rate of mobility events in idle and connected mode is more balanced when co-channel MLB is additionally employed. This is due to the optimization of the serving cell hysteresis value that necessarily increases the camping time on the picocell

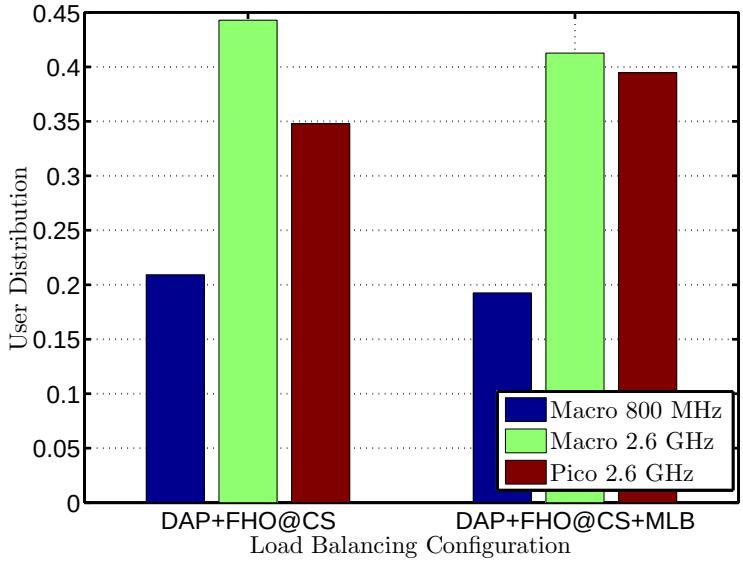
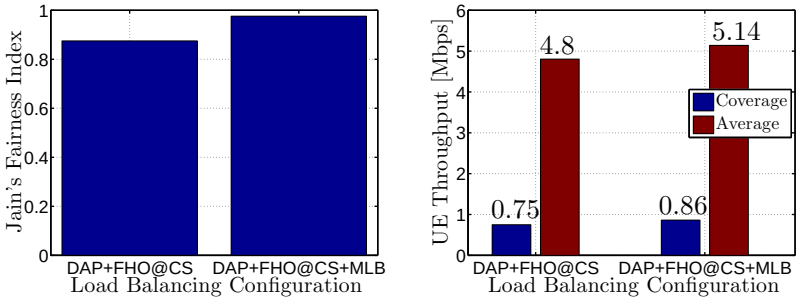


Fig. C.2: User distribution with and without MLB. An offered load of 40 Mbps per macrocell area is considered.

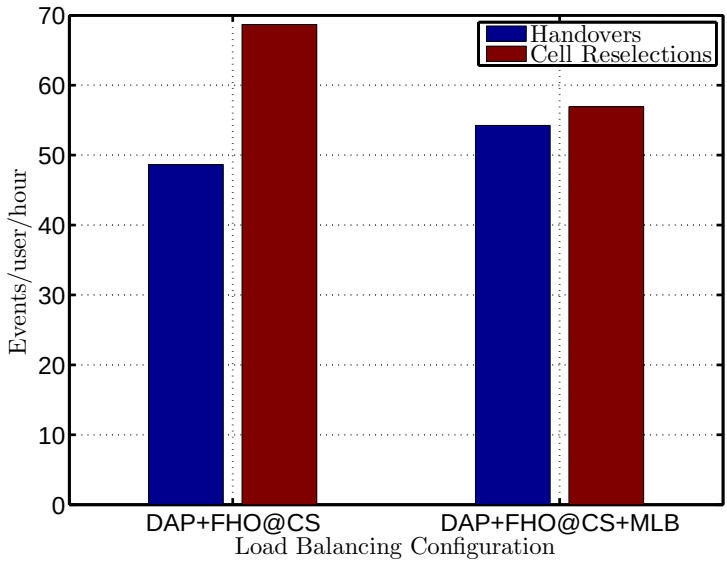
layer. This in turn reduces the amount of cell reselections performed between the two 2.6 GHz layers.

Finally, the estimated RLF rates for the simulated traffic steering policies are shown in Fig. C.4b. With regards to the DAP+FHO@CS case, no mobility problems are observed proving that the developed inter-frequency traffic

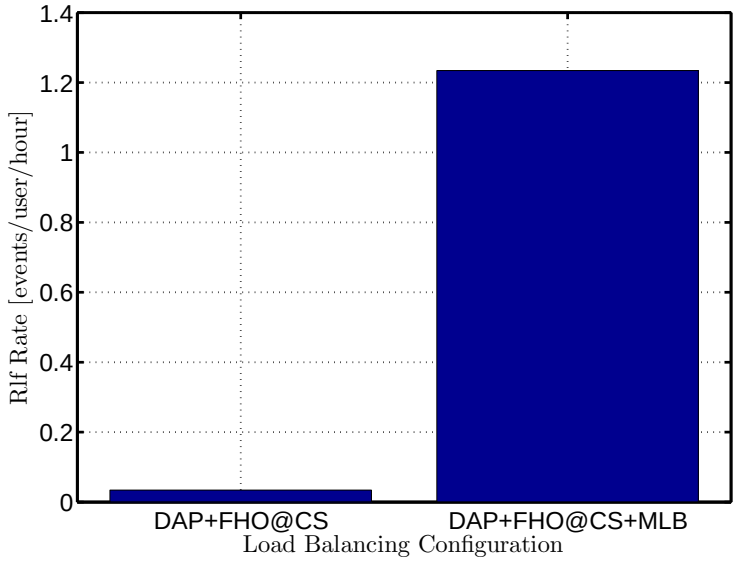


(a) Jain's Fairness Index **(b)** UE Throughput

Fig. C.3: MLB impact on the degree of achievable load balancing and UE throughput in scenario B. An offered load of 40 Mbps per macrocell area is considered.



(a) Handover and Cell Reselection Rates



(b) RLF Rates

Fig. C.4: Mobility performance with and without MLB. An offered load of 40 Mbps per macrocell area is considered.

steering solution does not jeopardize the mobility robustness at the redirected carrier. On the other hand, the co-channel load balancing naturally increases the RLF occurrence since the 2.6 GHz users no longer connect to the best cell in terms of signal strength. Nevertheless, the cost paid in RLFs is rather affordable owing to the additional coverage provided by the 800 MHz macrocell carrier. In this context, MRO becomes less relevant.

All in all, the joint operation of intra-frequency and inter-frequency load balancing is recommended for scenarios where a co-channel HetNet deployment is overlaid by a dedicated macrocell carrier. Albeit this approach balances better the network load, the additional capacity benefits provided by enabling intra-frequency MLB heavily depend on the experienced interference at the co-channel HetNet deployment. Hence, interference mitigation is still required. From a mobility performance perspective, the handover and cell reselection rates occurrence are balanced, while the estimated RLF rate remains within an acceptable range thanks to the overlaid macrocell carrier that is deployed at a dedicated frequency band.

Appendix D

Emulating Different Scheduling Policies with CA

This paper reprint presents a mathematical framework for emulating different packet scheduling policies in network environments with CA. Compared to the model shown in Appendix A, it can derive a stringent mathematical solution for calculating the virtual amount of CA users per CC subject to the targeted user fairness. Nevertheless, it is only valid for scenarios where the CA users can connect to all CCs and similar SINR conditions per CC are perceived. Hence, it is included in the dissertation for the sake of completion since it was not used in the simulation campaigns of Chapter 5 and Chapter 6.

Abstract Radio Resource Management Framework for System Level Simulations in LTE-A Systems

Panagiotis Fotiadis
Aalborg University[1], Denmark
Department of Electronic Systems
E-mail: paf@es.aau.dk

Ingo Viering
Nomor Research GmbH
Munich, Germany
Email: viering@nomor.de

Paolo Zanier[†] and Klaus I. Pedersen[1,2],
Nokia Solutions Networks
Munich, Germany[†]
Aalborg, Denmark[2]
Email: name.surname@nsn.com

***Abstract*—This paper aims at the theoretical modeling of different packet schedulers in Long Term Evolution-Advanced (LTE-A) systems. For that purpose, an abstract Radio Resource Management (RRM) framework has been developed, considering Proportional Fair (PF) and cross-Component Carrier scheduling. To validate our work, extensive system level simulations have been conducted for different ratios of users with Carrier Aggregation (CA) capabilities. The associated results confirm that the proposed model satisfactorily captures the main properties of the aforementioned scheduling metrics without any need for explicit design at a subframe resolution; hence, making it a promising candidate for a convenient scheduler implementation in simulators with simplified RRM modeling.**

***Index Terms*—LTE-A; Radio Resource Management (RRM); Carrier Aggregation; Modeling; Packet Scheduling**

I. Introduction

To meet the growing demand for higher data rates, Carrier Aggregation (CA) is introduced in Release 10 Long Term Evolution-Advanced (LTE-A) specifications. In CA mode, two or more Component Carriers (CC) are aggregated to support effectively wider transmission bandwidths [1], reaching up to 100 MHz. Fig. 1 illustrates the basic CA framework, considering N deployed CCs at the base station. Being designed to be backwards compatible with User Equipments (UE) without CA capabilities, each CC follows the Release 8 LTE numerology. Users are assigned onto one or more CCs according to several factors (terminal capabilities, CC load, radio channel conditions, Quality of Service, etc), and dynamic packet scheduling takes places either independently per CC or jointly across the deployed CCs. Finally, link adaptation and Hybrid Automatic Repeat Request (HARQ) management are solely performed at a CC independent basis.

However, the explicit design of the aforementioned framework on a subframe granularity is only mandatory for particular case studies. E.g. dynamic packet scheduling investigations [2-4] are typically performed with such detailed simulation tools. Scheduling decisions are taken at a millisecond basis at the expense of large simulation runtimes. For studies, where some of these system components are not so relevant for performance evaluation, but the simulations still need to be conducted in a LTE-A environment, the detailed design of all L1/L2 aspects would be exhaustive. E.g. investigations associated to Self-Organizing Networks (SON) [5] require a lighter approach in terms of modeling the LTE-A Radio Resource Management (RMM) framework, since long simulations are needed for convergence to be achieved.

For that purpose, this paper contributes a simple mathematical LTE-A RRM framework that realistically emulates the performance of 2 particular scheduling policies, also referred to as independent Proportional Fair (PF) per CC and cross-CC PF [2], without requiring an explicit subframe-based implementation. In particular, a set of fairness scaling factors is utilized for modifying the resource allocation decisions subject to the UE assignment onto the different CCs and the desired RRM policy to be applied. However, since no additional Signal-to-Noise plus Interference (SINR) considerations are included, it is valid for interference limited scenarios only, where the SINR distributions over the deployed CCs are similar and system performance depends mainly on how users are assigned onto the available CCs.

The remainder of the paper is structured as follows. Section II provides a brief overview of packet scheduling in LTE-A systems, while the proposed abstract RRM framework is presented in Section III. The simulation assumptions along with the corresponding results are available in Section IV and V respectively. Finally, section VI concludes the paper.

II. Dynamic Packet Scheduling in LTE-A

This section outlines the basic properties of different packet scheduling policies in LTE-A systems. More specifically, we

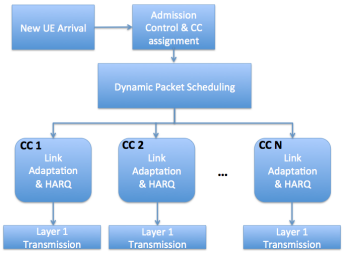


Fig. 1. Carrier Aggregation Framework

focus on the CC independent PF scheduler and the joint cross-CC scheduling approach.

A. Independent Proportional Fair per CC

The PF metric is a well-known example of packet scheduling that maximizes network utility (defined as the sum of logarithmic user throughput) by exploiting multi-user diversity [6]. In the long term, user fairness is guaranteed and resources are equally shared among UEs regardless of channel conditions and fading characteristics [7]. For independent PF scheduling per CC, the j^{th} Physical Resource Block (PRB) on carrier *i* will be assigned to the user $u_{i,j}$ according to:

$$u_{i,j} = \arg\max_{u}\{\frac{R_{u,i,j}}{\widetilde{R}_{u,i}}\} \tag{1}$$

where $R_{u,i,j}$ is the instantaneous throughput of user *u* at the j^{th} PRB of CC *i* and $\widetilde{R}_{u,i}$ is the past user perceived throughput on the same CC averaged over a specific time window. However, since non-CA users have limited access to the overall available spectrum, maximizing the network utility within the CC does not necessarily imply that the global utility over all CCs is maximized.

In fact, CA UEs will be significantly favored if PF scheduling is performed independently per CC, while non-CA users co-exist in the system. Let us consider a scenario with *n* CCs and equal split of non-CA and CA UEs. Assuming a load balancing mechanism that distributes evenly the non-CA users over the different CCs, then the total amount of PRBs allocated per CA terminal will be *n* times larger than the resource share of the non-CA UEs. Apparently, legacy non-CA devices will have limited access to the spectrum resources, a fact that may severely impact their perceived data rates.

B. Cross-CC Packet Scheduling

The joint cross-CC scheduling [3] overcomes the aforementioned resource allocation fairness problem by modifying the PF metric as follows:

$$u_{i,j} = \arg\max_{u}\{\frac{R_{u,i,j}}{\sum_{i=1}^{n}\widetilde{R}_{u,i}}\} \tag{2}$$

In (2) the aggregated past experienced user throughput over all CCs is used so the scheduling priority for CA UEs actually decreases. Hence, resource starvation for non-CA devices is avoided and significant cell edge throughput gains are achieved without any noticeable impact on the aggregated cell throughput [3]. As shown in [2], the cross-CC scheduler maximizes the global network utility also in cases when non-CA and CA UEs co-exist in the system. Nevertheless, a more flexible resource allocation is feasible by generalizing (2) as follows:

$$u_{i,j} = \arg\max_{u}\{\frac{\alpha \cdot R_{u,i,j}}{(\sum_{i=1}^{n}\widetilde{R}_{u,i})^{\beta}}\}, \tag{3}$$

where the a (*user type specific*) and b weighting parameters are utilized for adjusting the UE category fairness. E.g. by setting $a = 1$ for CA UEs and $a > 1$ for non-CA devices, the scheduling priority of non-CA devices will further improve, and vice versa.

III. Abstract LTE-A RRM Framework

In this section, the proposed abstract LTE-A RRM model is thoroughly described. A set of weighting factors is introduced that dynamically adjust resource allocation decisions subject to the targeted scheduling fairness between non-CA and CA devices.

A. Problem Formulation

Let us consider a LTE-A scenario with *n* CCs per macrocell area. Let N_i and N_{CA} denote the number of non-CA and CA users in the i^{th} CC. Obviously:

$$N_{nonCA} = \sum_{i=1}^{n} N_i, \tag{4}$$

where N_{nonCA} is the total number of legacy non-CA users in the cell area. Given that K_{tot} is the CC bandwidth, the allocated resource share per UE class, assuming independent PF scheduling per CC, will be:

$$K_i^{nCA} = \frac{K_{tot}}{N_i + N_{CA}} \tag{5}$$

$$K^{CA} = \sum_{i=1}^{n}\frac{K_{tot}}{N_i + N_{CA}} \tag{6}$$

Since CA UEs are allocated resources in all CCs, we can define the virtual amount of CA terminals in the i^{th} CC, $V_i^{N_{CA}}$, as follows:

$$V_i^{N_{CA}} = w_i N_{CA}, \tag{7}$$

where w_i are the virtual scheduling weights for the CA users on the i^{th} CC. Thus, based on (7), the modified resource shares per UE class can be expressed by:

$$K_i^{nCA} = \frac{K_{tot}}{N_i + w_i N_{CA}} \tag{8}$$

$$K^{CA} = \sum_{i=1}^{n}\frac{w_i K_{tot}}{N_i + w_i N_{CA}} \tag{9}$$

(8) and (9) indicate that resource allocation decisions could be adjusted by different w_i assignments. Therefore, given that α_u is the target Resource Share Ratio (RSR) between non-CA and CA UEs, the proper virtual scheduling weights need to be found, satisfying:

$$K_1^{nCA} = K_2^{nCA} = \ldots = K_n^{nCA} = \alpha_u \cdot K^{CA} \tag{10}$$

subject to:

$$\sum_{i=1}^{n} w_i = \frac{1}{\alpha_u} \tag{11}$$

$$w_i \in [0, 1/\alpha_u] \tag{12}$$

In such a manner, the fairness adjustments provided by the α, β parameters of the generalized cross-CC scheduler can be emulated by spanning α_u over different RSRs. Note that (11) and (12) are mandatory constraints for guaranteeing that the sum of $V_i^{N_{CA}}$ over all CCs along with the sum of total resources allocated per CC will not exceed N_{CA} and K_{tot} respectively.

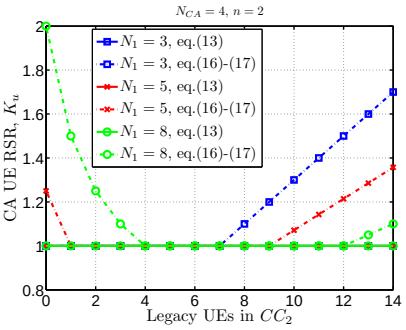


Fig. 2. K_u for different non-CA UE assignments onto the 2 CCs. 4 CA users are assumed and $a_u = 1$.

B. Solution for virtual scheduling weights

Based on (10), we can express all scheduling weights as a function of w_1:

$$w_i = w_1 + \frac{N_1 - N_i}{N_{CA}} \quad , i = 2, \ldots, n, \tag{13}$$

whereas by combining (4), (11) and (13), the set of virtual scheduling weights can be defined, as follows:

$$w_i = \frac{1}{n \cdot \alpha_u} + \frac{N_{nonCA} - n \cdot N_i}{n \cdot N_{CA}} \quad , i = 1, \ldots, n \tag{14}$$

In order to illustrate the behavior of (13), a simple case with 2 deployed CCs, 4 CA terminals and different non-CA UE distributions is assumed for $\alpha_u = 1$. Resource fairness performance between the 2 different user categories is evaluated by the obtained average CA UE RSR, K_u, defined as:

$$K_u = \frac{K^{CA}}{(\sum_{i=1}^{n} N_i K_i^{nCA})/N_{nonCA}} \tag{15}$$

Both K_u and the virtual CA weights on the 1^{st} CC ($w_2 = 1 - w_1$) are demonstrated in Fig. 2 and Fig. 3 respectively. We observe that ideal resource share fairness is guaranteed for all different UE assignments, as $1/\alpha_u = K_u = 1$. This model behavior is expected since (13) derives directly from (10). Note that $w_j > w_i$ for $N_i > N_j$; hence, CA UEs are allocated more resources in the CCs where fewer non-CA users are assigned, in order to maintain fairness. However, α_u might not always be feasible, simply due to the UE assignment onto the different CCs. In these cases, the weights solution violate (12), as it is clearly shown by Fig. 3. This limitation is overcome if the negative weights are truncated to zero and the remaining ones are normalized accordingly, such as to fulfill (12):

$$\widetilde{w_i} = \begin{cases} 0 & , \widetilde{w_i} < 0 \\ \frac{1}{n \cdot \alpha_u} + \frac{N_{nonCA} - n \cdot N_i}{n \cdot N_{CA}} & , \widetilde{w_i} \geq 0 \end{cases} \tag{16}$$

TABLE I
SYSTEM SIMULATION PARAMETERS

Parameter	**Value**
Scenario	CA Scenario 1 (19 sites, 3 cells per site)
Inter-Site Distance	500 m
Pathloss	$128.1 + 37.6 \cdot log_{10}(R)$
CC Information	4 CCs at 2GHz (10 MHz each)
Resources per CC	50 PRB
Number of UEs per cell	10
CA UE Ratio	0%, 20%, 50%, 80% and 100%
Antenna Configuration	1x2
Transmit Power	43 dBm per CC
Traffic Type	Full Buffer
CA Packet Scheduler	Independent PF versus Cross-CC

$$w_i = \frac{\widetilde{w_i}}{\alpha_u \sum_{i=1}^{n} \widetilde{w_i}} \tag{17}$$

The results related to solution (16)-(17) are also depicted in Fig. 2 and 3, as the regions of resource fair infeasibility are indicated by $K_u > 1$, while $w_i \in [0, 1]$ for any UE assignment onto the 2 CCs. Obviously, in the $K_u > 1$ regions, CA devices will be scheduled in a single CC.

IV. SIMULATION ASSUMPTIONS

The proposed abstract LTE-A RRM model is implemented in the system level simulator presented in [8][9]. The performance of the mathematical framework is assessed in a network layout consisting of 4 collocated CCs deployed at 2 GHz (CA deployment scenario 1 [10]). Statistics are collected from a sufficiently large number of snapshots for different user positions and several CA UE ratios. Full buffer traffic is simulated and non-CA UEs are assigned onto the different CCs based on the Least Load (LL) algorithm [11]. CA devices are assigned on all CCs. The major simulations assumptions are listed in Table I.

The UE perceived throughput derives directly from SINR-to-throughput mapping curves, calibrated by extensive link level simulations with explicit implementation of all packet

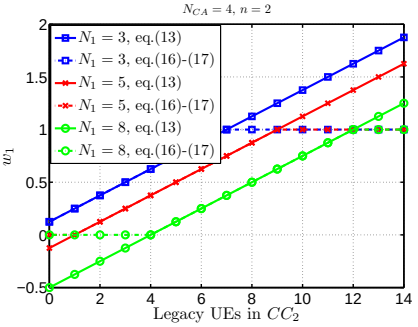


Fig. 3. Cross-CC weights solution for different non-CA UE assignments onto the 2 CCs. 4 CA users are assumed and $a_u = 1$.

scheduling, link adaptation and HARQ procedures. It is always assumed that the optimal Modulation and Coding Scheme (MCS) is selected for each user, depending on its experienced wideband SINR conditions. The impact of the Frequency Domain Packet Scheduling (FDPS) diversity gain [12] is also included in the system modeling by selecting the proper curve subject to the amount of users that are scheduled in the cell. Therefore, the experienced UE throughput, $r_{u,i}$, of user u on the i^{th} CC is calculated as follows:

$$r_{u,i} = R(SINR_{u,i}) \cdot f_{u,i}, \tag{18}$$

where $f_{u,i}$ is the number of resources allocated to user u on the i^{th} CC and $SINR_{u,i}$ the average SINR per PRB. The generalized cross-CC performance is approximated by calculating the scheduling weights according to (16)-(17), while $f_{u,i}$ derives directly from (8) and (9) depending on the user type. Similarly, (5) and (6) are utilized for the case of independent PF scheduling per CC.

V. Simulation Results

The Key Performance Indicators (KPI) for the conducted study are the network utility (sum of logarithmic throughput in Mbps), the 5%-ile user throughput (coverage) and the aggregated cell throughput over the deployed CCs.

A. Impact of target RSR α_u

Fig. 4 illustrates the normalized network utility, coverage and average cell throughput for different α_u settings and 50% CA UE ratio. All related KPIs are normalized by the values associated to $\alpha_u = 1$. Indeed, network utility is maximized for $\alpha_u = 1$, emulating the performance of the generalized cross-CC scheduler with $\alpha = \beta = 1$. This is case when the weights are calculated such as to provide the maximum possible fairness between non-CA and CA users. Consequently, the highest coverage throughput gains are derived as well. Note that for this particular simulation setting, the $\alpha_u = 0.25$ case actually represents the independent PF scheduler per CC. Therefore, no impact on the average cell throughput is observed compared to $\alpha_u = 1$, a fact that is aligned with other related studies in the literature [3]. Any other fairness adjustments performed by α_u replicate different combinations of the α, β parameters of the generalized cross-CC scheduler and consequently the network utility decreases.

The impact of different packet scheduling approaches on the experienced UE throughput per user category is shown in Fig. 5. As expected, CC independent PF scheduling favors significantly CA terminals, achieving *n* times higher throughput compared to non-CA UEs. Nevertheless, this performance gap is diminished when the joint scheduling approach is emulated by setting $\alpha_u = 1$.

B. Impact of CA UE Penetration

Since it has been confirmed that the performance of cross-CC scheduling can be emulated by $\alpha_u = 1$, the model is investigated for different ratios of CA devices. The coverage throughput for different CA UE ratios is depicted in Fig. 6.

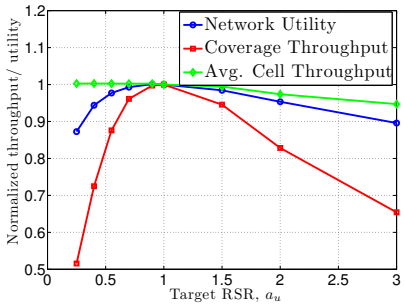


Fig. 4. Normalized network utility, coverage and avg. cell throughput for different target α_u values. 50% of CA UE ratio is assumed.

We observe that the cross-CC solution enhances significantly the cell edge throughput, whenever non-CA and CA users co-exist in the network, providing gains that are in the same range with the ones derived in [3]. More specifically, for the cases of 20%, 50% and 80% of CA UE ratio, the recorded coverage gains are 44%, 92% and 78% respectively. Note that for ratios above 50%, the performance of the independent PF improves, since the coverage throughput statistics are also biased by cell edge CA UEs that actually experience higher data rates.

The significant fairness enhancements between the 2 user categories is also highlighted by the network utility, that is provided in Fig. 7. In principle, higher utility is achieved by the cross-CC solution, an observation that is in very good agreement with the simulation results in [2].

Finally, Fig. 8 demonstrates the corresponding average cell throughput results. Once again, the abstract RRM model realistically replicates the subframe implementation, as similar

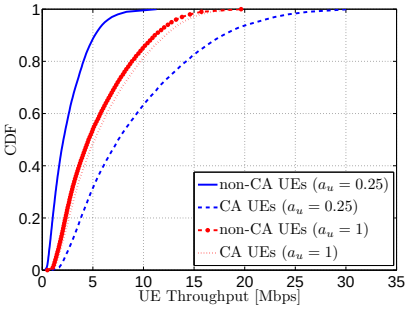


Fig. 5. CDF for non-CA and CA users for different α_u values. 50% of CA UE ratio is assumed.

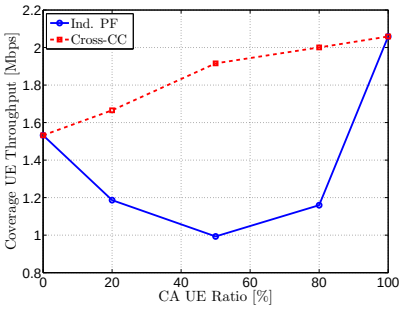


Fig. 6. Coverage throughput for different CA UE ratios

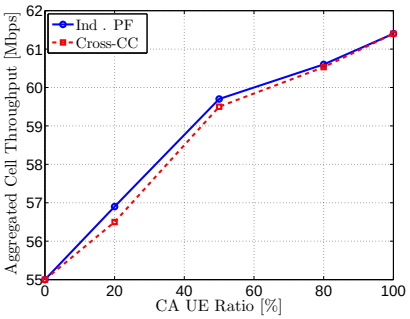


Fig. 8. Average cell throughput for different CA UE ratios

system performance is achieved, regardless of the CA UE penetration and the scheduling approach.

VI. Conclusions

In this paper, an abstract RRM framework for system level simulations in LTE-A systems has been developed. It allows to emulate different packet scheduling policies in scenarios with a mixture of non-CA and CA users. In particular, a set of weighting factors is introduced that adjusts resource allocation decisions depending on the fairness to be maintained between the different UE categories. The model behavior has been tested for 2 particular packet scheduling policies, also denoted as CC independent PF and cross-CC scheduler. Results have shown that the performance of the aforementioned metrics can be satisfactorily approximated in an abstract manner, without any requirement for explicit simulations at a subframe basis. Therefore, it could be considered as an excellent candidate for a scheduler implementation in simplified system level simulators, where the detailed modeling of all L1/L2 LTE-A aspects would have been a rather exhaustive approach.

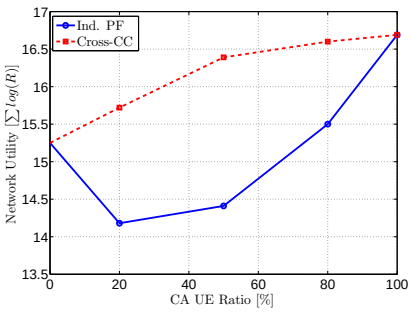


Fig. 7. Network utility for different CA UE ratios

References

[1] M. Iwamura, K. Etemad, M. Fong, Y. Wang, R. Nori and R. Love, *Carrier Aggregation Framework in 3GPP LTE-Advanced*, Communications Magazine, IEEE , vol.48, no.8, pp. 60-67, August 2010

[2] Y. Wang, K. I. Pedersen, and T. B. Sorensen, *Utility Maximization in LTE-Advanced Systems with Carrier Aggregation*, in IEEE Vehicular Technology Conference, May 2011.

[3] Y. Wang; K. I. Pedersen, T. B.Sorensen, P. E. Mogensen, *Carrier load balancing and packet scheduling for multi-carrier systems*, in Wireless Communications, IEEE Transactions on, vol.9, no.5, pp. 1780-1789, May 2010.

[4] Y. L. Chung, L. J. Jang and Z. Tsai, *An efficient downlink packet scheduling algorithm in LTE-Advanced systems with Carrier Aggregation*, in IEEE Consumer Communications and Networking Conference, pp. 632-636, January 2011.

[5] 3GPP TR 36.902, *Self-configuring and self-optimizing network cases and solutions (Release 9)*, v9.3.1, March 2011.

[6] F. Kelly, *Charging and rate control for elastic traffic*, European Transactions on Telecommunications, vol. 8, pp. 33-37, 1997.

[7] J. M. Holtzman, *Asymptotic Analysis of Proportional Fair Algorithm*, in IEEE Personal, Indoor and Mobile Radio Communication, pp. F-33- F-37, September 2001.

[8] I. Viering, M. Dottling and A. Lobinger, *A Mathematical Perspective of Self-Optimizing Wireless Networks* , IEEE International Conference on Communications, pp.1-6, June 2009.

[9] I. Viering, A. Lobinger and S. Stefanski, *Efficient Uplink Modeling for Dynamic System-Level Simulations of Cellular and Mobile networks*, EURASIP Journal on Wireless Communications and Networking, August 2010

[10] S. Zukang, A. Papasakellariou, J. Montoya, D. Gerstenberger and X. Fangli, *Overview of 3GPP LTE-Advanced Carrier Aggregation for 4G Wireless Communications*, IEEE Communications Magazine, vol.50, no.2, pp.122-130, February 2012.

[11] T. Dean and P. Fleming, *Trunking Efficiency in multi-carrier CDMA Systems*, in IEEE Vehicular Technology Conference proceedings, vol. 1, pp. 156-160, September 2002.

[12] A. Pokhariyal, T. Kolding and P. Mogensen, *Performance of the Downlink Frequency Domain Packet Scheduling for the UTRAN Long Term Evolution*, in IEEE Personal, Indoor and Mobile Radio Communication, pp. 1-5, September 2006.

Appendix E

Complementary Results for Chapter 5

This paper reprint provides another performance comparison between the developed DAP+FHO@CS solution and the RSRQ-based mobility framework. Unlike Chapter 4, the associated simulation campaign is conducted in Scenario C with and without intra-eNB CA. Albeit the trends are similar to the ones observed in Scenario B, it is shown that the RSRQ-based configurations become even more costly in terms of handover signaling in Scenario C. The reason is that this scenario considers an additional macrocell carrier deployed at the 1.8 GHz band. By enabling intra-eNB CA, the cost paid in handovers essentially reduces; nevertheless, DAP+FHO@CS still finds a better trade-off between the capacity gain and the associated signaling overhead.

Load-Based Traffic Steering in Multi-Layer Scenarios: Case with & without Carrier Aggregation

Panagiotis Fotiadis, Michele Polignano
Aalborg University[1], Denmark
Department of Electronic Systems
E-mail: {paf, mpo}@es.aau.dk

Klaus I. Pedersen[1,2], Preben Mogensen[1,2]
Nokia Solutions Networks[2]
Aalborg, Denmark
E-mail: name.surname@nsn.com

***Abstract*—This paper aims at evaluating different mechanisms for providing Inter-Frequency (IF) Load Balancing (LB) in multi-layer heterogeneous deployments. More specifically, the performance of IF mobility management based on signal quality measurements is compared against a load-dependent Traffic Steering (TS) framework that triggers IF mobility events only if load imbalance is detected. To evaluate the joint interaction of the aforementioned schemes with more advanced LB features, system level simulations have been conducted with and without Carrier Aggregation (CA) capable users. Results have shown that although quality-based handoff procedures can act as a passive TS mechanism, they are costly in handovers and measurements gaps. The developed TS scheme utilizes cell neighbor measurements more efficiently, achieving significant handover reduction. Finally, CA makes the proposed framework even more attractive, since its careful parameterization becomes less relevant and load imbalances can be tackled by the packet scheduler as well.**

Keywords*—*Load balancing; Mobility management; Radio Resource Control (RRC); Carrier Aggregation

I. Introduction

Multi-layer deployments are envisaged to be the necessary network evolution for meeting the future capacity and coverage requirements. Hence, cells with different characteristics will co-exist in the same environment, also denoted as Heterogeneous Network (HetNet), providing a common pool of resources to be efficiently utilized subject to the User Equipment (UE) capabilities, power consumption, load conditions, requested service, etc. This functionality is also denoted as Traffic Steering (TS) and its target is to properly distribute traffic such as to accommodate the optimum combination of the aforementioned factors based on the network operator use cases and performance indicators.

Focusing specifically on load-based TS schemes, the majority of the state-of-art literature investigates the potentials of Load balancing (LB) in co-channel deployments, where both small cells (e.g. pico/femtocells) and the overlay macro are deployed at the same frequency [1]. This involves dynamic range extension schemes that positively bias measurements from underutilized cells such as to virtually enlarge their power footprint and attract more users by means of handover (HO) executions. In the context of Inter-Frequency (IF) TS, different layer selection schemes are available in [2], where UEs are directed to the optimal cell during the Radio Resource Control (RRC) connection establishment. Nevertheless, the associated signaling cost of the proposed mechanisms is neglected. As shown in [3], the auto-tuning of mobility parameters can further be utilized for IF TS. However, unlike to the co-channel case, IF measurements are not always available. In principle, they should be kept at a reasonable level, since measurements gaps are required for the device to perform such measurements. High measurements rates could have an impact on the experienced data rates along with a potential UE power consumption increase [4]. To maintain IF mobility procedures tightly coupled with TS functionalities requires adequate measurement availability for mobility events to be triggered. Nonetheless, the overhead cost might be relatively high, jeopardizing UE power consumption due to excessive cell neighbor measurements.

On top of the applied TS policies, features such as Carrier Aggregation (CA) [5][6] can actually relax IF LB requirements. Being introduced as part of the LTE-Advanced standardization, CA allows terminals with multi-connectivity capabilities to simultaneously access the bandwidth of multiple carriers. Packet scheduling can be further utilized for inter-layer LB, while IF measurements become less relevant for CA users, as they may concurrently receive data on one carrier, while performing measurements on a different carrier [7].

This paper focuses on evaluating different solutions for IF HetNet LB with and without CA. To cope with HO overhead and UE power consumption, a low cost TS framework has been developed, that decouples IF mobility management from

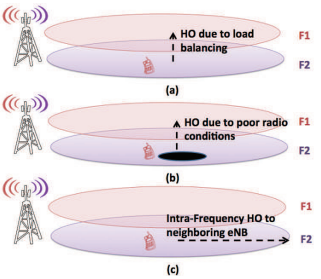


Fig. 1. Decoupling IF mobility management from TS in the RRC Connected. (a): TS-driven IF HO due to overload detection at F2, (b): Mobility driven IF HO due to coverage hole , (c): No need for any action if F2 is not overloaded

LB functionalities. IF measurements are explicitly requested by the network whenever overload is detected, while HOs are kept low by aligning the LB procedures in both RRC UE states. IF events – HOs and cell reselections – are classified into 2 different categories depending on the triggering cause. As it is shown in Fig. 1, an event is defined as mobility-driven, if it is performed due to poor radio conditions, while events triggered for LB purposes are categorized as TS-driven. The performance of the designed scheme is compared against the standardized quality-based handoff mechanisms, where IF HOs/cell reselections are triggered by exploiting the in-built load information that is available in the Reference Signal Received Quality (RSRQ) measurements. No other TS mechanism is applied and the distinction between mobility-driven and TS-driven events is not possible. System level simulations are conducted in a Long Term Evolution-Advanced (LTE-A) scenario consisting of a co-channel macro/ pico deployment at 2600 MHz, supplemented by 2 additional macrocell carriers at 1800 MHz and 800 MHz respectively. To investigate the CA impact on the aforementioned schemes, intra site CA is enabled, meaning that CA UEs can aggregate spectrum from multiple co-sited macro carriers

The remainder of the paper is organized as follows. Section II outlines the IF mobility management framework for non-CA and CA users, whereas the proposed TS scheme is thoroughly presented in Section III. Simulation assumptions and results are provided in Section IV and V respectively. Finally, Section VI concludes the paper.

II. Inter-Frequency Mobility Management Framework

This section outlines the standardized IF mobility management framework for both non-CA and CA terminals in the RRC Connected and RRC Idle state.

A. Physical Layer Measurements

Measurements in terms of Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ) are specified for mobility support. RSRP corresponds to the signal strength measurement and therefore is insensitive to load fluctuations. On the other hand, RSRQ is defined as:

$$RSRQ = \frac{RSRP}{RSSI}, \tag{1}$$

where RSSI is the Received Signal Strength Indicator and comprises the linear average of the total received power including co-channel serving and non-serving cells, adjacent channel interference, thermal noise, etc [8]. The contribution of RSSI in (2) makes RSRQ partly capture load information. Hence, if properly configured, RSRQ-based mobility management can operate as a passive TS mechanism by triggering IF HOs/ cell reselections due to the load variations between the serving and a target cell.

B. Non-CA Framework

1) RRC Connected State: RRC Connected mobility management is network-controlled and UE-assisted. UEs perform physical layer measurements and the associated reports are sent to the network, either periodically or whenever an event is triggered. Devices can be configured to initiate IF measurements only if the serving radio conditions become worse than a particular threshold, also denoted as A2 event [9]. As soon as the A2 event is reported, IF mobility is activated by configuring the corresponding A3 event [9] (*neighbor becomes offset better than serving*). Given that a target cell fulfills the A3 condition for a specific time duration, referred to as Time-To-Trigger (TTT) window, an IF HO is triggered.

2) RRC Idle State: RRC Idle UEs autonomously reselect to a neighboring cell based on the reselection rules that are broadcast in the system information. Similarly to the intra-frequency case, the cell selection S criterion along with the cell ranking R criterion [10] can be utilized for IF mobility management in the RRC state. Nevertheless, an alternative mechanism, referred to as Absolute Priorities (AP) [10], is available for prioritizing particular carriers during the cell reselection process. Carrier priorities are broadcast in the system information and a set of priority-based rules is evaluated for reselecting towards an IF cell. More specifically, devices camping on a lower-priority carrier reselect towards a higher-priority one once the target signal strength or quality exceeds the $Thresh^{AP}_{2High}$ threshold. On the other hand, reselecting towards a lower-priority carrier requires a more restrictive condition to be fulfilled, since the serving cell must drop below $Thresh^{AP}_{sLow}$ and the target to exceed $Thresh^{AP}_{High2Low}$. Note that such reselection rules are only valid for carriers with different priorities. In case of equal priorities being assigned to cells belonging to different frequencies, the conventional S and R criteria are applied for evaluating the cell reselection process.

C. CA Framework

1) RRC Connected State: CA devices can simultaneously access the bandwidth of multiple carriers. A set of serving cells is configured, and one of them is designated as the Primary Cell [5][6]. The PCell is responsible for all basic operations including mobility support and Radio Link Failure (RLF) supervision. HOs are solely performed at the PCell, following the non-CA handoff procedures. With regards to IF measurements, CA users may perform relaxed background measurements with a certain periodicity (i.e. 40 msec) [4]. Nonetheless, the A2 event can still be utilized for enabling PCell IF HOs.

Additionally configured cells are denoted as Secondary Cells (SCell), and they can be added, changed or removed depending on the UE measurements. Consequently, whenever an SCell event condition is met, the UE sends a measurement report via uplink RRC signaling for triggering the corresponding SCell action. An example of dynamic RSRQ-based PCell and SCell management is illustrated in Fig. 2. Situation (a) refers to the case when the RSRQ of the PCell is higher than the A2 threshold, $A2_{Thresh}$. Obviously, the PCell remains the same regardless of the SCell radio conditions, as the associated IF HO event is not configured yet. Unlike case (a), an IF HO is triggered for both situations (b) and (c). Cell j is now assigned as the PCell and cell i will eventually be configured as a SCell only in case (b), where the RSRQ of cell i is above the SCell addition threshold.

2) RRC Idle State: CA is not applicable in the RRC Idle and CA UEs follow the typical non-CA framework for the

cell reselection process. Thus, whenever CA users switch to the RRC Connected, the latest camping cell is assigned as the PCell, unless any TS action occurs.

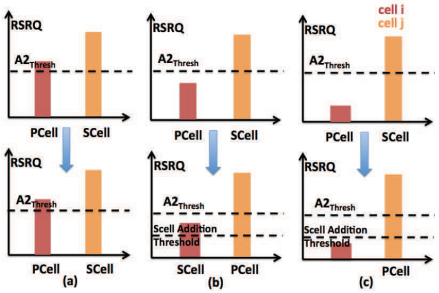


Fig. 2. Dynamic CA UE PCell and SCell management.

III. Proposed Load-Based Traffic Steering Framework

A LB framework is proposed, where IF HOs/cell reselections are primarily performed by TS-driven procedures. A relatively low $A2_{Thresh}$ is configured, while measurements are explicitly requested by the network whenever overload is detected. Consequently, measurements gaps are minimized and IF mobility management is decoupled from the TS functionality. Mobility-driven IF events can only occur only if the radio conditions degrade dramatically, such that the A2 condition is met and the IF handoff mechanisms are configured.

To minimize the impact of TS on signaling overhead and UE power consumption, the LB procedures in the RRC Idle and Connected state should be aligned. For that purpose, the developed framework applies the load-based decisions at the switching instances of the RRC UE state machine. UEs switching to the RRC Idle are provided with dedicated mobility parameters such as to camp at the appropriate carrier, while TS-driven HOs are employed for users switching to RRC Connected.

A. Load and Composite Available Capacity Formulation

In order to provide TS support, load information for neighboring target cells should be available at the base stations. Since high Physical Resource Block (PRB) utilization does not necessarily mean overload conditions [11], the resource share of user u, f_u, is scaled by the satisfaction ratio R_t/R_u, where R_t represents the desired data rate that should on average be achieved in the cell, and R_u is the actual rate that the device experiences. Hence, the load contribution ρ_u of user u to its serving cell is defined as follows:

$$\rho_u = \min\{\frac{f_u \cdot R_t}{R_u \cdot B}, \rho_{max}\}, \tag{2}$$

where B is the cell bandwidth and ρ_{max} specifies the maximum load that a user can contribute to the cell in order to avoid situations where a single UE in poor channel conditions could declare the cell in overload. Note that for CA devices, R_u represents the aggregated data rate that the UE experiences over the multiple carriers that is scheduled. Cells periodically monitor their own load conditions $\widetilde{\rho}_{own} = \sum_u \rho_u$ and the relevant information exchange is performed in terms of Composite Available Capacity (CAC) [12]. To control TS operation, a target operational cell load, ρ_t, is specified, and CAC is expressed as below:

$$CAC = 1 - \frac{\widetilde{\rho}_{own}}{\rho_t} \tag{3}$$

TS procedures are triggered whenever $\widetilde{\rho}_{own}$ exceeds a predetermined overload threshold. As load oscillations around ρ_t may repetitively trigger TS events, a hysteresis region is applied and the overload detection threshold is defined as $\rho_{high} = \rho_t + \rho_{hyst}$. Similarly, cells below the $\rho_{low} = \rho_t - \rho_{hyst}$ threshold are only willing to accept load.

B. TS upon RRC Connection Establishment

Whenever a UE switches to RRC Connected, it is requested to initiate IF measurements if overload is detected. Once the measurements reports are collected, the strongest RSRP-measured cell per carrier is selected, subject to the following constraints:

$$Q_{meas}^{RSRP} \geq A_{thresh}^{RSRP} \tag{4}$$

$$Q_{meas}^{RSRQ} \geq A_{thresh}^{RSRQ}, \tag{5}$$

where $Q_{meas}^{RSRP}, Q_{meas}^{RSRQ}$ are the performed measurements in terms of RSRP and RSRQ and $A_{thesh}^{RSRP}, A_{thesh}^{RSRQ}$ correspond to the respective thresholds that the target IF cells should satisfy. The final set of candidate LB targets is sorted in a descending CAC order and the cell with the highest value is selected. The load condition of the target cell is derived directly from CAC and if it is below the ρ_{low} threshold, a forced IF HO is initiated towards that cell for LB purposes.

Note that A_{thesh}^{RSRP} is set Δ dB higher than $A2_{Thresh}$ in order to ensure that the steered device will not perform IF measurements when is connected to the target layer. In such a manner, ping pong HOs [13] are less likely to occur and mobility performance is not compromised by the TS intervention. Finally, interference-related information for the target layer is provided via (5).

1) TS at RRC Connection Release: In the context of TS at the connection release, the dedicated priorities framework is applied, where carrier priorities are dynamically adjusted at a UE resolution, according to the exchanged CAC information [14]. Therefore, the highest priority is assigned to the least loaded carrier. Note that no additional RRC signalling is required since UE-dedicated Idle mode parameters can be provided to the device via the RRC CONNECTION RELEASE MESSAGE [9]. Dedicated priorities provide significant signalling gains, as the number of forced TS-driven HOs required for LB can be decreased. In particular, UE distributions in the RRC Idle are balanced and the probability of establishing the a new RRC connection at an overloaded cell decreases noticeably.

The developed dedicated priorities scheme follows the same logic in terms of radio conditions constraints, implying

that (4) and (5) must be fulfilled as well. Nevertheless, A^{RSRP}_{thesh} is replaced by $Thresh^{AP}_{2High}$, since $Thresh^{AP}_{2High}$ controls cell reselections towards higher priority carriers in the RRC Idle state. Hence, the algorithm ensures that the redirected UE will camp at the least loaded carrier, as it is the one being assigned with the highest priority.

IV. Simulation Assumptions

The TS framework is evaluated by means of extensive system level simulations for 0% and 50% CA UE ratio respectively. As a reference, the RSRQ-based IF mobility management framework is used, assuming 3 different $A2_{Thresh}$ values. Mobility management in both RRC states is explicitly modeled, along with the associated delays regarding the Idle-to-Connected (and vice versa) transition timers, HO executions, cell reselections and SCell additions, removals or changes. Finite buffer traffic is simulated and packet arrivals are modeled as a Poisson process. The payload is negatively exponentially distributed with a mean value of 3 Mbits. 2 high traffic areas (hotspots) are randomly generated per macrocell area and picocells are deployed concentrically. 2/3 of the users are confined within the hotspots and the remaining 1/3 is uniformly distributed in the macrocell area, moving at straight line trajectories. Low mobility at 3 km/h is assumed for all UEs. A detailed list of the key simulation parameters is provided in Table I.

TABLE I. System Simulation Parameters

Parameter	Value
Scenario	3GPP Hexagonal grid (7 sites, 3 cells per site)
ISD	500m
Carrier Frequencies	800MHz, 1800MHz, 2600MHz
Bandwidth	10 MHz, 10 MHz, 20 MHz
Number of UEs per macro area	100
Number of picocells per macro area	2
Hotspot Radius	40 m
Hotspot UE Density	2/3
CA UE Ratio	0%, 50%
Transmit Power	43 dBm (macro), 30 dBm (pico)
Shadowing Standard Deviation	8 dBm (macro), 10 dBm (pico)
Shadowing Correlation Distance	50 m (macro), 13 m (pico)
Antenna Configuration	1x2
Traffic Type	Finite Buffer
Packet Size	3 Mbits
Intra-Frequency HO	RSRP-based A3 event
$A2_{Thresh}$ (TS case)	-110 dBm
$A2_{Thresh}$ (RSRQ case)	-12, -14, -16 dB
IF mobility-driven HO (RSRQ case)	RSRQ-based A3 event
IF mobility-driven HO (TS case)	RSRP-based A3 event
A3 Offset	3 dB (Intra-HO), 4 dB (Inter-HO)
HO execution Timer	0.15 sec
SCell Addition Configuration Delay	0.05 sec
TTT window	0.4 sec (Intra-HO), 0.5 sec (Inter-HO)
Measurements Error	1 dB
L3 Filtering Factor	4
R_t	{3,6}Mbps for {0%,50%}CA UE ratio
ρ_t	0.8, 0.6, 0.4, 0.35, 0.2
ρ_{hyst}	0.1
Idle-to-Connected Transition Time	0.1 sec
Connected-to-Idle Transition Time	1 sec

The $A2_{Thresh}$ is set equal to $Thresh^{AP}_{sLow}$ in order to minimize the probability of RRC Idle to Connected (*and vice versa*) ping pong events. An idle-to-connected ping pong event is declared whenever a user that switches to RRC Connected, is immediately handed over to a different cell either due to radio conditions or LB purposes [15]. The RRC Idle priority assignment for the RSRQ-based LB simulations is fixed and prioritizes the 2600 MHz capacity carrier ($p_{2600} > p_{1800} = p_{800}$). Measurements towards higher priority carrier are always performed, in contrast to the ones towards a lower priority carrier, which are triggered whenever the serving quality/power drops below the $Thresh^{AP}_{sLow}$ threshold.

CA UEs support a single SCell. The associated RSRQ-based criteria for adding, removing or changing a SCell are outlined in Table II. In particular, a relatively low threshold of -16 dB is set for SCell additions in order to exploit CA as much as possible. If more than one cells meet the SCell addition criterion, the highest RSRQ-measured cell is selected. The SCell removal threshold is set 2 dB lower, avoiding repetitive additions and removals of the same SCell due to RSRQ flunctuations. Finally, a SCell change event is also defined, according to which, the serving SCell is changed whenever a 3 dB stronger neighbor IF cell is detected. Note that TS-driven actions are only applied on the PCell, while the SCell decisions are taken independently based on the aforementioned criteria. Scheduling across the macro carriers is performed jointly, by using a modified proportional fair metric, also denoted as cross-Component Carrier (CC) scheduling [16] that enhances fairness between legacy and CA users. Conventional proportional fair scheduling is applied, if CA is not supported.

The Key Performance Indicators (KPIs) for the conducted study are the average UE throughput and the overall HO rate, defined as the absolute number of HOs averaged over the simulation time and the number of users (*including both intra-frequency and IF HOs*). As IF measurements are more relevant for non-CA devices, 2 additional KPIs have been explicitly utilized for the simulation campaign with 0% CA terminal penetration. To provide an indication of the potential impact on measurement gaps and UE power consumption, the Cumulative Distribution Functions (CDF) referring to the measured RSRQ range and the network cell load are used.

TABLE II. SCell Event Definition

Scell Action	Event	RSRQ Value (dB)
Addition	A4: target becomes better than threshold	-16
Removal	A2: serving becomes worse than threshold	-18
Change	A6: target becomes offset better that SCell	3

V. Simulation Results

Fig. 3 illustrates the average UE throughput for the case when only non-CA devices exist in the network. More specifically, we observe that the $A2_{Thresh} = -12dB$ configuration outperforms any other simulated setup, since it provides adequate IF measurement availability for exploiting the in-built load information that RSRQ carries. Although $A2_{Thresh} = -16dB$ performs the worst for all offered traffic conditions, the performance gap between the -14 dB and -12 dB case increases at lower traffic demands. This effect is explained by the AP behavior in the RRC Idle. At lower load conditions, the $Thresh^{AP}_{sLow} = A2_{Thresh} = -14dB$ threshold is not high enough for triggering reselections towards lower priority carriers. Therefore, the 1800 MHz and 800 MHz carriers are gradually being underutilized. UEs camp at the prioritized 2600 MHz carrier and they establish their RRC connection at the same carrier, whenever they switch to the RRC Connected state. Although not presented, for values higher than $A2_{Thresh} = -12dB$, throughput gains saturate

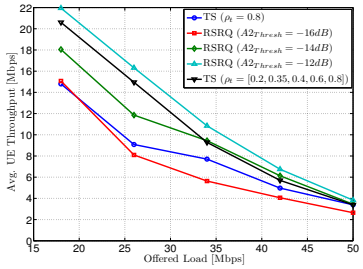


Fig. 3. Avg. UE throughput versus offered load for different LB configurations. 0% CA UE ratio is assumed.

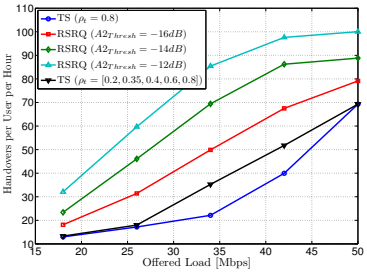


Fig. 4. HO rate versus offered load for different LB configurations. 0% CA UE Ratio is assumed.

and therefore should not be recommended due to the excessive cost in terms of HOs.

The load-based TS policy manages to follow the -12 dB RSRQ performance only if the operational target load, ρ_t, is set according to the offered load conditions (*capacity driven configuration*). This behavior is expected, as the number of TS-driven actions decrease at lower traffic demands, given that the high load $\rho_t = 0.8$ configuration is used. The R_t data rate requirement is met, and no overload is detected for triggering TS events. Nevertheless, the throughput gains of the $A2_{Thresh} = -12dB$ case over the capacity driven TS scheme are in the range of $\sim$7%-$\sim$15% depending on the offered traffic. Better performance can be achieved if TS is more carefully parameterized or an additional LB mechanism is applied, being triggered during the life time of the session.

The associated HO rates are presented in Fig. 4. As expected, there is a clear trade-off between the capacity gains and the derived HO rates. An $A2_{Thresh}$ of -12 dB is the most costly configuration due to the relatively high number of IF HOs that are triggered by the RSRQ sensitivity to the load fluctuations. The advantage of decoupling mobility management from LB is the fact that low HO rates can be achieved. In particular, such an approach results in a 30%-

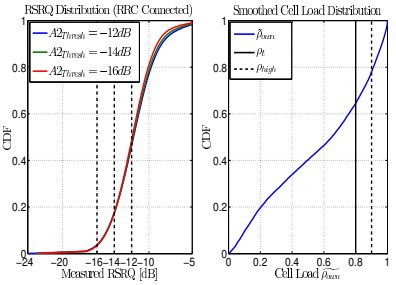


Fig. 5. CDFs for the measured RSRQ and network cell load, $\tilde{\rho}_{own}$. Depending on the selected configuration, different IF measurement availability is provided. 0% CA UE Ratio and traffic of 50 Mbps are assumed.

60% reduction in the HO rates compared to the -12 dB RSRQ case. Considering the 2 different load-based TS configurations, the capacity-driven results in more HOs since LB is triggered, validating the better UE throughput performance that Fig. 3 illustrated. Finally, no difference is observed at low traffic demands due to the fact that LB is provided via the RRC Idle state and the applied dynamic dedicated priority scheme.

Fig. 5 shows the measured RSRQ and network cell load CDFs for the 50 Mbps offered load case. Regarding the developed TS framework, recall that IF measurements are solely triggered whenever $\tilde{\rho}_{own}$ exceeds ρ_{high}. Compared to the proprietary RSRQ-based mobility, the proposed mechanisms not only maintain satisfactory data rates and decrease HO rates, but also achieve such a performance by utilizing IF measurements more efficiently. Although the presented KPIs refer to the RRC Connected state, trends are the same for the RRC Idle. In fact, dedicated priorities ensure that UEs are camping on the highest priority carrier, and therefore, no IF measurements are performed [14].

The CA impact on the investigated configurations is depicted in Fig. 6 with 50% CA UE ratio. Compared to the

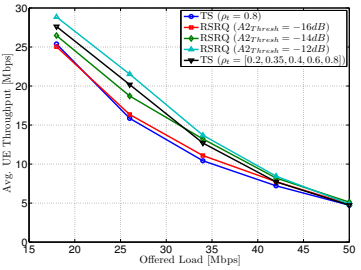


Fig. 6. Avg. UE throughput versus offered load for different LB configurations. 50% CA UE ratio is assumed.

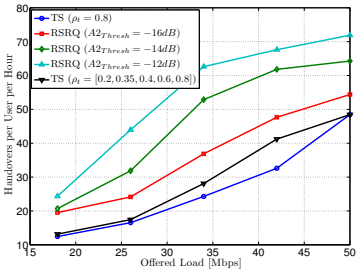


Fig. 7. HO rates versus offered load for different LB configurations. 50% CA UE ratio is assumed.

case without CA, the vast resources availability and the larger transmission bandwidth significantly boosts the system performance at low load. At high traffic demands, gains saturate and the benefits come primarily from the increased multi-user diversity. Moreover, CA makes system performance less sensitive to the mobility/ TS configuration. Fast access to an overlay IF cell is achieved by means of SCell additions for UEs with CA capabilities. Apparently, the $A2_{Thresh}$ effect is only visible at low traffic demands and derives primarily from non-CA users. At higher offered load conditions, all $A2_{Thresh}$ cases perform the same, since the scheduler improves the resource allocation fairness between the 2 different UE categories, maintaining an acceptable performance for non-CA devices. Additionally, CA UEs empty their buffers at a faster rate, releasing resources to be utilized by the legacy terminals. In such a manner, any potential lack of IF measurements for legacy UEs is compensated by the CA scheduler.

Finally, Fig. 6 shows the HO rates for the 50% CA UE ratio case, where it is rather visible that the same gains in terms of HO reduction are maintained by the TS framework. Compared to the corresponding 0% CA UE ratio results, lower rates are now observed. This behavior is an outcome of the finite buffer traffic model as the downlink buffers empty faster and the time that a UE spends at the RRC Connected state is reduced. Note that this plot does not include any SCell-related overhead. In principle, the RRC signaling is dominated by SCell events and minor differences between the different setups has been observed.

VI. Conclusions

Different solutions for inter-frequency load balancing in multi-layer HetNet deployments have been studied with and without carrier aggregation. A traffic steering framework has been developed that triggers inter-frequency handovers/cell reselections whenever load imbalance is detected. Its performance is compared against the standard LTE mobility management framework based only on RSRQ measurements. Results have shown that properly configured RSRQ-based mobility can perform as a passive traffic steering mechanism, exploiting the implicit load information carried in the RSRQ quantity. Albeit an easy approach, the cost in handovers and physical layer measurement rates is relatively high. The developed algorithms tackle this problem by using inter-frequency measurements more efficiently. Handover events can be reduced up to 30%-60% by aligning the load balancing decisions in both RRC states, while UE power consumption is not jeopardized by excessive cell neighbor measurements. If carrier aggregation is supported, system performance becomes less sensitive to the parameterization of the investigated mechanisms. The need for inter-frequency load balancing by means of handoff procedures is relaxed and load imbalances can be compensated by the scheduler as well. Nevertheless, the aforementioned benefits of the proposed traffic steering framework are maintained even in a carrier aggregation environment, making it an attractive low cost solution for inter-frequency load balancing.

References

[1] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis and J.G. Andrews, *User Association for Load Balancing in Heterogeneous Cellular Networks*, IEEE Transactions on Wireless Communications, Volume 12, Issue 6, June 2013.

[2] N. T. K. Jorgensen, D. Laselva and J. Wigard, *On the Potentials of Traffic Steering Techniques between HSDPA and LTE*, IEEE Vehicular Technology Conference, May 2010.

[3] P. Munoz, R. Barco, D. Laselva and P. Mogensen, *Mobility-Based Strategies for Traffic Steering in Heterogeneous Networks*, IEEE Communications Magazine, Volume 51, Issue 5, May 2013.

[4] A. Prasad, O. Tirkkonen, P. Lunden, O. Yilmaz, L. Dalsgaard, and C. Wijting, *Energy-Efficient Inter-Frequency Small Cell Discovery Techniques for LTE-Advanced Heterogeneous Network Deployments*, IEEE Communications Magazine, Volume 51, Issue 5, May 2013.

[5] M. Iwamura, K. Etemad, M. Fong, Y. Wang, R. Nori and R. Love, *Carrier Framework in 3GPP LTE-Advanced*, IEEE Communications Magazine, vol.48, no.8, pp. 60-67, August 2010.

[6] S. Zukang, A. Papasakellariou, J. Montoya, D. Gerstenberger and X. Fangli, *Overview of 3GPP LTE-Advanced Carrier Aggregation for 4G Wireless Communications*, IEEE Communications Magazine, vol.50, no.2, pp.122-130, February 2012.

[7] K.I. Pedersen, P. H. Michaelsen, C. Rosa and S. Barbera, *Mobility enhancements for LTE-advanced multilayer networks with inter-site carrier aggregation*, IEEE Communications Magazine, vol.51, no.5, pp. 64-71, May 2013.

[8] 3GPP TS 36.214, *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer Measurements*, v11.1.0, September 2012.

[9] 3GPP TS 36.331, *Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol Specification*, v11.4.0, September 2012.

[10] 3GPP TS 36.304, *Evolved Universal Terrestrial Radio Access (E-UTRA) User Equipment (UE) Procedures in Idle Mode*, v.10.3.0, September 2011.

[11] R. Kwan, R. Arnott, R. Patterson, R. Trivisonno, and M. Kubota, *On Mobility Load Balancing for LTE Systems*, IEEE Vehicular Technology Conference, September 2010.

[12] 3GPP 36.423 *Evolved Universal Terrestrial Radio Access (E-UTRA); X2 Application Protocol*, v10.5.0, March 2012.

[13] 3GPP TR 36.839, *Mobility Enhancements in Heterogeneous Networks*, v. 11.1.0, December 2012.

[14] P. Fotiadis, M. Polignano, L. Jimenez, I. Viering, C. Sartori, A. Lobinger and K. I. Pedersen, *Multi-Layer Traffic Steering: RRC Idle Absolute Priorities & Potential Enhancements*, IEEE Vehicular Technology Conference, May 2013.

[15] S. Hamalainen, H. Sanneck, and C. Sartori, *LTE Self-Organizing Networks (SON): Network Management Automation for Operational Efficiency*, Wiley 2012.

[16] Y. Wang, K. I. Pedersen, P. E. Mogensen and T. B. Sorensen, *Carrier Load Balancing and Packet Scheduling for Multi-Carrier Systems*, in IEEE Personal, Indoor and Mobile Radio Communication, Pages 370-374, September 2009.

Acronyms

2G Second Generation

3G Third Generation

3GPP Third Generation Partnership Project

4G Fourth Generation

AP Absolute Priorities

BB Basic Biasing

CA Carrier Aggregation

CAC Composite Available Capacity

CAGR Compound Annual Growth Rate

CBR Constant Bit Rate

CC Component Carrier

CIO Cell Individual Offset

CSFB Circuit Switched Fall-Back

DAP Dedicated Absolute Priorities

DC Dual Connectivity

DRX Discontinuous Reception

eICIC enhanced Inter-Cell Interference Coordination

FDPS Frequency Domain Packet Scheduling

FHO@CS Forced Handovers upon Connection Setup

FLC Fuzzy Logic Controllers

GERAN GSM EDGE Radio Access Network

GSM Global System for Mobile Communications

HARQ Hybrid Automatic Repeat Request

HetNet Heterogeneous Network

HOF Handover Failure

HSPA High Speed Packet Access

IMT-A International Mobile Telecommunications Advanced

LA Link Adaptation

LTE Long Term Evolution

M2M Machine-to-Machine

MLB Mobility Load Balancing

MRO Mobility Robustness Optimization

MSE Mobility State Estimation

NGMN Next Generation Mobile Networks

OFDM Orthogonal Frequency Division Multiplexing

PCell Primary Cell

PCI Physical Cell Identity

PF Proportional Fair

PRB Physical Resource Block

RACH Random Access Channel

RAT Radio Access Technology

RE Range Extension

RLF Radio Link Failure

RRC Radio resource Control

RRH Remote Radio Head

RRM Radio Resource Management

RSRP Reference Signal Received Power

RSRQ Reference Signal Received Quality

RSSI Received Signal Strength Indicator

SCell Secondary Cell

SINR Signal-to-Interference plus Noise Ratio

SON Self Organizing Networks

TA Tracking Area

TTT Time-To-Trigger

UE User Equipment

Bibliography

[1] International Telecommunications Union, *World Telecommunication/ICT Indicators database 2013.* 17th Edition, June 2013.

[2] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017," *Cisco VNI Forecast*, *White Paper*, February 2013.

[3] T. Halonen, J. Romero and J. Melero, *GSM, GPRS and EDGE Performance: evolution towards 3G/ UMTS.* John Wiley & Sons, 2002.

[4] E. Dahlman, S. Parkvall, J. Skold and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband.* Academic Press, 2nd Edition, 2008.

[5] H. Holma and A. Toskala, *WCDMA for UMTS: HSPA evolution and LTE.* John Wiley & Sons, 2007.

[6] D. Astely, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom and S. Parkvall, "LTE: The Evolution of Mobile Broadband," *Communications Magazine, IEEE*, vol. 47, pp. 44-51, April 2009.

[7] 3GPP, TS 36.300, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN): Overall Description," *Technical Specifications V8.12.0*, April 2010.

[8] H. Holma and A. Toskala, *LTE for UMTS: OFDMA and SC-FDMA Based Radio Access.* John Wiley & Sons, 2009.

[9] ITU-R, "Requirements related to technical performance for IMT-Advanced radio interface," Recommendation M.2134, November 2008.

[10] J. M. Chapin and W. H. Lehr, "Mobile Broadband Growth, Spectrum Scarcity and Sustainable Competition," TPRC, September 2011.

[11] Ericsson, "It all comes to backhaul," *White Paper*, February 2012.

[12] H. Liang, C. Coletti, H. Nguyen, P. Mogensen and J .Elling, "How Much Can Wi-Fi Offload? A Large-Scale Dense-Urban Indoor Deployment Study," *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, pp.1-6, May 2012.

[13] 3GPP, TS 32.501, "Self-Configuration of Network Elements; Concepts and Requirements", Technical Specifications, v9.2.0, Release 9, June 2010.

[14] Next Generation Mobile Networks (NGMN) Alliance," *Use Cases Related to Self Organizing Network, Overall Description*," *Deliverable*, Version 2.02, December 2008.

[15] Next Generation Mobile Networks (NGMN) Alliance," *Informative List on SON Use Cases*", *Deliverable*, Version 2.02, April 2007.

[16] S. Hamalainen, H. Sanneck and C. Sartori, *LTE Self Organizing Networks (SON)*. John Wiley & Sons, 2012.

[17] S. Zukang, A. Papasakellariou, J. Montojo, D. Gerstenberger and X. Fangli Xu, "Overview of 3GPP LTE-advanced carrier aggregation for 4G wireless communications," *Communications Magazine*, IEEE, vol.50, no.2, pp.122-130, February 2012.

[18] K.I. Pedersen, F. Frederiksen, C. Rosa, H. Nguyen, L.G.U. Garcia and Y.Wang, "Carrier aggregation for LTE-advanced: functionality and performance aspects," *Communications Magazine*, IEEE, vol.49, no.6, pp.89-95, June 2011.

[19] Ericsson, "Traffic and Market Report: On the Pulse of the Networked Society," *White paper*, June 2012.

[20] Nokia Solutions Networks, "Deployment Strategies for heterogeneous Networks," *White Paper*, October 2013.

[21] H. Huang, O. Alrabadi, J. Daly, D. Samardzija, C. Tran,R. Valenzuela and S. Walker, "Increasing throughput in cellular networks with higher-order sectorization," *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth ASILOMAR Conference on*, pp.630-635, November 2010.

[22] D. W. Kifle,B. Wegmann, I. Viering and A. Klein, "On the potential of traffic driven tilt optimization in LTE-A networks," *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, pp.2909-2913, September 2013.

[23] A. Damnjanovic, J. Montojo, W. Yongbin, J. Tingfang Ji, L. Tao, M. Vajapeyam, Y. Taesang, S. Osok and D. Malladi, "A survey on 3GPP heterogeneous networks," *Wireless Communications*, IEEE , vol.18, no.3, pp.10-21, June 2011.

[24] A. Damnjanovic, J. Montojo, C. Joonyoung, J. Hyoungju, J. Yang and Z. Pingping, "UE's role in LTE advanced heterogeneous networks," *Communications Magazine*, IEEE, vol.50, no.2, pp.164-176, February 2012.

[25] A.B. Saleh, S. Redana, B. Raaf, J Hamalainen, "Comparison of Relay and Pico eNB Deployments in LTE-Advanced," *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th* , pp.1-5, September 2009.

[26] C. Coletti, P. Mogensen and R. Irmer, "Deployment of LTE In-Band Relay and Micro Base Stations in a Realistic Metropolitan Scenario," *Vehicular Technology Conference (VTC Fall)*, 2011 IEEE, pp.1-5, Sept. 2011.

[27] J.G. Andrews, "Seven ways that HetNets are a cellular paradigm shift," *Communications Magazine*, IEEE, vol.51, no.3, pp.136-144, March 2013.

[28] Y. Qiaoyang, R. Beiyu, C. Yudong, M. Al-Shalash, C. Caramanis and J.G. Andrews, "User Association for Load Balancing in Heterogeneous Cellular Networks," *Wireless Communications, IEEE Transactions on*, vol.12, no.6, pp.2706-2716, June 2013.

[29] S. Corroy, L. Falconetti and R. Mathar, "Dynamic cell association for downlink sum rate maximization in multi-cell heterogeneous networks," *Communications (ICC), 2012 IEEE International Conference on*, pp. 2457-2461, June 2012.

[30] H. Holma and A. Toskala, *LTE-Advanced: 3GPP Solution for IMT-Advanced*. John Wiley & Sons, October 2012.

[31] Lopez-Perez, D.; Guvenc, I; de la Roche, G.; Kountouris, M.; Quek, T.Q.S.; Jie Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *Wireless Communications, IEEE*, vol. 18, no. 3, pp. 22-30, June 2011.

[32] 3GPP, TR 36.839, "Mobility Enhancements in Heterogeneous Networks," *Technical Report V11.1.0*, December 2012.

[33] M.P. Wylie-Green, T. Svensson, "Throughput, Capacity, Handover and Latency Performance in a 3GPP LTE FDD Field Trial," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE* , vol., no., pp.1-6, December 2010.

[34] T. Bonald, S.E. Elayoubi, A. El Falou and J.B. Landre, "Radio capacity improvement with HSPA+ dual-cell," *Communications (ICC), 2011 IEEE International Conference on*, pp. 1-6, June 2011.

[35] 3GPP, TR 36.842, "Study on Small Cell enhancements for E-UTRA and E-UTRAN; Higher layer aspects," *Technical Report V12.0.0*, December 2013.

[36] 4G Americas, "Self Organizing Networks: The benefits of SON in LTE," *White Paper*, July 2011.

[37] M. Amirijoo, P. Frenger, F. Gunnarsson, H. Kallin, J. Moe and K. Zetterberg, "*Neighbor Cell Relation List and Measured Cell Identity Management*

in LTE," *Network Operations and Management Symposium, IEEE*, pp.152-159, April 2008.

[38] L. Jaechan and D. Hong, "*Management of Neighbor Cell Lists and Physical Cell Identifiers in Self-Organizing Heterogeneous Networks*", *Journal of Communications and Networks*, vol.13, no.4, pp.367-376, Aug. 2011.

[39] J. Ramiro and K. Hamied, *Self Organizing Networks: Self-Planning, Self-Optimization and Self-healing for GSM, UMTS and LTE*. John Wiley & Sons, 2012.

[40] R. Barco, P. Lazaro and P. Munoz, "A unified framework for self-healing in wireless networks," *Communications Magazine, IEEE*, vol. 50, no. 12, pp.134-142, December 2012.

[41] Chu Eunmi, Bang Inkyu, Kim Seong Hwan and Sung Dan Keun, "Self-organizing and self-healing mechanisms in cooperative small-cell networks," *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, pp.1576-1581, September 2013.

[42] O. N. C. Yilmaz, J. Hamalainen and S. Hamalainen, "Self-optimization of Remote Electrical Tilt", *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on*, pp.1128-1132, September 2010.

[43] O. N. C. Yilmaz, J. Hamalainen and S. Hamalainen, "Self-optimization of Random Access Channel in 3GPP LTE", *Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International*, pp. 1397-1401, July 2011.

[44] S. Samulevicius, T. B. Pedersen, T. B. Sorensen and G. Micallef, "Energy Savings in Mobile Broadband Network Based on Load Predictions: Opportunities and Potentials", *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th* , pp 1-5, May 2012.

[45] P. Szilagyi, Z. Vincze and C. Vulkan, "Enhanced Mobility Load Balancing Optimisation in LTE", *Personal Indoor and Mobile Radio Communications (PIMRC), 2012 IEEE 23rd International Symposium on*, pp.997-1003, September 2012.

[46] A. Lobinger, S. Stefanski, T. Jansen and I. Balan, "Load Balancing in Downlink LTE Self-Optimizing Networks", *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, pp.1-5, May 2010.

[47] S. Mwanje and A. Mitschele-Thiel, "A Q-Learning strategy for LTE mobility Load Balancing," *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, pp. 2154-2158, Sept. 2013.

[48] I. Viering, A. Lobinger and M. Dottling, "A Mathematical Perspective of Self-Optimizing Wireless Networks," *Communications, 2*009IEEE International Conference on, pp.1-6, June 2009.

[49] N. T. K. Jorgensen, D. Laselva and J. Wigard, "On the Potentials of Traffic Steering Techniques between HSDPA and LTE," *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, pp 1-5, May 2011.

[50] I. Viering, B. Wegmann, A. Lobinger, A. Awada and H. Martikainen, "Mobility robustness optimization beyond Doppler effect and WSS assumption", *Wireless Communication Systems (ISWCS), 2011 8th International Symposium on*, pp.186-191, November 2011.

[51] A. Awada, B. Wegmann,I. Viering and A. Klein, "A SON-Based Algorithm for the Optimization of Inter-RAT Handover Parameters", *Vehicular Technology, IEEE Transactions on*, vol.62, no.5, pp.1906-1923, June 2013.

[52] R. Romeikat, H. Sanneck and T. Bandh, "Efficient, Dynamic Coordination of Request Batches in C-SON Systems," *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, pp.1-6, June 2013.

[53] 3GPP, TR 36.814, "Further advancements for E-UTRA physical layer aspects," *Technical Report V9.0.0*, March 2010.

[54] J.E.V. Bautista, S. Sawhney, M. Shukair, I. Singh, V.K. Govindaraju and S. Sarkar, "Performance of CS Fallback from LTE to UMTS", *Communications Magazine*, IEEE, vol.51, no.9, pp.136-143, September 2013.

[55] Northstream, 'LTE and the 1800 MHz Opportunity', *White Paper*, March 2012.

[56] Qualcomm, "Standby Time Analysis for Mobile Devices in Multi-carrier WCDMA Networks," *White Paper*, May 2010.

[57] D. Flore, C. Brunner, F. Grilli and V. Vanghi, "Cell Reselection Parameter Optimization in UMTS," *2nd International Symposium on Wireless Communication Systems, 2005*, September 2005.

[58] Ren-Huang Liou, Yi-Bing Lin and Shang-Chih Tsai, "An Investigation on LTE Mobility Management," *Mobile Computing, IEEE Transactions on*, vol.12, no.1, pp.166-176, January 2013.

[59] P. Fotiadis, M. Polignano, L. Chavarria, I. Viering, C. Sartori, A. Lobinger and K. I. Pedersen, "Multi-Layer Traffic Steering: RRC Idle Absolute Priorities & Potential Enhancements," *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, pp.1-5, June 2013.

[60] 3GPP, TS 36.331, "Radio Resource Control (RRC); Protocol specification ", *Technical Specifications V11.5.0*, September 2013.

[61] 3GPP, TS 36.214, "Physical layer; Measurements", *Technical Specifications V11.1.0*, December 2012.

[62] 3GPP, TS 25.304, "User Equipment (UE) procedures in idle mode and procedures for cell reselection in connected mode", *Technical Specifications V8.6.0*, June 2009.

[63] A. J. Fehske, I. Viering,J. Voigt, C. Sartori, S. Redana and G.P Fettweis, "Small-Cell Self-Organizing Wireless Networks," *Proceedings of the IEEE*, vol. 102, no. 3, pp.334-350, March 2014.

[64] 3GPP, TS 36.423, "X2 application protocol (X2AP)", *Technical Specifications V12.2.0*, June 2014.

[65] R. Kwan, R. Arnott, R. Paterson, R Trivisonno and M. Kubota, "On Mobility Load Balancing for LTE Systems," *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, pp.1-5, September 2010

[66] B. Soret, Hua Wang, K. I. Pedersen and C. Rosa, "Multicell cooperation for LTE-advanced heterogeneous network scenarios," *Wireless Communications, IEEE*, vol.20, no.1, pp.27-34, February 2013.

[67] A. F. Cattoni, H. T. Nguyen, J. Duplicy, D. Tandur, B. Badic, R. Balraj, F. Kaltenberger, I. Latif, A. Bhamri, G. Vivier, I.Z Kovacs and P. Horvath, "Multi-user MIMO and Carrier Aggregation in 4G systems: The SAMURAI approach," *Wireless Communications and Networking Conference Workshops (WCNCW), 2012 IEEE*, pp.288-293, April 2012.

[68] 3GPP, TR 25.942, "Radio Frequency (RF) system scenarios", *Technical Report V11.0.0*, September 2012.

[69] P. Mogensen, Wei Na, I. Kovacs, F. Frederiksen, A. Pokhariyal, K. I. Pedersen,T. Kolding, K Hugl and M. Kuusela, "LTE Capacity Compared to the Shannon Bound," *Vehicular Technology Conference, VTC2007-Spring, IEEE 65th*, pp.1234-1238, April 2007.

[70] A. Pokhariyal, T. E. Kolding and P. E. Mogensen, "Performance of Downlink Frequency Domain Packet Scheduling for the UTRAN Long Term Evolution," *Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on*, pp.1-5, September 2006.

[71] R. Jain, D. Chiu and W. Hawe, "A Quantitative Measure of Fairness And Discrimination for Resource Allocation in Shared Computer System," *DEC Technical Report 301*, September 1984.

[72] Y. Wang, B. Soret and K.I. Pedersen, "Sensitivity study of optimal eICIC configurations in different heterogeneous network scenarios," *Communications (ICC), 2012 IEEE International Conference on*, pp. 6792-6796, June 2012.

[73] Y. Wang and K.I. Pedersen, "Performance Analysis of Enhanced Inter-Cell Interference Coordination in LTE-Advanced Heterogeneous Networks," *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, pp. 1-5, May 2012.

[74] D. Lopez-Perez, C. Xiaoli and I. Guvenc,"On the Expanded Region of Picocells in Heterogeneous Networks," *Selected Topics in Signal Processing, IEEE Journal of*, vol.6, no.3, pp.281-294, June 2012.

[75] Qiaoyang Ye, Beiyu Rong, Yudong Chen, C. Caramanis and J.G. Andrews, "Towards an optimal user association in heterogeneous cellular networks," *Global Communications Conference (GLOBECOM), 2012 IEEE*, pp.4143-4147, December 2012.

[76] S. Barbera, P. H. Michaelsen, M. Saily and K. Pedersen, "Mobility performance of LTE co-channel deployment of macro and pico cells," *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*, pp.2863-2868, April 2012.

[77] K. Zetterberg, P. Ramachandra, F. Gunnarsson, M. Amirijoo, S. Wager and T. Dudda, "On Heterogeneous Networks Mobility Robustness", *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*,pp. 1-5, June 2013.

[78] J. Puttonen, N. Kolehmainen, T. Henttonen and J. Kaikkonen, "On Idle Mode Mobility State Detection in Evolved UTRAN," *Information Technology: New Generations, 2009. ITNG '09. Sixth International Conference on*, pp. 1195-1200, April 2009.

[79] N. Kolehmainen, J. Puttonen and J. Kaikkonen, "Performance of idle mode mobility state detection schemes in Evolved UTRAN," *Wireless Pervasive Computing (ISWPC), 2010 5th IEEE International Symposium on*, pp.584-588, May 2010.

[80] S. Barbera, P. H. Michaelsen, M. Saily and K. Pedersen, "Improved mobility performance in LTE co-channel hetnets through speed differentiated enhancements", *Globecom Workshops (GC Wkshps), 2012 IEEE*, pp. 426-430, December 2012.

[81] D. Lopez-Perez, I. Guvenc and C. Xiaoli, "Mobility enhancements for heterogeneous networks through interference coordination," *Wireless Communications and Networking Conference Workshops (WCNCW), 2012 IEEE*, pp.69-74, April 2012.

[82] P. Mugen, L. Dong,W. Yao, L. Jian, and C. Hsiao-Hwa, "Self-configuration and self-optimization in LTE-advanced heterogeneous networks," *Communications Magazine, IEEE*, vol.51, no.5, pp.36-45, May 2013.

[83] P. Munoz, R. Barco and I. de la Bandera, "On the Potential of Handover Parameter Optimization for Self-Organizing Networks," *Vehicular Technology, IEEE Transactions on*, vol.62, no.5, pp.1895-1905, June 2013.

[84] S. S. Mwanje and A. Mitschele-Thiel, "Minimizing Handover Performance Degradation Due to LTE Self Organized Mobility Load Balancing," *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, pp.1-5, June 2013.

[85] J.M. Ruiz-Aviles, S. Luna-Ramirez, M. Toril and F. Ruiz, "Fuzzy Logic Controllers for Traffic Sharing in Enterprise LTE Femtocells," *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, pp. 1-5, May 2012.

[86] C. Werner, J. Voigt, S. Khattak and G. Fettweis, "Handover Parameter Optimization in WCDMA using Fuzzy Controlling," *Personal, Indoor and Mobile Radio Communications, IEEE 18th International Symposium on*, pp.1-5, Sept. 2007.

[87] P. Munoz, R. Barco, J. M. Ruiz-Aviles, I. de la Bandera and A. Aguilar, "Fuzzy Rule-Based Reinforcement Learning for Load Balancing Techniques in Enterprise LTE Femtocells," *Vehicular Technology, IEEE Transactions on*, vol.62, no.5, pp.1962-1973, June 2013.

[88] 3GPP, TR 36.902, "Self-configuring and self-optimizing network (SON) use cases and solutions," *Technical Report V9.3.1*, March 2011.

[89] A. Awada, B. Wegmann, I. Viering and A. Klein, "Cell-pair specific optimization of the inter-RAT handover parameters in SON," *Personal Indoor and Mobile Radio Communications (PIMRC), 2012 IEEE 23rd International Symposium on*, pp.1168-1173, Sept. 2012.

[90] A. Awada, B Wegmann, I. Viering and A. Klein, "A location-based self-optimizing algorithm for the inter-RAT handover parameters," *Communications (ICC), 2013 IEEE International Conference on*, pp.6168-6173, June 2013.

[91] T. Dean and P. Fleming, "Trunking efficiency in multi-carrier CDMA systems," *in Proceeding of IEEE Vehicular Technology Conference*, vol. 1, pp. 156–160, September 2002.

[92] A. He, "Performance comparison of load balancing methods in multiple carrier CDMA systems," *in Proceeding of IEEE Personal Indoor and Mobile Radio Communications (PIMRC)*, vol. 1, pp 113–118, September 2000.

[93] Hua Wang, C. Rosa and K. Pedersen, "Performance Analysis of Downlink Inter-Band Carrier Aggregation in LTE-Advanced," *Vehicular Technology Conference (VTC Fall), 2011 IEEE*, pp.1,5, 5-8 September 2011.

[94] J. Kurjenniemi, T. Henttonen and J. Kaikkonen, "Suitability of RSRQ measurement for quality based inter-frequency handover in LTE," *Wire-*

less Communication Systems, ISWCS '08, IEEE International Symposium on, pp. 703-707, October 2008.

[95] M. Kazmi, O. Sjobergh, W. Muller, J Wierok and B. Lindoff, "Evaluation of Inter-Frequency Quality Handover Criteria in E-UTRAN," *Vehicular Technology Conference, 2009 VTC Spring, IEEE 69th*, pp.1,5, 26-29 April 2009.

[96] P. Munoz, R. Barco, D. Laselva and P. E. Mogensen, "Mobility-Based Strategies for Traffic Steering in Heterogeneous Networks," *IEEE Communications Magazine*, Volume 51, Issue 5, May 2013.

[97] S. Luna-Ramirez, F. Ruiz, M. Toril and M. Fernandez-Navarro, "Inter-System Handover Parameter Auto-Tuning in a Joint-RRM Scenario," *Vehicular Technology Conference, VTC Spring 2008 IEEE*, pp.2641-2645, May 2008.

[98] I. De La Bandera, S. Luna-Ramiirez, R. Barco, M. Toril, F. Ruiz and M. Fernandez-Navarro, "Inter-system cell reselection parameter auto-tuning in a joint-RRM scenario," *Broadband and Biomedical Communications (IB2Com), 2010 Fifth International Conference on*, pp.1-6, December 2010.

[99] A. Prasad, P. Lunden, O. Tirkkonen and C. Wijting, "Energy-Efficient Flexible Inter-Frequency Scanning Mechanism for Enhanced Small Cell Discovery," *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, pp.1-5, June 2013.

[100] A. Prasad, P. Lunden, O. Tirkkonen and C. Wijting, "Energy Efficient Small-Cell Discovery Using Received Signal Strength Based Radio Maps," *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, pp.1-5, June 2013.

[101] A. Prasad, P. Lunden, O. Tirkkonen and C. Wijting, "Enhanced small cell discovery in heterogeneous networks using optimized RF fingerprints,"*Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, pp. 2973-2977, September 2013.

[102] A. Prasad, O. Tirkkonen, P. Lunden, O.N.C. Yilmaz, L. Dalsgaard and C. Wijting, "Energy-efficient inter-frequency small cell discovery techniques for LTE-advanced heterogeneous network deployments," *Communications Magazine, IEEE*, vol.51, no.5, pp.72-81, May 2013.

[103] Y. Wang; K. I. Pedersen, T. B.Sorensen, P. E. Mogensen, "Carrier load balancing and packet scheduling for multi-carrier systems", *Wireless Communications, IEEE Transactions on*, vol. 9, no. 5, pp. 1780-1789, May 2010.

[104] Y. Wang, K. I. Pedersen, T. B. Sorensen and P. E. Mogensen, "Utility Maximization in LTE-Advanced Systems with Carrier Aggregation," *Vehic-*

ular Technology Conference (VTC Spring), 2011 IEEE 73rd, pp.1-5, May 2011.

[105] K. I. Pedersen, P. H. Michaelsen, C. Rosa and S. Barbera, "Mobility enhancements for LTE-advanced multilayer networks with inter-site carrier aggregation," *Communications Magazine, IEEE*, vol.51, no. 5, pp.64-71, May 2013.

[106] K. Yagyu, T. Nakamori, H. Ishii, M. Iwamura, N. Miki, T. Asai and J. Hagiwara, "Investigation on Mobility Management for Carrier Aggregation in LTE-Advanced," *Vehicular Technology Conference (VTC Fall), 2011 IEEE*, pp. 1-5, September 2011.

[107] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33-37, 1997.

[108] J. M. Holtzman, "Asymptotic analysis of proportional fair algorithm," *Personal, Indoor and Mobile Radio Communications, 2001 12th IEEE International Symposium on*, vol.2, pp. F33-F37, September 2001.

[109] Y. Wang, K. I. Pedersen, T. B. Sorensen and P. E. Mogensen, "Resource allocation considerations for multi-carrier LTE-Advanced systems operating in backward compatible mode, *"Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*, pp.370-374, September 2009.

[110] Liu Liu, Mingju Li, Juejia Zhou, Xiaoming She, Lan Chen, Y. Sagae and M. Iwamura, "Component Carrier Management for Carrier Aggregation in LTE-Advanced System," *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, pp.1-6, May 2011.

[111] S. Barbera, K. Pedersen, P. H. Michaelsen and C. Rosa, "Mobility Analysis for Inter-Site Carrier Aggregation in LTE Heterogeneous Networks," *Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th*, pp. 1-5, September 2013.

[112] M. Karabacak, D. Wang, H. Ishii and H. Arslan, "Mobility Performance of Macrocell-Assisted Small Cells in Manhattan Model," *Vehicular Technology Conference (VTC Spring), 2014 IEEE 79th*, pp.1-5, May 2014.

[113] Hua Wang, C. Rosa and K.I. Pedersen, "Dedicated carrier deployment in heterogeneous networks with inter-site carrier aggregation," *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, pp. 756-760, April 2013.

[114] M. Lauridsen, Hua Wang and P. Mogensen, "LTE UE Energy Saving by Applying Carrier Aggregation in a HetNet Scenario," *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, pp.1-5, June 2013.

[115] 3GPP, TS 36.413,"S1 Application Protocol (S1AP)", *Technical Specifications V12.2.0*, June 2014.

[116] I. Viering, M. Dottling and A. Lobinger, "A Mathematical Perspective of Self-Optimizing Wireless Networks," *Communications, 2009. ICC '09. IEEE International Conference on*, pp.1-6, June 2009.

[117] C. Coletti, *Heterogeneous Deployment Analysis for Cost-Effective Mobile Network Evolution: An LTE Operator Case Study.* PhD Thesis, Aalborg University, September 2012.

[118] 3GPP 30.03,"Selection Procedures for the choice of radio transmission technologies of the UMTS", *Technical Specifications V9.0.0*, April 1998.

[119] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics Letters*, vol.27, no.23, pp.2145-2146, Nov. 1991.

[120] 3GPP, TS 36.133, "Requirements for support of radio resource management," *Technical Specifications V12.4.0*, July 2014.

[121] J. Huang, F. Qian, A. Gerber, Z. Morley, M. Subhabrata and S. O. Spatscheck, "A close examination of performance and power characteristics of 4G LTE networks," *Proceedings of the 10th International Conference on Mobile systems, applications and services*, pp. 225-238, June 2012.