# SDN-based Flexible Resource Management and Service-Oriented Virtualization for 5G Mobile Networks and Beyond

Rudraksh Shrivastava

Ph.D.

University of York

Electronics

December 2016

# Abstract

This thesis examines how Software Defined Network (SDN) and Network Virtualization (NV) technologies can make 5G and beyond mobile networks more flexible, scalable and programmable to support the performance demands of the emerging heterogeneous applications. In this direction, concepts like mobile network slicing, multi-tenancy, and multi-connectivity have been investigated and their performance is analyzed. The SDN paradigm is used to enable flexible resource allocation to the end users, improve network resource utilization and avoid or rapidly solve the network congestion problems. The proposed network architectures are 3rd Generation Partnership Project (3GPP) standards compliant and integrate Open Network Foundation (ONF) SDN specifications to ensure seamless interoperability between different standards and backward/forward compatibility. Novel mechanisms and algorithms to efficiently manage the resources of evolving 5G Time-Division Duplex (TDD) networks in a flexible manner are introduced. These mechanisms enable formation of virtual cells on-demand which allows diverse resource utilization from multiple eNBs to the users. Within the scope of this thesis, SDN-based frameworks to enhance the QoE of end user applications considering Time Division-Long Term Evolution (TD-LTE) small cells have also been developed and network resource sharing scenarios with Frequency-Division Duplex (FDD)/TDD coexistence has been studied.

In addition, this thesis also proposes and investigates a novel service-oriented network slicing concept for evolving 5G TDD networks which involve traffic prediction mechanisms and includes user mobility. An analytical model is also introduced that formulates the network slice resource allocation as a weighted optimization problem. The evaluations of the proposed solutions are performed using 3GPP standard compliant simulation settings. The proposed solutions have been compared with the state-of-the art schemes and the performance gains offered by the proposed solutions have been demonstrated. Performance is evaluated considering metrics such as throughput, delay, network resource utilization etc. The Mean Opinion Score (MOS) metric is used for evaluating the Quality of Experience (QoE) for end-user applications.

With the help of SDN-based network management algorithms investigated in this work, it is shown how 5G+ networks can be managed efficiently, while at the same time provide enhanced flexibility and programmability to improve the performance of diverse applications and services delivered over the network to the end users.

# Contents

# List of Figures

# List of Tables

# Acknowledgements

The research work for this doctoral thesis has been carried out jointly at the NEC Laboratories Europe, located in Heidelberg, Germany and at the Department of Electronics of The University of York, York, United Kingdom. This work is a part of European Marie Curie Innovative Training Networks (ITN) called CROSSFIRE (unCooRdinated netwOrk StrategieS for enhanced interFerence, mobIlity, radio Resource, and Energy saving management in LTE-Advanced networks). Firstly, I would like to express my deepest gratitude to my academic supervisor, Prof. David Grace and my industrial supervisor, Dr. Konstantinos Samdanis for their infinite patience, continuous motivation and immense knowledge. Their guidance, support and encouragement helped me in all the time of this research work and writing of this thesis. Their invaluable comments helped in improving the quality of the work presented in this thesis. It has been a great pleasure, honor and privilege to work with them during the course of my doctoral studies. I could not have imagined having better advisors and mentors for my Ph.D. study. I would also like to thank Dr. Paul Mitchell at The University of York for being my thesis advisor. For continuous support, enlightening discussions and invaluable comments during my work at NEC Laboratories, I would like to express special thanks and gratitude to Dr. Athul Prasad (Senior Specialist, 5G Radio Research at Nokia Bell Labs) for his support and collaboration which helped in completing the work in this thesis. I would like to sincerely thank Dr. Xavier Costa-Perez (Head of 5G Networks R&D at NEC Laboratories Europe), Dr. Vincenzo Sciancalepore (Research Scientist, NEC Laboratories Europe), Dr. Peter Rost (Senior Researcher, Nokia Networks) and Dr. Nikos Passas, Mr. Salvatore Costanzo, Dr. Dionysis Xenakis and Prof. Lazaros Merakos from the Communication Networks Laboratory, Dept. of Informatics & Telecommunications University of Athens for the productive collaboration that resulted in many publications. Additionally, I would like to thank all the people and project partners involved in the EU Marie Curie CROSSFIRE ITN project for all the fruitful discussions, collaborations and project meetings.

I would also like to thank all my former colleagues at NEC Lab. Europe, especially Dr. Juergen Quittek, Dr. Heinrich Stuttgen and Prof. Tarik Taleb (Aalto University, Finland) for their support during my time in the lab. I would also like to express my gratitude towards my colleagues at Fraunhofer IIS, Mr. Bernhard Niemann (Head of Broadband and Broadcast Department) and Mr. Thomas Heyn (Group Manager, Applied Communication Systems) for providing encouragement during the writing phase of my thesis. Last but not the least, I would like to sincerely thank my parents and my brother for their support and encouragement on my path towards the doctoral degree and in my life in general. Their continuous support, encouragement and motivation were invaluable.

The research work leading to this thesis has been partially funded by the European Commission ITN FP7 Marie Curie project called CROSSFIRE (MITN 317126).

# Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. Some of the research work presented in this thesis have resulted in publications, in peer-reviewed journals, conference proceedings and EU research project deliverables. To the best knowledge of the author, all the work presented in this thesis is original. References and acknowledgements to other researchers have been given as appropriate.

**Conference Papers**

K. Samdanis, R. Shrivastava, A. Prasad, P. Rost and D. Grace, "Virtual Cells: Enhancing the Resource Allocation Efficiency for TD-LTE," IEEE Vehicular Technology Conference (VTC 2014-Fall), Vancouver (BC), Canada, 2014.

R. Shrivastava, S. Costanzo, K. Samdanis, D. Xenakis, D. Grace and L. Merakos, "An SDN-based framework for elastic resource sharing in integrated FDD/TDD LTE-A HetNets," IEEE International Conference on Cloud Networking (CloudNet-2014), Luxembourg, 2014.

S. Costanzo, R. Shrivastava, D. Xenakis, K. Samdanis, D. Grace and L. Merakos, "An SDN-based virtual cell framework for enhancing the QoE in TD-LTE pico cells," IEEE International Workshop on Quality of Multimedia Experience (QoMEX-2015), Pylos-Nestoras, Greece, 2015.

S. Costanzo, R. Shrivastava, K. Sarndanis, D. Xenakis, X. Costa-Perez and D. Grace, "Service-oriented resource virtualization for evolving TDD networks towards 5G," IEEE Wireless Communications and Networking Conference (WCNC-2016), Doha, Qatar, 2016.

V. Sciancalepore, K. Samdanis, R. Shrivastava, A. Ksentini and X. Costa-Perez, "A Service-tailored TDD Cell-Less Architecture," IEEE Personal, indoor and Mobile Communications Conference (PIMRC-2016), Valencia, Spain, 2016.

**Journal Articles**

K. Samdanis, R. Shrivastava, A. Prasad, D. Grace, and X. Costa-Perez, "TD-LTE Virtual Cells: An SDN architecture for user-centric multi-eNB elastic resource management," Comput. Commun. 83, C (June 2016).

R. Shrivastava, K. Samdanis, V. Sciancalepore, D. Grace, "5G TDD Network Slicing as a Service: Analysis and Design for practical Solutions," (Planned to be submitted to IEEE Transactions).

**Book Chapter**

R. Shrivastava, K. Samdanis, D. Grace, "Towards Dynamic Service-oriented TDD for 5G systems," Access, Front-haul and Backhaul Networks for 5G & Beyond, IET Book, 2016 (Accepted).

**Patents**

Rudraksh Shrivastava, Konstantinos Samdanis et.al., "Network Slicing in 5G TDD networks", Pat. PCT/GB2016/050914, 2016.

Rudraksh Shrivastava, Khaled Hassan et.al., "Full Duplexing Downlink and Uplink Directions, Pat. (EP17171718.4), 2016.

# Chapter 1

# Introduction

## 1.1 Overview

The rapid proliferation of new types of devices such as smartphones, tablets, wearable electronics combined with the widespread evolution of high speed 5G mobile networks, have led to an evolution of diverse mobile services, generating huge amounts of data traffic [1]. Some emerging 5G applications need faster, higher-capacity networks that can deliver video and other content-rich services while some applications have heterogeneous requirements which may create conflicts while co-existing with other type of applications. Mobile applications that involve social media and cloud services have changed the way humans communicate and acquire information from the Internet, being also more interactive due to "always on" features, with asymmetric uplink/downlink demands [2]. Applications such as Machine Type Communication (MTC) and IoT have varying traffic demands in both UL and DL directions. MTC applications need to upload or exchange high data volumes raising certain Quality of Service (QoS) or Quality of Experience (QoE) demand issues at different times. The MTC applications have varying downlink traffic demands but more diverse uplink traffic patterns. Hence, such emerging mobile applications and services require an increased degree of network resource elasticity to effectively guarantee different QoS/QoE demands. Internet of Things (IoT) applications are fostering and seeding the need for massive connectivity of devices that require ultra-reliable, ultra-low-latency (URLLC) connectivity over Internet Protocol (IP) (for e.g. in industrial control systems (from sensor to actuator, very low latency is needed for some applications), real time control of vehicles, road traffic, accident prevention (location, vector, context, low Round Trip Time (RTT)) etc. [3]). A number of applications have been identified where in the existing mobile network infrastructure may struggle to deliver the expected performance demands of the emerging 5G applications. For example, such applications include V2X (vehicle-to-vehicle) communications, industrial automation and utility applications, consumer and business applications, augmented reality applications, smart city services such as smart homes and mobile broadband everywhere etc. Given the use cases and service requirements that 5G networks are expected to support, it is imperative that the network operators should upgrade their network. This however may lead to increase in CAPEX and OPEX for the network operators. Thus, mobile network operators need to efficiently manage the network resources and create new revenue areas such as providing services to different tenants that includes vertical markets and Mobile Virtual Network Operators (MVNOs). The new features offered by 5G networks should also cultivate new innovations and support various business models with the ability to rapidly test and deploy new services. These objectives are considered while designing solutions for 5G networks and are reflected in the overall evolution of networks towards 5G. 5G TDD has the potential to support the emerging heterogeneous services and their corresponding diverse Service Level Agreements (SLAs) with diverse UL/DL requirements. It employs among others, state of the art enabling technologies such as Software Defined Networking (SDN) and Network Virtualization that also support multi-tenant operations, i.e. multi-tenancy [4]. On-demand resource provisioning is essential to support multi-tenancy therefore resource flexibility is necessary and essential. Network programmability and SDN are being increasingly considered as potential solutions in this direction. SDN with the help of standard interfaces allows application and service providers including vertical markets and MVNOs to communicate their service requirements to the infrastructure provider. The infrastructure provider can then program the network resources to meet the Quality of Service (QoS) constraints following the SLAs established with the service providers and vertical market players [5] [6]. Network programmability refers to the ability to dynamically configure the operational network parameters and network resource slicing policies for a particular tenant.

Furthermore, it is essential for mobile networks to simultaneously support multiple generations of mobile services (i.e., 3G and 4G) along with a range of user services such as VoIP, streaming media, and messaging which results in widely varying traffic properties. This support should be done in a cost-effective manner at a time when rapidly increasing capacity demands far exceed the growth in revenues. It is also required that the budgets should address these new demands. To deliver services seamlessly across different technologies is another challenge that the network operators face today. They need to enforce increasingly complex, granular policies to ensure the right access for the right service and control handovers between access types. Fast rollout of new mobile services and rapid adoption of new technologies is needed in a business environment that is dynamic and competitive. Existing mobile network operators not only face competition from each other but also from the Over The Top service providers (OTT) and established internet giants. The challenges and competition faced by the mobile network operators today is causing decrease in voice revenues. As a result the mobile operators are seeking new business opportunities and developing innovative business models for different markets. The operators are also interested to generate revenues from a variety of new data services, such as location, e-commerce, and analytics. Supporting such services requires highly scalable and agile network capabilities despite the large capital outlays. Therefore, to achieve this, it is necessary to lower the cost of hardware wherever possible, utilize maximum hardware assets while reducing operational costs by embracing new techniques and technologies such as SDN, cloud computing and automation.

In cellular networks, the RAN (Radio Access Network) provides wide area connectivity to the mobile devices. One of the primary problems in existing networks is how to manage limited network resources to best serve the users. The latest 3GPP LTE-A release [7] [8], defines a heterogeneous environment consisting of E-UTRAN NodeBs (eNodeBs) and small cells (i.e pico eNBs, Relay Nodes (RNs) and Home eNodeBs (HeNodeBs). In such a dense and heterogeneous environment with limited network resources the task to efficiently allocate and manage radio resources, implement handovers, manage interference, balance load between cells, etc. becomes even more challenging. Interference is a critical issue, if left unmanaged it could significantly degrade the network capacity. There are two main reasons that have been identified for interference especially in dense network deployment scenarios. First, due to the limited spectrum availability, cells may have a high frequency re-use factor and second is the broadcast nature of the wireless networks. Also, in such networks traffic load fluctuates more rapidly due to high user mobility. Consequently radio resource allocation, handovers and cell association have to be managed at each base station in coordination with their neighbors to maximize the network capacity by carrying out tasks such as interference management, resource optimization, load balancing, etc. A conventional radio access network consists of a group of base stations, each base station usually making independent control plane decisions at the radio layer with some loose distributed coordination via mechanisms such as SON (self-organizing networks), Inter-cell Interference Coordination (ICIC), etc. However, in a dense network deployment scenario, coordinated control plane decisions have to be made across several neighboring base stations simultaneously. These control plane decisions should be made with as low latency as possible. Due to the large number of base stations these distributed algorithms often do not scale well especially in terms of latency. This leads to performance degradation and significant capacity reduction due to the inability to efficiently manage load and interference. Furthermore, the distributed coordination algorithms tend to be more complex and computationally intensive as often they require iterative and periodic adjustments of RRM decisions that are difficult to scale.

3GPP's existing LTE/LTE-A architecture with distributed control plane is suboptimal for efficiently utilizing network resources while providing high speed and robust services to the end user. Existing networks are static, difficult to scale and inflexible in nature, therefore unable to fully address the high bandwidth required to keep up with the rising user demands. Conventional wireless networks are difficult to manage. Currently the networks rely on OSS (Operation Support Subsystems) and management systems that require significant expertise and platform resources to operate the network. These systems are manually intensive and hence prone to misconfiguration errors and lengthy delays in provisioning and troubleshooting.

The networks today are becoming increasingly costly due to inefficient and inflexible use of network bandwidth and ever-increasing complexity, burdening the CAPEX and especially the OPEX for the network operators and the service providers. Due to the inflexible network architecture, existing networks require weeks or even months to introduce new services because of the manually intensive processes for service activation, delivery, and assurance. Multi-tenancy and traffic isolation are limited to such network architectures as WLANs and tunnels, with limited policy management mechanisms. Considering these issues, it is necessary that we fundamentally re-think the network design and architecture of the existing mobile networks to be able to cope with rapidly increasing traffic and operators' requirements.

## 1.2 Purpose

Emerging 5G and beyond networks are envisioned to support higher data volumes and a plethora of heterogeneous services/applications with diverse and often conflicting requirements. To accommodate such service requirements, a cost efficient and flexible network architecture considering different service types is desired. Existing mobile networks have a monolithic composite network architecture that may not offer the desired flexibility and scalability necessary to support the emerging 5G services/applications. Thus we need a holistic approach to transform the way networks are managed based on real time traffic behavior utilizing the strength of Software Defined Networking (SDN) and network virtualization (NV) technologies.

The work in this thesis aims to design service-oriented, modular end-to-end network architectures that simplify network management and provide flexibility to enable new services that the 5G networks and beyond are expected to deliver. To achieve this, SDN framework offers a logically centralized control model, unprecedented programmability, and a flow-based paradigm that is suited for highly scalable and flexible telecommunication networks-including access, backhaul to core. The basic idea behind the SDN paradigm is to provide a separation between data and control plane to have more control and flexibility over the network which can help in improving the overall performance of the network. The purpose of the work in this thesis is to make the evolving 5G mobile networks more flexible and scalable by proposing a number of ideas, algorithms and techniques that improves the network resource management. The proposed techniques and ideas in this thesis, helps in realizing concepts such as multi-connectivity, network slicing, multi-tenancy and network resource sharing etc. while at the same time enhance the overall network performance in terms of Key Performance Indicators (KPIs) such as delay, throughput, QoE, capacity etc. The primary aim is to support the performance demands of the evolving heterogeneous applications that the 5G networks are expected to deliver. The work in this thesis explores various aspects of SDN and network virtualization (NV) as enabling technologies in 5G and beyond telecommunication networks for different duplexing techonlogies such as

Time Division Duplex (TDD) and frequency division duplex (FDD). SDN with the help of network virtualization (NV) technology has the potential to provide forward looking solutions for efficient and cost effective radio resource management and network optimization including load and interference management for different scenarios such as heterogeneous dense network deployments to enhance the overall network performance while reducing the operational costs.

Whilst the main focus of this thesis is to investigate SDN based flexible network resource management mechanisms to support 5G applications/services and enhance the overall network performance, the concepts developed in this thesis will also help in harvesting underlying business opportunities and provide the following benefits:

- Foster rapid innovation
- Shorter time to market
- Easy testing and faster deployment of new concepts and services
- Reduced OPEX
- Support multi-tenancy which involves on-demand resource provision where, resource flexibility is an essential parameter.

Thus, it is envisioned that SDN enabled 5G and beyond communication systems will provide significant flexibility with autonomous re-configurable capabilities to satisfy the varying needs of the end users while delivering high performance.

## 1.3 Thesis Outline

The rest of the thesis is outlined as follows:

Chapter 2 provides the overview of the state of the art and background information related to the work done in this thesis. A literature review on mobile network virtualization is presented first that summarizes the latest work and research being done in this field. Following this, comprehnsive information on 5G enabling technologies such as SDN and NFV is provided. Finally the concepts related to Time Division Duplex (TDD) and Dynamic TDD are explained that form the basis of this thesis, including state of the art research work relevant to the work done in the thesis.

Chapter 3 introduces the concept of virtual cells in Time-Division Long Term Evolution (TD-LTE) systems, which enables users residing in overlapping cells' coverage regions to utilize resources from multiple base stations. The benefits of virtual cells are realized through efficient resource utilization, via adapting the network resource availability with the traffic demand taking also into account the impact of cross-slot interference. Besides the increased resource flexibility, virtual cells also resolve pseudo congestion, and introduce a customized, user specific, resource utilization. Such a feature enables mobile users to utilize sub-frames from different base stations creating an on demand virtual frame that is comprise of subframes derived from the frame configurations of multiple base stations. In addition, this chapter also introduces mechanisms and algorithms for efficiently managing the resources of 5G evolving TDD networks in a flexible manner enabling (i) dynamic frame alternation at each evolved Node B (eNB), (ii) algorithm to form virtual cells with selective neighbor cells allowing diverse resource utilization from multiple eNBs for users residing within certain overlapping cell regions and (iii) analytical logic to allocate resources to multi-eNB connected users, that are served by virtual frame configuration.

Chapter 4 proposes an SDN-based framework for enhanced Quality of Experience (QoE) in the presence of Time Division-Long Term Evolution (TD-LTE) pico-cell hotspots. The SDN based framework will enable users to utilize radio-resources from multiple base stations with the aim to enhance the QoE of the applications. It will also help in preventing situations where QoE degradation of individual applications and service interruptions may occur.

This is performed by utilizing the concept of virtual cells (presented in Chapter 3) and by dynamically customizing the TD-LTE frames at the picocell stations. The aim of the proposed SDN based framework is to prevent or resolve the pseudo congestion problem incase it occurs and enable elastic and efficient radio-resource management to match the individual application traffic demands at the end users, in this way supporting the overall purpose of this thesis. In addition, in this chaper a novel SDN-based framework that enables efficient and elastic spectrum utilization among multiple operators in 3GPP LTE-A HetNet scenario is presented. Assuming a multi-operator environment of Frequency Division Duplex (FDD) macrocells complemented by multi-tenant Time Division Duplex (TDD) pico cells, an SDN-based architecture is presented that allows efficient resource sharing among the TDD and FDD systems in a dynamic way. A TDD frame re-configuration mechanism is also employed, to optimize the ratio of uplink and downlink slots in the TDD frame of picocells. Performance evaluation analysis of the proposed concepts and algorithms have also been discussed within this chapter in detail.

Chapter 5 introduces the concept of network slicing in 5G TDD networks. First, a static service-oriented network resource slicing scheme is introduced for a Time Division Duplex (TDD)network that, for a pre-defined time duration, forms service specific network slices based on traffic prediction. Following this, a more advanced, flexible and adaptive 5G TDD slicing concept with dynamic UL/DL frame reconfiguration based on traffic prediction is proposed. To program the network and configure the network slices as per the application/service demands an SDN based architecture is discussed. An analytical model is introduced that formulates the network slice resource allocation as a weighted optimization problem and helps periodically in monitoring and allocating appropriate capacity to the slices. A Novel metric for UL/DL configuration selection for each application specific slice is also proposed which involves traffic prediction and spatial throughput of a particular region within the RAN. The proposed metric determines the probability that a user visits a particular location at a certain known point in time. This metric is then used to adapt UL/DL TDD frames, instead of the conventional UL/DL buffer status and past average throughput. SLAW mobility model is employed to model the mobility of the users and model traffic variations within the network. The performance evaluation of the proposed solutions are also analyzed in this chapter.

In chapter 7 future work is discussed. In particular, this chapter discusses some of the ideas for taking forward the research work presented in this thesis. Finally, the conclusion and the novel contributions of the work done in the framework of this thesis are summarized in chapter 6.

# Chapter 2

# Literature-Review

## 2.1 Mobile Network Virtualization

Since the beginning, mobile communication networks have been designed, built and operated as vertically integrated systems, this means that the operator owns the network infrastructure and spectrum, applications are well defined for example, voice or sms and devices are built with closed architectures. This type of architecture exhibits very tight coupling between all network elements. As a result existing cellular networks suffer from expensive and rigid equipment, complex control plane protocols and vendor specific configuration interfaces. One of the most indubitable challenges for the existing cellular infrastructure is to address the rapid and variable growth in mobile data traffic due to the massive increase in the number of mobile devices, data service usage and dynamic traffic patterns. Therefore designing, developing and standardizing new network architectures to support the rising demand in mobile data services without increasing the CAPEX is not just a need but a necessity. Software Defined Network (SDN) with virtualization as an enabling technology offers a potential solution to address the issues being faced by the existing mobile networks. OpenFlow is considered as one of the enabling communication protocols for SDN that allows access to the forward plane of the network elements over a network [9]. The Open Networking Foundation (ONF) has defined an SDN architecture model that is discussed in [10] with compelling business cases for mobile and wireless networks. Here, two use cases have been discussed that illustrate the value proposition: wireless network control for inter-cell interference and mobile traffic management.

Over the past few years, network virtualization and SDN have received significant attention especially in the mobile network virtualization domain. Virtualization in the field of computer science is a well-known technique and has been studied and used for several years. For example, virtual memory, virtualizing computer hardware, virtualized operating system etc. In next generation broadband wireless networks and future Internet, one possibility is to have multiple co-existing mobile operators, in which each is designed and customized to fit one type of network with specific requirements. Network Virtualization will play a vital role in diversifying next generation high speed communication infrastructure into for example; separate virtual networks that are isolated from each other and can run different services within. For designing SDN enabled flexible, re-configurable and high performance next generation networks, a critical review and analysis of the existing wireless network architectures with state of the art research and developments in this area is essential. Hence the details regarding the state of the art research and developments in this area will be discussed in the subsequent sections of this chapter.

The basic idea is to virtualize the physical elements in a mobile network to ensure maximum flexibility in order to efficiently manage the network resources and adapt to the changing traffic behavior. It creates a system in which the user and control plane are decoupled by an abstraction layer from the services that run on top of it and allows network elements to be controlled, operated and optimized by other entities in the higher layers. This approach supports backward compatibility and has the potential to dynamically deploy innovative services in a very short time. One of the key motivations behind mobile network virtualization is to enable dynamic and on demand network sharing for both core and access networks. Network virtualization provides operators with more independent control over their share of the common physical network which further improves sharing by making it more efficient and by providing greater flexibility.

Network consolidation and network slicing are the two different aspects of network virtualization that have [11]. Network consolidation aims to consolidate separate physical networks into one physical infrastructure as shown in figure 2-1. While network slicing refers to partitioning the network resource into several isolated networks with the help of virtualization technology as shown in figure 2-2. Each of these separate networks then could be configured and optimized to deliver specific services and it can be possible to have a multitude of service specific networks.

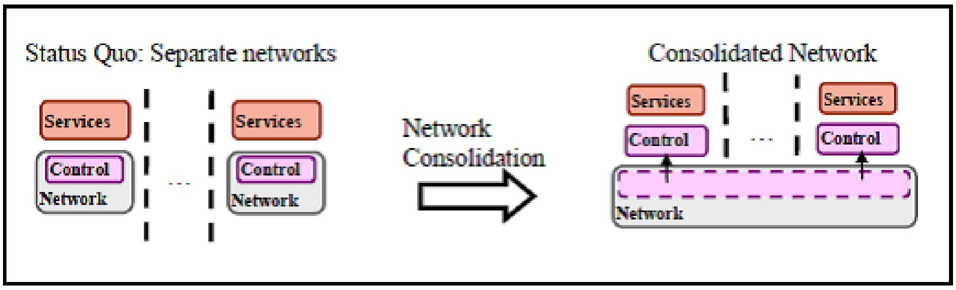


Figure 2-1: Network Consolidation (Reproduced from [11])

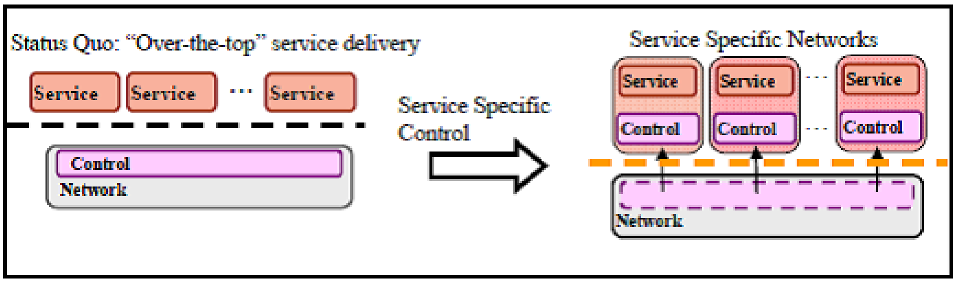


Figure 2-2: Network Consolidation (Reproduced from [11])

The concept of network virtualization in mobile networks can be applied in several layers of the network for example in the access, backhaul and core network. However, in this chapter we will mainly focus on the Radio Access Network (RAN) virtualization for LTE-A. SDN with network virtualization enables efficient management and sharing of network resources, providing the possibility to implement a wide range of new and innovative business models within a short time. A number of virtualization solutions for RAN have been proposed to support efficient sharing of network resources. Efficient spectrum sharing can be potentially enabled via virtualization in LTE. The term eNodeB virtualization refers to the case where multiple virtual network functions (VNF) responsible for performing different tasks within the network, share physical resources of the same physical eNodeB. These VNFs can be responsible for carrying out tasks such as spectrum access, mobility, multi-connectivity, cell discovery etc.

A preliminary approach towards eNodeB virtualization included the proposal of the scheduling of the physical Resource Blocks (PRBs) between the different virtual eNodeBs [12]. This means the splitting of the frequency spectrum between the different eNodeBs of the different operators. A controlling entity called a hypervisor was proposed in order to make use of prior knowledge (e.g. user channel conditions, virtual operator contract, load etc.) to schedule the PRBs [13]. The frequency spectrum among the different operators has to be scheduled. This is the most challenging part because of the additional degree of freedom that has been added due to the fact that the scheduling could be based upon different criteria (such as bandwidth, data rates, power, interference, predefined contract, channel conditions, traffic load or a combination of them). According to these requirements, a framework is proposed where the hypervisor has to convert these criteria into a number of PRBs to be scheduled for each operator, but also need to make sure that the allocated PRBs will be fair and satisfy operators' requirements. Fig. 2-3 presents the proposed virtualized eNodeB protocol stack [12]

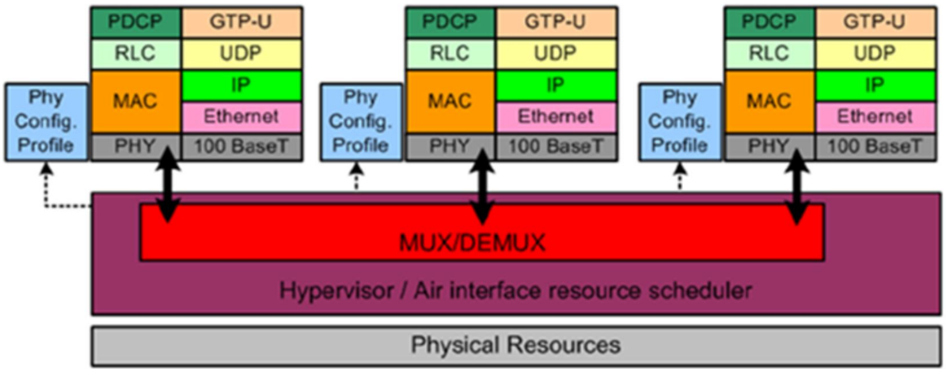


Figure 2-3: Virtualized eNodeB Protocol Stack (Reproduced from [12])

The virtualization of physical hardware in the eNode B that is responsible for transmission and reception of data from an LTE user can be performed by adding a hypervisor layer on top of the physical resources similar to any other node virtualization. This layer is then responsible for allocating and scheduling resources such as air interface or LTE spectrum between different virtual base station instances running on top of it. OFDMA is used as air interface in the downlink in LTE, which implies that the spectrum is divided into a number of sub-bands. Air interface resources are actually the physical resource blocks (PRBs) which are the smallest unit an LTE MAC scheduler can assign to the user. Different policies and scheduling schemes can be applied by the hypervisor to share the resources amongst virtual eNBs. The information such as channel conditions, QoS, priorities, traffic load, contract of each virtual operator etc. is collected by the hypervisor from each virtual eNode B and is used to schedule resources. This type of technique requires some mechanisms or guidelines to be defined, to guarantee fair and on-demand allocation of resources to the operators. The time frame that the hypervisor operates in order to guarantee the predefined requirements is crucial. For this purpose, two different types of scheduler algorithms for spectrum allocation are proposed, namely static and a dynamic. The first algorithm divides the spectrum between different virtual operators in the beginning and each operator will get a fixed share of spectrum that does not change over time. The second algorithm allocates certain amounts of resources that are fixed and doesn't change so that the operators can serve some users without interruption and a certain amount of resources are flexible and shared amongst different operators in real time and allocated to the operators based on their instantaneous traffic requirements. This work could be considered as an initial contribution in the wireless network virtualization field but may not be efficient for application in LTE-A systems as it handles PRBs in a static way.

Two use cases are considered here each with different versions of the hypervisors as explained in [12]. One use is the multiplexing use case, which exploits multiplexing gain achieved through eNB virtualization and spectrum sharing amongst different virtual network operators. Potential for achieving multiplexing gain via spectrum sharing arises from the fact that different operators experience their peak traffic load at different times. Two different versions of hypervisors could be employed in the multiplexing use case. First, the static Hypervisor, where, equal numbers of PRBs are allocated by the hypervisor just once in the beginning to each virtual operator and the allocated PRBs remains fixed regardless of whether all the PRBs are utilized by the virtual operator or not. Second, the dynamic Hypervisor, where, PRBs are dynamically allocated to different virtual operators at equal time intervals according to the load that each operator experienced over the last time instance.

The other use case is the multi-user diversity use case. This use case arises from the fact that channels are usually frequency selective i.e. each user experience different channel conditions on different PRBs. Multi-user diversity case employs a channel aware scheduler which tries to take advantage of the user channel conditions and assigns the PRBs accordingly. As virtual operators share the spectrum, this means that the scheduler has a larger pool of PRBs to exploit the multi-user diversity. This use case provides an opportunity to achieve gains by employing dynamic spectrum allocation based on user channel conditions, even when no multiplexing is allowed and each operator has a fixed amount of spectrum.

A RAN virtualization solution based on spectrum sharing is presented in [14]. An entity called The Network Virtualization Substrate (NVS) can be natively implemented in base stations. It is a feature for managing and sharing the radio spectrum and eNodeB processing resources. NEC's eNodeB product line includes the NVS feature which manages sharing of the radio spectrum and eNodeB processing resources. In the backhaul, multiple VLANs are operated and traffic shaping is performed in the eNodeB (for uplink) and the gateway (for downlink). The OAM server allows each operator's virtual network to be separately configured and managed. NVS performance is evaluated in an LTE network by means of simulation, showing that it can meet the needs of future virtualized mobile carrier networks in terms of isolation, utilization, and customization [14]. In addition to this, the authors propose a scheduler in order to efficiently assign the spectrum resources. This is basically a slice scheduler which works in collaboration with the MAC scheduler. The slice scheduler monitors the amount of resources that the MAC scheduler assigns to each slice and dynamically adjusts the bearer priorities in the MAC scheduler to maintain the required resource allocation for each operator. In this way, all operators have access to the whole system bandwidth. This work presents an efficient scheme for managing and scheduling the network resources.

A dynamic way of slicing the network resources to improve the overall network efficiency is presented in [15]. Cell slicing is one proposed dynamic framework to achieve active RAN sharing. The goal of this technique is to remotely control the scheduling decisions ensuring that each operator receives its share of the wireless resources. It does not require any modification in the BS schedulers but it controls the BS scheduling decisions from a remote gateway. Slicing can be done with either a base station-level solution or a gateway-level solution. Compared to BS-level solutions, remotely slicing wireless resources makes the solution easily deployable, enables easier network-wide resource reservations for the slices and enables multi-vendor BS operations. The BSs may or may not support virtualization. Moreover, the work focuses on remote slicing of uplink resources, since wireless resource reservation requests from the clients for enabling uplink transmissions terminate at the BS and are not visible to the gateways. Fig. 2-4 presents the proposed framework by [15].

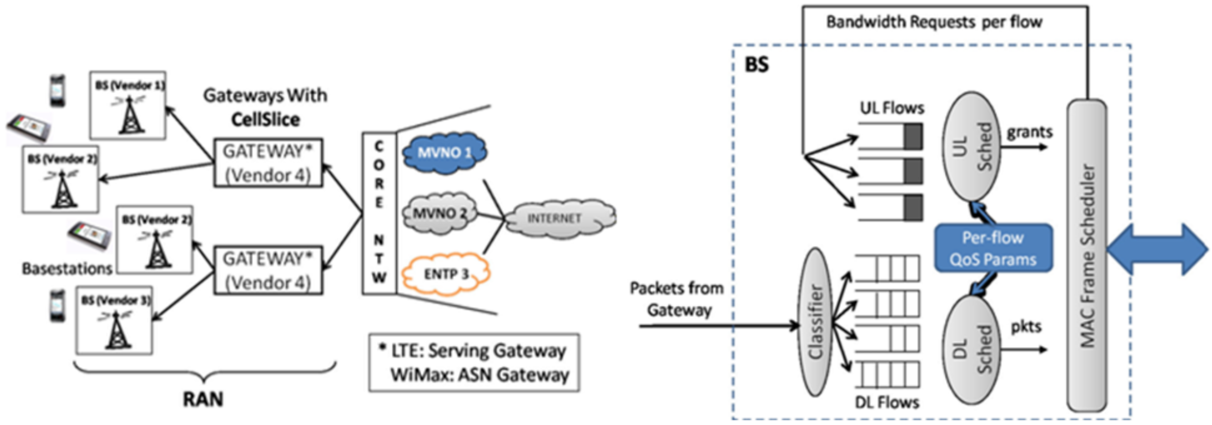


Figure 2-4: Cell Slice Proposed Framework (Reproduced from [15])

## 2.2 5G Enabling Technologies

### SDN and NFV

SDN and Network Function Virtualization (NFV) are some of the key enabling technologies for 5G systems that leverage flexibility and on-demand network programmability. Virtualization facilitates SDN and NFV operations however, virtualization have a slightly distinct meaning in SDN and NFV communities [16]. Network virtualization in an SDN paradigm involves abstraction of particular underlying network resources and their selective allocation to a particular client, application or services. While in the NFV context, virtualization refers to a software entity in a container, which is typically understood to be a Virtual Machine (VM) over a hypervisor on a Commercial Of-The-Shelf (COTS) server or dedicated hardware, i.e. a switch or base station. The term hypervisor refers to a hardware virtualization technique that allows logical functions or otherwise software instances to share a single hardware platform while appearing as individual elements with their own hardware resources such as processor, memory and hard disk.

Currently, mobile networks are rigid in nature formed by monolithic network functions that enable firm services, which are difficult to be adjusted and customized. The application of virtualization in the evolving 5G networks aims to transform the way networks are managed and operated, enabling rapid deployments and testing of innovative services. Specific network entities or elements may also be virtualized and implemented onto industry standard high volume servers, switches and storage, which could be dynamically placed in different network cloud locations, Network Point of Presence (N-PoP), Network Nodes and in the end user premises. The SDN architecture as defined in [6] [17]consists of three different sets of functional layers illustrated in Fig. 2-5. The SDN application layer allows SDN applications such as analytics, optimization algorithms, management mechanisms and network control etc. or 3rd parties to communicate service and resource requirements to the SDN controller via the Application-Controller Plane Interface (A-CPI). The control layer contains the SDN controller, a logical entity with a global network view, which allows SDN applications and 3rd parties to control and program the network via the Data-Controller Plane Interface (D-CPI) based-on an abstracted network resource view that hides insight information. The SDN controller separates the control from the data plane, i.e. decoupling the control from traffic forwarding and processing and collect network statistics. The infrastructure layer involves the network devices that control the forwarding and data processing, i.e. the data plane.

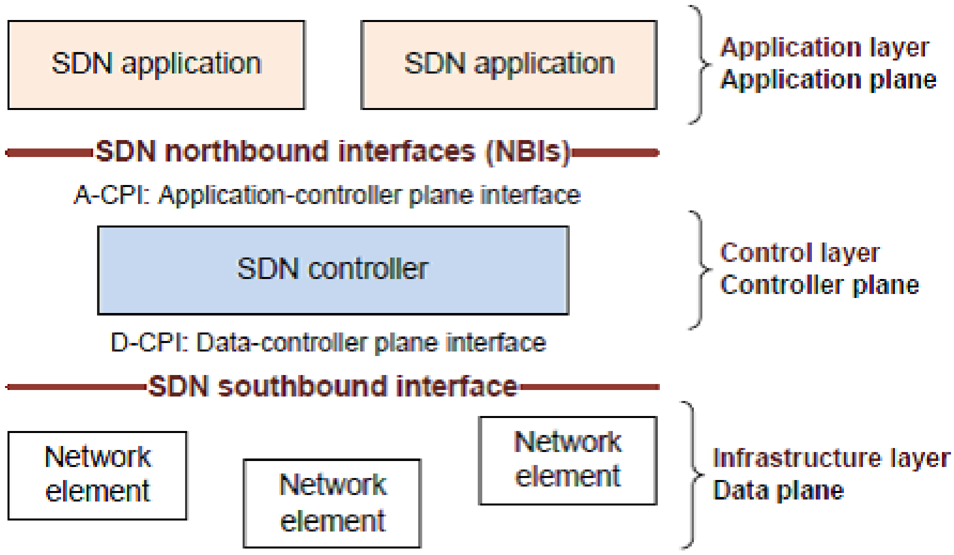


Figure 2-5: Overview of the SDN Architecture [17])

SDN enables software based flexible control and on-demand programming of the mobile network infrastructure which includes business relationships, geographic span and everything from service creation and delivery to operation and maintenance. SDN also provides standard interfaces that enable application and service providers including vertical markets and Mobile Virtual Network Operators (MVNOs) to communicate their service requirements to the infrastructure provider. This in turn can program the network resources to meet the Quality of Service (QoS) constraints [5] [17].

According to [6] and [17], SDN is based on the following three principles:

- Separation of the control and data planes including decoupling of control from traffic forwarding and processing
- Logically centralized control and governance of the network considering a global view and over all implications on the network of the applied policies or configuration
- Open interfaces between control and data plane entities
- Programmability of network services and the network by external entities via standard interfaces.

The use of SDN allows resource flexibility supporting customized services and logical network instances on the top of a common network infrastructure, meeting the service constraints of application providers, vertical segments and virtual operators. In the context of mobile networks, SDN can assure a unify control across heterogeneous radio infrastructures and dense RAN deployments [5], while providing a join control of mobile and transport network layers [5]. It is anticipated that the adoption of the NFV paradigm for both access and core networks have the potential to enhance further the service flexibility by scaling appropriately Virtual Network Functions (VNFs) for supporting efficiently the 5G services [18]. For an SDN controller, VNFs appear as another type of resource [16], which forms a network graph that the SDN controller can chain together forming a particular service [19]. Hence, SDN and NFV are complementary technologies that can be used together forming tailored services considering evolving traffic conditions, mobility patterns and target QoS demands.

To manage and optimize the RAN in a flexible way, as previously discussed in this chapter, SDN and NFV seems promising as an enabling technology. SDN can be used to dynamically program the network in real time to match the traffic demand and support multiple heterogeneous applications/services with often conflicting requirements in a cost efficient way. The SDN paradigm may also be used for monitoring network resource utilization and allowing applications or services to request resources based on certain service level agreements (SLAs). The SDN concept aims to create an open, programmable network architecture that separates network control and data plane. Within the SDN paradigm, the network elements can be virtualized and software based network functions referred to as network function virtualization (NFV) technology are created that allows carriers to easily deploy them on demand and easily manage the network. This provides additional flexibility and allows carriers to program the network to support the requirements of various applications and services. SDN, with the help of virtualization, masks the lower layers of the protocol stack. This reduces complexity and allows efficient configuration of upper layers. It is faster to test and deploy new services and business models with the help of SDN technology.

The evolution of the 5G mobile network architecture aims to integrate different technologies to enable a myriad of diverse use cases. Some of the use cases may have requirements that may cause conflict with the functional requirements of other use cases. This necessitates the need to provide the required functionality at the right place and time within the network. It is anticipated that the adoption of the NFV paradigm for both access and core networks has the potential to provide the flexibility required to support the 5G use cases. NFV enables the mobile network functions to be decomposed into smaller functional blocks that may be instantiated flexibly on-demand without the need for installation of new physical network equipment. NFV refers to a network architecture concept where in virtualization technologies are applied to virtualize entire classes of network node functions into building blocks that may connect, or chain together, to create communication services and support a number of use cases.

NFV offers several benefits which include [20]: Reduction in costs of the equipment and reduced power consumption as physical equipment is consolidated with the help of virtualization technology and the specific tasks associated with particular equipment is realized via software implementation over the virtualized platform. This Shortens the time to market, minimizing the conventional network operator cycle of innovation. Capital investments in hardware-based functionalities that also contribute in capital expenses (CAPEX) are not required in case of the software based development. Network Functions Virtualization enables network operators a significant reduction in time of the maturation cycle of a particular technology. It also helps in rapid test and deployment of new network services and business applications. NFV helps in supporting and facilitating multi-tenancy, which allows use of a common infrastructure platform for different applications, users and tenants. This allows network operators to share resources across different services and across multiple customers including vertical markets. It allows rapid service or application deployment and testing which includes scaling up and down certain services as and when required. The targeted delivery of services to a particular area or group of users can also be achieved with the help of SDN and NFV. It enables the development and deployment of a wide variety of innovative business models and ecosystems and opens up market for vertical players such as M2M, IoT services, smart city, V2X etc. and Mobile Virtual Network Operators (MVNOs). It also encourages rapid development of new services and revenue streams at a lower risk.

NFV enhances SDN service agility and improves its ability to rapidly create, scale or relocate virtual resources. As per [16], Virtualized Network Functions (VNF) appear as resources to an SDN controller that involves, functions in a network graph with known connectivity points and known and a controllable transfer function. Both SDN and NFV technologies can be deployed independently however; both technologies can be combined and used together to achieve greater benefits. SDN and NFV support multi-tenancy that enables network operators to provide tailored services to several users or different applications, MVNOs are referred to as tenants. It also allows co-existence of multiple services over the same physical infrastructure with secure isolation among administrative domains. The flexibility offered by SDN and NFV technologies allows optimization of network configuration and/or topology in near real time based on the actual traffic conditions/user mobility patterns, for individual services to match their desired performance demands. SDN and NFV help in reducing overall network energy consumption by adopting advanced power management techniques to help leverage efficient workload management and location optimization. Consequently, the benefits offered by SDN and NFV improves the overall operational efficiency of the system and facilitates inter-operability and backward compatibility among multiple technologies.

The use of SDN/NFV can also enable network slicing, defining logical self-contained networks that can accommodate different business requirements, on the top of a common physical network infrastructure [21]. The SDN controller can play the role of mediator facilitating admission control and network slice allocation offering programmability to third parties [22]. Network slicing can assure isolation between different services and tenants, allowing an efficient co-existence of heterogeneous applications with often conflicting requirements. Currently, 3GPP is performing a study named NexGen [23] within the Architecture Group SA 2, considering various slice operations in the data and control plane for the next generation of mobile networks.

### Network Slicing

The concept of network slicing refers to the deployment of multiple dedicated logical mobile networks with varying levels of mutual isolation over the same physical infrastructure. A network slice may comprise of one or multiple mobile virtual network functions (that may also be grouped together) and specific Radio Access Technologies (RATs) (or specific RAT settings) necessary to operate an end-to-end logical mobile network in an independent way. The network functions and the corresponding settings are grouped in a way that the control and data plane functionality associated with a specific slice may be adapted to support various services, users and business cases. Thus, as a result, network slicing technology is an enabler of multi-tenancy and service tailored mobile networks.

An Innovative 3GPP EPS mobile network architecture and the need to develop a flexible architecture that integrates different technologies and enables diverse use cases is discussed in [24]. Furthermore, the authors introduce and explain various concepts related pre-defined grouping of network functions in a flexible way, software-defined mobile network control, orchestration and management. The relevance of different standards defining organizations has been outlined and their roadmap is discussed. A comprehensive overview of the 5G RAN design guidelines, key design considerations, and functional innovations as identified and developed by key players in the field is discussed in [25]. The key functional design considerations for the 5G RAN, highlighting the difference to legacy systems such as LTE-A and the implications of the overall RAN design is also explained. The article also explains the logical architecture, mapping between logical and physical network architecture and methods to support network slicing. In addition, the authors propose functional paradigm changes to concepts such as a beam-centric design, lower-layer service prioritization, traffic steering in 5G, and a novel RRC state model etc. A radio resource slicing framework called Hap-SliceR for Haptic communications is proposed in [26]. The authors use a resource learning approach in a virtualized environment to derive a network-wide radio resource strategy. To improve efficiency, a Post-Decision State Learning (PDSL)-based approach is used. System level simulations have been performed to evaluate the performance of the proposed algorithm with the State-of-the Art (SoA) solutions. A network slicing concept applied to a multi-cell RAN shared among multiple tenants is analyzed in [27]. Network slicing in RAN is challenging due to varying radio channel characteristics. In case of a multi-cell, multi-tenant environment, RAN slicing can be challenging and requires isolation among tenants. The article discuss two perspectives of isolation i.e. traffic isolation and the radio-electrical isolation. The authors have addressed the RAN slicing problem from a comprehensive perspective incorporating these two concepts and have presented four possible RAN slicing approaches that differ in the RRM functions used as a support for splitting the radio resources among slices. The presented alternatives have been compared along different dimensions (for example in terms of offered traffic in one tenant vs SINR status of the other tenant), both qualitatively and quantitatively. [28] presents an overview of the 3GPP standard evolution which includes network sharing principles, mechanisms and architectures to future on-demand multi-tenant systems. The article reviews the latest standardization efforts and discusses the open challenges for enabling network slicing considering the allocation of virtualized network functions based on ETSI NFV, the introduction of shared network functions, and flexible service chaining. In addition, the authors introduce the concept of 5G Network Slice Broker in 5G systems, which allows mobile virtual network operators, over-the-top providers, and industry vertical market players to request and lease resources from infrastructure providers dynamically via signaling means. A survey on Resource Slicing in Virtual Wireless Networks is presented in [29]. The authors define and discuss, in detail, the problems and challenges associated with the network slicing in wireless networks. A review of the latest work in this area is presented that analyzes the relation of SDN and NFV with network slicing. The authors highlight the network slicing problems in relation with the wireless networks such as virtualization of wireless resources, resource isolation and allocation. Following this, the authors review existing proposals for predominant wireless technologies. Challenges and improvement required for realizing network slicing are also discussed.

Thus, from the state-of-the-art, it is clearly reflected that technologies such as SDN, network virtualization (NFV), network slicing, multi-connectivity etc., would be key enablers in developing flexible and programmable 5G networks. Such adaptive future networks are required to support high performance, efficient and independent business-driven logical networks on a common physical infrastructure.

## 2.3 TD-LTE Overview

3GPP LTE has been designed to accommodate both paired spectrum via Frequency Division Duplexing and unpaired spectrum via Time Division Duplexing (TDD). As per the current 3GPP standards, TDD utilizes the same subframe structure and configuration protocols as the FDD. Compared to FDD, the main difference lies in the fact that TDD supports unpaired spectrum, where, transmissions in uplink (UL) and downlink (DL) are separated in time domain. The fact that unpaired spectrum is easier to acquire makes the TD-LTE deployment even more feasible. As per [30], the ITU is considering a spectrum allocation for IMT-Advanced in the 698-803MHz, 2300-2400MHz, 2500-2600MHz and 3400-3600MHz bands, where a large chunk of unpaired spectrum is expected to be allocated for TDD.

Each TDD frame is 10ms long and consists of DL, UL and special (S) sub-frames with 1ms duration each as illustrated in Fig. 2-6. The S sub-frame is used for switching from DL to UL transmission direction and contains a conventional DL part named the Downlink Pilot Time Slot (DwPTS), a Guard Period (GP) of a blank gap that assists the User Equipment (UE) to switch from the DL to UL and an Uplink Pilot Time Slot (UpPTS) part which carries the synchronization information, that facilitates the UE in establishing UL connectivity. The S- sub-frame is included at least once within each 10ms frame or in certain UL/DL configuration cases this is included twice, i.e. once every 5 ms.

3GPP has defined 7 different TDD frame configurations as shown in table 2.1 [7], with the UL/DL portion of each frame configured accordingly, reflecting estimated traffic demands. Such a TDD feature allows resource configuration flexibility that helps in supporting various asymmetric applications such as sensor measurements and IoT, social applications with high UL demands, etc., besides the conventional video

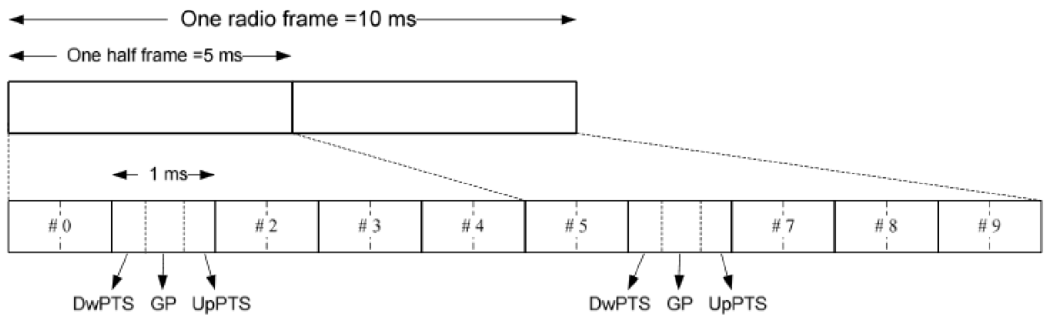


Figure 2-6: Overview of the TDD Frame Structure

streaming and Voice over IP (VoIP).

Table 2.1: TD-LTE Uplink-downlink Frame Configurations

| UL/DL Configuration | Sub-frame Number | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 0 | DL | S | UL | UL | UL | DL | S | UL | UL | UL |
| 1 | DL | S | UL | UL | DL | DL | S | UL | UL | DL |
| 2 | DL | S | UL | DL | DL | DL | S | UL | DL | DL |
| 3 | DL | S | UL | UL | UL | DL | DL | DL | DL | DL |
| 4 | DL | S | UL | UL | DL | DL | DL | DL | DL | DL |
| 5 | DL | S | UL | DL | DL | DL | DL | DL | DL | DL |
| 6 | DL | S | UL | UL | UL | DL | S | UL | UL | DL |

TD-LTE offers the capability to handle asymmetric UL/DL traffic in a dynamic manner, which is also referred to as dynamic TDD. It allows eNBs to select an UL/DL frame configuration out of the seven frame configurations defined by 3GPP as specified in [7], which supports best their corresponding UL/DL traffic load at that time. This way, each eNB may have a different UL/DL frame configuration according to their UL/DL traffic demands at a particular time. Having different UL/DL frame configurations across the RAN may give rise to different forms of interference therefore, interference management is a key functionality for efficient operation of a dynamic TDD system. Beside its inherent flexibility, the first generation of TD-LTE networks supported a static synchronous configuration where, a set of base stations serving a greater geographical area follow a TDD configuration with a common UL/DL ratio that suits the overall average long term traffic demands. Such synchronous operation simplifies interference control assuring no cross-slot interference among UEs served by neighboring base stations or direct base station interference as illustrated in Fig. 2-7.

Static TD-LTE operates in this way mainly to avoid base station-to-base station interference or cross-slot interference that prohibits the usage of transmissions in opposite direction within different cells in conventional macro cell deployment scenarios. Hence in this case there are only two type of interference, namely user equipment (UE) to eNB and eNB to UE. The UE-eNB interference occurs as a result of the UL transmission of UEs served by the neighbor eNBs on the UL transmission of UEs of the serving eNB. The eNB-UE interference affects the DL transmission of the serving eNB due to the DL transmission of the neighbor eNBs. The primary reason for synchronous operation of the TD-LTE system is to avoid base station-to-base station interference. Since interference could severely degrade the overall system performance, [31] specifies strict synchronization requirements for TD-LTE between base stations with overlapping coverage areas.

## 2.4 Dynamic Time Division Duplex Networks

To alleviate the rigid attributes of the static TD-LTE deployments, 3GPP initiated a study in [32], to investigate dynamic TDD focusing on small cells considering scenarios of isolated small cells, various small cell deployments with limited interference or by using almost blank frames. Such a study lead to the specification of the TDD enhanced Interference Management and Traffic Adaptation (eIMTA) [7], based-on power control. Other efforts for achieving dynamic TDD also took place around the same time, with the Cell Specific Dynamic Reconfiguration [33] being one of the most significant.

Cell Specific Dynamic Reconfiguration in TD-LTE, allows customized UL/DL frame selection to satisfy the traffic demands in UL and DL respectively. In other words, each base station or evolved Node B (eNB) in the 3GPP terminology, can dynamically change the UL/DL configuration at specific timescales, in order to match best the instantaneous traffic demands. Thus, in this case, the interference scenario becomes more complex as in addition to UE-eNB and eNB-UE there is also UE-UE and eNB-eNB interference. Each eNB is allowed to dynamically change its UL/DL configuration at specific timescales, in order to best match its instantaneous traffic demands. Therefore, the proportion of numbers of UL/DL timeslots varies across the entire RAN. As a result, regions of cross UL/DL transmission may occur among neighbor cells leading to an interference called cross-slot interference or base station-to-base station (eNB-eNB) interference. The UE-UE interference is due to the UL transmission of neighbor cell UEs over the DL transmission of the serving eNB. eNB-eNB interference is caused by the DL transmission of the neighbor eNB when the UEs in the serving eNB are in their UL transmission cycle as shown in 2-7. In Fig. 2-7, the DL transmission in eNB A causes interference on the UL transmission of eNB B and the UL transmission in eNB B causes interference on the DL transmission of eNB A, where the cross slot interference scenario is depicted by red arrows.

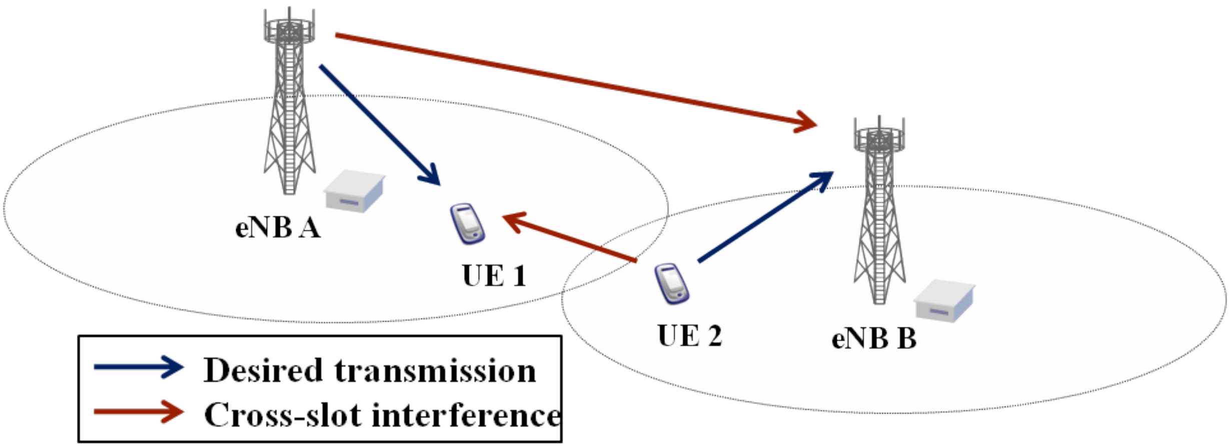


Figure 2-7: Cross-slot Interference

Among the UE-UE and eNB-eNB interference the eNB-eNB interference is considered more serious as the DL transmission power of eNB is much higher than the UL transmission power of the UEs. Therefore in case of the cell specific dynamic UL/DL reconfiguration scenario the additional UE-UE and eNB-eNB interference may lead to performance degradation and hence limit the flexibility offered by the of cell specific dynamic UL/DL reconfiguration. The issue of interference caused by neighboring cells can be handled via deployment and cell clustering means or by the use of almost blank frames and power control. Almost blank sub-frames allows the base station downlink power to be completely blanked or muted in specific sub-frames where cross slot interference is likely to occur, but this may impact the spectral efficiency.

As interference could severely degrade system performance, therefore to minimize the loss in system performance [31] specifies strict synchronization requirements for TD-LTE, between eNBs that may have overlapping coverage areas. The length of the guard period defined within special subframe [34] in TD-LTE can be dynamically adjusted to ensure that any base station within close proximity does not interfere with the uplink reception in the serving cell. This affects the UL/DL frame reconfiguration process and the uplink-downlink frame reconfiguration is selected based on long term traffic demands across the entire TD-LTE network and does not change frequently. To further address the issues related to interference, advanced interference control and management schemes have been proposed and interference management in dynamic TDD systems is being studied to further improve the overall performance and enhance the flexibility. For example, to manage interference in a dynamic TD-LTE system a hybrid approach is presented in [35], which combines the static and dynamic resource allocation area. The dynamic allocation area is negotiated among the neighbor eNBs by selecting and indicating UL/DL sub-frames, while the cross slot interference is avoided by introducing power control techniques. Other techniques include cell clustering and the use of almost blank sub-frames [33]. A cell cluster may consist of one or more cells and the isolation is guaranteed based on certain parameters to avoid the interference between multiple cell clusters. The transmission directions within the cells associated to different cell clusters maybe different. Cell clustering provides an additional benefit to control UE-UE interference as UEs connected to serving cells belonging to different clusters are likely to be isolated from each other. The use of an almost blank subframe (ABS) is an extreme method in which the base station downlink power is completely blanked or muted in specific subframes where cross slot interference is highly likely to occur. Such a method is useful in co-channel heterogeneous network deployment scenarios, e.g. where macrocell and small cells are deployed on the same carrier frequency. In this scenario, several subframes are muted in the macro cells with the aim of mitigating the interference on the small cells caused by the macrocells. Thus, the performance of the small cells can be improved but this may impact the spectral efficiency.

Alternatively, traffic offloading from macro to the small cells can also provide overall gain in the system performance. Asymmetric UL/DL subframe allocation is also investigated in [31] which exploits the backhaul resources in an efficient way with the aim of reshaping the interference on the channels. In [36] and [37], the authors considered the buffer status to flexibly switch between different TD-LTE frame configurations, introducing extra UL/DL resource diversity for the eNB's inner cell regions. The UL/DL sub frames can be reconfigured by updating the broadcast system information. This process depends on the general system information update procedures, in which a paging indication is used to notify users about an upcoming configuration change. As explained in [33], the system information update procedure has some drawbacks when the UL/DL frames are modified to support the instantaneous traffic conditions. These limitations include:

- According to the standards, the modification time to successfully notify all the users about the upcoming changes is restricted. As a consequence, the reconfiguration timescale is restricted and cannot be done more frequently than every 640 ms [38].
- An ambiguity period is present, from the system information modification boundary until the time when a particular UE has successfully received and decoded the new system information. The ambiguity period causes UL/DL frame reconfiguration knowledge to become unsynchronized between the network node and the UE. Consequently, this leads to a discontinuity of the Hybrid Automatic Repeat reQuest (HARQ) processes and hence a degradation in throughput.

To perform the UL/DL frame reconfiguration more frequently and flexibly, alternative methods need to be explored. For example, UE-specific higher layer signaling can be used instead of a system information update to indicate a new UL/DL frame configuration which can be initiated by the scheduler. This involves sending a downlink control message, i.e. a radio resource configuration (RRC) message or medium access control (MAC) element to the target UE which then updates it UL/DL configuration. RRC and MAC messages can be transmitted more frequently depending on the system requirements, potentially as frequently as every few tens of milliseconds. Therefore, this approach opens up the possibility of having different configurations for different users. It also allows for faster UL/DL reconfiguration timescales which helps in improving adaptation to the instantaneous traffic demands. The HARQ discontinuity may still operate in the same way as with the system broadcast reconfiguration mechanism. However, since the RRC and MAC messages are acknowledged, the ambiguity period is shorter compared to the system broadcast reconfiguration mechanism.

The ambiguity period in this case is mainly caused by the feedback delay to transmit the acknowledgement. Some additional methods and ideas to perform UL/DL frame reconfiguration have been presented in [33].

The HARQ timeline is also affected when the UL/DL reconfiguration is proposed for 5G TDD networks. HARQ is an effective physical layer mechanism that is used to improve the reliability of the transmission in the wireless environment that varies with time. It involves explicit feedback indication, when the data transmitted has been successfully received. Usually, multiple HARQ processes are active in parallel to the successfully transmitted data in each TTI. One HARQ process includes transmission of data in one TTI and its corresponding acknowledgement signaling. In TD-LTE, asynchronous HARQ is used for downlink data transmission over physical downlink shared channel (PDSCH) and synchronous HARQ for uplink data transmission on the physical uplink shared channel (PUSCH). With different UL/DL configurations in 5G TD-LTE, the number of HARQ processes and the HARQ timeline also varies. Hence, as the UL/DL frames are reconfigured, the HARQ timeline is also changed. Some of the active HARQ processes may be disrupted at the boundary where the HARQ timeline changes and cause data rate degradation due to reduced transmission opportunities for HARQ feedback or scheduling grant. In existing TD-LTE systems, the cell specific dynamic UL/DL reconfiguration occurs at a slower rate or longer timescales, hence the impact of the HARQ timeline change on performance of the system is negligible. With faster reconfiguration timescales of the UL/DL subframes and enhanced flexibility in 5G TDD, the impact of the HARQ timeline change may become significant. It is therefore imperative to ensure that the faster UL/DL reconfiguration timescales do not affect every HARQ process. This is necessary to maintain the performance benefits offered by the dynamic UL/DL reconfiguration mechanism. To achieve this, a totally new HARQ timeline may be designed, however it is not feasible to design a totally new HARQ timeline as it requires significant effort in terms of implementation, specification and standardization. Therefore [33] suggests exploring the possibility of using one of the existing HARQ timelines irrespective of the UL/DL frame configuration. They conclude that since the UL/DL configuration # 0 or # 5 is the most uplink or downlink favored configuration, its corresponding PUSCH or PDSCH HARQ timeline is suitable as the unique PUSCH or PDSCH HARQ timeline for TD-LTE cell specific dynamic frame reconfiguration systems with faster reconfiguration timescales.

Recent advancements on TD-LTE investigate mechanisms that aim to optimize the UL/DL ratio selection for each eNB, considering a variety of different constraints. An asymmetric assignment of UL and DL is investigated in [36], which aims to reshape interference channels exploiting efficiently the available backhaul resources. In [39] a cooperative decentralized mechanism is introduced that provides a local optimal solution considering instantaneous data rate conditions and traffic demands, which are exchanged via reliable low-rate signaling among neighboring cells. An alternative approach that aims to minimize the information exchange among neighboring cells by enabling eNBs to perform autonomously an UL/DL ratio optimization based-on a game theoretic method is introduced in [40], considering UL/DL delays in relation with traffic load, interference and flow-level dynamics. Optimizing further the selection of the TDD frame considering the evolving Quality of Service (QoS) demands in terms of bit rate guarantees and packet delay is elaborated in [41]. The potential gains in throughput and reduced packet delays once each TDD slot is freely assigned as UL or DL instead of using one of the 7 pre-determined TDD frames is explored in [42], assuming the use of Interference Rejection Combining (IRC) capable receivers to handle cross-slot interference. A link-proportional dynamic channel assignment (LP-DCA) scheme is proposed in [43]. It uses directional antennas to alleviate the impact of cross-slot interference in a TDD/code division multiple access (CDMA) system. In this scheme, the users within a sector are grouped according to their received link quality and the LP-DCA assigns the time slots with higher potential cross-slot interference, to the group of users with better link quality, or vice versa. The LP-DCA scheme aims to reduce the cross-slot interference and improve the system capacity. In [44] the authors study the performance of Internet access using the TCP protocol in different downlink-heavy asymmetries. Simulations are performed to show that the performance depends on many factors such as the transferred file size, the control channel errors and the downlink/uplink traffic mix. The authors explain that the potentially higher downlink capacity of the TDD may not be fully used when the file size is small. This is due to longer uplink access delays as well as shortage of uplink resources in the chosen configurations. However, as the file size increases, this effect disappears and TDD provides higher bit rates in the downlink than FDD. But in this case, the realized increase in bit rate is not as high as the calculated increase in available downlink resources.

Such proposals consider individual eNBs focusing on the means of providing a dynamic UL/DL ratio in an autonomous way considering a various optimization parameters. However, unlike these cell centric approaches, the concepts and techniques developed within the framework of this thesis stretch beyond a single cell, i.e. local optimization, exploiting efficiently the resource diversity within overlapping cell areas as elaborated in [45]. This enables UEs to utilize sub-frames from multiple eNBs, i.e. forming virtual cells that offer customized TDD frames, which match best their UL/DL traffic demands. Performing radio resource management that combines dynamic TDD re-configuration with virtual cells in a distributed manner is suboptimal. Hence, centralized intelligence is recommended for efficient interference mitigation considering traffic dynamics. The detailed analysis of the Virtual Cell concept is provided in chapter 3.

A similar approach referred to as V-Cell, offers a combination of heterogeneous radio resources, i.e. macro-cells, picos and femtos, as a resource pool to UEs, which perceive such an access as a logical single macro-cell [46]. The network resource management is performed by an SDN controller, which maintains a logical global view of the underlying network in order to efficiently schedule resources across the entire pool of physical radio elements. An equivalent SDN-based Radio Access Network (RAN) management architecture that relies on abstracting the RAN resources and using them as a single virtual wireless access is also analyzed in [47], which also provides more details about the SDN controller including the main associated functions and the control plane mechanisms. Such components enable RAN programmability as considered in [5], which analyzes the SDN impact of separating the control and data planes in simplifying the management of heterogeneous networks, considering mobility and QoS-aware network operation. In fact, the SDN architecture offers an API called N-API to allow communication between application providers and the infrastructure providers to facilitate on-demand QoS provision. In this thesis, a similar SDN paradigm is employed offering RAN programmability, which is realized by enabling a unified control for provisioning a dynamic TDD frame configuration at selected eNBs with the potential of forming virtual cells based-on the user resource demand. Whilst our focus is on network programmability based-on network measurements and resource utilization, the proposed SDN architecture may potentially offer a customized UL/DL ratio upon a request via the N-API, allowing applications to program the network and provision resources for particular services within certain times. Such a feature may advance the network customization offering flexibility, scalability and ease deployment of new services and enhance performance of TD-LTE systems.

The virtual cell concept is also employed in [48], which adopts the SDN paradigm to enhance the user QoE considering the application performance characteristics in a pico-cell environment. In particular, the QoE assessment function at the eNB contacts the SDN controller when it identifies users that suffer a QoE degradation, which can benefit from utilizing resources from multiple eNBs forming virtual cells. Existing telecommunication networks have intrinsic Quality of Service (QoS) provisioning and measurement mechanisms to evaluate the network performance. However, as the telecommunication networks are evolving with the aim and inevitable need to support emerging future mobile applications/services, it has been identified that the system-centric view of QoS provisioning is no longer sufficient [49]. Thus, the QoS provisioning and measurement mechanisms needs to be replaced or complemented with more user-centric approaches. In this direction, QoE seems a potential solution that complements the QoS mechanisms. It helps in assessing the acceptability of an application or service, as perceived by the end-users, delivered via the network to the end-users. This way, the QoE can help in assessment of the performance of both the network as well as the application or service provided to the end-users. Further details related to this idea are discussed in chapter 4.

QoE can be monitored by using subjective or objective techniques. As discussed in [50], quality of service (QoS) is the connection between network performance and application requirements, while QoE is the connection between application performance and the user. One of the most common methods for subjective determination of the QoE consists of computing the Mean Opinion Score (MOS) in line with the International Telecommunication Union (ITUT) Rec. P.800. Nonetheless, measuring the MOS of speech quality is rather complicated, given that MOS is averaged over a large number of user opinions. MOS can be measured using objective techniques as well, such as the E-Model and the Perception Evaluation of Speech Quality (PESQ) model

[50]. The PESQ model estimates MOS by comparing a reference signal with the received (degraded) real signal. On the other hand, the E-Model can be readily used to estimate MOS in real-time. The work done in this thesis, adopts the E-Model to assess the QoE of individual users in real-time and adapt the UL/DL frame ratio depending on the current user requirements. Aiming to meet the individual QoS requirements of the end users (i.e. in terms of throughput and packet delay), the authors in [41] have proposed a dynamic cell specific UL/DL frame reconfiguration mechanism.

In [51] the authors propose a downlink scheduling method to improve the QoE for VoIP traffic in LTE networks. To achieve this, they consider that the scheduling decision is made based on feedback provided by the users. The authors in [52] present a QoE-based framework for network planning and propose an analytical method based on queueing theory to dimension access networks such as LTE. Different from [51] and [52], we propose an SDN-based network management architecture and solution based on the formation of virtual cells and the flexible adaptation of the UL/DL ratio at the TD-LTE picocell. Besides, as shown in the numerical results, the proposed framework improves the QoE of users suffering from service quality degradation without affecting the performance of others. In [53], the authors propose an architecture that uses a central QoE server that collects performance indicators from different network elements and takes actions to improve the QoE of particular users. Decentralized data offloading mechanisms are proposed in [54], to improve end-to-end delay and enhance the MOS required for service continuity.

In [55] the authors propose a radio resource management (RRM) strategy, which dynamically reallocates resources assigned to users with adverse channel conditions towards other users aiming to improve their MOS. To accomplish this they consider two expulsion criteria: (i) the direct expulsion of unsatisfied users that underutilize radio resources and (ii) the expulsion of best-effort users that can tolerate service delays. In HetNets nowadays, the dense yet unplanned deployment of small cells results in unbalanced utilization of the physical resources among the cellular infrastructure. Even though a specific subset of cellular stations can be overloaded, other nearby cellular stations may underutilize their dedicated resources, due to the irregular spatiotemporal variations of the user traffic or the selected service provider. In such situations, Radio Access Network (RAN) and spectrum sharing can considerably improve the area spectral efficiency of the current cellular networks. For example, multiple mobile network operators can share their infrastructure so as to reduce the amount of active network equipment, e.g. base stations, or reduce the capital investments for setting up the LTE-A RAN infrastructure. In two recent studies, 3GPP has overviewed the service and business requirements for realizing the so-called network sharing paradigm [56] and highlighted the required architectural enhancements [57].

Current literature includes a notable amount of frameworks, architectures, and mechanisms that allow for effective sharing of radio-resources or physical network components that can reside in either the access or the core network. The solution in [58] aims at providing on-demand infrastructure and spectrum sharing among different operators in 3GPP LTE networks. The authors propose mechanisms for virtualizing the evolved Node B (eNB) hardware by creating logically independent base stations, a.k.a. virtual eNBs (VeNBs). Among others, the proposed mechanisms are shown to balance the traffic load among the eNBs involved in the sharing process. The authors in [59] propose a multi-tenant solution that enables resource isolation and coexistence of independent policies among different eNB instances. To achieve this, they propose a two-layer resource-scheduler composed of a global and a local resource scheduler that permits implementation of different scheduling policies to different eNBs. A similar approach is discussed in [12], which modifies the scheduler of a shared eNB to isolate the traffic between multiple operators, while achieving a multiplexing gain. The key component in [12] is an entity called the Hypervisor, which virtualizes the eNB into a number of slices allocating the Physical Resource Blocks (PRBs) among multiple virtual operators according to existing agreements. Although the results in [58]-[60] are promising, further work is required to adapt the presented solutions in a HetNet environment, where different functional capabilities are supported by the cellular base stations.

An SDN-based framework is proposed in [61], to enable efficient on-demand sharing of base stations that belong to different operators. The key idea is to allow the cellular users to attach to the nearest base station. The OpenRAN architecture is introduced in [62], to leverage the convergence benefits of HetNets and achieve, at the same time, customization via network programmability. Network virtualization (NV) is realized at different levels including the application, spectrum, and network level, achieving a higher granularity of the combined pool of network resources. The use of the SDN paradigm as a tool to simplify cellular network management is analyzed in [63], where a hierarchical architecture of a local and a global SDN controller is proposed. The local controller manages the processes inside a single network, while the global controller handles the events throughout the HetNet and coordinates the RANs through the backhaul.

The problem of re-configuring the TDD frames of pico eNBs has also been considered in [33] [64] [36]. It is envisioned that both FDD and TDD technologies will co-exist in the future mobile networks. Considering this, the design and implementation challenges of FDD and TDD systems and the comparison of their pros and cons are discussed in [65]. It presents a number of advantages and flexibilities that a TDD system can offer compared to the FDD system in 4G systems. The authors identify some major challenges in deploying the TDD in cellular networks including also the cross-slot interference. A quantitative analysis is also provided that discusses the impact of cross-slot interference on co-channel and adjacent channel interfering cells. Finally, the authors conclude that the application of sectored antennas and time slot grouping are very effective to alleviate cross-slot interference. An SDN framework for elastic resource sharing among a Frequency division Duplex (FDD) macro cell network and TDD pico cell network that aims to enhance flexibility and network resource management efficiency is described in [66]. This paper also adopts an SDN network management paradigm, but instead it concentrates on network programming aspects, i.e. forcing a TDD ratio re-configuration for providing TDD virtual-cells, enhancing the user service performance and the network resource utilization. This work has been further discussed in more detail in chapter 4.4 .

# Chapter 3

# TD-LTE Virtual Cells: An SDN Architecture for User-centric Multi-eNB Elastic Resource Management

## 3.1 Introduction

Time Division Duplex (TDD) is one of the two variants of 3rd Generation Partnership Project (3GPP) LTE. TDD utilizes the same radio access scheme as the Frequency Division Duplex (FDD), and uses the same sub-frame format as well as the same configuration protocols [7]. The main difference compared with FDD, is the support of unpaired frequency bands, where downlink and uplink are separated in the time domain. In TDD each frame is composed from downlink (DL), uplink (UL) and special (S) sub-frames, which are used to switch from DL to UL and they are included at least once within each frame. The UL/DL portion of each frame may be dynamically configured to adapt to the instantaneous traffic conditions according to the specification provided in [7], which defines 7 different UL/DL configurations as shown in table 2.1. Such resource flexibility and diversity is particularly useful for emerging applications such as video streaming, V2X or Machine Type Communication (MTC), etc. that are highly asymmetrical.

Despite such UL/DL resource flexibility, the initial deployment of TD-LTE involves a synchronous frame configuration across certain network regions with all the cells offering an identical UL/DL ratio in order to avoid cross-slot interference [33]. Such network arrangements effectively limit the advantage of UL/DL resource flexibility since it imposes restrictions among neighboring cells and therefore prevents operators from fully exploiting the benefits offered by the TD-LTE. Such limitations are associated mainly with (i) interference when neighboring cells adopt a different UL/DL configuration, restricting the degree of resource flexibility and (ii) admission control when particular overloaded slots may introduce pseudo congestion which arises in the case when adequate resources exist but in the opposite transmission direction [67]. In response, the 3GPP focused on a set of enhancements to address interference and accommodate traffic adaptation [32]. These effort introduce cell specific mechanisms in the sense that UL/DL configuration change takes place in each cell individually in a distributed manner with the goal to optimize uplink and downlink network resources to best serve the local traffic demands, assuming that user equipment (UE) is associated with a single cell at a given time.

Thus, the proposed concepts in this chapter, aims to address the limitations and resolve the issues that are emphasized in the state-of-the art literature analysis, by introducing the novel concept of virtual cells to enhance the resource allocation efficiency. Such a virtual cell concept is aligned with the flexible TDD frame configuration study carried-out by 3GPP Radio Access Network (RAN) Working Group (WG) 1 [32], providing also a set of enhancements for regions where neighbor cells overlap. Virtual cells are formed within overlapping regions enabling residing UE to use sub-frames from different base stations or evolved NodeBs (eNBs) in 3GPP terminology. Expecting a TD-LTE deployment with neighboring cells employing a different TDD frame configuration as in [32], virtual cells can offer a unique type of frame that consists of different sub-frames that belong to the cells that form such an overlapping region. In this way a new notion of a cell is realized, which is virtual since there is no physical infrastructure. TD-LTE deployments with virtual cells may allocate resources with a higher degree of flexibility enabling mobile operators to offer more diverse UL/DL configurations, which can match more effectively the customer needs close to certain geographical areas. The adoption of virtual cells could prove particularly useful for densely deployed networks with diverse traffic demands per geographic area, while its efficiency relies on mechanisms that reflect evolving traffic conditions. Operations, Administration and Management (OAM) mechanisms need to monitor the current network traffic and forecast future demands, considering also history data. The OAM should also perform frame reconfiguration for specific eNBs, assisting the virtual cell formations.

## 3.2 Virtual-Cell Concept

Virtual cells offer a novel way of resource management, which allows UEs to utilize sub-frames from multiple eNBs matching better their traffic demands. UEs are no longer restricted to use the frame of a single eNB, but can utilize specific sub-frames from more than one eNBs. The concept of virtual cell allow users to utilize subframes from multiple eNBs, enabling the formation of a customized virtual frame by deriving specific subframes from different eNBs that can reflect best the users' UL and DL transmission demands [45]. The flexibility offered by this feature allows the resolution of pseudo congestion, while enhancing user performance. Pseudo congestion refers to the phenomenon wherein adequate resources exist in a TD-LTE eNB, but in the opposite transmission direction than the user demand. Such a technique could also be beneficial for the users that are outside the virtual cell region as it could free up additional resources for them to achieve their desired QoS. In addition, this technique not only exploits the spatial domain of conventional load balancing but also the time domain, to dynamically configure the cell setup.

Although virtual cells can be configured under a homogeneous environment, with all cells following the same UL/DL ratio, higher resource flexibility is envisioned when neighboring cells adopt a different UL/DL configuration. In this case conventional power control methods, similar to [35], can be employed to resolve cross-slot interference. Hence, power control coordination among neighboring eNBs ensures that sub-frames with the opposite transmission direction are not provided at the same time within the virtual cell region. The use of advanced power control mechanisms and interference mitigation techniques could further improve the efficiency and performance of the virtual cell scheme in TD-LTE systems. For associating UEs to virtual cells, the UL and DL resources should be considered separately, allowing UEs to utilize UL and DL sub-frames from different eNBs. The adjacent cooperating eNBs in the virtual cell concept appear as one logical cell with each eNB offering a different UL/DL configuration. This provides the capability to support multiple and diverse applications within smaller geographical regions. It is worth noting that the UEs cannot utilize UL and DL sub-frames within the virtual cell region at the same time because of device and hardware restrictions.

A simple example of the virtual cell concept is illustrated in Fig. 3-1 where, a UE residing within the virtual cell region is utilizing UL resources from eNB A and DL resources from eNB B to match its traffic demand. Specifically, the UE is utilizing UL resources from eNB A and DL ones from eNB B, creating in this way a new virtual frame as which is capable to resolve pseudo congestion [67] that would otherwise be caused if the UE would solely utilize the resources from either eNB A or eNB B.

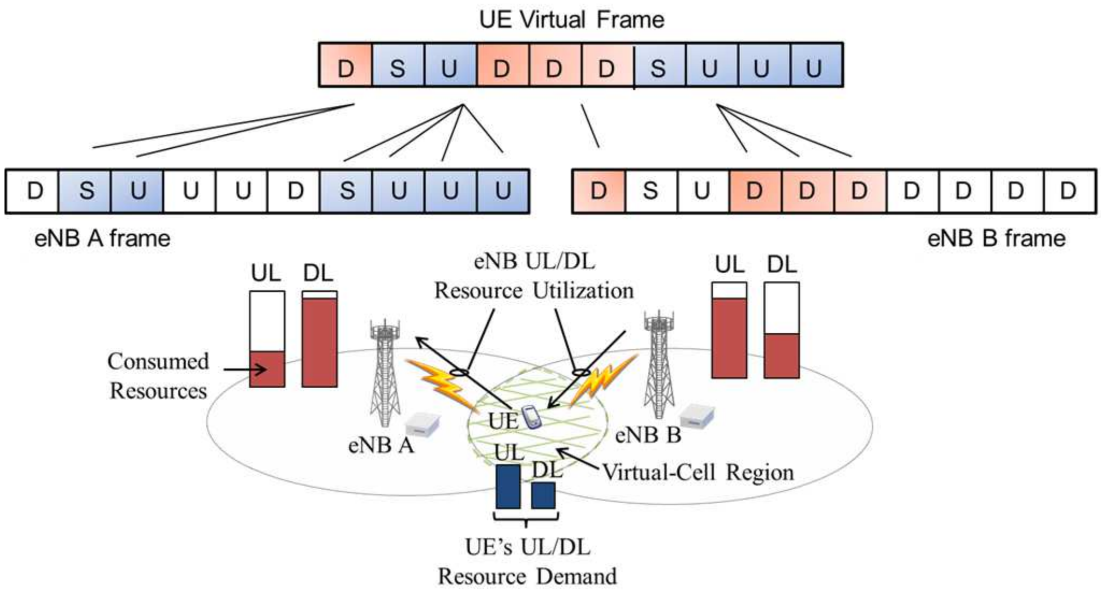


Figure 3-1: A simple example of the virtual cell concept

Specifically, a UE with a high UL and relatively low DL resource demand cannot be served by any of the depicted eNBs solely without experiencing and causing congestion. By assigning the UE to both eNBs and utilizing resources from both, pseudo congestion can be avoided. For instance a UE associated with eNB A to take advantage of the UL resources may switch to the DL of eNB B in order to fulfill its DL demand, instead of remaining on the UL of eNB A. Such operation creates a customized or virtual frame for these particular UEs, which is composed of the shaded UL slots from eNB A and the shaded DL slots from eNB B as shown in Fig. 3-1. Fig. 3-1, shows one of the examples of various traffic scenarios that may occur in the real world. The virtual cell concept and adaptive UL/DL frame reconfigurations can be most effective under low and medium traffic load settings especially when the traffic demands in UL and DL directions are highly asymmetry. While under high traffic load conditions the gain in performance could be relatively lower.

The process of utilizing sub-frames from different eNBs requires enhanced mechanisms to synchronize UEs and align their transmit/receive modes accordingly. Therefore synchronization is required between the eNBs involved in the virtual cell formation to ensure that the data towards and from the UE appears as a single stream hence the virtual cell requires additional signaling mechanisms for control purposes. In general, the process of transmission and reception via multiple eNBs should be synchronized in order to ensure that the data towards and from the UE appears as a single stream. This may be achieved by:

- All eNBs receiving replicated data, which is handled accordingly for transmitting in selected sub-frames,
- A single eNB receiving all data and using the X2 to transfer it towards other eNBs involved in the transmitting process,
- Splitting data sessions in the PDN/S-GW (Packet Data Network/Serving Gateway) before it arrives to the appropriate eNBs

It should be noted that session transmission and reception via multiple eNBs should be transparent to the user. Once virtual cells are configured, mechanisms to perform management and maintenance are essential to reflect evolving traffic demands. The aim is to assess the current UL/DL configuration in combination with virtual cell formation to decide whether potential alternations may enhance the system's performance. A key feature is to consider the UL and DL load separately. The reconfiguration process is envisioned to be hybrid, executed partly on eNBs and on a centralized controller. The process of managing the formation of virtual cells should be handled by the corresponding eNBs in order to reflect short term traffic alterations, while the controller performs the UL/DL frame reconfiguration for eNB based-on long term traffic statistics. In particular, eNBs that cannot ensure adequate resources for residing UEs should notify the controller to check for alternative UL/DL configurations, considering UL/DL traffic statistics from the network-wide eNBs and cell planning data. The core of the reconfiguration function consists of an optimization algorithm that aims to match the input constraints to a specific UL/DL configuration.

## 3.3 SDN-based Network Management Architecture

The SDN-based system aims to perform network resource management and TD-LTE programming considering adjustments on selected TDD frames including formation and deactivation of virtual cells. The objective is to abstract the control plane from individual eNBs and logically centralize it, resulting in a collective resource management and control of eNBs' resources, as described in [47] . To accommodate such a vision, the OAM accompanied by the Data-Controller Plane Interfaces (D-CPI), can facilitate periodic or on demand RAN state updates in order to help the SDN controller to form a global network view [68] . In particular, the OAM can provide the SDN controller with RAN topology information, UL and DL load and Key Performance Indicators (KPIs), e.g. handover failures, latency, throughput, etc. as specified in [69] . The D-CPI, may additionally provide the SDN controller with certain information related to specific rules including monitoring, such as interference levels and the TD-LTE frame configurations per eNB [10] .

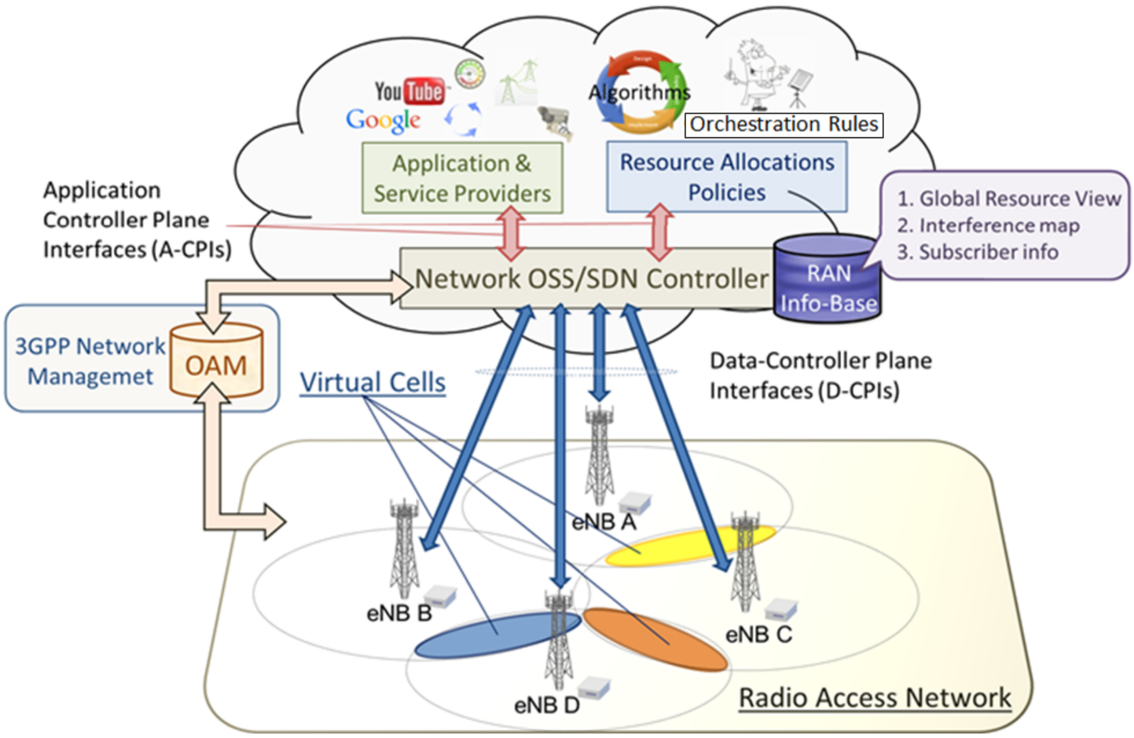

Figure 3-2: SDN-based network architecture

With a global visibility across the RAN, the SDN controller can enhance the resource allocation, enabling virtual cells by adjusting the power and sub-carrier allocation profile of each eNB in a dynamic manner. The SDN controller can assess the impact of a TDD frame re- configuration in the overall performance of the entire network and selectively enforce certain TDD frame changes at specific RAN locations, which otherwise will not be performed through the use of the local adaptive TD-LTE [33] . In addition, the scalability is improved since TD-LTE re-configuration algorithms that may require a significant amount of data can be executed at the SDN controller rather than at eNBs, which have limited computational capacity, avoiding also extensive distributed signaling among multiple eNBs and backhaul elements, while assuring stability. When the SDN controller is notified or it identifies the need for a change in the TD-LTE arrangement, it tries to determine if a TD-LTE frame re-configuration on particular eNBs and/or the use of virtual cells, can enhance the resource utilization. To accomplish this, the SDN controller makes use of the global network state provided by the RAN info-base and executes the provided algorithm and network orchestration policy. It should be noted that the algorithmic logic and orchestration policy can potentially be programmed via the N-API or Application-CPI (A-CPI), where operators have this degree of freedom and flexibility. Once, the SDN controller determines the new TD-LTE resource allocation solution it communicates the essential changes on the TD-LTE configuration back to the corresponding eNBs via the C-DPIs. Effectively, this may alter the transmission power associated with particular TDD sub-frames to provide interference mitigation among neighboring eNBs and facilitate the creation of virtual cells. The SDN controller may also allow application providers to program the RAN via A-CPI, enabling QoS provision for particular services, however, this is left for further study and only considered here for the architecture completeness. An overview of the SDN architecture that elaborates the main elements of the elastic TD-LTE mechanism including their interaction is provided in Fig. 3-2.

Since network traffic interference conditions may fluctuate significantly even for short time periods, especially within the relatively small virtual cell regions, due to user mobility, additional distributed mechanisms should be considered. Such mechanisms may relax the workload on the SDN controller, allowing longer time scale tolerance on the TD-LTE configuration decisions provided to the RAN. Hence, local radio resource adjustments should complement the ones provided by the SDN controller as long as no neighbor eNBs are affected.

## 3.4 SINR Analysis and Effective Capacity Allocation for TD-LTE Virtual Cells

TD-LTE systems are particularly sensitive to interference, especially when neighbor eNBs follow a different UL/DL ratio, e.g. in the case of cell specific adaptive reconfiguration and virtual cells, due to cross-slot interference. Cross-slot interference is caused by eNBs that directly interfere or among UEs in close proximity, which communicate in the opposite transmission direction [35]. Due to physical limitations of the radio frequency frontend at eNBs and UEs, there is the need to avoid such cross-slot interference. In our proposal we achieve this via the use of power control and ensure that the interference introduced cause negligible degradation in the user performance by computing the SINR considering the aforementioned interference phenomena.

The SINR is computed for each user associated with a particular eNB over the RBs assigned to it for transmitting the data. The expression for SINR for the UL and DL directions is given by:

$$\gamma = \frac{P_{rx}}{I + N_0} \tag{3.1}$$

where, $P_{rx}$ is the received power in UL or DL direction respectively, I is the interference power and $N_0$ is the noise power. In the UL direction, UEs that transmit data towards their serving cell over the RBs assigned by the MAC scheduler, may possibly employ the same subset of RBs that are utilized in the neighbor cell at the same time. Therefore, the interference experienced by the received signal at the serving cell over the RBs should be considered while computing the SINR expression in the UL direction. In the DL, all transmissions on the same subset of RBs coming from other eNBs are considered while computing SINR expression. The $P_{rx}$ in UL or DL is given by:

$$P_{rx} = P_t.\tau.(\frac{d_0}{d})^{\phi}.\psi \tag{3.2}$$

where, $P_t$ is the transmit power in UL or DL, $\tau$ is a unitless constant, which depends on the antenna characteristics and average channel attenuation, $d$ is the distance between the transmitter and receiver and $d_0$ is a reference distance for the antenna far field [35] [70]. The pathloss exponent is given by $\phi$ and $\psi$ is a Gauss-distributed random variable representing the shadowing effects in propagation with mean zero and variance $\sigma_{\psi}^2$. The interference power is given by:

$$I = \sum\nolimits_{l \neq o} P_{t,l}.\tau_l.(\frac{d_{0,l}}{d_l})^{\phi_l}.\psi \tag{3.3}$$

While deriving the SINR expression for the cell specific adaptive reconfiguration and virtual cell considering the UL and DL directions, the cross-slot interference, that may arise, should also be taken into account. The SINR expression for both cases in the UL direction for a user $i$ connected to a cell $l$ with respect to a set of neighboring cells $J_l$ can be expressed as:

$$\gamma_{i,l}(UL) = \frac{P_{UL}}{\sum_k^{J_l} I_{DL}(l,k) + \sum_{k,k \neq l}^{J_l} I_{UL}(l,k) + N_0} \tag{3.4}$$

where, $I_{DL}(l,k)$ is the interference power of the DL signal in the neighbor cell $k \in J_l$ observed at the serving cell $l$, $I_{UL}(l,k)$ is the interference power of the UL signal in the neighbor cell $k \in J_l$ observed at the serving cell $l$ and $N_0$ is the noise power. The equivalent SINR expression in the DL direction is:

$$\gamma_{i,l}(DL) = \frac{P_{DL}}{\sum_{k,k \neq l}^{J_l} I_{DL}(l,k) + \sum_k^{J_l} I_{UL}(l,k) + N_0} \tag{3.5}$$

where, $I_{DL}(l,k)$ is the interference power of the DL signal from the neighbor cell $k \in J_l$ measured at the UE $i$ in the cell $l$ and $I_{UL}(l,k)$ is the interference power of the UL signal from an active UE operating in the neighbor cell $k \in J_l$ at the UE operating in the serving cell $l$. The likelihood of forming virtual cells depends on traffic load conditions in the UL and DL directions and on the resource availability from neighboring cells, i.e. similarly to the likelihood of providing load balancing. A virtual cell is merely a more advanced means of performing load balancing in a TD-LTE network, since a user can utilize partial resources from a neighbor cell in a particular transmission direction.

For enabling an efficient virtual cell formation, there is a need for a mechanism to ensure that the resource gain for a particular cell is does not result in starving the users in another cell, i.e., there is no negative impact of virtual cell configuration on the users of neighbor eNB from where the resources are taken. To ensure this capacity gain and loss calculations are performed considering the effective bandwidth model presented in [71]. Regarding the size of a virtual cell, there is no specific size in terms of a geographical area. Virtual cells are user-centric, with the number of users served via a virtual cell varying depending on the traffic load and the users' SINR at particular locations and times, influencing in this way the physical size. For calculating the throughput on the serving link as well as the potential virtual neighbor cell. The capacity of the link $S$ [b/s/Hz] in UL and DL is:

$$S_{\mathrm{i,k}} = \min\left(B_{\mathrm{eff}} \cdot \log_2\left(1 + \frac{\gamma_{\mathrm{i,k}}}{\gamma_{\mathrm{eff}}}\right), S_{\mathrm{eff}}\right), \tag{3.6}$$

where $B_{\mathrm{eff}}$ is the bandwidth efficiency, $\gamma$ is the link-level SINR, and $\gamma_{\mathrm{eff}}$ the SINR efficiency, for a system with maximum spectral efficiency $S_{\mathrm{eff}}$. For borrowing resources from neighbor eNBs and creating virtual cells, we consider a capacity gain and loss metric, similar to the one considered in [72]. Let $R_{\mathrm{k}}$ be the total available resource blocks at a neighbor eNB $k$, $\beta$ be the percentage of available resources that can be borrowed, and $\gamma_{\mathrm{i,k}}$ be the SINR experienced by the user $i$ with eNB $k$. The capacity gain by borrowing resources and creating a virtual cell, $C_{\mathrm{g}}$ [b/s] in UL/DL is given by:

$$C_{\mathrm{g}} = R_{\mathrm{k}} \cdot \beta \log_2\left(1 + \gamma_{\mathrm{i,k}}\right) \tag{3.7}$$

and the capacity loss $C_{\mathrm{l}}$ [b/s] for the mean user $m$ of eNB $k$, having SINR $\gamma_{\mathrm{m,k}}$ (UL/DL), due to the user $i$ borrowing the resources is given by:

$$C_{\mathrm{l}} = R_{\mathrm{k}} \cdot \beta \log_2\left(1 + \gamma_{\mathrm{m,k}}\right) \tag{3.8}$$

The virtual cell is created with eNB $k$ for user $i$, if $C_{\mathrm{g}} > \delta C_{\mathrm{l}}$. This condition ensures that the resource borrowing is done only when there are some capacity gains for the congested cell of UE $i$, and is limited by the factor $\delta \in [0,1]$. The mean user SINR, $\gamma_{\mathrm{m,k}}$ for eNB $k$ having $N_{\mathrm{k}}$ UEs, is given by:

$$\gamma_{\mathrm{m,k}} = \frac{\sum_{u=1}^{N_{\mathrm{k}}} \gamma_{\mathrm{u,k}}}{N_{\mathrm{k}}} \tag{3.9}$$

The virtual cell creation decision is taken per UE, depending on its link quality with its serving cell having congestion, as well as the strongest neighbor cell. For a user $i$, with eNB $o$ as its own cell, i.e. the cell that experience congestion, having $R_{\mathrm{i,o}}$ resource allocated to it and eNB $k$ as the virtual cell with $R_{\mathrm{i,k}}$ resources borrowed from the cell, the total link level capacity of the user, $C_{\mathrm{i}}$ [b/s] in UL/DL is given by:

$$C_{\mathrm{i}} = \sum_{c=i,k} R_{\mathrm{i,c}} \cdot S_{\mathrm{i,c}} \tag{3.10}$$

If the capacity gain and loss condition is satisfied only then the required resources $\beta$ are borrowed to resolve the congestion, otherwise the users are served via the standard cell specific adaptive reconfiguration process. It should be noted that our virtual cell proposal assumes the use of two base stations since the proposed virtual cell concept can be built on the top of dual connectivity as documented in [7], which introduces the bearer split concept involving two base stations.

## 3.5 Elastic Resource Management Algorithms for TD-LTE

The resource management algorithm aims to perform an elastic capacity allocation in the UL or DL direction regulating the formation of virtual cells in order to enhance the users' performance and resolve potential pseudo-congestion problems. The algorithm also intends to maximize the throughput of low SINR users without causing negative effects on the QoS of other users, residing within the region of the serving cell or the surrounding neighboring cells.

The algorithm is executed at the SDN controller taking the global view of the network and the set of congested cells, i.e. cells that experience congestion for a time duration $t > T_c$, where $T_c$ is the time period beyond which a cell with continuously limited resources is declared as congested. The notion of congestion for the best effort traffic, where there is no strict service quality requirement, is accounted for by considering the minimum achieved throughput. The minimum throughput thresholds in the UL and DL directions are used to detect congestion and trigger the algorithm in order to assess and resolve the situation. The throughput thresholds on the UL and DL are selected considering the service delay, which should be upto 300 ms according to [73] for the best effort traffic.

Our algorithm considers the effective capacity, i.e. capacity gain and capacity loss constraints, which are dependent on users' location, so it takes into account the user position and potentially it can take into account user movements. Considering user mobility we clarify that virtual cells do not indtend to serve users with high mobility but rather stationary and low mobility such as pedestrian users. The algorithm initially examines the set of congested cells with the objective of forming virtual cells, selecting the optimal neighboring cell, which can offer the desired amount of resources. During this process the algorithm may also enforce a TD-LTE frame re-configuration to resolve potential pseudo-congestion, if that allows adequate resources for forming virtual cells.

Table 3.1: List of variables used in the pseudo-algorithms

| **Notation** | **Description** |
|---|---|
| $l \in L$ | Cell belonging to a set of total network cells $L$ |
| $o \in O$ | Cell belonging to a set of congested cells $O \in L$ |
| $l[tp]$ | UL or DL average active user throughput in cell a $l \in L$ |
| $tp(i)$ | UL or DL active user $i$ throughput |
| $tp_{th}$ | UL or DL throughput threshold |
| $T_c$ | Time duration beyond which a cell with continuously limited resource is declared as congested |
| $k \in J_o$ | Cell belonging to a set of neighboring cells $J_o$ of a congested cell $o$ |
| $x_{(J_o)}$ | Cell belonging to $J_o$ with maximum $R_k$ in UL or DL direction |
| $R_k$ | Total available resources in a cell $k$ |
| $R_{req}$ | Amount of resources needed to resolve congestion in $o$ |
| $\beta_l$ | Available UL or DL resources of cell $l$ that can potentially be used to form a virtual cell |
| $F[l]$ | Current TD-LTE frame of cell $l \in L$ |
| $S_F$ | Set of sub-frames that belong to a TD-LTE frame $F[l]$ |
| $F_n[l]$ | New TD-LTE frame for cell $l \in L$ |
| $i \in U_o$ | User belonging to a sorted set of active users $U_o$ in an incremental $\gamma_{i,o}$ order, residing in a congested cell $o$ |
| v$Cell_{UE}$ | Set of UEs assigned to the virtual cell region |
| $r_{x_{(J_o)}}(i)$ | Resources allocated to user $i$ once associated with a virtual cell borrowed from the selected neighbor cell $x_{(J_o)}$ |

Table 3.1 continued..

| **Notation** | **Description** |
|---|---|
| $\gamma_{i,o}$ | SINR experienced by an active user $i$ residing in a congested cell $o$ |
| $\gamma_{i,k}$ | SINR experienced by an active user $i$ from a neighbor cell $k$ |
| $\gamma_{m,k}$ | SINR experienced by the mean active user of a neighbor cell $k$ |
| $C_g$ | Capacity gain |
| $C_l$ | Capacity loss |

Once the cells that comprise the virtual cell region are selected, the algorithm examines which users should be associated with such a virtual cell region considering the capacity gain and capacity loss. The variables used throughout the proposed algorithm are summarized in table 3.1.

**Algorithm 1** Forming Virtual Cells: Neighbor Cell Selection

1: $O \Leftarrow$ set of cells $l \in L$ with $l[tp] \leq tp_{th}$ for time duration $t \geq T_c$;<br>
2: **foreach** $o \in O$<br>
3: $J_o \Leftarrow$ neighbor cells $l \in L$;<br>
4: $x_{(J_o)} \Leftarrow$ neighbor cell $k \in J_o$ with $\mathbf{max}(R_k)$ in UL or DL congestion direction;<br>
5: **if**$\beta_{x_{(J_o)}} \geq R_{req}$<br>
6: call **Algorithm 2**;<br>
7: **else**<br>
8: $x_{(J_o)} \Leftarrow$ neighbor cell $k \in J_o$ with $\mathbf{max}(\sum_{UL/DL} R_k)$;<br>
9: *//check and enforce TD-LTE frame re-configuration on* $x_{(J_o)}$<br>
10: $F[x_{(J_o)}] \Leftarrow$ current $x_{(J_o)}$ TD-LTE frame configuration;<br>
11: **while** $x_{(J_o)}[t_p] > tp_{th}$<br>
12: $F_n[x_{(J_o)}] \Leftarrow F[x_{(J_o)}]$ with $\mathbf{min}(S_F)$ re-configured in congestion direction;<br>
13: **if** $x_{(J_o)}[t_p] \leq tp_{th}$<br>
14: **break**;<br>
15: **else**<br>
16: $F[x_{(J_o)}] \Leftarrow F_n[x_{(J_o)}]$;<br>
17: **continue**;<br>
18: **end**<br>
19: **end**<br>
20: call **Algorithm 2**;<br>
21: **end**<br>
22: **end**

The pseudo-code of the algorithm that concentrates on the cell selection to form virtual cells is illustrated in algorithm 1. In line 1, the algorithm collects the set of congested cells $O$. For each congested cell $o \in O$ it identifies its neighbor list $J_o$, from which it selects the neighbor cell referred to as $x_{(J_o)}$ with the maximum resource availability $R_k$ towards the congestion direction, which may either be on the UL or DL as shown in lines 3 and 4 respectively. Here, the goal is to identify a neighboring cell that can accommodate adequate resources in the desired transmission direction allowing the creation of a virtual cell, which fulfils both UL and DL demands. In this way, the algorithm tries to resolve congestion, while making the best use of the current network formation performing no changes to the TD-LTE network configuration.

If the selected cell $x_{(J_o)}$ is able to offer adequate potential resources, $\beta_{x_{(J_o)}}$, it can be used to form a virtual cell satisfying the resource request $R_{req}$ as shown in line 5. As a next step, the algorithm selects the users to associate within the virtual cell region, allocating resources based on their location and interference levels, as elaborated in algorithm 2. Otherwise, the algorithm examines the entire per cell resources irrespective of the transmission direction $\sum_{UL/DL} R_k$, with the goal to identify a neighbor cell with the maximum total resource availability and stores its TD-LTE frame as $F[x_{(J_o)}]$ according to lines 8 and 10 respectively. It should be noted that such a neighbor cell even in cases where it cannot fully accommodate the resource request $R_{req}$, can still provide the best solution toward resolving congestion, enhancing the average user performance depending on the location and interference conditions.

For such neighbor cell $x_{(J_o)}$, the algorithm in lines 11 to 19 tries to investigate whether enforcing a potential TD-LTE frame re-configuration may enhance the resource allocation towards the virtual cell, i.e. $\beta_{x_{(J_o)}}$, without compromising the average user performance, ensuring $x_{(J_o)}[t_p] > tp_{th}$. In particular, in line 12 a TD-LTE frame re-configuration is performed, with the new TD-LTE frame $F_n[x_{(J_o)}]$ selected considering the current one, i.e. $F[x_{(J_o)}]$, with the minimum number of sub-frames $S_F$ re-configured towards the congestion direction. For example if the current configuration $F[x_{(J_o)}]$ employs a DL/UL ratio of 8:1, i.e. configuration 5 in table 2.1, and there is a need to enhance the potential of UL resources towards the virtual cell region, then re-configuring the minimum number of sub-frames towards the UL direction would result in a new TD-LTE frame $F_n[x_{(J_o)}]$ with an UL/DL ratio 7:2, i.e. configuration 4 in table 2.1, while in the following iteration if the intention is to enhance UL resources even further would result in a $F_n[x_{(J_o)}]$ with an UL/DL ratio of 6:3, i.e. configuration 3 in table 2.1.

A new frame is adopted by the system becoming the current one, continuing such an iterative process provided that the throughput change associated with the average user is still beyond the performance target threshold, i.e. $x_{(J_o)}[t_p] > tp_{th}$, otherwise it ceases, breaking the iterative process, as shown in lines 13 to 18. As stated before, once a neighbor cell is selected, potentially with a re-configured TD-LTE frame, a virtual cell is formed and then algorithm 2 allocates resources towards specific users from the overloaded cell based on their location (i.e. the users residing within the virtual cell region) and interference conditions. The resource allocation is performed based on the user's SINR levels and throughput based on equations (7) and (8).

**Algorithm 2** Allocating Users to the Virtual Cell Region

1: v$Cell_{UE} \Leftarrow \emptyset$;<br>
2: $U_o \Leftarrow$ sorted set of active users in an incremental $\gamma_{i,o}$ order;<br>
3: **foreach** $i \in U_o$ starting from $i$ with **min**$(\gamma_{i,o})$<br>
4:   **if** $tp(i) < tp_{th}$<br>
5:     $C_g \Leftarrow R_k \cdot \beta_{x_{(J_o)}} \log_2(1+\gamma_{i,k})$;<br>
6:     $C_l \Leftarrow R_k \cdot \beta_{x_{(J_o)}} \log_2(1+\gamma_{m,k})$;<br>
7:     **if** $C_g > C_l$<br>
8:       *//allocate user $i$ in the virtual cell region*<br>
9:       v$Cell_{UE} \Leftarrow$ v$Cell_{UE} \cup i$;<br>
10:       $\beta_{x_{(J_o)}} \Leftarrow \beta_{x_{(J_o)}}$ - $r_{x_{(J_o)}}(i)$;<br>
11:     **elseif** $C_g \leq C_l$<br>
12:       *//do not interrupt user $i$*<br>
13:       **continue** with next the user from $U_o$;<br>
14:     **end**<br>
15:   **end**<br>
16:   **if** $\beta_{x_{(J_o)}} \leq 0$<br>
17:     **break foreach** loop;<br>
18:   **end**<br>
19: **end**

In particular, algorithm 2 initiates a set v$Cell_{UE}$ to keep a record of the users assigned to the virtual cell region and creates a set of active users residing in a congested cell $U_o$, which is sorted in an incremental order according to the SINR experienced, $\gamma_{i,o}$, as shown in lines 1 and 2 respectively. For each active user starting from the one with the minimum $\gamma_{i,o}$, line 3, the algorithm checks if the user throughput $tp(i)$ is below or equal with the pre-determined threshold $tp_{th}$ in line 4. The rational for assessing users with the minimum $\gamma_{i,o}$ first is to try to improve the performance of users that are more in need since their SINR level is the lowest. For a user with throughput lower than the performance target, i.e. $tp(i) < tp_{th}$, the algorithm examines whether it is beneficial to allocate such a user to the virtual cell region, in lines 4 to 14. To accomplish this, initially the algorithm in lines 5 and 6 calculates the capacity gain $C_g$ and capacity loss $C_l$, as elaborated in section 3.4.

In case the capacity gain is greater than the capacity loss, the user is allocated in the virtual cell region. The algorithm adds that user to the v$Cell_{UE}$ set and subtracts the allocated resources $r_{x_{(J_o)}}(i)$ from the potential resources $\beta_{x_{(J_o)}}$ that can be used within the virtual cell region. Otherwise, the user remains constant, i.e. uninterrupted, and the algorithm continues with the next user until all users $i \in U_o$ are considered or the $\beta_{x_{(J_o)}}$ resources are exhausted. The algorithm returns the set of users v$Cell_{UE}$ that should be allocated in specified virtual cell regions, the neighbor cells involved and any enforced TD-LTE frame re-configuration associated with a particular neighbor cell. It is also worth noting that the users allocated resource from the virtual cell region could potentially free up resources for other users residing outside the virtual cell region and help towards resolve congestion in the problematic cell, thereby improving the overall network performance.

## 3.6 Simulation Results and Analysis

### 3.6.1 Simulation Results and Analysis (Phase-1)

In order to initially test the idea of virtual cell concept, we considered a simulation topology with two macro eNBs having an overlapping region as depicted in Fig. 3-1. UEs are uniformly distributed in the service area and can access the system in a sequential manner following the Poisson traffic model, with an arrival rate $\lambda$, adopting the evaluation methodology defined in [74]. Users accessing the system transmit or receive a file of size 0.5 MB, assuming that traffic is generated randomly and independently in the uplink and downlink directions. The system load is controlled by varying the user arrival rate $\lambda_D$ and $\lambda_U$ that represent the average number of users accessing the system for transmitting or receiving a file in UL and DL direction respectively. We allocate users with an increased DL demand on one eNB and users with an increased UL demand on the other, while edge users, within the overlapping region are assumed to have equal UL/DL requirements. In this way we create a scenario with high per region traffic diversity in the UL and DL transmission direction. The results of the state-of-the art static TDD configuration and Cell Specific Dynamic UL/DL Frame Re-configuration (CSDR) have been reproduced in a slightly different deployment scenario, however, it follows the similar pattern as in [33], [64]. For UL/DL frame re-configuration at a specific re-configuration timescales, a methodology similar to [64] is adopted. The simulation parameters summarized in table 3.2 are based-on the 3GPP TD-LTE system specification [74]. Users not allowed to access the system due to resource limitations, are queued and attempt to gain access at a later time. Separate simulations were performed for the uplink and downlink. For calculating the throughput, we considered the effective bandwidth model considered in [71].

Table 3.2: System Simulation Parameters

| **Parameter** | **Value** |
|---|---|
| eNB ISD | 500m |
| System Bandwidth | 10 MHz |
| Duplexing Scheme | TDD |
| eNB Max Tx Power | 46 dBm |
| eNB Antenna Gain | 15 dBi |
| UE Total Tx Power | 23 dBm |
| UE Antennal Gain | 0 dBi |
| Path loss Model | 128+37.6 $log_{10}$(R), R in Km |
| UE-UE path loss | If R $\leq$ 50m, PL=98.45+20$log_{10}$(R), R in km<br>If R >50m, PL= 55.78+40$log_{10}$(R), R in Km |
| Spectral Efficiency, Seff | 4.0 |
| Number of RBs, NRB | 50 |
| PRB size, RBs | 180 kHz |
| Bandwidth efficiency, Beff | 0.65 |
| SINR efficiency, SINReff | 0.95 |
| File size (FS) | 0.5 MB |
| $tp_{th}$ | (0.5/1) Mbps |

The capacity of the link S [bps/Hz] is:

$$S = \min\left(B_{\text{eff}} \cdot \log_2\left(1 + \frac{\text{SINR}}{\text{SINR}_{\text{eff}}}\right), S_{\text{eff}}\right) \tag{3.11}$$

where, $B_{\text{eff}}$ is the bandwidth efficiency and $SINR_{\text{eff}}$ is the Signal to Interference plus Noise Ratio (SINR) for a system with maximum spectral efficiency $S_{\text{eff}}$. The parameters used for estimating the throughput are also shown in the table 3.2, based on the values used in [75], with same values assumed for downlink and uplink. The TD-LTE frame configuration scenarios adopted in the evaluation consider two fixed

TDD frame schemes, i.e. Config-0 and Config-5 based on table 2.1, employed by both eNBs for the uplink and downlink simulations. These schemes were compared against the adaptive and virtual cell ones. The adaptive scheme distinguishes inner and outer cell users based on the estimated path loss, allocating UL/DL sub-frames in a flexible manner with inner cell users having access to all the available sub-frames, while fixed subframes are allocated for the outer cell users according to [33]. The adaptive scheme adopts Config-0 for one eNB and Config- 5 for the other in both uplink and downlink simulations. The virtual cell adopts the same configuration arrangement as the adaptive one for the inner cell users but serves the overlapping cell users with virtual frames composed from sub-frames of the respective inner cell configuration according to the UL or DL traffic demand.

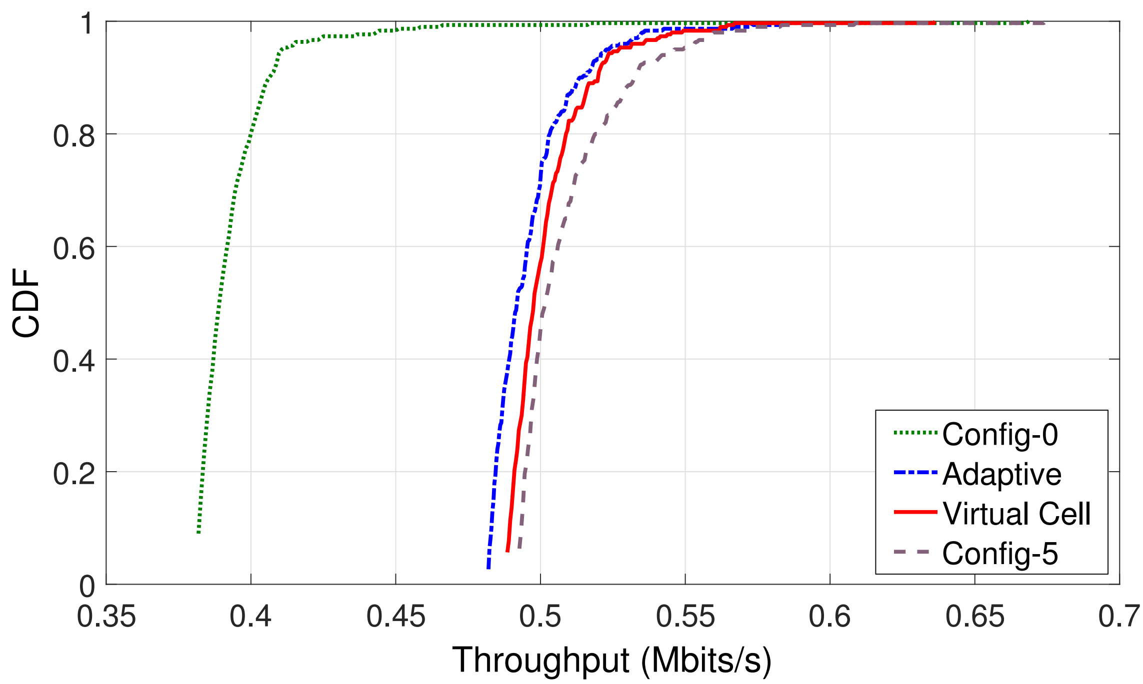


Figure 3-3: CDF for DL direction

TD-LTE Config-0 and Config-5 provide an indication of the maximum performance for the UL and DL respectively, demonstrating for each transmission direction the relative performance deviation of the adaptive and virtual cell schemes. The evaluation study compares the Cumulative Distribution Functions (CDFs) and the blocking probability of the described TD-LTE schemes considering the UL and DL directions separately as already mentioned. The blocking probability is defined as the statistical probability of a connection, which cannot be established due to insufficient system resources. In that case a blocked user re-tries to access the system after a random uniform interval with mean $\lambda$, hence the blocking probability represents the portion of users that experience high delays in accessing the system.

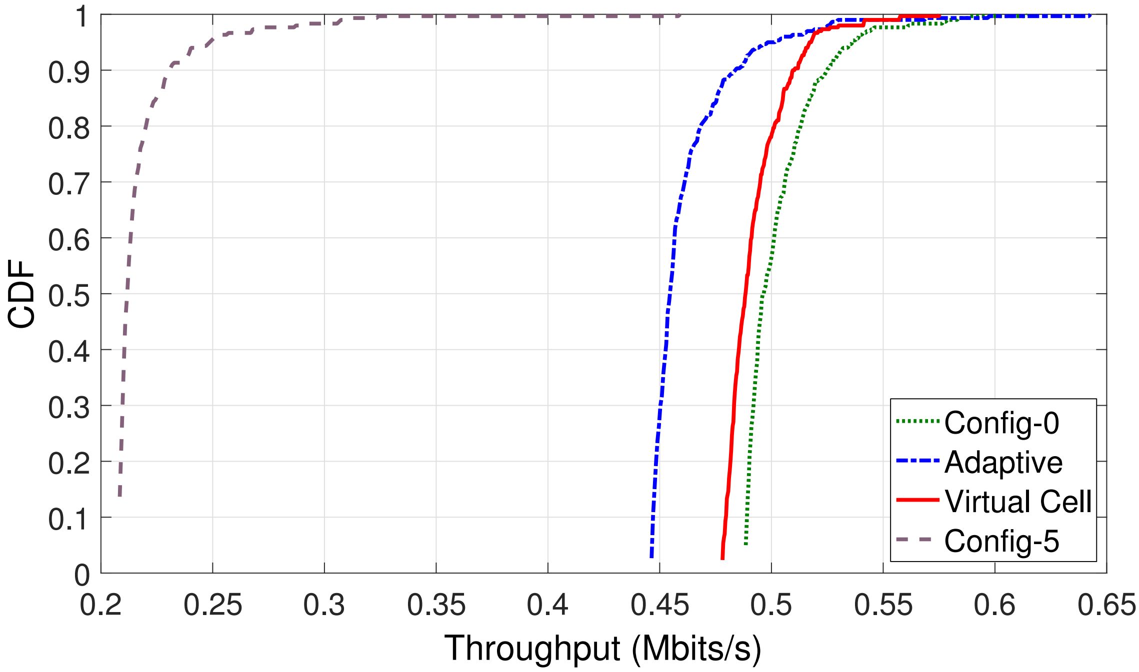


Figure 3-4: CDF for UL direction

The CDFs of the throughput for the Downlink (DL) and Uplink (UL) scenarios are depicted in Fig. 3-3 and Fig. 3-4 respectively, considering the arrival rate of $\lambda_D$=$\lambda_U$=10 for the DL and UL transmission direction respectively. It is clear that the schemes that provide a flexible UL/DL configuration among neighboring cells, i.e. the adaptive and virtual cell, result in higher overall throughput outperforming the fixed Config-0 and Config-5, which only offer a high throughput for the UL or DL respectively, limiting the other transmission direction. In particular, the virtual cell scheme may potentially provide a summarized UL and DL improvement in throughput up to 28% and 57% compared to Config-0 and Config-5 respectively. Compared with the adaptive scheme, the virtual cell increases the user's throughput up to 14%, performing close to Config-5 in the DL direction and close to Config-0 in the UL one. The reason for such an improvement is the fact that the virtual cell scheme allocates sub-frames in UL or DL direction in a dynamic manner, to the users present in the overlapping regions within the neighboring cells'. This is done based-on real time traffic demands. This provides enhanced flexibility increasing the performance of the system, because it addresses the user demands residing in the virtual cell region, resolving in this way the problem of pseudo-congestion. Since TD-LTE configurations that can serve a maximum UL demand may cause performance degradation for users with high DL requirements and vice-versa, having the ability of creating a number of smaller regions, which can offer more diverse UL/DL resources, can improve the TD-LTE system performance.

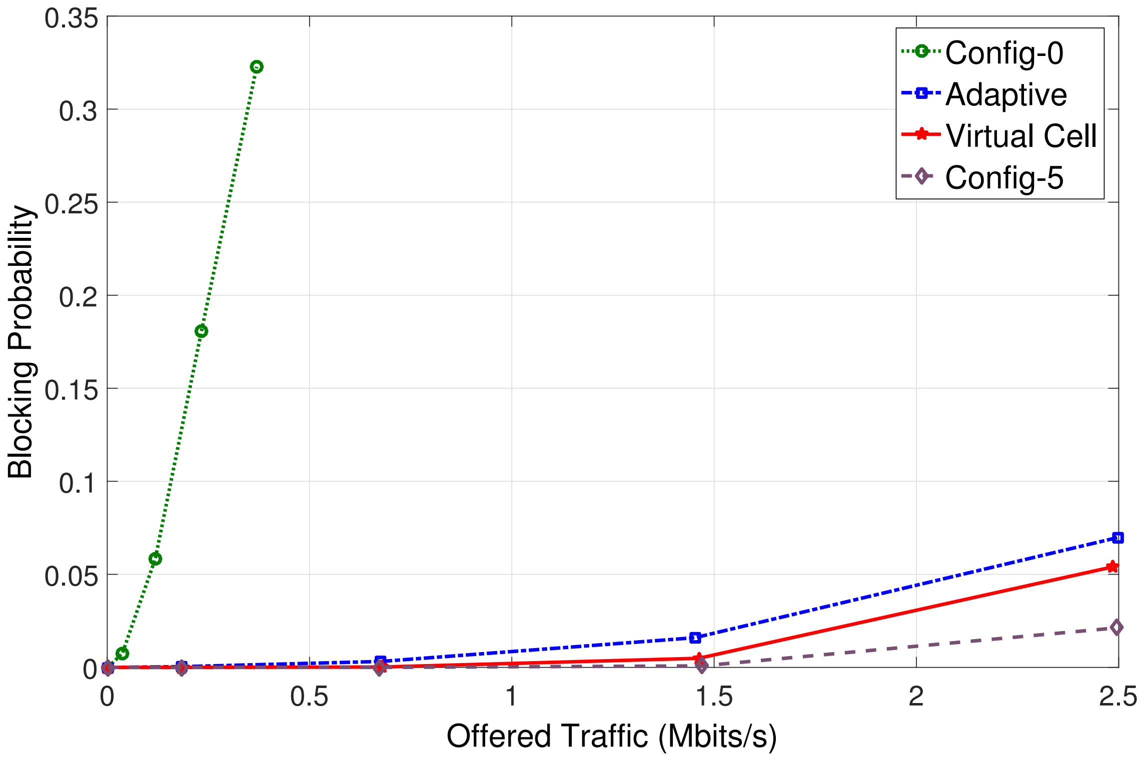


Figure 3-5: Blocking probability vs Offered load for DL

The blocking probability as a function of the offered load is illustrated in Fig. 3-5 and Fig. 3-6 for the downlink and uplink scenario respectively. For the DL direction, i.e. Fig.1-5, the blocking probability of the adaptive, virtual cell and Config-5 is similar for lower values of offered load. In contrast, Config- 0 introduces a very high blocking probability, since high DL demand users are starving for resources. As the offered load increases, it is observed from Fig. 3-5, a relatively small improvement for the virtual cell against the adaptive, both being relatively close to Config-5 especially until the offered load reach 1.5 Mbps. Considering the UL direction in Fig. 3-5, it is obvious that Config-5 introduces an increased blocking probability from early low offered load values, because of the limited amount of UL resources. The adaptive scheme follows, resulting in much less blocking compared to Config-5. However, even for low offered load, i.e. less than 0.25 Mbps, the adaptive scheme introduces a 0.1 blocking probability, which is a significant value. As the offered load increases the blocking probability of the adaptive scheme further increases, while the blocking probability experienced with the virtual cell and Config-5 remain zero accommodating more than four times higher offered load. Even when the offered load is high, the blocking probability of the virtual cell is close to the Config-0 one, illustrating that the virtual cell is able to serve more traffic than the adaptive scheme, performing close to the maximum indicator.

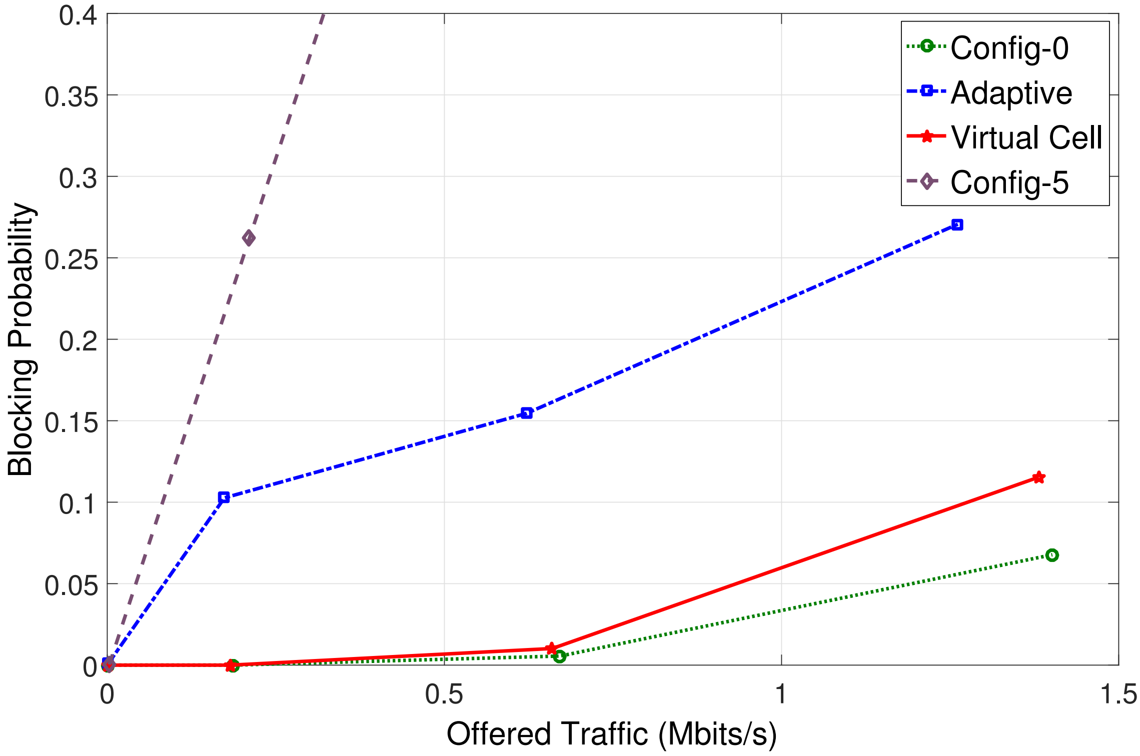

Figure 3-6: Blocking probability vs Offered load for UL

Simulation results show that the virtual cell scheme is able to serve 15% more traffic than the adaptive one considering both the DL and UL direction. Virtual cells also improve the dynamic resource allocation by combining resources from different overlapping cells enabling users to achieve higher throughput. Such a percentage may vary depending on the traffic conditions and TD-LTE configuration employed. Virtual cells may enhance the signal quality at the cell edge, by helping to manage the interference in the service region. The evaluation results demonstrate that virtual cells may further improve the performance of TD-LTE systems and resolve the problem of pseudo congestion by providing virtual customized frames for UEs on finer geographical locations, and improved mapping of their UL/DL transmission demands. Hence, virtual cells have the potential to offer an improved overall system performance by utilizing available resources from multiple eNBs according to the users' demand.

### 3.6.2 Simulation Results and Analysis (Phase-2)

After the initial evaluation of the virtual cell concept, it was further developed and extensive event based system level simulations were performed in Matlab to evaluate the performance of the proposed SDN based elastic resource management solutions/algorithms that introduce virtual cells in TD-LTE networks. For the evaluation of proposed solutions and algorithms, we considered a standard 19-site and 3-sector hexagonal network layout, altogether forming 57 cells and adopted the evaluation methodology defined in [74].The motivation behind our experimentation is to compare our virtual cell proposal with the existing standards study performed in [32] and the basic TDD static UL/DL configuration. For this reason we adopted 3GPP TR 36.828 experimentation scenario and parameters, which are commonly used for driving standards contributions in 3GPP RAN Working Groups.

UEs are randomly distributed in the service area and can access the system following a Poisson traffic model, with a mean arrival rate $\lambda$. Each UE accessing the system is capable of transmitting a file of size 0.5MB in both UL and/or DL transmission direction at different Transmission Time Intervals (TTIs), assuming that the traffic portion per transmission direction is generated randomly. The system load is controlled by varying the user arrival rate that represents the average number of users accessing the system for transmitting and/or receiving a file. The traffic load of a cell is measured based on the number of active users at a given time and the number of resources used out of the total available resources. The detailed simulation parameters summarized in table 3.2 are based-on the 3GPP LTE system specification [74].

We considered a scenario where at any given time, any random cell in the network may experience very high traffic in UL or DL leading to pseudo-congestion, while the other regions in the network carry relatively medium/low traffic. Considering this traffic scenario, we compared the following TD-LTE frame configuration schemes including:

- **Static configuration 1**, where all eNBs employ the same UL/DL configuration, with a subframe ratio of 60% DL and 40% UL.

- **Cell specific adaptive reconfiguration**, where the UL/DL subframe ratio is dynamically selected from the set of seven potential TD-LTE frame configurations. The selection of a suitable UL/DL configuration for individual cells is based on estimations of the uplink and downlink traffic demands as detailed in [33].

- **Virtual cell**, that utilizes resources from more than a single cell. Virtual cells are created by the SDN controller, which may optionally enforce an UL/DL reconfiguration to a particular cell in order to secure adequate resources for the virtual cell region. Under the virtual cell scheme eNBs can still perform locally a cell specific adaptive reconfiguration.

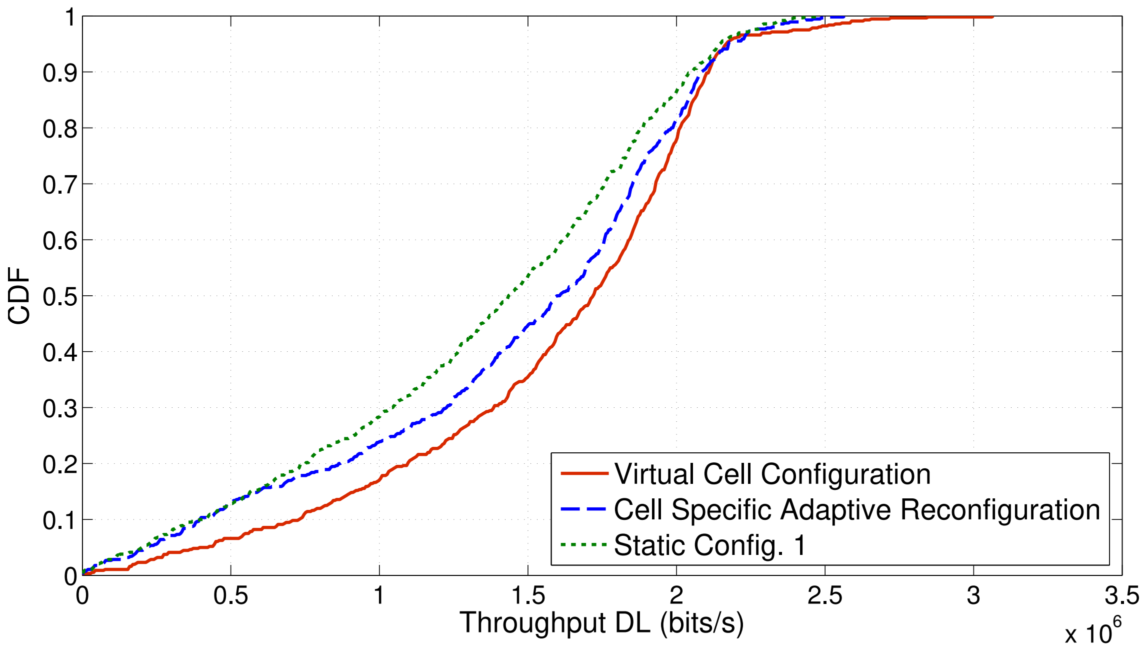


Figure 3-7: Downlink Throughput Distribution

The gain in the throughput performance can be observed in Fig. 3-7 for DL and Fig. 3-8 for the UL, which shows the cumulative distribution function (CDF) of throughput comparing the three aforementioned configuration schemes. From both figures it is obvious that the Virtual Cell Configuration achieves significant gains in throughput compared to Cell Specific Adaptive Reconfiguration and Static Config.1. This gain is achieved considering both users with low SINR conditions that may reside at the cell edge and users with relatively higher SINR levels, that may reside in the inner cell and surrounding regions. The throughput of high SINR inner cell users in the DL and UL are presented distinctly in Fig. 3-9 and Fig. 3-10 respectively.

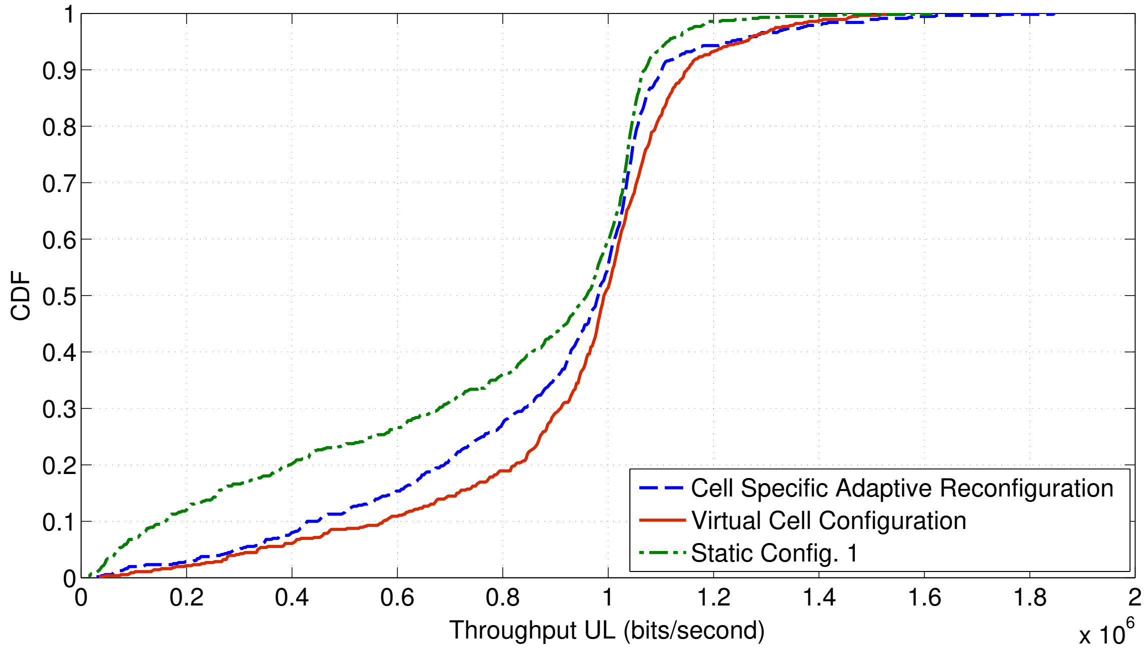


Figure 3-8: figure
Uplink Throughput Distribution

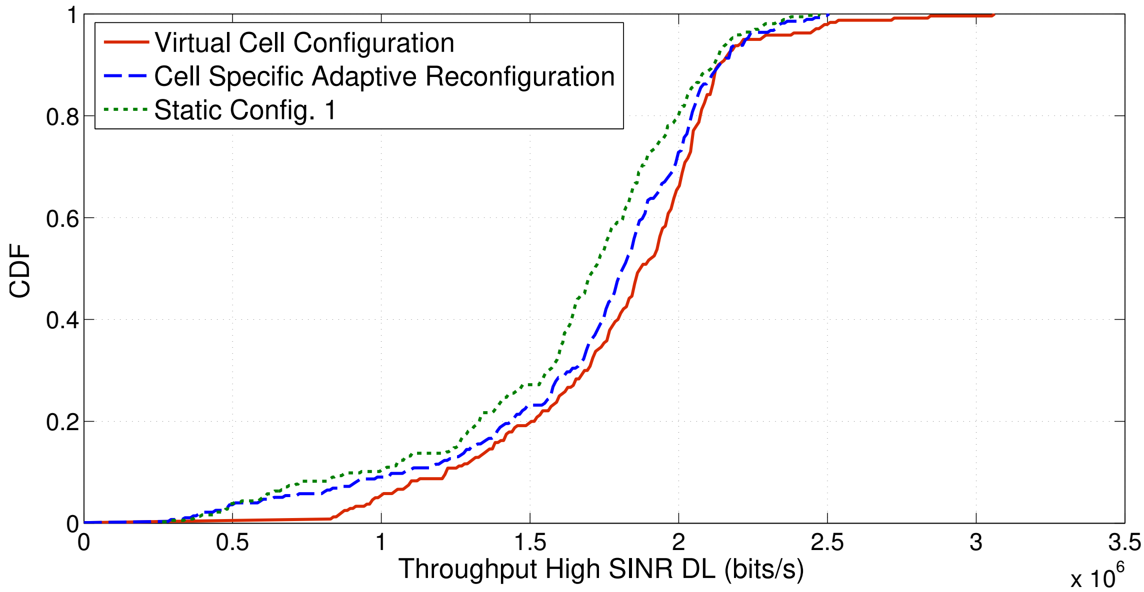


Figure 3-9: figure
Throughput Distribution High SINR Inner Cell Users DL

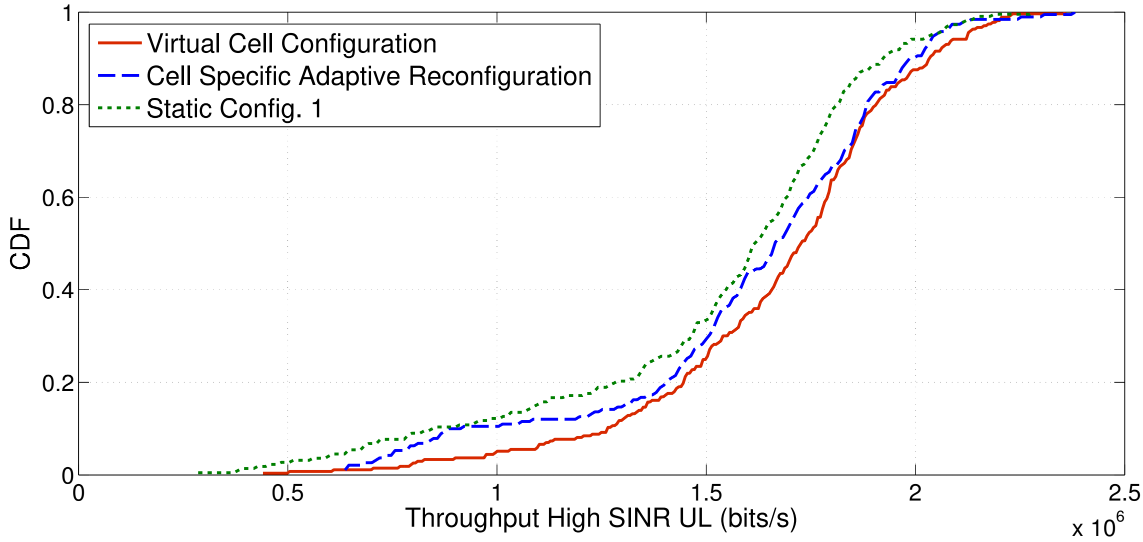


Figure 3-10: Throughput Distribution High SINR Inner Cell Users UL

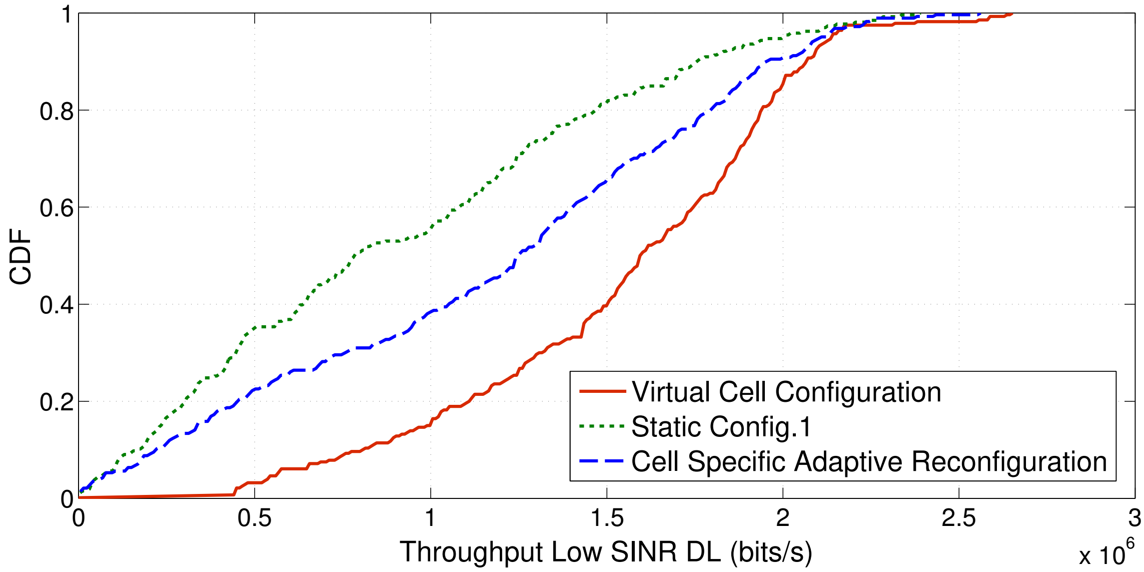


Figure 3-11: Throughput Distribution Low SINR Edge Users DL

From Fig. 3-9 and Fig. 3-10, it can be observed that high SINR inner cell users do not exhibit loss in throughput for the virtual cells. This clearly indicates that virtual cells have no negative impact on the system performance.

Fig. 3-11 and Fig. 3-12 show solely the low SINR edge users' throughput for DL and UL directions. It can be clearly observed that the virtual cell configuration offers significant gains for the low SINR edge users compare to the Cell Specific Adaptive Reconfiguration and Static Config. 1. The gains offered by the Virtual Cell Configuration are summarized in Fig. 3-13 and Fig. 3-14, which shows the comparison between the throughput of low SINR edge users, the mean user throughput and the throughput of high SINR inner cell users of the three schemes in DL and UL directions respectively.

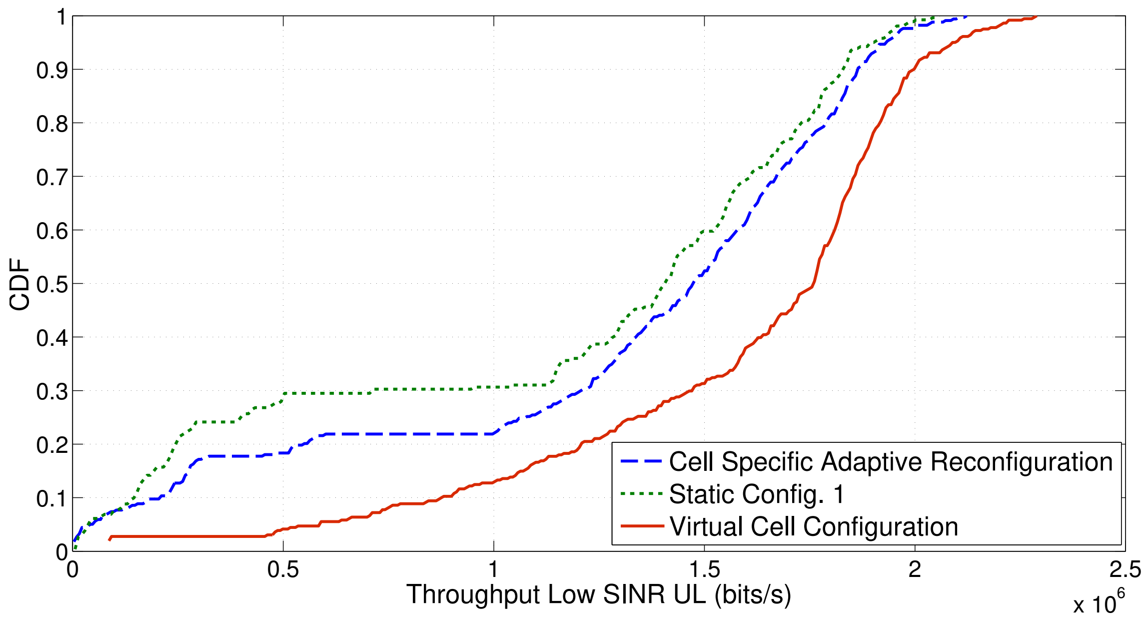


Figure 3-12: figure
Throughput Distribution Low SINR Edge Users UL

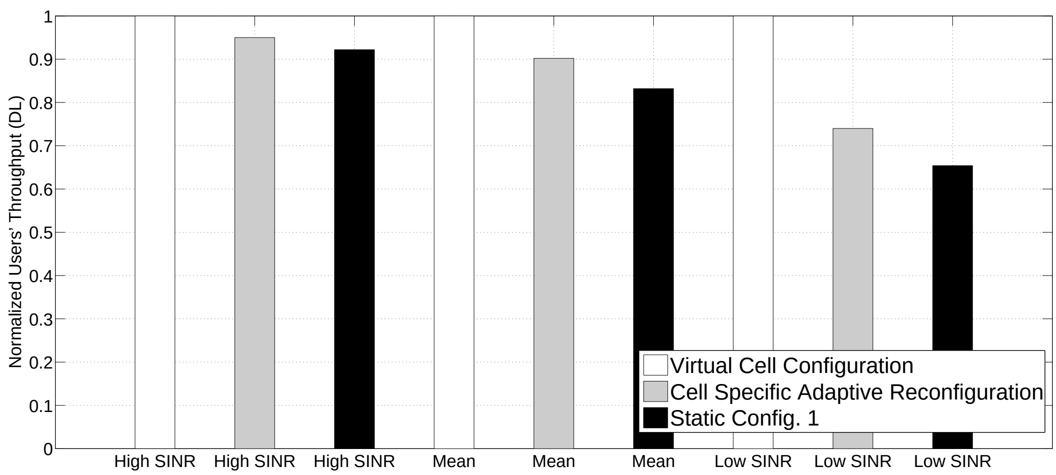


Figure 3-13: Downlink Mean and Low SINR Users' Throughput Normalized to the Virtual Cell Configuration

The throughput of Cell Specific Adaptive Reconfiguration and Static Config. 1 is normalized to the Virtual Cell Configuration which helps in visualizing the gains offered by the Virtual Cell Configuration in comparison with the Cell Specific Adaptive Configuration and Static Config. 1.

Clearly, both schemes that provide a flexible UL/DL configuration, i.e. the Cell Specific Adaptive Reconfiguration and Virtual Cell Configuration, result in higher overall throughput, thereby outperforming Static Config. 1, improving the performance of low SINR users. In particular, we can observe from Fig. 3-13 and Fig. 3-14 that the

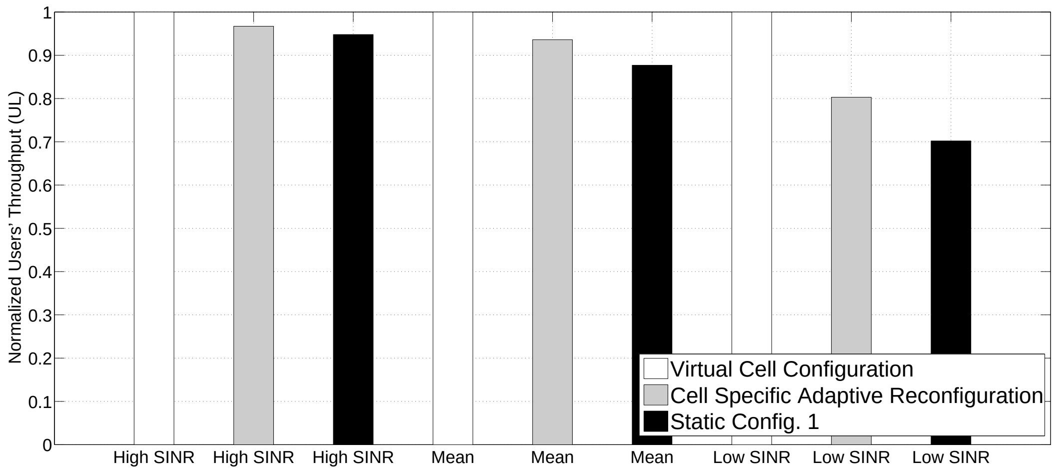


Figure 3-14: Uplink Mean and Low SINR Users' Throughput Normalized to the Virtual Cell Configuration

Virtual Cell Configuration provides around 25% improvement in the low SINR user throughput in DL and around 20% in the UL compared to the Cell Specific Adaptive Reconfiguration, while 35% and 30% compared to Static Config. 1, in the DL and UL respectively. For the mean user throughput, the Virtual Cell Configuration shows an improvement of 10% in the DL and around 6% in the UL compared to the Cell Specific Adaptive Reconfiguration, while around 16% and 12% compared to the Static Config. 1 in DL and UL respectively. The high SINR inner cell users do not experience any negative impact on their performance but instead experience a small gain of 5% and 3% compared to the Cell Specific Adaptive Reconfiguration and a gain of around 7% and 6% compared to the Static Config.1 in the DL and UL directions respectively. The reason behind this is due to the fact that some of the low SINR edge users that are served by the virtual cell configuration free up certain resources that were previously allocated by the serving eNB. These freed up resources can be allocated to the remaining users.

Fig. 3-15 and Fig. 3-16 illustrate the delay CDFs of the aforementioned three schemes for the DL and UL respectively.

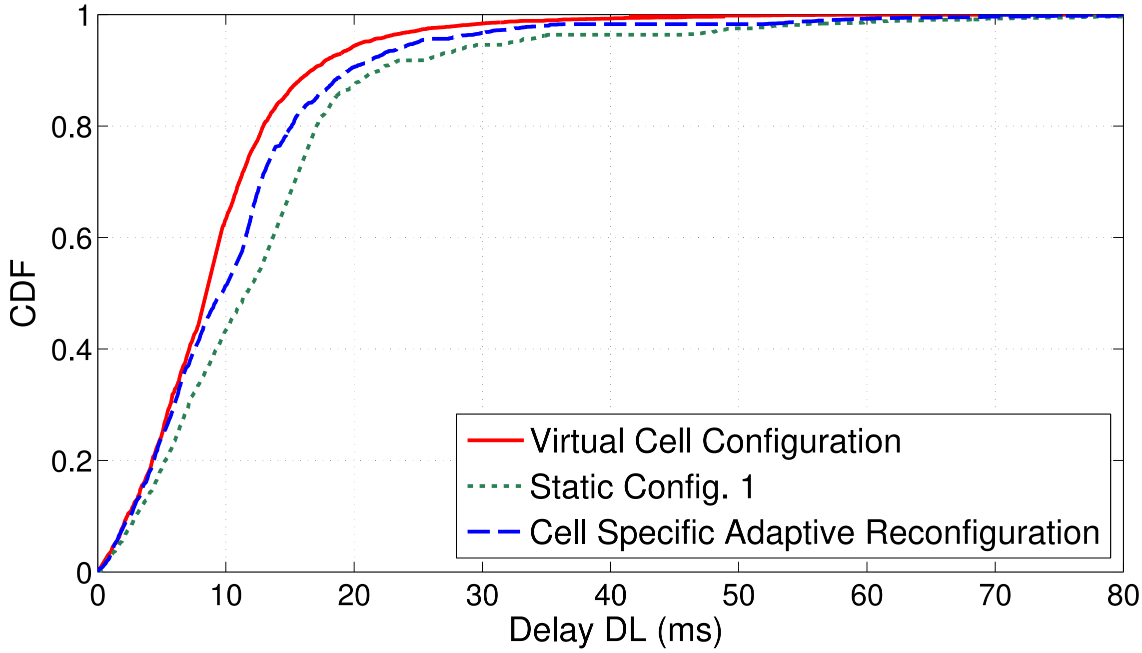


Figure 3-15: Downlink Delay CDF

It can be observed that the Virtual Cell Configuration achieves significant gains in transmission delay compared to the Cell Specific Adaptive Reconfiguration and Static Config. 1 both for low SINR users, which may reside at the cell edge and other users who experience relatively higher SINR levels, that may reside in the inner cell and surrounding regions.

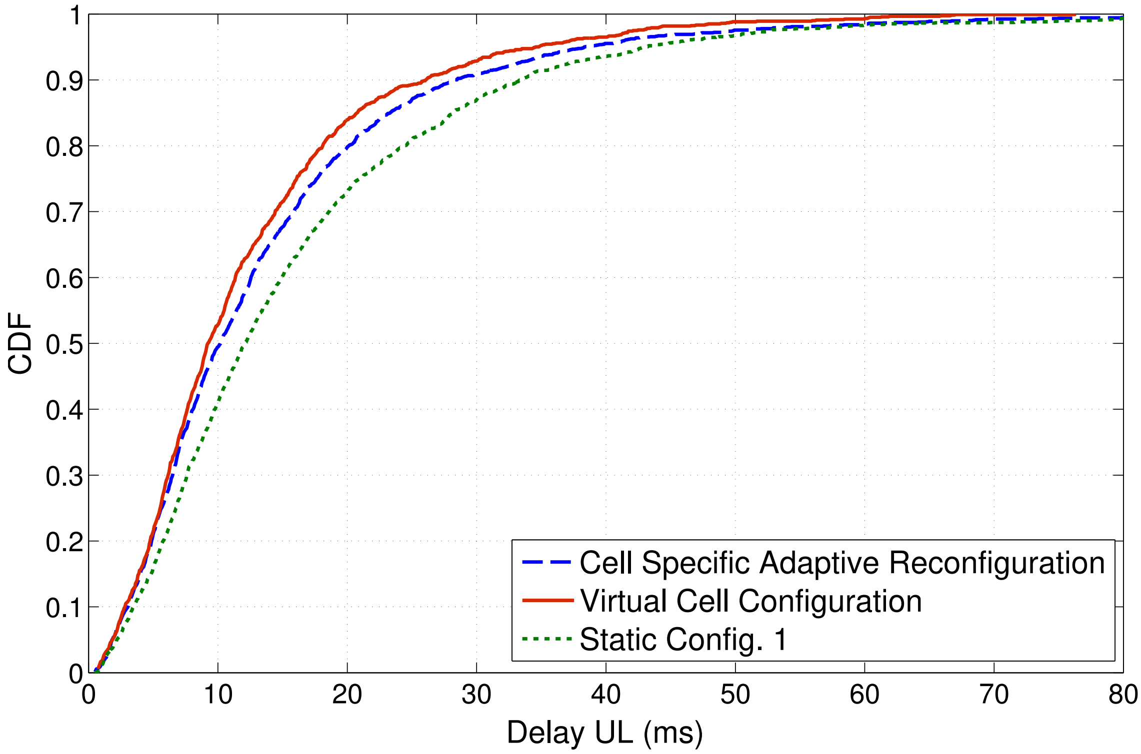


Figure 3-16: Uplink Delay CDF

The delay of high SINR inner cell users in DL and UL are distinctly presented in Fig. 3-17 and Fig. 3-18.

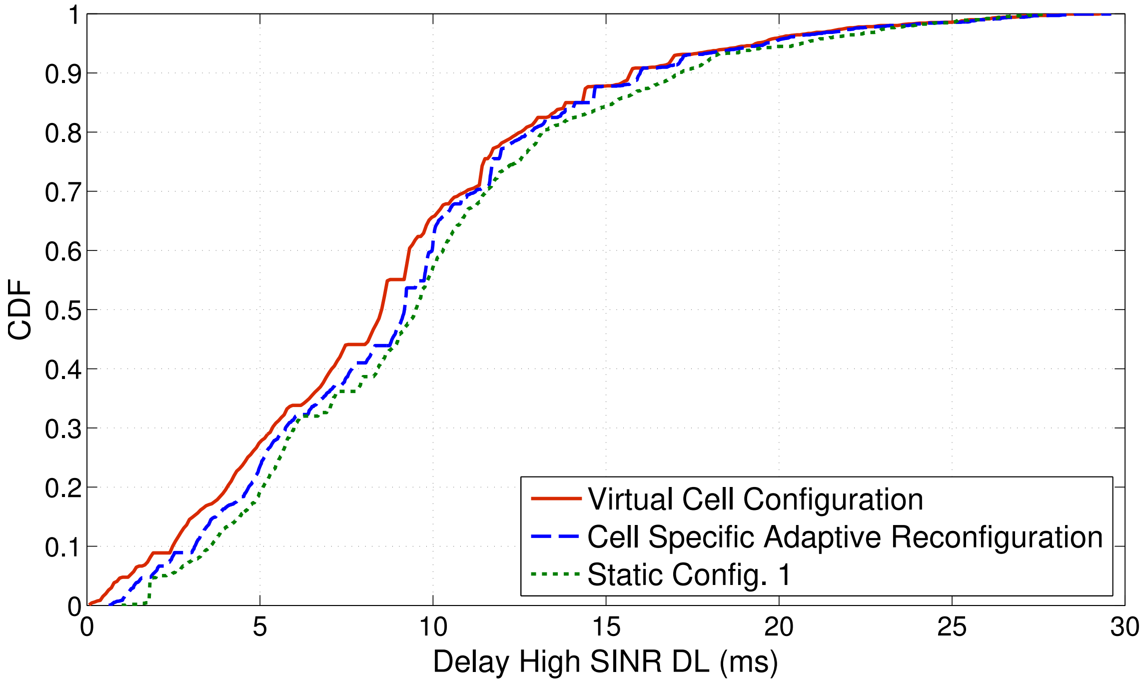


Figure 3-17: Downlink High SINR Inner Cell Users' Delay CDF

From the figures, it can be observed that high SINR inner cell users do not experience an increased delay in case of the virtual cell configuration. This confirms that the virtual cell configuration has no negative impact on the overall system performance.

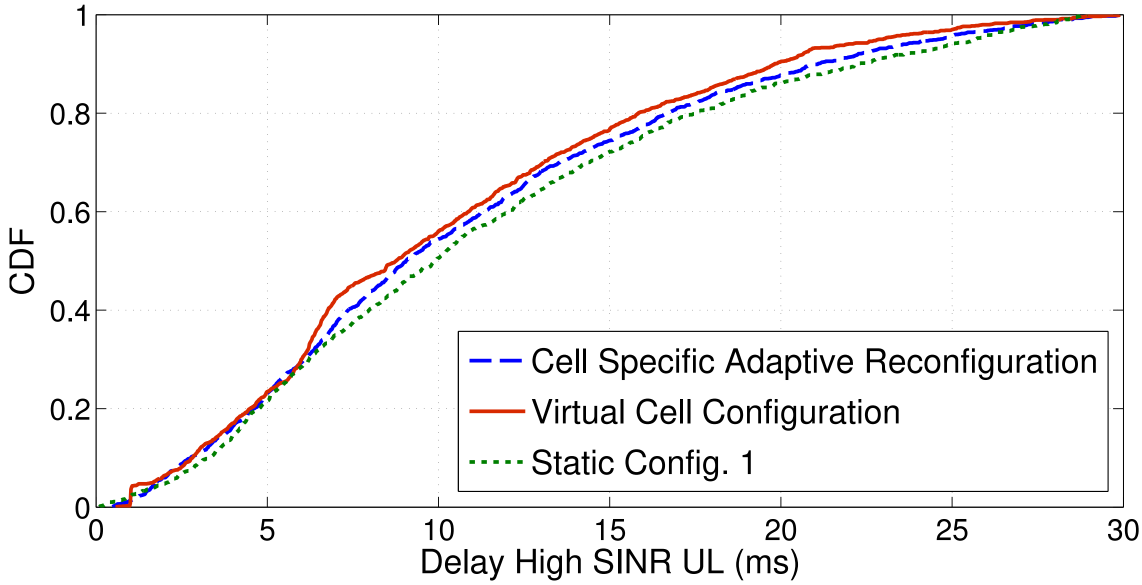


Figure 3-18: Uplink High SINR Inner Cell Users' Delay CDF

Fig. 3-19 and Fig. 3-20 show the low SINR edge users' delay for DL and UL directions. It can be clearly observed that the virtual cell configuration offers significant improvements in terms of delay for the low SINR edge users. The delay of low SINR edge users in the case of virtual cell configuration is lower compared to the Cell Specific Adaptive Reconfiguration and Static Config. 1 illustrating performance improvement offered by the virtual cell configuration.

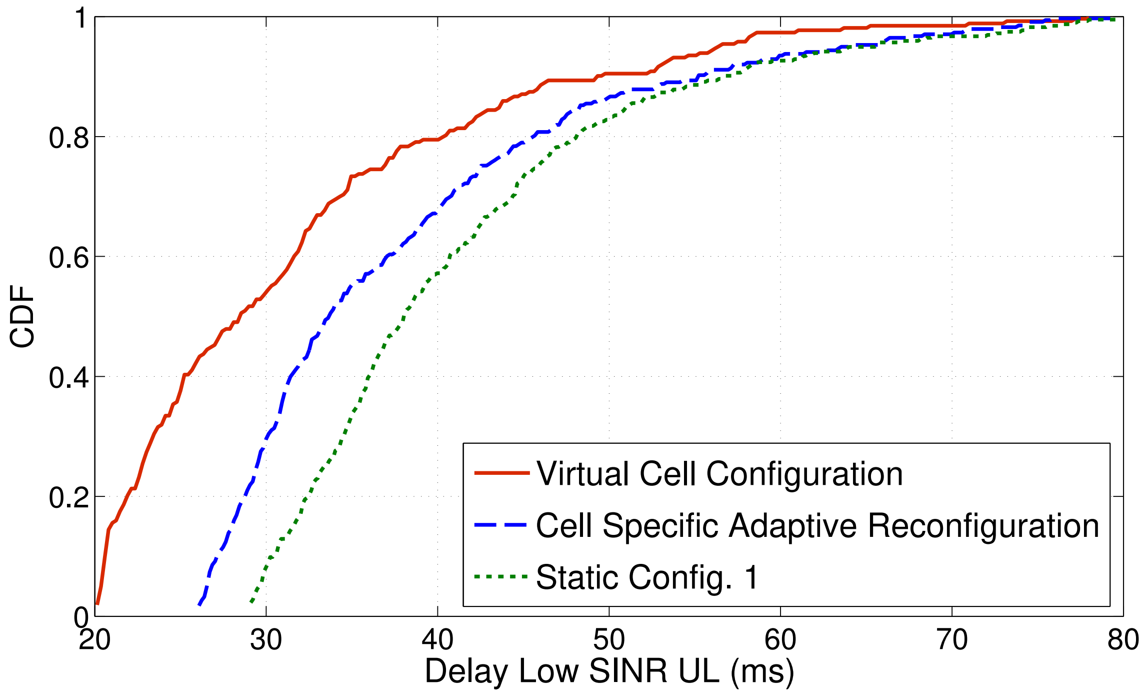


Figure 3-20: Uplink Low SINR Inner Cell Users' Delay CDF

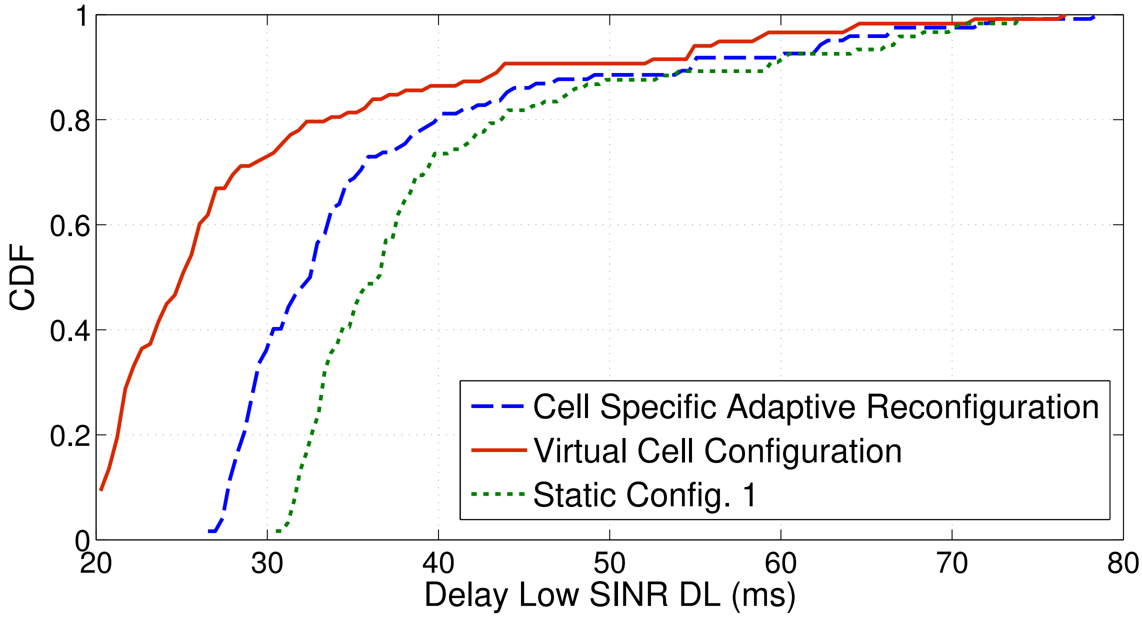


Figure 3-19: Downlink Low SINR Inner Cell Users' Delay CDF

The gain in the transmission delay can also be visualized in From Fig. 3-21 and Fig. 3-22 for the DL and UL direction respectively. As expected, the results are aligned with the throughput gains, as throughput and delay are closely related metrics. To assess the gain in delay offered by the Virtual Cell Configuration, the delay measures of Cell Specific Adaptive Reconfiguration and Static Config. 1 are normalized to the delay of the Virtual Cell Configuration. From Fig. 3-21 and Fig. 3-22 it can be observed that low SINR users in the Virtual Cell Configuration introduce around 25% and 20% less delay than the low SINR users in cell Specific Adaptive Reconfiguration, while 35% and 30% compared to the Static Config. 1 in DL and UL respectively. It is also noted that there is no increase in delay of high SINR inner cell users. In fact slight reduction in delay of about 5% and 3% compared to the Cell Specific Adaptive Reconfiguration while around 7% and around 6% compared to the Static Config. 1 are observed in DL and UL directions. Also, the mean transmission delay for Virtual Cell Configuration is reduced slightly by 10% and 6% in the DL and UL direction compared to the Cell Specific Adaptive Reconfiguration and around 16% and 12% compared to the Static Config. 1.

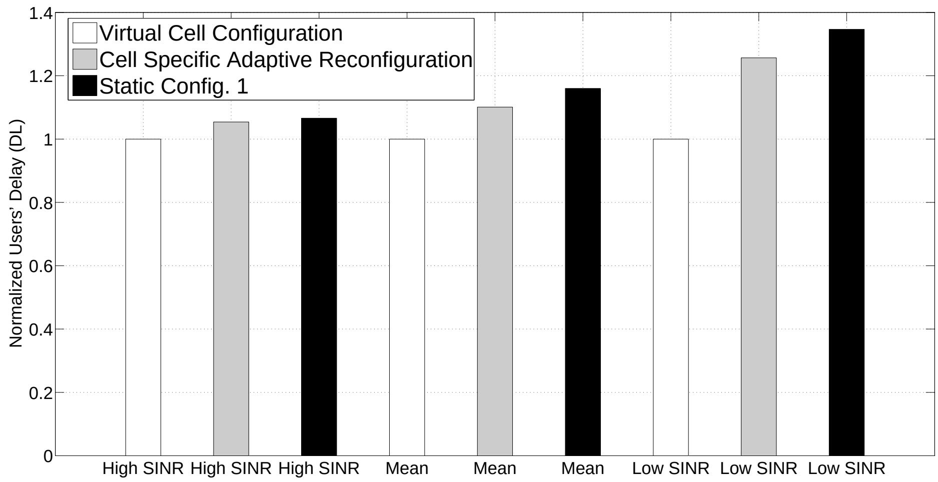


Figure 3-21: Downlink Mean and Low SINR Users' Delay Normalized to the Virtual Cell Configuration

In summary, it is evident from the performance analysis that the Virtual Cell Configuration outperforms the state of the art mechanisms mainly because it can dynamically allocate sub-frames within neighboring cells' overlapping regions for the UL or DL directions according to the real time traffic demands. This provides enhanced flexibility by allowing the system to program the network resources on-demand considering a global network view addressing user demands in particular cell areas and resolving pseudo congestion. It is also worth noting, that the users residing in the virtual cell region and served via multiple eNBs potentially free up resources for other users in the cell, which may experience enhanced SINR improving the overall network performance. With the help of SDN based resource management and the use of virtual cells, operators can maintain a tight control over the network with the ability to flexibly allocate resources on-demand, not only to the end users but also to the OTT applications considering the user subscription plans and SLAs. In particular, mobile operators may dynamically program the network to address the traffic needs while considering UL and DL traffic separately, resolving situations that could lead to congestion. In addition, mobile operators can handle efficiently the service elasticity requirements of cloud providers enhancing the quality of experience taking full advantage of the network resource availability.

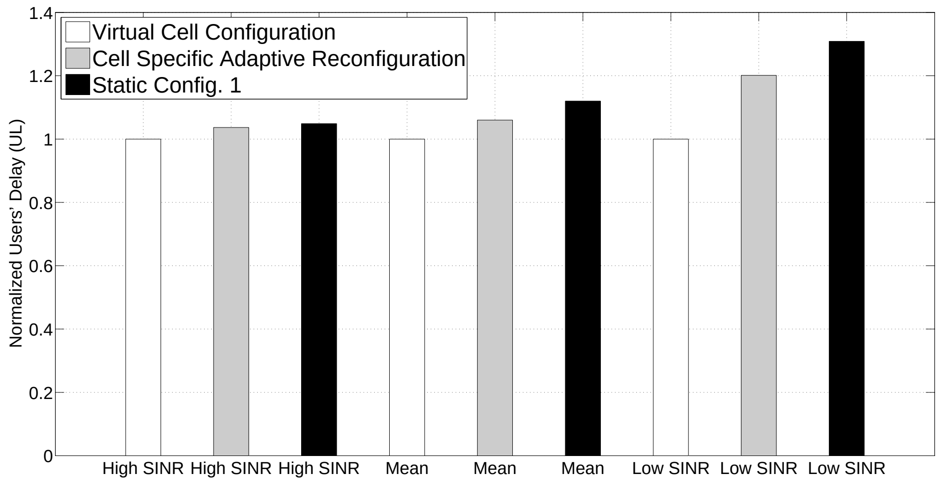


Figure 3-22: Uplink Mean and Low SINR Users' Delay Normalized to the Virtual Cell Configuration

## 3.7 Summary and Conclusion

This chapter has introduced the concept of virtual cell that enables UEs to utilize resources from multiple eNBs taking full advantage of the TDD resource flexibility and providing a means for efficient resource management. Among the several benefits of virtual cells, the most significant concentrate on resolving or avoiding the pseudo congestion and on the provision of a virtual, customized frames from multiple eNBs for UEs residing in the virtual cell region. Virtual cells offer a distributed approach that exploits both spatial and time domain for enhanced load balancing and efficient resource sharing in real time to address varying traffic needs in UL and DL directions. In addition to this, an SDN-based network management architecture and control mechanisms to provide resource elasticity in a TD-LTE system have also been introduced. Such a resource management flexibility introduced by this proposal can enhance the UL/DL diversity in a RAN deployment increasing performance gains, while providing a key enabler for the network operators to support a broad range of OTT applications and cloud services with a wide variety of UL/DL traffic demands. The proposed mechanisms can address flexibly UL and DL traffic requirements in an autonomous manner addressing effectively pseudo-congestion by forming virtual cells enabling users to utilize resources from multiple eNBs allowing customized frames that resolve pseudo-congestion. An algorithm to manage the network resource providing flexibility and enhanced user performance has been introduced at the SDN controller to assess congestion situations and provide resolution via virtual cell provision and by enforcing TD-LTE frame re-configuration at selected eNBs. The results obtained via system level simulations (presented in Phase-1 section 3.6.1 and Phase-2 section 3.6.2) show significant improvements in the average user performance including both edge users that reside within the virtual cell region and inner cell users, without compromising the overall system performance. Further research is envisioned to extend the proposed SDN based mechanisms considering split bearers and towards dynamic adjustments in the mobile backhaul provisioning resources for lower layer transport mechanisms.

# Chapter 4

# SDN-based Elastic Resource Management and QoE Enhancements in 5G Networks

## 4.1 Introduction

In this chapter, an innovative SDN framework that aims at enhancing the QoE perceived by the users running OTT applications is proposed. A scenario is considered where the LTE network is composed of TD-LTE picocells that can better handle symmetric traffic and effectively match instantaneous traffic demands by adapting the UL/DL frame ratio [33]. In this direction, we consider the presence of an SDN Controller with a global view of the TD-LTE picocell status and the QoE requirements of the end users. Based on the available knowledge, the SDN Controller centrally orchestrates the instantiation of virtual cells and adapts the UL/DL frame ratio in neighboring picocells when needed. This is different from the state-of-the art where, for example, in [41] the authors have proposed a dynamic cell specific UL/DL frame reconfiguration mechanism to meet the individual QoS requirements at the end users (i.e. in terms of throughput and packet delay). In contrast to this, the framework proposed in this chapter, enables customized TD-LTE frame ratio to satisfy the QoE demands at the end users. Besides, in this work we utilize the SDN paradigm [17] to achieve efficient synchronization of the TD-LTE pico cells in order to enable an SDN-enhanced formation of virtual cells. In this manner, we allow interaction between the OTT application providers and network operators. A downlink scheduling method for improving the QoE for VoIP traffic in LTE networks is proposed in [51] where scheduling is performed based on users' feedback. a QoE-based framework for network planning and propose an analytical method based on queuing theory to dimension access networks such as LTE is presented in [52]. Different from [51] and [52],in this chapter, A novel SDN-based network management architecture and solution based on the formation of virtual cells and the flexible adaptation of the UL/DL ratio at the TD-LTE pico cell is presented. The numerical analysis shows, that the proposed framework improves the QoE of users suffering from service quality degradation without affecting the performances of other regions of the RAN. In [53], the authors propose an architecture in which a central QoE server collects the performance indicator from different network elements which is then used to improve the QoE of the users. In [54], Decentralized data offloading techniques are proposed which aims to improve end-to-end delay and provide the required MOS for service continuity. Different from [53] and [54], in this chapter, the proposed concept adopts a hybrid approach where, even though the users' QoE is estimated locally at the pico-eNBs, a centralized SDN-based resource management is used to enhance the QoE of specific users by forming virtual cells. In [55] a radio resource management (RRM) strategy is proposed, which dynamically reallocates resources assigned to users with adverse channel conditions towards other users aiming to improve their MOS. Different from this work, in this chapter, the proposed novel concept provides additional resources to users with MOS below a certain level depending on the type of application. To achieve this, we allow cell-edge users to utilize resources from multiple pico eNBs with the help of the virtual cell concept. Such a strategy provides sufficient resources to users with bad channel conditions while enhancing individual MOS of the corresponding user applications.Thus, the proposed framework in this chapter, extends the state of the art by dynamically programming the UL/DL ratio of TD-LTE pico eNBs, while introducing virtual cells considering the requirements of the OTT applications to satisfy QoE of end users. Furthermore, the proposed framework aims at resolving the so-called pseudo congestion problem, where a picocell has a surplus of resources in the opposite direction as compared to the desired one. Besides, the proposed framework aims at creating virtual cells in regions with overlapping network coverage in order to enable cell-edge users to utilize resources from multiple pico-cells. We envisage that the SDN-enhanced employment of UL/DL frame reconfiguration in conjunction with the instantiation of virtual cells can enhance service quality and improve resource utilization [45].

Additionally, in section 4.3, this chapter also propose a novel SDN-based architecture that enables dynamic resource sharing among the different operators and employ a dynamic re-configuration mechanism that enables the TDD pico eNBs to adapt their

UL/DL ratio in order to release or embody a set of radio resources to/from the FDD macro eNBs operators. System level simulation results demonstrate that the combination of these solutions in the scenario of resource transfer from a TDD pico eNB operator to a FDD macro eNB operator achieves notable performance gains in terms of communication delay in the application layer. This is different from the state-of-the art, where, for example in [61], an SDN-based framework is proposed to enable efficient on-demand sharing of base stations that belong to different operators. The OpenRAN architecture, to leverage the convergence benefits of HetNets and achieve, at the same time, customization via network programmability is introduced in [62]. In [63], The SDN paradigm as a tool to simplify cellular network management is discussed. Here, network resources are pooled and Network virtualization (NV) is realized at different levels (e.g. application, spectrum, network level) to have more granular control over the network. Here, a hierarchical architecture with a local and a global SDN controller is also considered. The local controller manages the activities inside the network, while the global controller handles the events throughout the HetNet and coordinates the RANs through the backhaul. Different from [61]-[63], in this chapter, we focus on the resource sharing problem in a multi-operator integrated FDD macrocell and TDD picocell LTE-A system, employing network programmability through re-configuration of the TDD frames at the pico eNBs. The problem of re-configuring the TDD frames of pico eNBs has also been considered in [33] [64] [36]. However, different from these approaches, in this chapter, we additionally consider the amount of resources shared in the FDD system and the requirements of the users associated with the TDD picocells. Focusing on TDD systems, the idea of flexible spectrum sharing is investigated in [45]. The authors in [76] consider a mix of TDD and FDD, where the TDD occupies the guard band spectrum between the FDD UL and DL. In contrast to [45] [76], in this chapter, we consider elastic resource sharing in a multi-tenant multi-operator LTE-A HetNet environment. The detailed discussion on state-of-the art related to the novel ideas presented in this chapter can be found in the literature review chapter 2.

This chapter is organized as follows: Section 4.2 introduces the proposed SDN framework for QoE-aware flexible network management. It also includes a brief introduction of the virtual cell concept (that has also been discussed in detail in chapter

3) and QoE Aware Flexible Network Management. Section 4.2.4 includes system-level simulation results on the performance of the proposed framework. Section 4.4 presents the concept of elastic resource sharing in LTE-A FDD/TDD HetNets. It describes the proposed SDN-based architecture for elastic resource sharing among multiple operators and analyzes the resource sharing logic in combination with the TDD UL/DL frame re-configuration mechanism. Section 4.4.3 discuss the performance evaluation results and analyzes the gains obtained by employing the proposed SDN-based elastic resource sharing solution. Finally, section 4.5 concludes the chapter.

## 4.2 SDN Framework for QoE Enhancement

This section describes the main features of the proposed SDN-based framework for enhanced QoE in the presence of TD-LTE picocells. In section 4.2.1, the virtual cell concept is briefly introduced. This is followed by the discussion on how the virtual cell concept can be applied to enhance the QoE of end user applications in a TD-LTE picocell deployment scenario. In section 4.2.2, an SDN-based reference architecture for flexible network management that is accompanied with the logic signaling flow for enhanced QoE management on a per user and per application type basis has been presented. Accordingly, in the same section, an algorithm for assessing the QoE of the users locally at the pico eNBs and triggering the instantiation of the virtual cell function (which is a logical entity residing in the SDN Controller) is discussed. The details of the QoE-Assessment algorithm are described in section 4.2.3.

### 4.2.1 Virtual Cell Concept

The concept of virtual cell has been introduced and explained in detail in chapter 3. Virtual cells provide efficient and flexible resource management for TD-LTE networks. An overview of the virtual cell concept is shown in Fig. 4-1. The SDN based solution proposed in this chapter uses the virtual cell concept to avoid or counter any congestion related problems and aims to enhance the QoE of the end users.

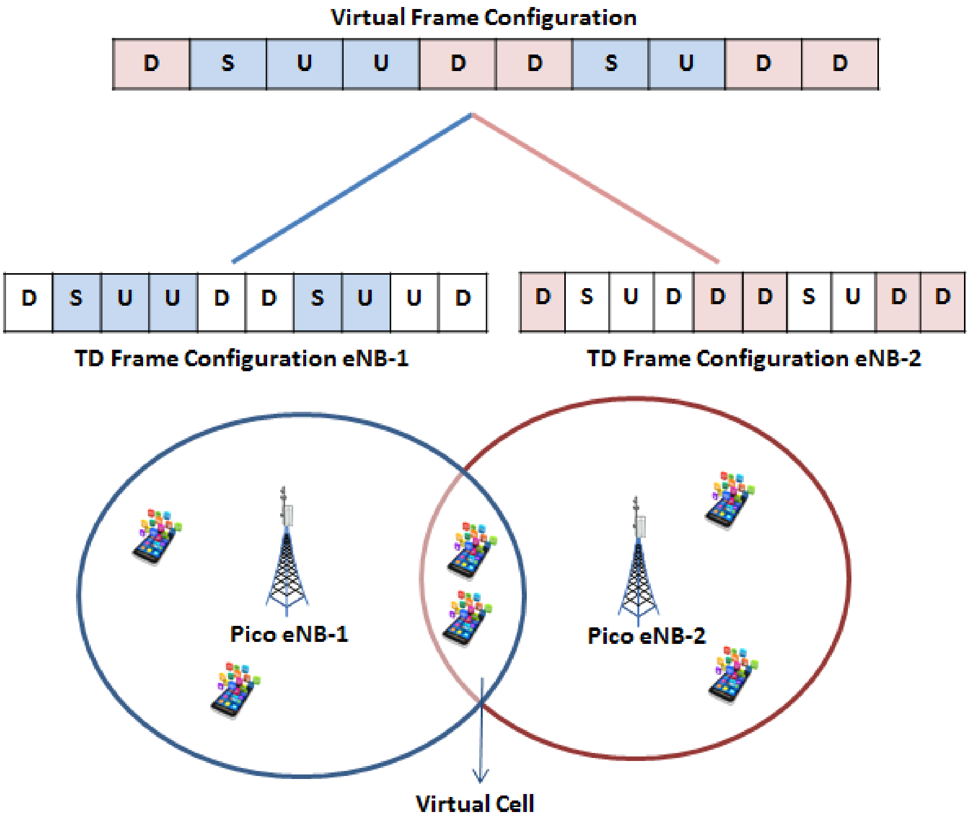


Figure 4-1: Virtual Frame Configuration

In certain traffic conditions, cell-specific adaptive reconfiguration is unable to meet UL-DL traffic requirements. In such situations, the proposed framework applies the virtual cell concept to further enhance the QoE of certain user flows that suffer from QoE degradation due to varying traffic conditions. This is done by providing additional resources via virtual cell formation with the neighboring eNBs.

### 4.2.2 QoE Aware Flexible Network Management

Fig. 4-2 illustrates the proposed SDN-based flexible network management architecture for enhanced QoE. In this scenario, the virtual cells are created on demand for specific user applications between adjacent pico eNBs. Each pico eNB has a logical function for QoE assessment that works in cooperation with the TD-LTE cell specific adaptive re-configuration function. Virtual cells are created by the SDN-based virtual cell function as shown in Fig. 4-2. According to this architecture, OTT providers specify their service requirements and configuration needs to the SDN controller via the Application Controller Plane Interfaces (A-CPIs). This provides a degree of freedom and flexibility to network operators to program their network in real-time in order to satisfy the users' QoE requirements in an efficient fashion. The SDN controller attains a global network view by interacting with the 3GPP OAM network management plane [77]. The 3GPP OAM subsystem assists the SDN Controller to acquire on-demand (or periodically) information on the current state of the RAN. The RAN state includes information regarding the UL/DL traffic load measurements, the interference conditions and performance measurement KPIs e.g. throughput, delay, handover failures etc. This information can be then be forwarded to the SDN controller and used by the virtual cell function to suggest reconfiguration actions to the pico-eNBs in order to resolve issues related to QoE degradation and pseudo congestion. Such reconfiguration instructions are communicated from the SDN-controller to the pico-eNBs via the Data-Controller Plane Interface (D-CPI). The D-CPI also monitors QoE related KPIs e.g MOS scores of different applications, in order to feed the SDN controller with performance information related with virtual cells directly.

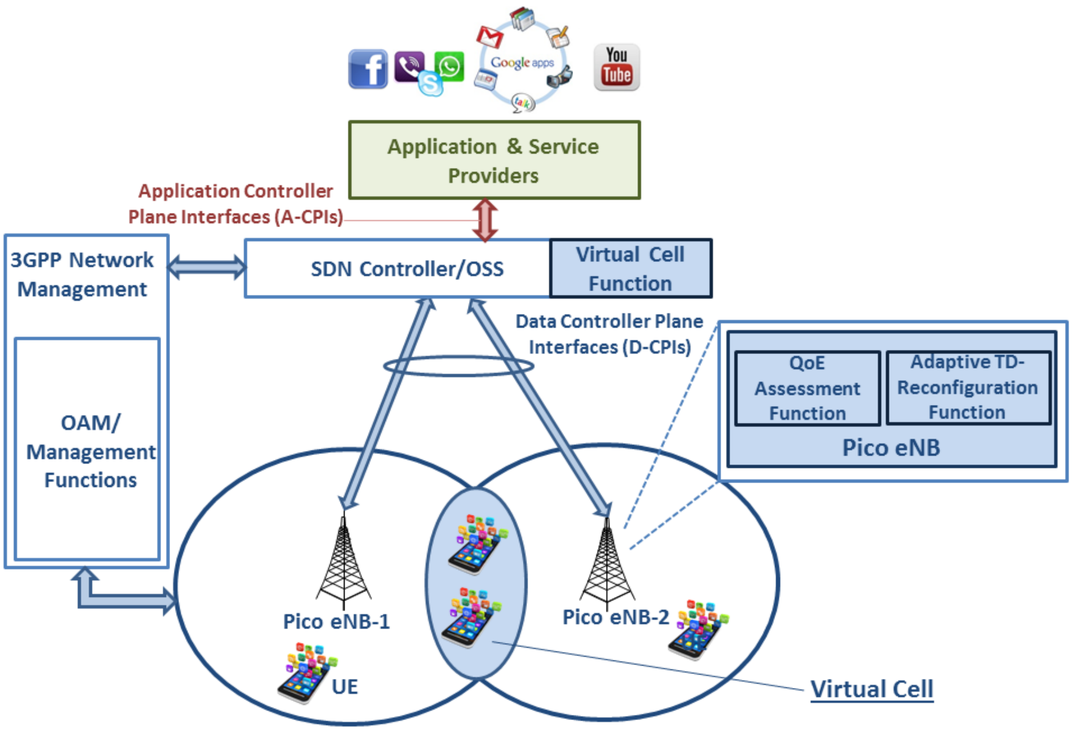


Figure 4-2: SDN Network Management Architecture

Fig. 4-3 shows the logical signaling flow required to support the proposed two-phase QoE-enhancement approach. As a first step, each user sends a measurement report to the serving pico eNB periodically.

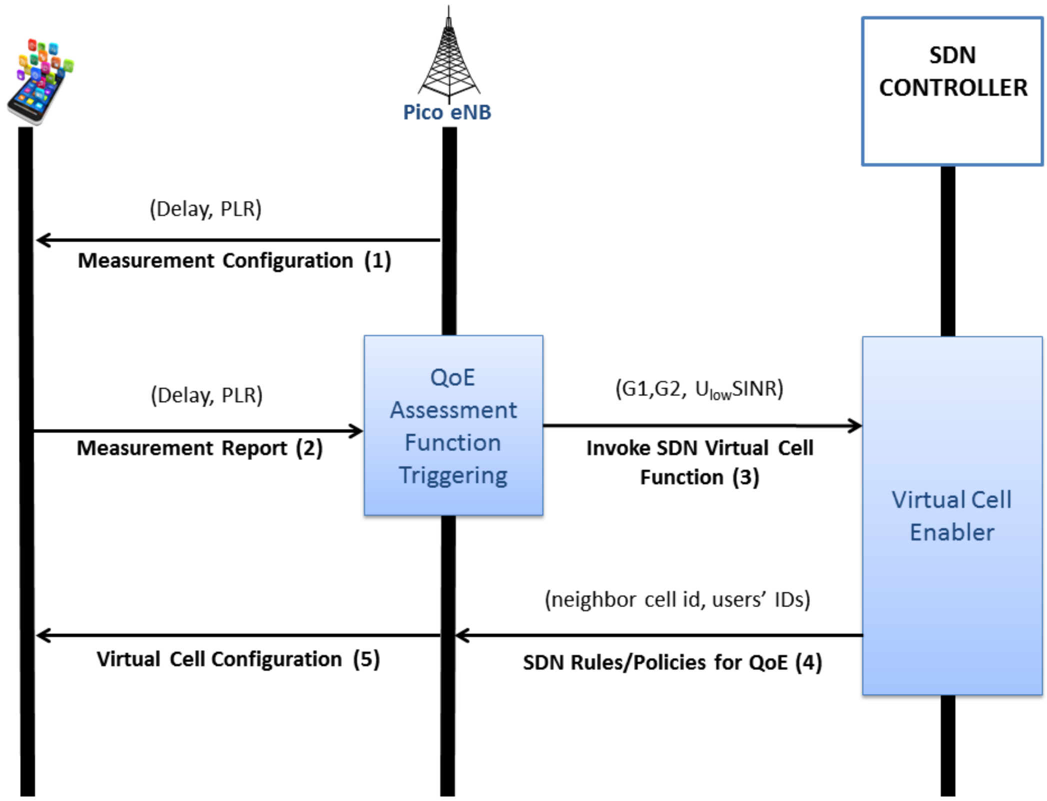


Figure 4-3: Signaling Flow Mechanism Logic

This thesis contains information regarding the end-to-end delay and the Packet Loss Rate (PLR) of each application flow running at the users' device. The PLR and the end-to-end delay measurements are subsequently used as inputs for the QoE-Assessment algorithm that runs at the pico eNB. If the performance of a particular application drops below a predefined limit, the QoE-Assessment algorithm triggers the SDN controller to create a virtual cell and indicate appropriate actions to resolve the problem of QoE degradation for specific applications or users. The triggering conditions are described in more detail in section 4.2.3. Once the SDN Controller receives such a request from the pico eNB, it initiates the virtual cell configuration function. In more detail, it collects information about the status of neighboring cells (e.g. traffic load, SINR, users' location) and identifies suitable neighbor cells for creating virtual cells in proximity with the tagged pico eNB. Accordingly, the SDN Controller reports back to the serving pico eNB the respective set of neighbor cell identities (IDs) to create a virtual cell. In addition it helps in identifying the list of users to be served via the virtual cell configuration, allocating the desired resources without causing any negative impact on the performance of other users in the neighbor cell.

### 4.2.3 QoE Assessment Algorithm

The parameters and variables used in the proposed QoE-Assessment algorithm are listed in table 4.1. A pseudo-code version of the proposed QoE-Assessment algorithm is presented in algorithm 3. The QoE-Assessment algorithm runs in all the pico eNBs in a distributed fashion. In the following, we consider a scenario where the users may host different applications on a single device. Each pico eNB calculates the MOS of each application flow running at the served users (step 4). The MOS is calculated according to the E-Model by taking into account the average delay and PLR of each flow over a specific time interval.

Table 4.1: List of Parameters and Variables

| Notation | Description |
|---|---|
| $C$ | Set of cells |
| $c, (c \in C)$ | Index for a cell |
| $u, (u \in U)$ | Total number of users in a cell |
| $f, (f \in F)$ | Set of application flows of a user $u$ |
| $MOS_{Thr}(f)$ | Application specific MOS Threshold for the flow $f$ |
| $MOS_{dist}(f)$ | Distance of $MOS(f)$ from the $MOS_{Thresh}(f)$ |
| $G_{1mos}$, $G_{1mos} \in F$ | Set of flows with $MOS_{dist}(f) \in [0, 1.5]$ |
| $G_{2mos}$, $G_{2mos} \in F$ | Set of flows with $MOS_{dist}(f) \in [1.5, 2.5]$ |
| $\gamma_{u,c}$ | SINR of user $u$ with cell $c$ |
| $\gamma_{Thresh}$ | SINR Threshold |
| $U_{lowSINR}$, $U_{lowSINR} \in U$ | Set of users with $\gamma < \gamma_{Thresh}$ |
| $P_{G0}$ | Probability of a flow $f$ to be in group $G_{0mos}$ |
| $P_{G1}$ | Probability of a flow $f$ to be in group $G_{1mos}$ |
| $P_{G2}$ | Probability of a flow $f$ to be in group $G_{2mos}$ |
| $P_{G0out}$ | Outage Probability for group $G_{0mos}$ |
| $P_{G1out}$ | Outage Probability for group $G_{1mos}$ |
| $P_{G2out}$ | Outage Probability for group $G_{2mos}$ |

In the next step, the proposed algorithm calculates the MOS distance ($MOS_{dist}$) between the MOS of a specific application flow and the application specific MOS threshold ($MOS_{Thr}$) that we consider to be fixed and known. For a flow f, the $MOS_{dist}$ is defined as the difference between the MOS of flow f and the application specific $MOS_{Thr}$. Note that a different MOS threshold can be used for each application (VoIP, Video, Best effort etc.) running at the end devices. For example, the MOS threshold for VoIP flows can be 3.5, whereas the respective one for video applications can be up to 4.0.

**Algorithm 3** :QoE Assessment Function

**for** $\forall\ c \in C$ **do**
  **for** $u$=1:$U$ **do**
    **for** $f$=1:$F$ **do**
      Calculate $MOS(f)$ according to the E-Model
      $MOS_{dist}(f)$=CalcDistFunc($MOS(f)$,$MOS_{Thr}(f)$)
      **if** $MOS_{dist}(f) \leq 0$ **then**
        Add $f$ to $G_{0mos}$
      **end if**
      **if** 0<$MOS_{dist}(f)$ <1.5 **then**
        Add $f$ to $G_{1mos}$
        Update $P_{G1}$
      **end if**
      **if** $1.5 < MOS_{dist}(f) < 2.5$ **then**
        Add $f$ to $G_{2mos}$
        Update $P_{G2}$
      **end if**
    **end for**
    **if** $\gamma_{u,c} < \gamma_{Thresh}$ **then**
      Add $u$ to $U_{lowSINR}$
    **end if**
  **end for**
  **if** $P_{G0} = P_{G0out}\ for\ t{=}\triangle T$ **then**
    **if** $P_{G1} = P_{G1out}$ & $P_{G2} \neq P_{G2out}\ for\ t{=}\triangle T$ **then**
      Invoke SDNVirtualCellFunc($G1$,$U_{lowSINR}$)
    **end if**
    **if** $P_{G2} = P_{G2out}$ & $P_{G1} \neq P_{G1out}\ for\ t{=}\triangle T$ **then**
      Invoke SDNVirtualCellFunc($G2$,$U_{lowSINR}$)
    **end if**
    **if** $P_{G1} = P_{G1out}$ & $P_{G2} = P_{G2out}\ for\ t{=}\triangle T$ **then**
      Invoke SDNVirtualCellFunc($G1$,$G2$,$U_{lowSINR}$)
    **end if**
  **end if**
**end for**

Once the $MOS_{dist}$ distance has been calculated for each flow, the flows are grouped in three different groups named $G_{0mos}$, $G_{1mos}$ and $G_{2mos}$ according to their MOS distance. Following this, the probabilities $P_{G0}$,$P_{G1}$ and $P_{G2}$, that is, the probability of flow f to be in group $G_{0mos}$, $G_{1mos}$ and $G_{3mos}$ are updated. More specifically, the variable $P_{G0}$ is defined as the probability of an application flow of a specific user to be in group 0 with $MOS_{dist} \leq 0$, $P_{G1}$ is defined as the probability of an application flow of a specific user to be in group G1 with $MOS_{dist}$ between 0 and 1.5 and $P_{G2}$ is the probability of an application flow of a specific user to be in group G2 with $MOS_{dist}$ between 1.5 and 2.5. An outage is considered when the probability of VoIP flows of users having MOS

$\leq MOS_{VoIP_Thresh}$ becomes equal to or greater than the MOS probability threshold of their respective groups. More specifically, the outage in the groups G0, G1 and G2 occurs when:

$$P_{G0out} = P[P_{G0} < P_{G0thresh}] \tag{4.1}$$

In group 1, outage occurs when:

$$P_{G1out} = P[P_{G1} \geq P_{G1thresh}] \tag{4.2}$$

Finally in group 2, the outage occurs when:

$$P_{G2out} = P[P_{G2} \geq P_{G2thresh}] \tag{4.3}$$

Without loss of generality, in the sequel for $P_{G0thresh}$ we use an arbitrary value of 90% i.e. outage is considered if more than 10% of VoIP flows have a $MOS_{dist}$ higher than 0. Note, that in this thesis we considered only VoIP application flows with an application (VoIP) specific MOS threshold of 3.5. This is because it has been observed that for VoIP applications, the minimum MOS for an acceptable QoE is 3.5. In addition to this, we identify and sort the users with low SINR, i.e. users with SINR lower than a specific SINR threshold and add them to the $U_{lowSINR}$ group. To invoke the SDN virtual cell function located in the SDN controller, we compare $P_{G1}$ and $P_{G2}$ with their outage probabilities $P_{G1out}$ and $P_{G2out}$. An arbitrary value of 70% of $P_{G0out}$ as $P_{G1thresh}$ and 30% of $P_{G0out}$ as $P_{G2thresh}$ may be used to obtain $P_{G1out}$ and $P_{G2out}$ using equations 2.2 and 2.3 respectively.

The SDN Virtual Cell function is invoked only at specific events and only for the specific application flows associated with groups $P_{G1}$ and $P_{G2}$ respectively, when they reach their outage probability as shown in algorithm 3. The necessary inputs, for the users in $U_{lowSINR}$ group and the flows in the G1 and G2 group are passed to the SDN virtual cell function. The SDN virtual cell function suggests an appropriate action such as enabling virtual cell configuration, to improve the QoE of the low MOS users based on a global network view. For creating the virtual cells, the SDN virtual cell function considers various parameters like users' location that is estimated using SINR, the traffic load of the serving pico eNB and the neighbor pico eNBs, considering UL and DL traffic separately similar to [45].

### 4.2.4 Simulation Results and Analysis

In this section we present system-level simulations using LTE-Sim [78] to analyze the gains achieved by the proposed framework. We considered a cluster of seven outdoor TD-LTE pico cell hotspots uniformly distributed in a small region. A variable number of users are uniformly distributed within each cell (see table 4.2).

Table 4.2: Simulation Parameters-I

| | |
|---|---|
| pico eNBs | 7 |
| pico eNB bandwidth | 5 MHz |
| TX power | 30 dBm |
| Path Loss Model | L = 140.7 + 36.7 log10(R), R in km |
| Fading Model | Jakes model |
| Scheduler Downlink | EXP Rule |
| Application Traffic Model | G.729 VoIP |

In the simulation, we have evaluated and compared the performance of the proposed framework, with two competing schemes: the Static Configuration (SC) scheme and the Cell Specific Adaptive Reconfiguration (AC) proposed in [33]. The SC scheme considers static configuration of the TDD frame for all cells: the TDD configuration 1. This configuration matches mobile broadband deployment use cases with adequate resources to serve traffic in both UL and DL directions. The AC scheme considers a dynamic frame re-configuration of the TDD frame with an UL/DL reconfiguration timescale of 10 ms given the seven UL/DL configurations available for TD-LTE. The details of the AC scheme are included in [33]. The proposed framework adopts the same configuration with the AC scenario but additionally serves users by employing virtual frame configuration (VC) according to the real time UL and DL user traffic demands. The remaining simulation parameters are in accordance with [78] and are listed in table 4.2.

Fig. 4-4 shows the MOS outage probability that we denote by $P_{G0out}$ and we define as the probability of having VoIP flows with MOS lower than 3.5. In Fig. 4-4 we observe that the proposed framework outperforms both the SC and AC schemes, especially for medium/high load traffic. In low traffic conditions, the improvement in

MOS outage probability is relatively low. Nonetheless, in high traffic conditions, the proposed framework is shown to reduce the MOS outage probability by up to 15% (in absolute values) compared to the SC scenario and up to 10% (in absolute values) compared to AC scenario.

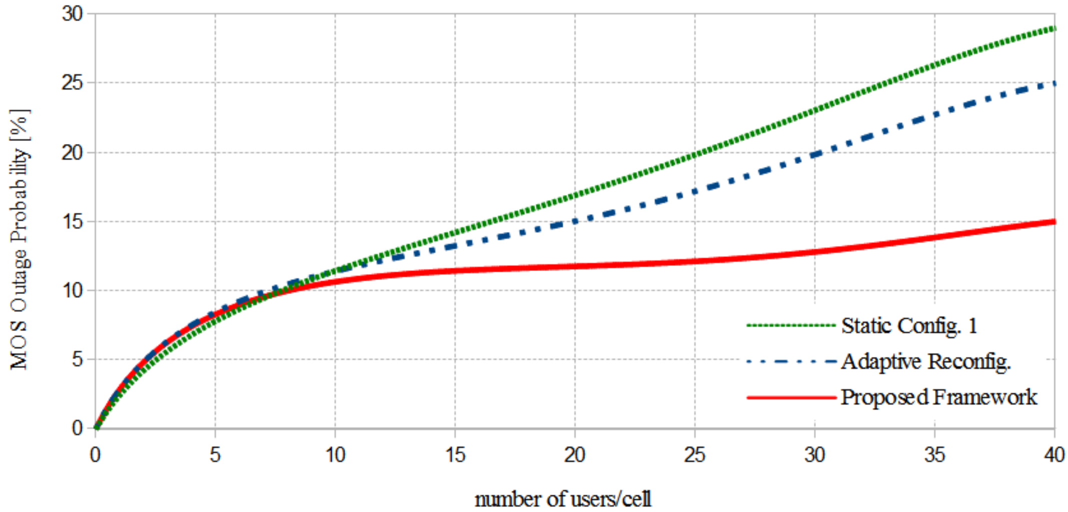


Figure 4-4: MOS Outage Probability

This performance can be explained as follows. In low traffic load conditions, the scheduler at the pico eNB is capable of satisfying the target MOS of VoIP flows and thus, the VC configuration is not utilized extensively. However, when traffic increases, the percentage of VoIP flows with a MOS less than the specific target threshold increases. In such conditions, the proposed framework is employed leading to substantial performance gains. In fact, since each pico eNB has limited resources, an increased traffic load increases the PLR, delay and reduces throughput, leading to QoS and QoE degradation. In such occasions, the outage probability threshold in algorithm 3 is reached and the proposed framework enables users in overlapping regions to dynamically share radio-resources from multiple cells. It is worth noting that the use of the proposed framework improves the QoE of both cell-edge users and users close to the center of the cell. This is because the users served by the virtual cell configuration would release resources that have been previously allocated to them by the scheduler. Accordingly, the scheduler can re-allocate those resources to users in other regions of the cell.

Fig. 4-5 shows the average end-to-end packet delay as a function of the user density (i.e. in terms of number of users per cell). The proposed framework is shown to reduce the end-to-end delay by up to 28% compared to the SC scheme and up to

20% compared to the AC scenario. The reduction of the delay under the proposed framework mainly follows from the reduced processing time at the packet scheduler. In fact, when the proposed framework is activated, the scheduler at the serving pico eNB is enabled to provide additional resources as a consequence of enabling virtual cell configuration. This results in a decrease in the queueing delay which finally reduces the overall end-to-end packet delay. In more general setups with other types of real-time traffic, the reduction of the end-to-end delay is expected to further reduce the playback time and the jitter observed in multimedia content, e.g. video services. Besides, in VoIP applications, reducing the end-to-end delay offer more opportunities for flexible radio-resource management at the serving eNBs and scheduling gaps for enhanced Discontinuous Reception (DRX) at the end terminals.

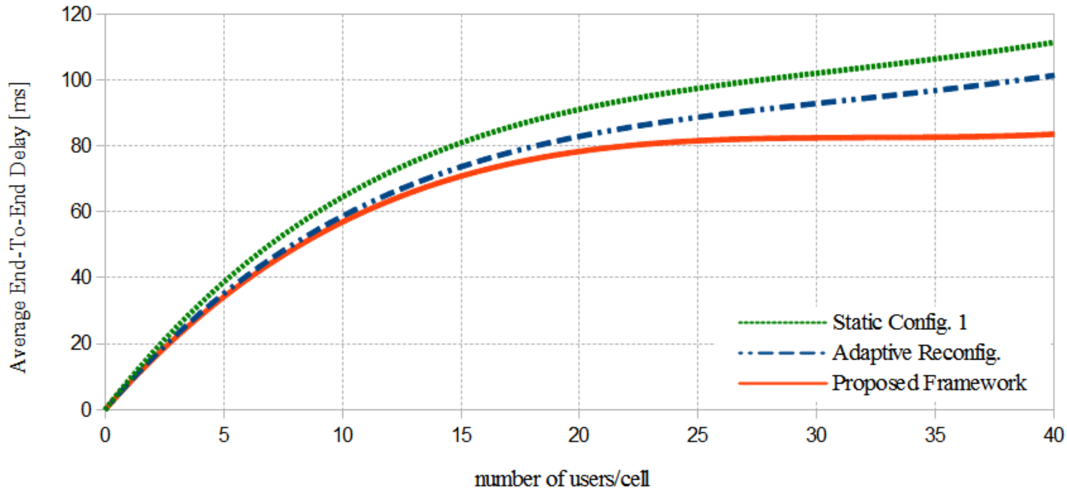


Figure 4-5: Average End-to-End Delay

Fig.4-6 illustrates the CDF of the PLR for the specific scenario with 40 users per cell. Considering that the target PLR for VoIP applications is 1%, the proposed framework is shown to attain an improvement of 15% compared to the SC scenario and 10% compared to the AC scenario (in absolute values).

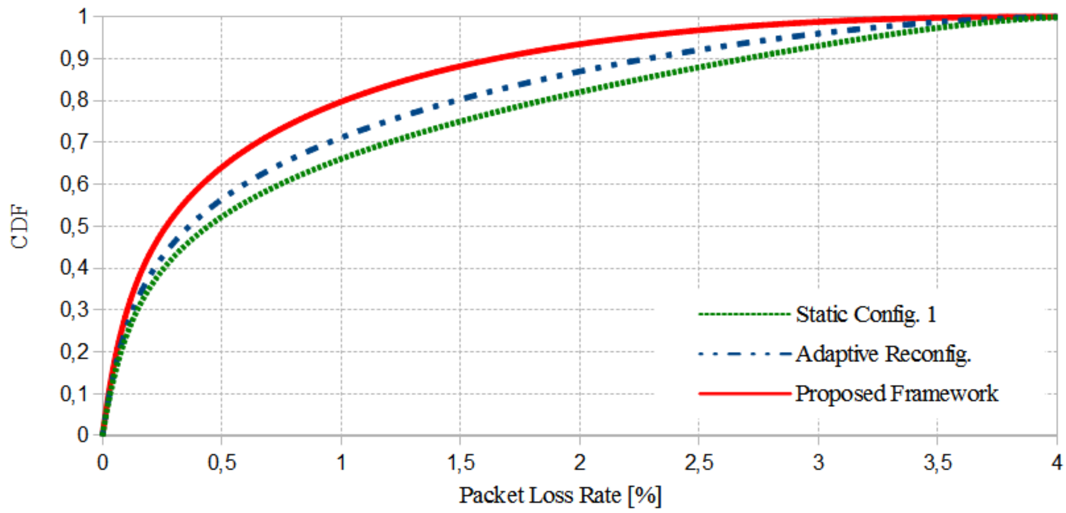


Figure 4-6: Packet Loss Rate CDF

This implies that under the proposed framework, 15% more users have PLR below the target PLR of 1% compared to the SC scenario, whereas, compared to AC, 10% more users are able to have a PLR below the target PLR of 1%. The PLR is one of the key parameters for QoE evaluation and the gains in PLR can provide improvement in the overall QoE for the end users.

## 4.3 SDN Framework for Elastic Resource Sharing in Integrated FDD/TDD LTE-A HetNets

One of the key challenges faced by the mobile network operators today is to support the growing demand for mobile data traffic in a cost effective manner [1]. In this direction, the installation of small-sized base stations into the macrocellular network layout, a.k.a. small cells, has recently drawn significant attention. The integrated cellular network infrastructure of macrocells and small cells is widely termed as Heterogeneous Networks (HetNets). Small cells can boost the area spectral efficiency in the licensed spectrum and bring the cellular network closer to the end user in a cost-effective manner. The support of small cells is integral part of the Long Term Evolution-Advanced (LTE-A) system, which also enables the end users to deploy small cells in an unplanned fashion [79]. Among others, small cells feature edge-based intelligence that enables them to adapt their uplink (UL) and downlink (DL) transmissions in order to avoid cross-tier interference with the macrocell network and support the Quality of Service (QoS) requirements of the associated users. Under this viewpoint, Time Division Duplex (TDD) is considered as a key enabler for achieving flexible and on-the fly adaptation of the UL and DL resources in the small cells. In parallel, the mobile network operators have typically access to a fixed set of network resources, a.k.a. dedicated resources, which can be portions of the licensed spectrum or a set of physical network components, e.g. base stations. Aiming to avoid the under utilization of the physical resources and enable efficient resource sharing among multiple network operators, recently, there has been a surge of interest for leveraging the benefits of SDN in mobile cellular networks [80]. Among others, SDN can reduce network provisioning, enhance network flexibility, and open the road ahead for innovative, dynamic, and cost-effective solutions based on the concept of network re-programmability. Besides, such flexibility creates new business opportunities for the mobile operators and enables on-the-fly network-tuning with respect to the applications or services accessed by the users.

In this section, we mainly focus on the LTE-A HetNet where a common infrastructure provider owns a set of macrocell evolved Node Bs (eNBs), called macro eNBs, and a set of picocell eNBs, called pico eNBs. All the macro eNBs are assumed to use Frequency Division Duplex (FDD), whereas all pico eNBs are assumed to use Time Division Duplex (TDD). The infrastructure provider leases its infrastructure to multiple macrocell operators and multiple picocell operators. Under this model, we address the problem of elastic resource sharing between the FDD macro eNB operators and the TDD pico eNB operators in a dynamic fashion. To achieve this, we propose an innovative SDN-based architecture that enables such dynamic resource sharing among the different operators and employ a dynamic re-configuration mechanism that enables the TDD pico eNBs to adapt their UL/DL ratio in order to release or embody a set of radio resources to/from the FDD macro eNBs operators. System level simulation results demonstrate that the combination of these solutions in the scenario of resource transfer from a TDD pico eNB operator to a FDD macro eNB operator achieves notable performance gains in terms of communication delay in the application layer.

## 4.4 FDD/TDD Elastic Resource Sharing Concept

TDD and FDD operation is an integral part of the baseline functionality of the LTE-A system. Nevertheless, FDD is widely identified more suitable for applications that generate symmetric traffic, e.g. voice-centric services, while TDD is more suitable for serving bursty/asymmetric data traffic, e.g. social media services and machine-to-machine (M2M) communications. In this section, we consider a heterogeneous network with overlaid FDD macro eNBs and underlying TDD pico eNBs that may belong to different network operators. All cellular stations are managed by a common infrastructure provider that leases the infrastructure to the FDD/TDD operators based on Service Level Agreements (SLAs). The infrastructure provider also allocates the frequency bands to the FDD and TDD systems. The use of frequency bands for FDD and TDD transmission modes depend on the geographic region and the SLAs between the infrastructure provider and the network operators. In the scenario considered in this case, the FDD (or TDD) system of a tagged macro (or pico) operator requires more radio-resources to efficiently support the ongoing services of the associated users. To address this requirement, we focus on the scenario where the TDD pico system is capable of leasing a part of its allocated resources to the FDD macro system in a highly efficient yet scalable manner. Accordingly, the FDD macro users can employ carrier aggregation to co-utilize the (potentially distant) set of radio resources of the TDD pico system. This transfer of radio resources, a.k.a. elastic resource sharing, can also take place in the opposite direction, i.e. from the FDD to the TDD system. The FDD and TDD systems can co-exist and complement each other serving different types of traffic provided that interference mitigation is assured [33].

To achieve elastic resource sharing among the two systems, we consider the presence of a centralized SDN Controller, which acts as a resource brokering entity with global resource knowledge. In addition, aiming to efficiently handle the transfer of resources in the case of the TDD system, we also consider that the TDD pico eNBs can re-configure the UL/DL frame ratio in the emitted TDD frames. The aforementioned process is performed in relation to the current resource share and the demands of the cellular users, while it allows for efficient re-adaptation of the TDD resources at the pico eNBs in a timely manner.

### 4.4.1 SDN-based Network Resource Management for FDD/TDD HetNets

Fig. 4.4.1 depicts the proposed SDN-based network management architecture. Without the loss of generality we focus on the scenario where radio-resources are leased from the TDD system to the FDD one, and consider a LTE-A HetNet of one macro eNB operator, coined as Operator A, and one pico eNB operator, coined as Operator Z. The macro eNBs of Operator A use FDD, the pico eNBs of Operator Z use TDD, while all types of eNBs are assumed capable of communicating with the central SDN controller (NV-aware eNBs). The discussion below can be readily extended to the scenario of multiple FDD macro eNB operators and multiple TDD pico eNB operators.

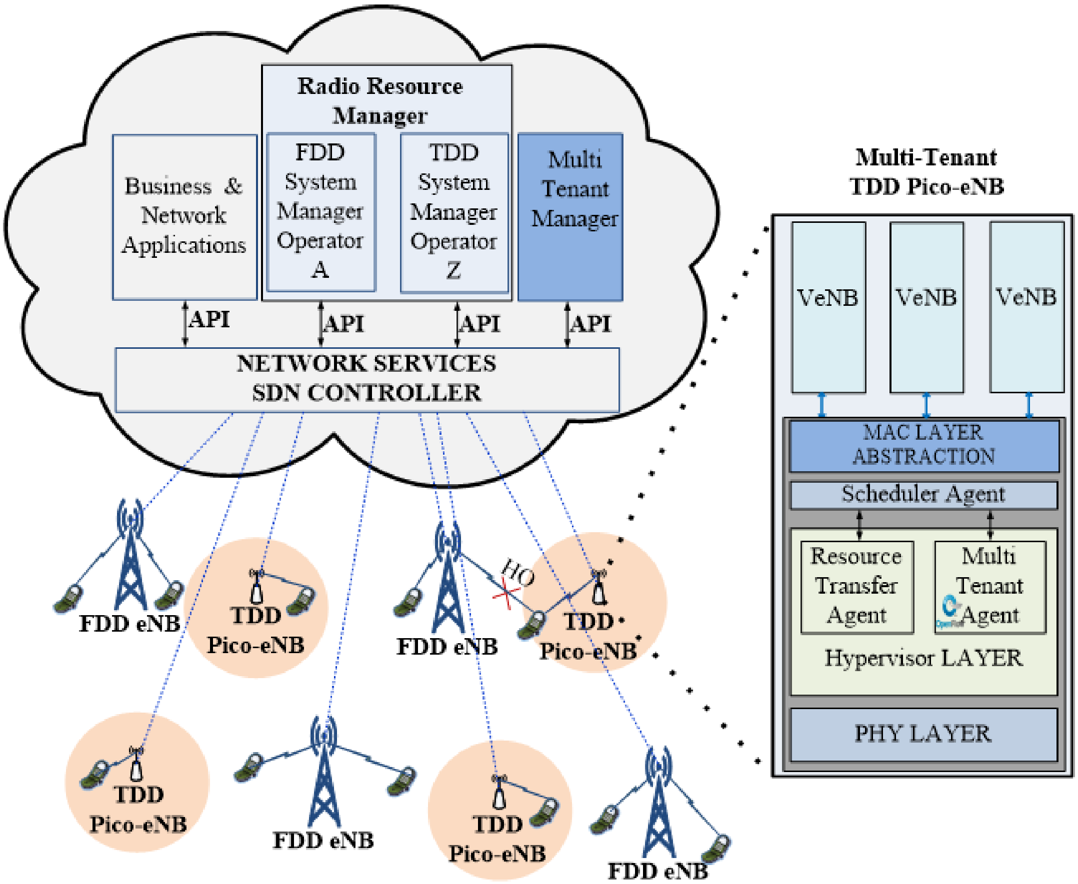


Figure 4-7: SDN-based network resource management architecture

The FDD and TDD network operators have to establish SLAs prior to the employment of elastic resource sharing. Apart from the dynamic sharing of radio-resources, the FDD and TDD network operators may also share the common base station infrastructure so as to enable their users to access the closest base station in proximity [61]. The latter functionality is termed as multi-tenant operation. Aiming to cover both these functionalities we extend the base station virtualization model in [61] by additionally enabling on-demand network reconfiguration capabilities for resource sharing of the underutilized spectrum resources between the FDD and TDD systems.

As shown in Fig. 4.4.1, the intelligence for resource management resides at the SDN Controller, which also provides application programming interfaces (APIs) for over-the-top (OTT) or business applications. Each transmission mode, i.e. FDD or TDD, is managed by a different control application that is capable of acquiring the knowledge of the network state by means of periodic information exchange with the FDD eNBs and TDD pico eNBs. The multi-tenant TDD pico-eNB architecture consists of a Hypervisor that is capable of virtualizing the physical resources to enable multiple operators to share the available bandwidth while remaining isolated from each other. To this end, two agents at the pico eNB are remotely assisted by two distinct manager applications at the SDN Controller: the Multi-Tenant Manager (MTM) and the Radio Resource Manager (RRM).

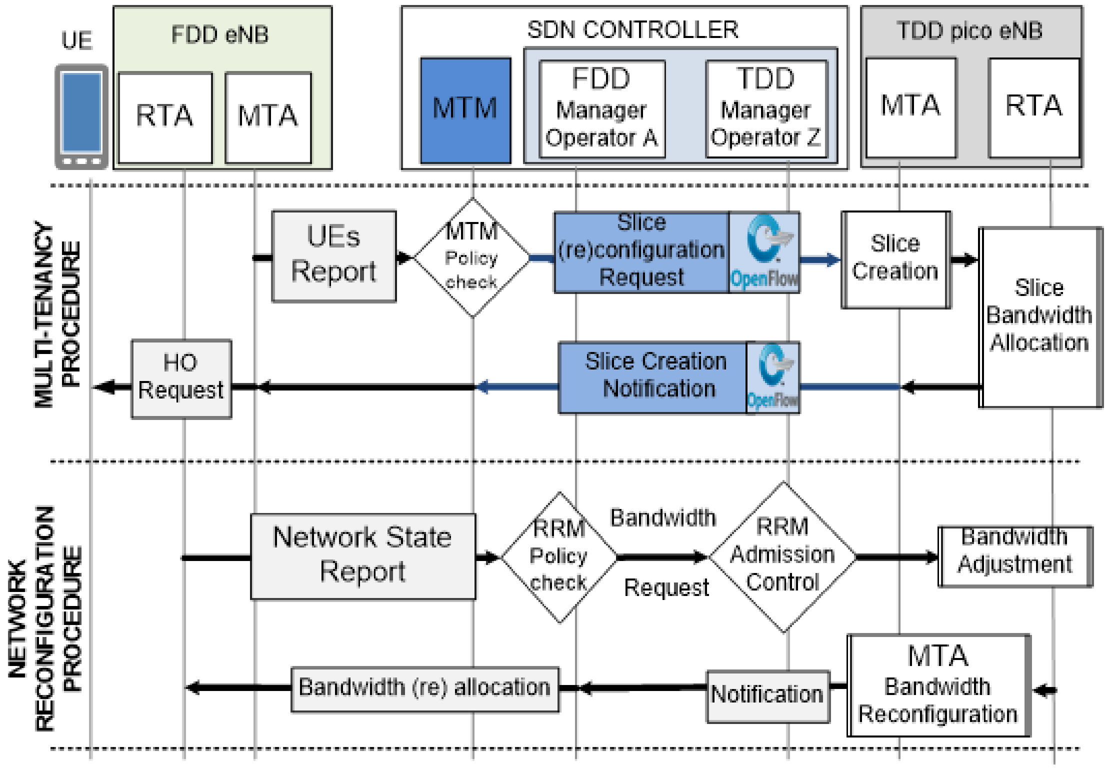

Figure 4-8: Logic Signaling Flow Mechanism

More specifically the MTM is the place where the multitenancy policy resides and where the handover (HO) decision towards the TDD pico eNBs is taken. It is the entity that instructs an agent into the Hypervisor at the pico eNBs, referred to as the Multi-Tenant Agent (MTA), providing it with real-time information to enable efficient sharing of the multitenant pico eNBs into a prescribed number of slices. Note that a slice is represented by the list of users belonging to a tenant operator and the amount of resources that the multitenant pico eNB shares. Another agent, referred to as Resource Transfer Agent (RTA), is responsible for implementing the radio resource transfer procedure. To achieve this, a MAC scheduler agent cooperating with the RTA and MTA provides the tenant operators an abstraction of the MAC layer. The MAC scheduler allocates to each operator a number of the available PRBs in a transparent way, i.e. transparent handling of all the operations related to the real-time configuration of each slice. The upper layers of the protocol stack are emulated in a pool of software applications, referred to as (Virtual eNB) VeNBs. In other words, VeNBs are virtualized instances of an eNB, where each VeNB is managed by a different tenant operator and is logically connected with the core network of the tenant operator.

The logic signaling flow for both the multi-tenancy procedure and the network re-configuration for elastic resource sharing is illustrated in Fig. 4-8. The multi-tenancy policy that enables HO of the users to multi-tenant pico eNBs, takes into account periodic physical measurements that give the MTM a global view of the network state. Once the HO decision is made, the MTM selects the more suitable target multi-tenant pico eNB and sends a slice re-configuration request. It also initiates the HO procedure to the MTA.

The communication between the SDN Controller and each NV aware base station is based on a combination of the OpenFlow protocol [80] and an appropriate high layer protocol, i.e. based on UDP, as in [61]. The MTM uses OpenFlow to instruct the MTA agents, sending them appropriate rules to dynamically configure each slice and enable the delivery of the packets from the users to the appropriate VeNB. Moreover, a high layer protocol is used to permit the exchange of messages between the entities involved in the network re-configuration procedure (RRM and RTA agents of the TDD pico eNBs). The RRM forecasts the resource availability and enables the transfer of resources between the TDD and FDD systems. To this end, the RRM is periodically informed about the network state, (i.e. the bandwidth utilization of each system) by the RTA agents of both FDD and TDD NV-aware base stations. In such a way it can acquire a global knowledge of the network state and distribute the available resources in an efficient way.

The RTA agent of each NV-aware base station is responsible for collecting the bandwidth requests delivered by the tenant operators and to distribute the available bandwidth among them. Such requests are sent to the FDD manager, if such request is performed by an FDD eNB, and forwarded them to the RRM module of the hosting system, e.g. the TDD manager. In this scenario, the TDD manager performs admission control and TDD frame re-configuration by the policy described in section 4.4.1. If the TDD manager is capable of supporting the requests for radio-resources from the FDD tenant operator, it transfers underutilized resources from the TDD pico eNB operator to the FDD macro eNB one. This operation is performed as follows: the

TDD manager instructs the MTA agents of the involved TDD pico eNBs how to adjust the bandwidth availability so as to make part of its resources to the FDD macro eNB operator, while the FDD manager of the respective macro eNB operator allocates the acquired resources among the overloaded FDD eNBs.

### 4.4.2 Dynamic Resource Sharing Among FDD/TDD HetNets

When transferring or sharing resources among different operators there is a need for a mechanism to ensure that the resource gain of a certain operator does not result in starving the users of the other. To regulate such resource transfer we adopted an approach, similar to the Distributed Fair Capacity Based Channel Allocation model elaborated in [72]. Without loss of generality, we focus on the scenario of transferring resources from the TDD operator to the FDD one. Such a process is based on the average capacity gain considering the additional resources provided to the FDD macrocells against the capacity lost on the TDD pico eNBs from where the resources are taken. The actual capacity gain and the capacity loss are estimated using the modified Shannon formula included in [71]. In particular, the capacity gain is calculated considering the mean SINR of all users in the FDD macrocell, where additional resources are transferred, while the capacity loss is estimated on TDD pico eNBs, from where resources are borrowed, taking into account only the experience of the users with the worst SINR, i.e. the users having SINR $< 0$.

Let $C_{FDD,n}$ be the capacity gain for the FDD macrocell when it borrows a specified set of resources from the TDD pico eNBs. The average capacity increase $G_{FDD}$ for the FDD macrocell users is given by:

$$G_{FDD} = \frac{1}{N_{FDD}} \sum_{n=1}^{N_{FDD}} C_{FDD,n} \tag{4.4}$$

where, $N_{FDD}$ are the total number of users in FDD macrocell. If the capacity loss in the TDD pico eNBs is less than the average increase in capacity gain for the FDD macrocell users, then the requested resources can be transferred, while assuring the desired user QoS for the TDD pico eNBs. In other words, the resources are transferred only if $G_{FDD} > L_{FDD}$, where $L_{FDD}$ is the capacity loss experienced by the worst SINR users in the TDD pico eNB region. It should be noted that the capacity gain and the capacity loss calculation for transferring resources in the opposite direction, i.e. from FDD to TDD, follows a similar process.

We envision that the resource transferring condition described above assists the SDN controller to take the decision regarding resource sharing among different operators provided that the amount of desired resources, i.e. PRBs, is communicated or estimated by the SDN Controller. Communicating the desired resources could easily be accomplished via the north bound API from where application providers can ask the SDN Controller regarding particular QoS, e.g. capacity. Alternatively, the FDD macrocell can signal the SDN controller requesting additional resources as soon as it starts experiencing congestion, which can be provided by exploiting the maximum amount of resources that can borrowed from the overlapping TDD pico eNBs. Since the SDN Controller has knowledge of the network state, i.e. load associated with macrocells and pico eNBs and the respective interference levels, it can estimate the maximum amount of resources it can transfer from the TDD pico eNBs that can satisfy the capacity gain and capacity loss resource transfer condition.

The SDN controller may also have knowledge of the specific TDD UL and DL resource utilization and hence can further optimize the UL/DL ratio associated with particular TDD pico eNBs with the objective to increase the amount of resources that can be transferred towards the FDD macrocell. The rationale here is to determine an appropriate UL/DL frame with respect to the residing user UL/DL demands, which can free more resources to be transferred towards the FDD macrocell. The mechanism for determining the corresponding UL/DL frame can be based on the same heuristics as the ones used in [76]. It should be noted that the amount of resources that can be shared among the FDD and TDD operators may vary depending on particular QoS demands and traffic conditions.

### 4.4.3 Simulation Results and Analysis

In this section, some system level simulations results are presented, which helps verify and assess the performance gains attained from the proposed SDN-based framework for elastic resource sharing in LTE-A HetNets. We consider a LTE-A HetNet composed by of FDD macro eNBs and TDD pico eNBs of a different operator. Two different operators are assumed to operate the macro and the pico eNB infrastructure, respectively.

The TDD pico eNB operator is willing to share part of its resources. We focus our analysis to the scenario where the to the FDD macro eNB operator, depending on the resource availability in both systems. The elastic resource sharing is performed based on SLAs that defines the maximum bandwidth that can be transferred from the TDD pico eNB operator to the FDD macro eNB operator and vice-versa.

In more detail, we consider a LTE-A HetNet with seven FDD macro cells and seven pico eNBs that are uniformly dropped within the FDD macro area with a minimum separation distance of 120m. The users of both the FDD macro and TDD pico eNB systems are uniformly distributed within the cellular coverage, whereas two different types of traffic are assumed between them. The FDD macro users are considered to have a H.264 encoded video streaming traffic at 440 kbps, while the TDD pico users host Poisson-distributed FTP traffic that is modeled in line with the methodology in [74]. The macro eNBs are assumed to utilize the EXP scheduler, which gives priority to real time DL packets which are buffered for more than a target delay threshold of 0.1 sec [78]. On the other hand, the traffic at the pico eNBs is randomly and independently generated in the UL and DL directions. Since we are not interested in the impact of user mobility in the TDD pico eNB system, we consider uncorrelated slow fading. The path-loss model for both systems are adapted based the 3GPP case 1 model defined in [74]. The remainder parameters of our simulation model are summarized in table 4.3.

Table 4.3: System Simulation Parameters-II

| Radio Configuration Parameters | |
|---|---|
| **Parameter** | **Value** |
| Macro eNB Inter Site Distance | 500 m |
| Pico eNB Minimum Separation Distance | 120 m |
| Shadowing Standard Deviation-Macro Cell | 8 dB |
| Shadowing Standard Deviation-Pico Cell | 10 dB |
| Spectram Allocation-Macro Cell | 10 MHz UL/DL |
| Spectram Allocation-Pico Cell | 10 MHz UL/DL |
| Max Tx Power-Macro Cell | 46 dBm |
| Max Tx Power-Pico Cell | 30 dBm |
| Duplexing Scheme | TDD |
| Antenna Gain-Macro/Pico Cell | 15/5 dBi |
| eNB Antenna Gain | 15 dBi |
| UE Total Tx Power | 23 dBm |
| UE Antennal Gain | 0 dBi |
| UE-Macro eNB Path loss Model | 128+37.6 $log_{10}$(R), R in Km |
| UE-Macro eNB Path Loss | 140.7+36.7 $log_{10}$(R), R in Km |
| Macro eNB Fading | Jakes Model |
| Macro eNB Scheduler (DL) | EXP Rule-target delay=0.1 s |
| Number of PRBs, NRB | 50 |
| PRB size, RBs | 180 kHz |
| Traffic Model-Macro Cell | H.264 Video Streaming |
| Traffic Model-Pico Cell | FTP (File Size = 0.5 MB) |

Since the TDD re-configuration scheme is used for optimizing the UL/DL ratio in the TDD pico eNB system only, in Fig. 4-9 we plot the cumulative distribution function (CDF) of delay for the H.264 video traffic in the DL of the FDD macro eNB system under two different schemes. The first scheme, coined as the FDD baseline scheme, corresponds to the performance of the macro eNB system without using the proposed

SND-based framework. On the other hand, the FDD with SDN scheme corresponds to the performance of the FDD macro eNB system under the proposed SDN-based framework. As expected, above the target delay threshold of 0.1 seconds, the performance of both schemes the baseline and the SDN-based approach is the same. However, a notable delay gain is observed for the SDN-based scheme below this threshold, i.e. the CDF delay of the SDN-based scheme is higher than that of the baseline scheme. Interestingly, the employment of SDN-based elastic resource sharing reduces the H.264 traffic delay in the DL direction by up to 21% compared to the baseline scenario where no SDN based sharing is applied. This improvement follows from the utilization of additional resources provided by the TDD pico eNB system.

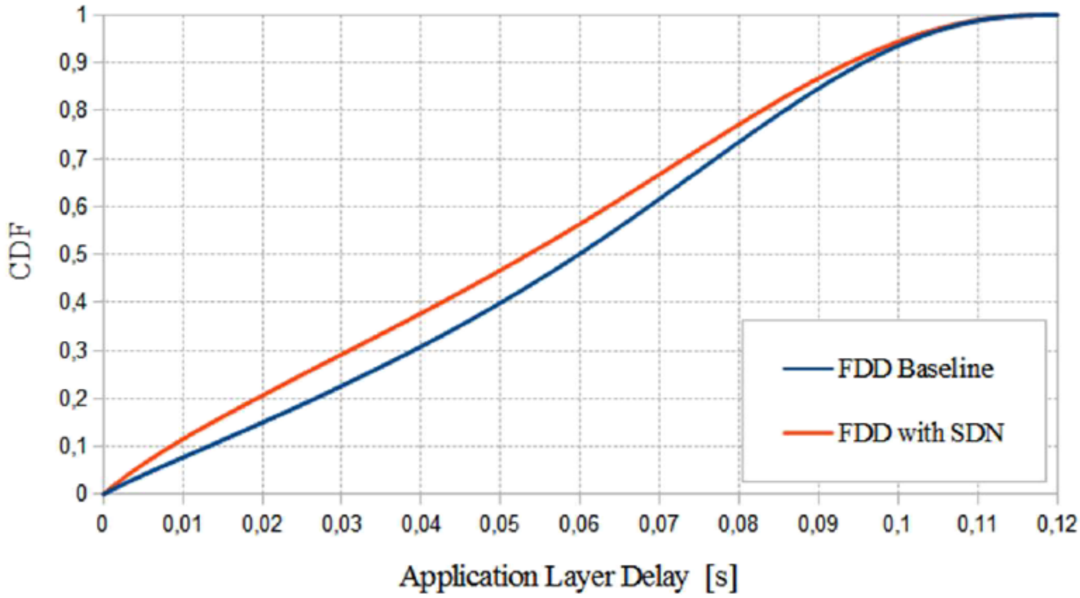

Figure 4-9: CDF-Application layer delay in the FDD macro eNB system

Although the reduction of the application layer delay in the FDD macro eNB system is prominent in Fig. 4-9, the negative impact of utilizing less radio-resource in the TDD pico eNB system, i.e. the ones that have been shared with the FDD macro eNB system, should also be investigated.

In Fig. 4-10, we plot the CDF of the application layer delay in the TDD pico eNB system under three different schemes: a) no sharing of resources with the FDD system and no reconfiguration of the TDD frames (blue line), b) SDN-based elastic resource sharing with the FDD system and no reconfiguration of the TDD frames (green line), and c) SDN based elastic resource sharing with the FDD system and reconfiguration of the TDD frames (red line). In the first two schemes, the reference UL/DL configuration-1 (60% downlink and 40% uplink) has been considered for the TDD frames at the pico eNBs [33]. In the third scheme, we consider an UL/DL reconfiguration timescale of 10ms and the seven UL/DL configurations available for TDD-LTE [33].

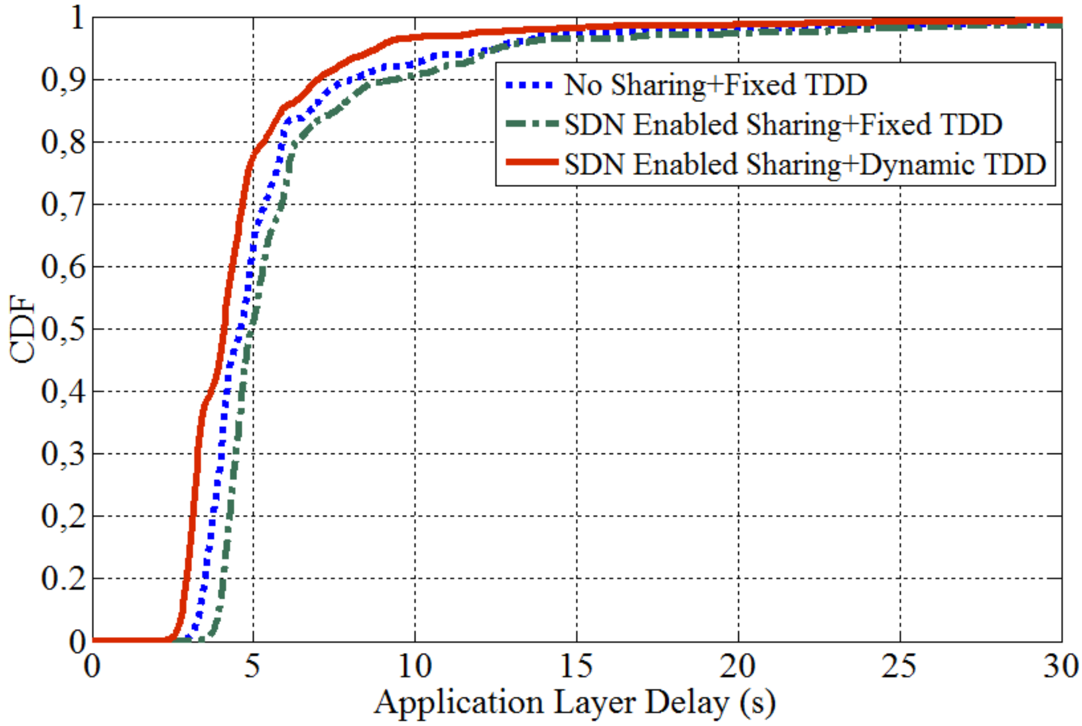


Figure 4-10: CDF-Application layer delay in the TDD macro eNB system

As shown in Fig. 4-10, the sharing of resources with the FDD macro eNB network degrades the performance of the TDD pico eNB system, if a static UL/DL configuration is applied, i.e. compare the blue and the green lines. On the contrary, if the elastic resource sharing between the two systems is combined with the dynamic reconfiguration of the TDD frames on a per pico eNBs basis (red line), the employment of the proposed SDN-based framework is shown to attain notable performance gains for the TDD pico eNB system as well. This performance improvement mainly follows from the efficient adaptation of the UL/DL ratio with respect to the ongoing user services and the resource availability in the TDD pico eNB network.

## 4.5 Summary and Conclusion

This chapter presented two main concepts: First, a novel SDN-based framework that enables QoE-aware network management for OTT mobile services. Building on-top of existing approaches for dynamic UL/DL frame reconfiguration, this chapter introduces

QoE-centric SDN-based framework that increases user satisfaction in TD-LTE networks with picocells. Within this framework, the virtual cell concept that is discussed in detail in chapter 3, has been applied to improve the performance of the overall system and the QoE of the end users. Among other benefits, the proposed framework has been shown to enable elastic radio resource management and significantly reduce the probability of a user being in MOS outage, enhancing thus the user perceived QoE. Considerable reductions in end-to-end packet delay and packet loss rate have also been observed, as compared to other competing schemes, when the virtual frame configuration is activated.

The second concept presented and discussed in this chapter includes: A novel SDN-based framework for elastic resource sharing in LTE-A HetNets with multiple network operators. Here, the framework considered focuses on a scenario with a FDD macro eNB operator and a TDD pico eNB operators with multiple eNBs and proposed a novel SDN-based framework that enables efficient on-the-fly elastic sharing of the radio resources between the FDD macro eNB and the TDD pico eNB systems. The proposed framework consists of a novel SDN-based architecture and a dynamic TDD frame reconfiguration for pico eNBs. The SDN-based architecture enables multi-tenant operation for the common physical network infrastructure as well as elastic resource sharing between the TDD and FDD systems in a dynamic way. The proposed TDD frame re-configuration algorithm within the SDN-based framework dynamically adjusts the UL/DL ratio of the pico eNB frames, so as to utilize the minimum possible radio-resources while preserving the bandwidth requirements of the pico eNB users. System-level simulation results have shown that the combination of the SDN-based architecture and the TDD frame re-configuration algorithm can significantly reduce the application layer delay in the DL of both the FDD macro eNB and the TDD pico eNB systems.

# Chapter 5

# 5G-TDD Network Slicing Solutions

## 5.1 Introduction

To support multi-tenancy, on-demand resource provision and resource flexibility is essential. In this direction, an increasing body of work considers network programmability and SDN control as an effective solution. SDN enables application and service providers considering vertical markets and MVNOs to communicate their service requirements to the infrastructure provider, which in turn can program the network resources to meet the Quality of Service (QoS) constraints [5] [17]. Network programmability typically involves the configuration of operational parameters and network resource slicing towards particular tenants.

Network resource slicing ensures isolation and customization of resources allocated for each tenant, allowing dedicated use that can accommodate specific application requirements. This feature enhances resource utilization, provides means to control network congestion and enables diverse services to be supported even with conflicting QoS requirements in each isolated slice [81]. However, at the same time, slicing reserves the network resources for an explicit use and allocating them to a particular tenant for a fixed time duration might lead to notable losses in multiplexing gain [82]. A survey on Resource Slicing in Virtual Wireless Networks is presented in [29]. While, A network slicing concept applied to a multi-cell RAN shared among multiple tenants is presented in [27]. Further discussion in detail on state-of-the art related to the novel ideas presented in this chapter can be found in the literature review chapter 2. In contrast to the state-of-the art, this chapter introduce the network slicing concepts in the emerging 5G TDD networks and evaluate the performance of the proposed concepts with the state of the art schemes. In particular, different mechanisms for allocating network resources considering adaptation of UL/DL ratio and network slicing are discussed and their performances are analyzed. An analytical model formulating the network slice resource allocation as a weighted optimization problem is also presented. To optimize the capacity allocated to each application specific slice and adapting the corresponding UL/DL ratio, a newly introduced resource allocation metric is considered in the analytical model. Additionally, this chapter also elaborates the SDN architecture that allocates and adjusts network slice resources including the corresponding TDD frame re-configuration.

The rest of this chapter is organized as follows. Section 5.2 introduces the service oriented network slicing concept and the proposed network resource virtualization framework, including the SDN-based architecture, resource slicing process and TDD UL/DL ratio re-configuration. Section 5.2.4 and 5.3.4 describes the simulation environment and analyzes the performance evaluation results. Section 5.3 first discusses the TDD network slicing concept. Following this, the traffic forecast based adaptive 5G TDD network slicing with help of SDN architecture and the UL/DL resource re-configuration schemes is introduced. The analytical model for creating application-oriented slices together with UL/DL frame-reconfiguration is elaborated in section 5.3.3. Finally, section 5.4 concludes this chapter.

## 5.2 Service Oriented Network Resource Virtualization Framework

This section presents the proposed framework for service oriented network virtualization. Initially, it introduces the network architecture and the associated building blocks and later the proposed mechanisms and solutions which are used for the performance evaluation.

### 5.2.1 SDN-based Network Architecture

The SDN-based network architecture adopted in the proposed framework facilitates network resource allocation for MVNOs and vertical market players, and provides on-demand network resource slicing addressing the requirements of diverse services. Fig. 5-1 illustrates the main building blocks and the corresponding interfaces considering backward compatibility with the 3GPP LTE system.

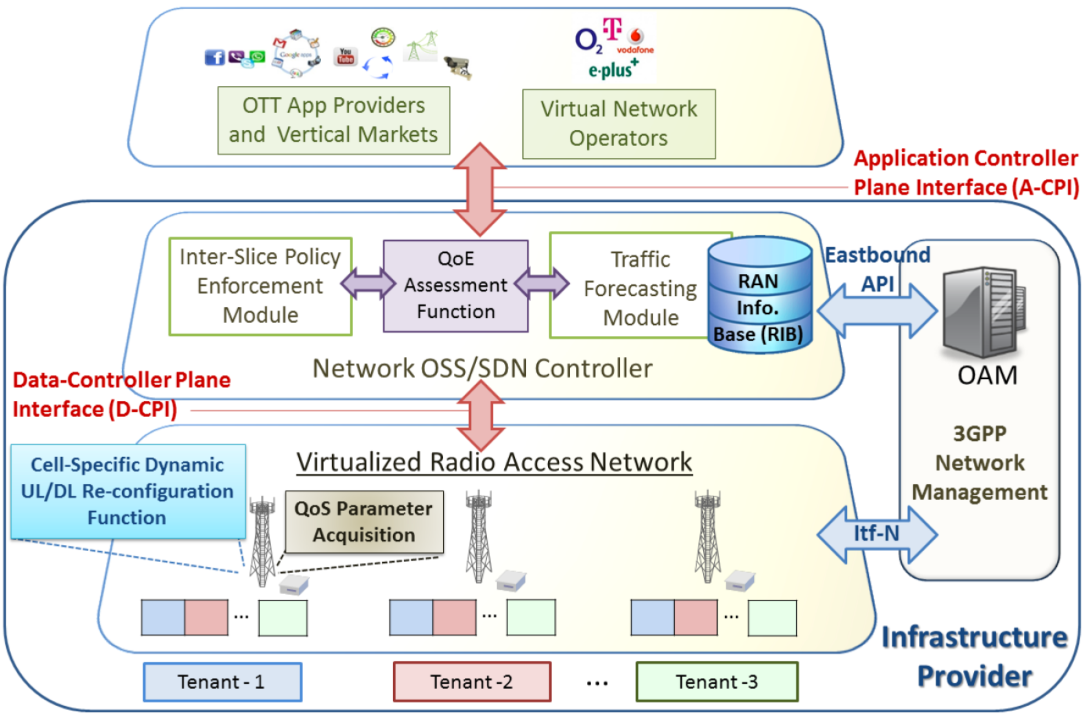


Figure 5-1: SDN-based network virtualization architecture

The main component of the proposed network architecture is the SDN controller, which receives service requests, specifying the desired SLA and time duration, from MVNOs and third parties via the Application-Controller Plane Interface (A-CPI). The SDN controller processes such incoming requests and establishes a corresponding network resource slice if the desired SLA is fulfilled providing also a QoE assessment control service through the Data-Controller Plane Interface (D-CPI). The SDN controller interacts with the legacy 3GPP Operations Administration and Maintenance (OAM) system via the Eastbound Application Program Interface (API), being able to retrieve network management information. Such information is obtained from the RAN via the legacy Northbound Interface (Itf-N) and includes the UL/DL load and the interference map of the entire network as well as the status of a number of Key Performance Indicators (KPIs), e.g. handover failures, throughput, etc. [69]. The collected network information feeds the RAN Information Base (RIB) facilitating a global network view to the SDN controller, which supports the following functions:

- Inter-slice policy enforcement: ensures that each slice receives the corresponding network orchestration policy, traffic prioritization and UL/DL frame re-configuration.

- QoE assessment function: collects application specific QoS information from base stations through D-CPI and provides the corresponding QoE computation and assessment.

- Traffic forecasting module: uses the RIB global network load view to forecast application specific traffic demands at regular short-term time intervals.

Once the SDN controller receives an incoming service request via the A-CPI, it consults the traffic forecasting module to ensure that the inquired resources are available for the time duration of the request. It then converts the requested SLA to slice related policy, using the inter-slice policy enforcement module and at the same time also provides the TDD UL/DL ratio and the frame re-configuration timescales. The SDN controller uses the D-CPI to configure the network slice accordingly. It also installs a QoS parameter acquisition function at each base station that collects performance information such delay, loss, throughput, etc., to feed the QoE assessment function, which makes the application performance visible to the SDN controller.

Each base station supports network virtualization by adopting a hypervisor, which ensures isolation among different service slices or tenants. To effectively reflect the evolving short term traffic patterns, base stations maintain a cell-specific dynamic UL/DL re-configuration function. An instance of such a function operates on each slice separately based-on the corresponding traffic variations, providing in this way a means to adjust the UL/DL ratio for each slice independently, meeting more accurately the evolving traffic demands. By providing dynamic UL/DL frame re-configuration explicitly for individual service-specific slices within each cell, we can reduce the losses in multiplexing gain and maximize the network resource utilization.

### 5.2.2 Service-oriented Cell-specific Dynamic Re-configuration

As mentioned previously, the state-of-the art CSDR [33] was proposed to improve the resource utilization in TDD networks allowing base stations to adopt different UL/DL ratios. This scheme is considered as the baseline for evaluating the proposed framework. Typically, CSDR is adopted in a network arrangement without considering network virtualization, taking into account the evolution of the aggregate traffic load, with the objective to enhance the resource utilization of the system. In CSDR, dynamic TDD frame re-configuration is performed at a timescale of 10 ms to match the instantaneous UL/DL traffic demands. Different timescales such as 100ms and 640ms may also be selected depending on the UL/DL traffic variations and traffic asymmetry.

The selection of appropriate frame configuration is performed according to the estimation of the required UL/DL sub-frames and as per the following equation [64]

$$N_{sub}^{DL} = \frac{\sum_{i=1}^{N_{UE}^{DL}} b_i^{DL}(t)/R_i^{DL}}{\sum_{i=1}^{N_{UE}^{UL}} b_i^{UL}(t)/R_i^{UL} + \sum_{i=1}^{N_{UE}^{DL}} b_i^{DL}(t)/R_i^{DL}} \cdot 10 \tag{5.1}$$

$$N_{sub}^{UL} = 10 - N_{sub}^{DL} \tag{5.2}$$

Where $b_i^{DL}(t)$ and $b_i^{UL}(t)$ is the buffer size for user $i$ at time $t$, while $R_i^{DL}$ and $R_i^{UL}$ is the past average data rate for the user $i$ at time $t$ in DL and UL directions respectively. Note that 10 is the total number of subframes in a TD-LTE frame. The frame configuration that matches closest calculated UL/DL ratio, is employed. The resources are then allocated to each application flow in UL/DL direction according to the policy adopted by the scheduler. In our scenarios we use the EXP scheduler [78], which prioritizes real time flows by using weights that are calculated as per the following equation.

$$w_{i,j}^{RT} = \alpha_i \cdot \exp\left(\frac{\alpha_i D_{HOL_i} - x}{1 + \sqrt{x}}\right) \cdot \frac{r_{i,j}}{R_i} \tag{5.3}$$

$$where \qquad x = \frac{1}{N_{RT}} \cdot \sum_{i=1}^{N_{RT}} \alpha_i D_{HOL_i} \tag{5.4}$$

Where, $w_{i,j}^{RT}$ is the weight of flow $i$ for the resource block $j$. We assume that only one application runs on each user's device, such that the application flow $i$ corresponds to the individual user $i$. $\alpha_i$ is a factor that takes into account the packet loss rate probability of flow $i$. $D_{HOL_i}$ is the head of line delay for flow $i$, $r_{i,j}$ is the instantaneous data rate of flow $i$ for resource block $j$ and $R_i$ is the past average data rate of flow $i$ respectively. The weight for non-real time traffic, $w_{i,j}^{NRT}$ is given by:

$$w_{i,j}^{NRT} = \alpha_i \cdot \frac{r_{i,j}}{R_i} \tag{5.5}$$

It is worth noting that in CSDR, even though the scheduler may prioritize certain application flows as per equations 3-5, the application flows are scheduled following a common UL/DL TD-LTE frame configuration, selected out of the 7 different TD-LTE configurations defined in [32]. The adopted TD-LTE frame configuration, including subsequent re-configurations, consider all application flows in UL/DL as an aggregate traffic load. Hence, it may not match accurately the UL/DL requirements of individual applications.

To resolve this issue, we propose an application-oriented CSDR, where the scheduler is modified to consider the corresponding load of a specified single application only, instead of the aggregate traffic load when selecting the UL/DL configuration. In such a scheme, a selected application is assigned the highest priority in the resource allocation process. The scheduler prioritizes the tagged application flows by adjusting the EXP scheduler weights as follows:

$$w_{i,j} = \varphi_i \cdot w_{i,j}^{RT} \tag{5.6}$$

Where $\varphi$ is the adjusting factor, $\varphi \in [0,1]$. A higher $\varphi$ indicates a higher priority for the respective application flow type. The performance of the application-oriented CSDR are evaluated in section IV and compared with the proposed framework.

### 5.2.3 Service-oriented Slice-explicit Dynamic TDD

The proposed service-oriented slice explicit dynamic TDD solution permits a service/application independent configuration of the UL/DL ratio. In other words, each network slice may adopt a different UL/DL ratio and separate CSDR, in order to accommodate best the QoS requirements of the corresponding service. A simple example, illustrated in 5-2, provides an overview of the proposed concept highlighting the difference to the baseline CSDR scheme. In the baseline CSDR, where network slicing is not applied (referred to as state of the art in 5-2), two neighboring base stations can adopt a different UL/DL ratio, but with each base station employing the same CSDR across the range of the 3 sub-carrier groups (SCG). In the proposed solution, each sub-carrier represents a different network slice allocated to a certain tenant, which can support a different CSDR.

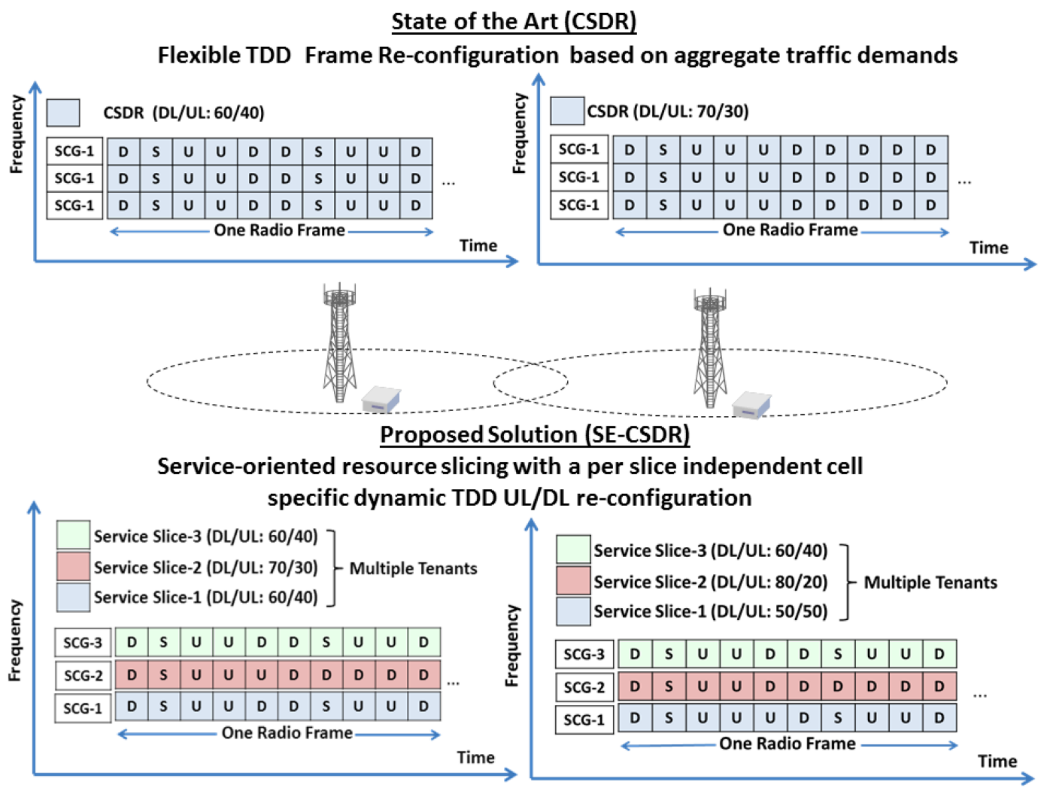


Figure 5-2: A simple example comparing the baseline CSDR scheme with the proposed service-oriented slice explicit CSDR (SE-CSDR)

The proposed Slice Explicit CSDR (SE-CSDR) solution, avoids increasing the cross-slot interference among users served by different network slices, i.e. inter-slice users, by allocating the same sub-carriers for each slice across the entire RAN. Hence, inter-slice interference is avoided due to the utilization of distinct frequencies, ensuring isolation. To establish a network slice after receiving a service request with a specified SLA and time duration, the SDN controller relies on the traffic forecasting module that provides information for performing admission control. The traffic forecasting adopted in this work is based on the Holt-Winter's exponential smoothing mechanism [83] [84] and performs a service/application specific traffic forecasting.

Once a network slice is created, customization is provided by allowing the corresponding SLA requirements to be converted into an inter-slice policy. Isolation among different network slices allows the operation of an independent scheduler and CSDR, improving the resource utilization efficiency, while reducing the loss of multiplexing gains, as resource scheduling is performed considering the instant UL/DL traffic load per slice. In addition, isolation ensures that traffic variations within a particular network slice cannot impact negatively the performance of other applications allocated into different network slices.

### 5.2.4 Simulation Results and Analysis (Phase-1)

In this section we present system-level simulations to assess the performance of the proposed framework using LTE-Sim as the simulation tool [78]. We considered a scenario with seven TD-LTE macro eNBs with three sectors each, thereby altogether forming 21 macro cells. In each macro eNB users are uniformly distributed and users' mobility is considered following the SLAW mobility model described in [85] The detailed list of simulation parameters are given in 5.1 [78] [74].

Table 5.1: Simulation Parameters-I

| | |
|---|---|
| eNBs | 7 |
| eNB inter site distance | 500 m |
| eNB bandwidth | 10 MHz |
| TX power | 30 dBm |
| Path Loss Model | L = 140.7 + 36.7 log10(R), R in km |
| Fading Model | Jakes model |
| Users | 40 Users/eNB |
| Application Traffic Model | G.729 VoIP, H.264 Video, BE Web browsing |
| Traffic Prediction Model | Holt-Winter's exponential smoothing |

In the simulations, we have considered the joint support of three different traffic models. In particular, we have considered that 30% of the users employ Voice over IP (VoIP), 30% of the users stream video and 40% of the users generate best effort (BE) traffic. The VoIP traffic is modeled by an ON/OFF Markov model, while Video traffic is modeled using H.264 foreman traces encoded at 440 kbps [78]. The BE traffic follows the 3GPP web-browsing traffic model as per TR 25.848, with Pareto distributed packet size. In the simulation scenario we assume that the three aforementioned application services are offered by three different OTT providers. The proposed SE-CSDR solution is compared against the baseline CSDR, and two variations of the application-oriented

CSDR solution, named the VoIP-oriented CSDR and the Video-oriented CSDR, which prioritize VoIP and video applications respectively.

To measure the impact on QoE for the users on a per application/service basis, we measured the Mean Opinion Scores (MOS). MOS is used to evaluate QoE of specific applications. Models to map QoS parameters to MOS scores are created based on subjective and objective experiments. Specific applications are evaluated in different ways using subjective methods under controlled QoS values, which are then mathematically mapped to MOS scores of the respective application. Further details on methodologies to create QoE evaluation models for specific applications and the performance analysis of QoE models of different applciations are discussed in [86] [87]. For measuring the VoIP MOS, an ITU-T objective model known as the E-model is used [88], while for the video application flows, a simplified video MOS model is employed [89]. Video MOS is measured as the Mean-Squared-Error (MSE) averaged over all frames of the video sequence as a function of distortion. It has two components namely the source distortion $D_S$ and loss distortion $D_L$ [89][90].

$$MSE = D_S + D_L = \eta \cdot R^{\xi} + \Gamma \cdot PEP \tag{5.7}$$

Where $\eta$,$\xi$ and $\Gamma$ are model parameters that have different values for different types of video sources. PEP is the packet error probability. A user data rate sensitivity factor ,$\gamma \in [0,1]$, can be introduced to emulate dissimilar users' sensitivity to the data rate as follows:

$$MSE = D_S + D_L = \gamma \cdot \eta \cdot R^{\xi} + \Gamma \cdot PEP \tag{5.8}$$

The Peak Signal-to-Noise Ratio (PSNR) is a function of MSE and a widely used objective quality indicator for video. PSNR can be expressed as:

$$PSNR(dB) = 10 \cdot \log_{10} \frac{255^2}{MSE} \tag{5.9}$$

The PSNR and MOS relation can be mapped as:

$$MOS = \begin{cases} 1 & 0 < PSNR < 20 \\ 1 + \frac{3.5}{20} \cdot (PSNR - 20) & 20 < PSNR < 40 \\ 4.5 & PSNR > 40 \end{cases} \tag{5.10}$$

For the VoIP application, a minimum target MOS of 3.5 is considered, which is sufficient to provide an acceptable QoE to the end users. For video applications acceptable MOS is between 3 to 3.5, while a MOS above 3.5 is considered as excellent [90].

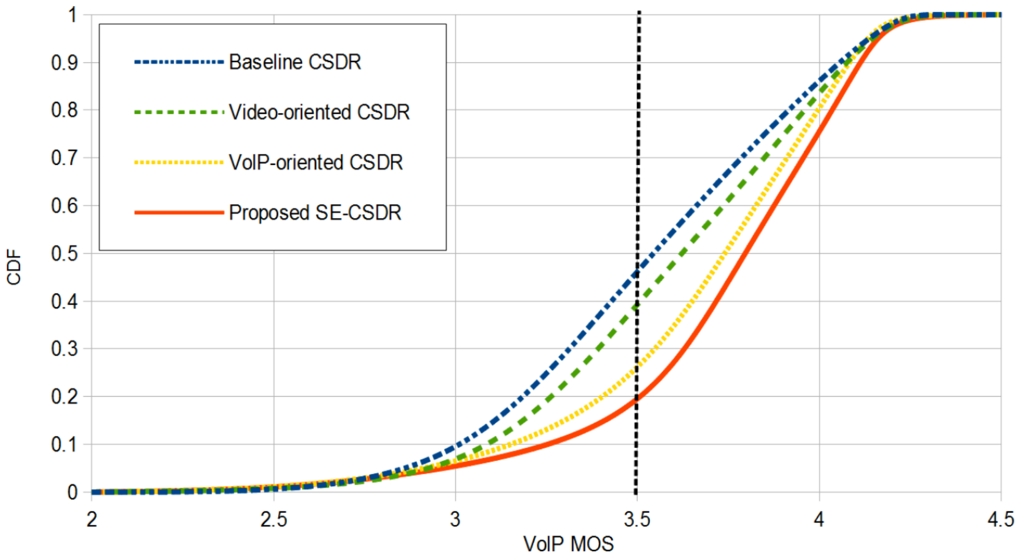

Figure 5-3: VoIP MOS CDF

Fig. 5.2.4 shows the MOS cumulative distribution function (CDF) for the VoIP traffic. From Fig. 5.2.4, it can be observed that the proposed SE-CSDR solution outperforms the baseline CSDR showing 26% gains considering the VoIP target MOS. Even compared with the VoIP-oriented CSDR solution, the gain of the proposed SE-CSDR is 8%, while the Video-oriented CSDR shows only 5% improvement compared to the baseline CSDR. The performance gains of the SE-CSDR are higher than the VoIP-oriented CSDR, despite the fact that this scheme prioritizes VoIP flows and adapts the UL/DL frame configuration to match VoIP traffic. The reason behind this is based on the fact that with a VoIP-oriented CSDR, all applications use a common scheduler and share the same CSDR. Hence the entire available spectrum has to be shared among all the users, unlike the proposed SE-CSDR that employs network slicing, which ensures resource isolation and allows the use of a different CSDR per slice.

The Video-oriented CSDR works similar to the VoIP-oriented CSDR, but this time video traffic flows are prioritized instead of VoIP, while the CSDR process is based-on video real time demands. Consequently, Video-oriented CSDR favors the performance of the video flows, keeping the MOS performance of the VoIP flows close to the baseline

CSDR. The Video MOS CDF is shown in Fig. 5-4. The proposed SE-CSDR outperforms the other solutions, with the Video-oriented CSDR performing closer to the proposed SE-CSDR, but with a difference of approximately 10% in performance gains.

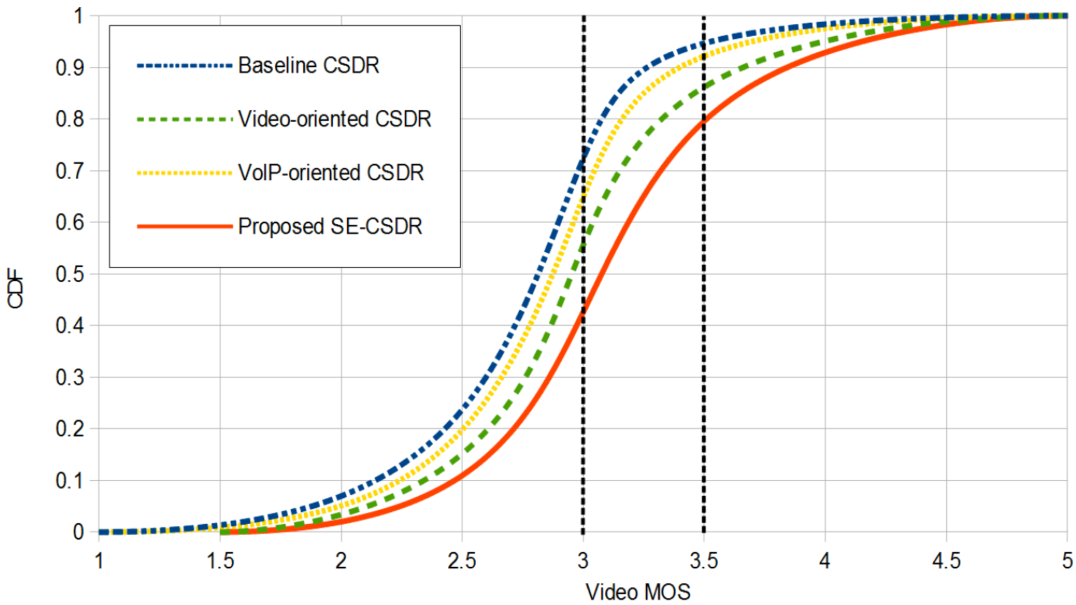


Figure 5-4: Video MOS CDF

As previously mentioned, the SE-CSDR assures isolation enhancing in this way the performance of specific applications/services with incompatible, conflicting and diverse QoS requirements. However, it comes at the cost of loss in multiplexing gain, which affects low priority traffic, i.e. the BE traffic in our scenario.

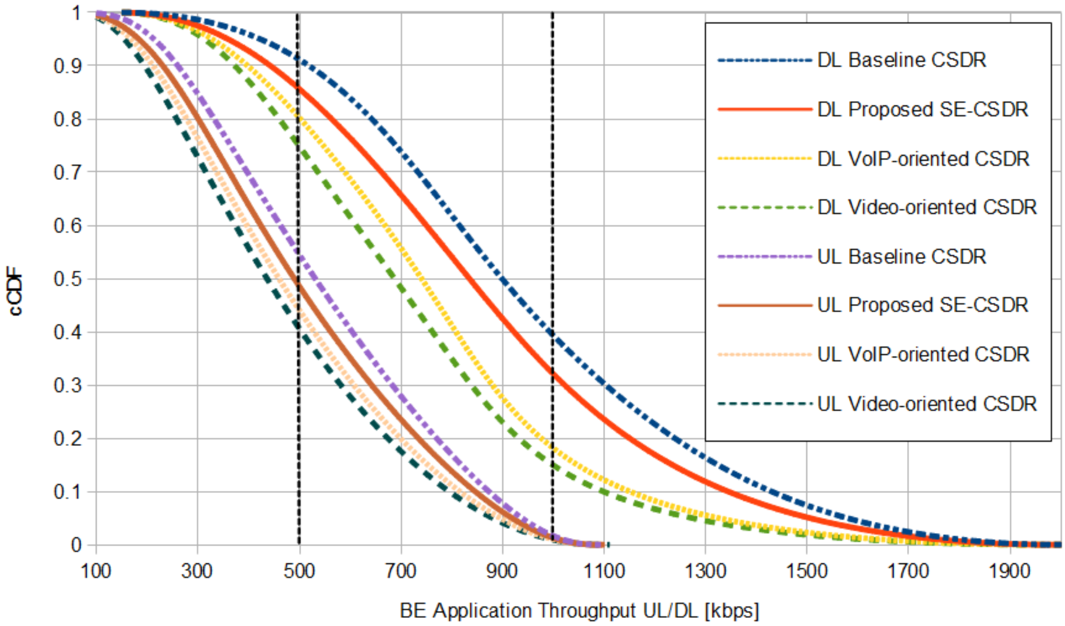


Figure 5-5: UL/DL Application Throughput cCDF BE Traffic (kbps)

The loss in performance considering the application throughput can be observed in

5-5, which shows the complementary CDF (cCDF) of BE throughput traffic in both UL and DL directions. Overall, the proposed SE-CSDR scheme provides significantly less throughput loss compared to the equivalent VoIP-oriented CSDR and Video-oriented CSDR. In particular, compared to the baseline CSDR, the proposed SE-CSDR results in 9% and 8% less throughput in the DL and UL direction, whereas the VoIP-oriented CSDR results in 21% and 12% less throughput considering a target limit of 1Mbps in DL and 0.5 Mbps in UL. The reason behind this result is the fact that SE-CSDR is capable of providing better resource scheduling to the flows of different slices, allowing also slice specific CSDR operation. In contrast, for application-oriented CSDR the adoption of a common scheduler and CSDR operation, results in bigger throughput losses for the application with lower service prioritization.

Fig. 5-6 shows the DL Throughput CDF for the Video traffic. From Fig. 5-6, it can be observed that the proposed SE-CSDR solution outperforms the baseline CSDR showing 35% gains considering the target throughput of 400 kbps. That is, 35% fewer users are below the target throughput. Even compared with the VoIP-oriented CSDR solution, the gain of the proposed SE-CSDR is 25%, while compared to the Video-oriented CSDR, SE-CSDR shows 12% improvement. The performance gains of the SE-CSDR are higher than the Video and VoIP-oriented CSDR, despite the fact that these schemes prioritises VoIP or Video flows and adapts the UL/DL frame configuration to match the respective Video or VoIP traffic.

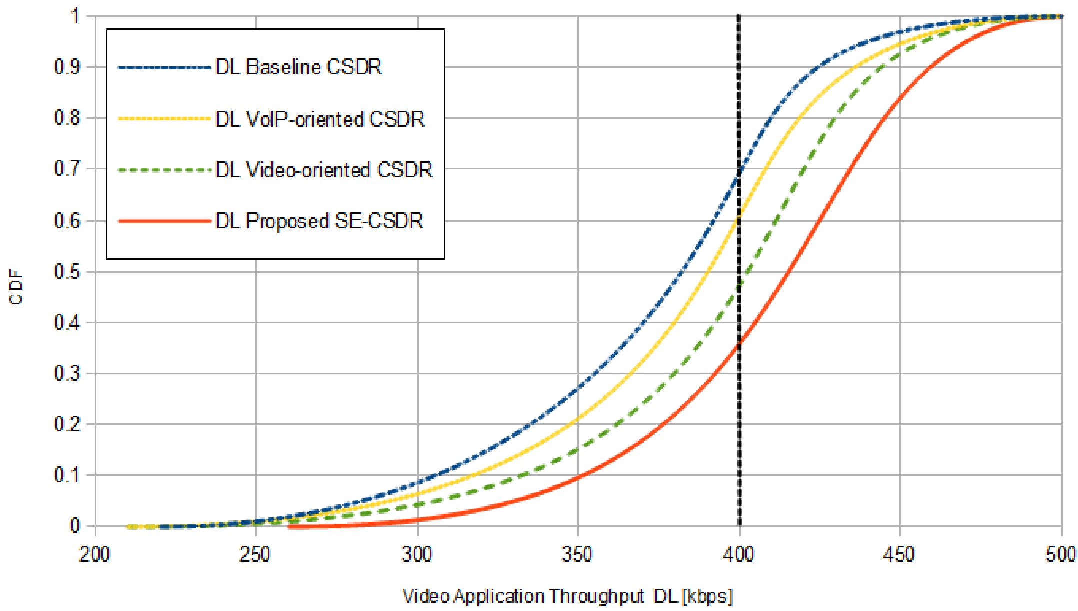


Figure 5-6: Video Application Throughput DL (kbps)

Fig. 5-7 shows the UL Throughput CDF for the Video traffic. From Fig. 5-7, it can be observed that the proposed SE-CSDR solution outperforms the baseline CSDR showing 27% gains considering the target throughput of 400 kbps. Even compared with the VoIP-oriented CSDR solution, the gain of the proposed SE-CSDR is 20%, while compared to the Video-oriented CSDR, SE-CSDR shows 10% improvement. The performance gains of the SE-CSDR are higher than the Video and VoIP-oriented CSDR, despite the fact that this scheme prioritizes VoIP or Video flows and adapts the UL/DL frame configuration to match the respective Video or VoIP traffic

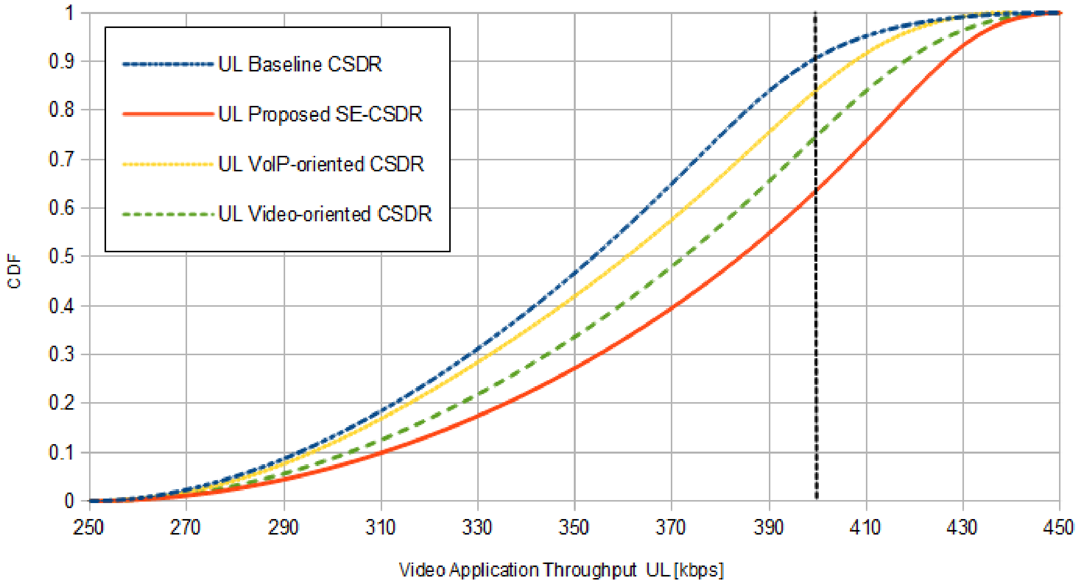


Figure 5-7: Video Application Throughput UL (kbps)

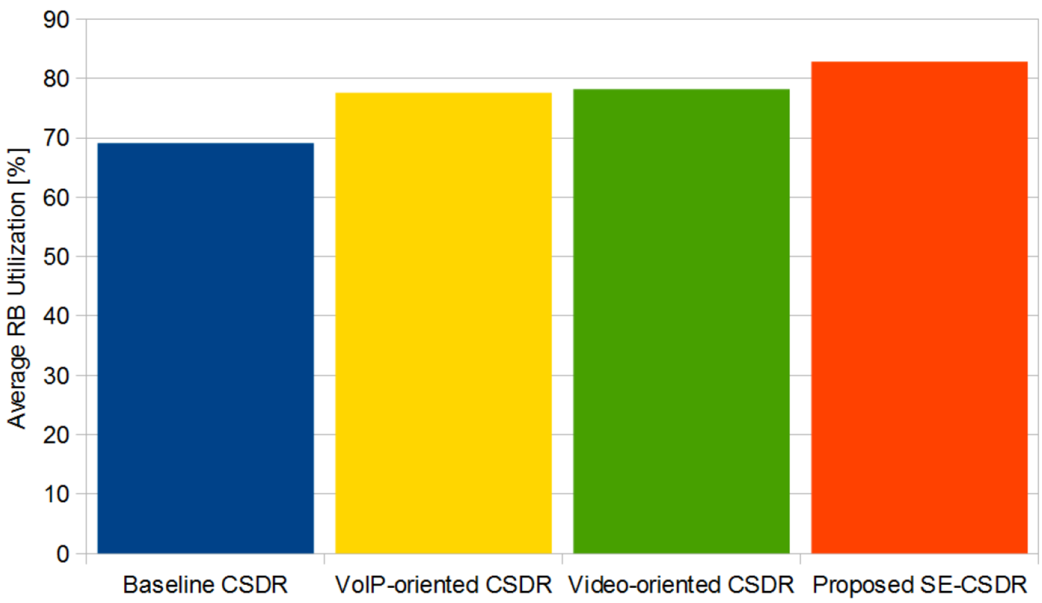


Figure 5-8: Average RB Utilization

Fig. 5-8 shows the average system resource block (RB) utilization. The proposed

SE-CSDR shows 14% improvement in resource utilization compared to the baseline CSDR and 8-10% compared to the Video oriented-CSDR and VoIP oriented-CSDR respectively. VoIP-oriented CSDR and Video-oriented CSDR provide higher average resource utilization compared to the baseline CSDR, but lower than the proposed SE-CSDR. For VoIP-oriented CSDR and Video-oriented CSDR, as the flows are prioritized, they get the required resources to satisfy the particular application demands at the cost of high losses on the low-priority BE traffic. Also, the use of common scheduling policies and a common CSDR operation do not favor the low-priority traffic. In general, SE-CSDR allows for a finer granularity when allocating UL/DL resources, which can be better tailored to specific application UL/DL patterns. This results in higher resource utilization efficiency.

## 5.3 Service-Tailored Adaptive TDD Network Slicing for Emerging 5G Networks

Section 5.2 introduced the concept of network slicing in the 5G TDD networks which is a type of static network slicing where the isolation is achieved via utilization of distinct frequencies as explained in 5.2.3. This section introduces the adaptive TDD network slicing concept in emerging 5G networks and elaborates the SDN architecture that allocates and adjusts network slice resources and the corresponding TDD frame re-configuration. Different from the static network slicing concept discussed in 5.2.3, in this section an advanced, SDN based adaptive slicing concept is developed using traffic forecast techniques to creat slices on-demand and reconfigure the UL/DL sub-frames to match the traffic requirements of the respective slices. In addition, this section elaborates four different mechanisms for allocating network resources considering UL/DL ratio adaptation and network slicing.

### 5.3.1 5G TDD Network Slicing Concept

Network slicing enables a logical isolated network experience over a common physical network infrastructure which is customized to support the performance requirements of a particular service. It allows multiple services with diverse and often conflicting service requirements to efficiently utilize network resources concurrently, enabling application based programmability and context aware optimization. In TDD networks, slicing can enable a service-oriented resource allocation allowing:

(i) Each slice to adopt a distinct TDD frame configuration that corresponds to a different UL/DL ratio, which can be adjusted independently reflecting slice specific traffic variations.

(ii) A slice independent re-configuration timescale, i.e. the time window that an eNB can dynamically adjust a TDD frame, which depends on the traffic type and load variation patterns.

(iii) A flexible selection of spectrum resources that can satisfy the QoS of a particular service, considering resource blocks from the entire available spectrum band; a process, which can reflect dynamic radio conditions.

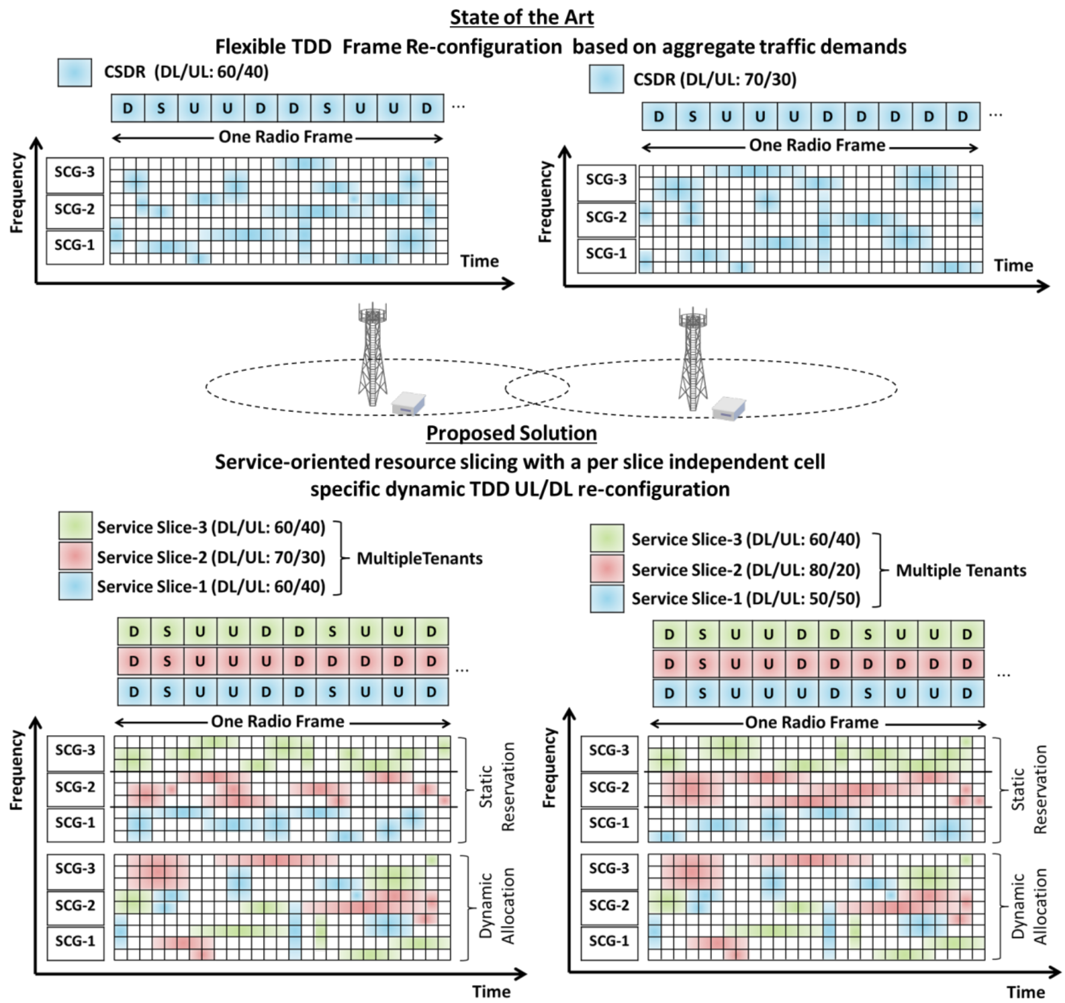


Figure 5-9: An example comparing the baseline dynamic TDD scheme with the proposed service-oriented network slicing TDD

A simple example that provides an overview of the proposes scheme with respect to the state of the art is illustrated in Fig. 5-9. In particular, the baseline scheme (named as state of the art in Fig. 5-9) aligned with the CSDR approach, allows neighboring base stations to adopt a TDD frame with a different UL/DL ratio across the entire spectrum band, i.e. the 3 Sub-Carriers Groups (SCG) in Fig. 5-9, without employing the concept of network slicing. The proposed TDD network slicing solutions allows the support of multi-tenancy enabling the following two distinct flavors:

- Service Explicit-CSDR (SE-CSDR) where each sub-carrier out of the 3 Sub-Carriers Groups (SCG) in Fig. 5-9, represents a separate network slice allocated to a certain tenant supporting a different TDD frame configuration, which can be adjusted independently. SE-CSDR simplifies the coordination for cross-slot interference among inter-slice users by allocating to each slice the same sub-carriers across the entire RAN in a static manner allowing conventional interference control per slice. However, SE-CSDR limits the number of resource blocks scheduled per slice restricting potentially the offered QoS. Detailed analysis on SE-CSDR is presented in section 5.2.3.

- Dynamic Service Explicit-CSDR (DSE-CSDR) also provides separate network slices that adopt a different TDD frame configuration, which is adjusted based on slice specific traffic condition independently. Unlike SE-CSDR, this solution enhances the flexibility in allocating resources across different slices. The flexibility is enhanced by enabling the network slices corresponding to different applications to have a variable slice size/capacity which can then be periodically or on-demand adapted in a dynamic manner as the application traffic load changes across different eNBs in the RAN. Following this, the scheduling over the resources corresponding to individual application specific slices can be performed in an abstract manner across all the eNBs in the entire RAN, i.e. all the Sub-Carriers Groups (SCG) as shown in Fig. 5-9. Interference is monitored and controlled by using conventional techniques such as power control and coordinated with the help of SDN controller. Resource isolation is a crucial characteristic of network slicing. In this case, the isolation is maintained via the RRM policies and SLAs between the tenant and the network operator. This thesis introduces three variations of DSE-CSDR namely Joint DSE-CSDR-10ms, Joint DSE-CSDR-640ms and Distributed DSE-CSDR, which are discussed in detail in section 5.3.3.

Once a network slice is allocated, customization is facilitated by matching the corresponding SLA requirements to the UL/DL frame ratio selecting an appropriate TDD frame configuration, and allocating the resource blocks that cause less interference considering the potential selection range. In DSE-CSDR and its variants, the RAN wide network slices are created, monitored and adapted, periodically (as explained in section 5.3.3), for example in every 640ms or 10ms, based on adaptive service-oriented network slicing algorithms considering network information such as interference conditions, UL/DL traffic variations and user mobility etc. The UL/DL ratio across the entire application specific RAN slice is also customized as per the instantaneous application traffic demand. Isolation among different network slices allows the operation of an independent scheduler improving resource utilization efficiency, while avoiding situations where traffic load variations of a particular slice impact negatively the performance of other services that operate on different slices.

### 5.3.2 SDN-based Network Slicing Architecture

The proposed TDD network slicing solution is facilitated via the means of SDN that allows service providers, vertical segments and Mobile Virtual Network Operators (MVNOs) to acquire and negotiate the provision of network slices on-demand. The SDN architecture proposed in this thesis assumes a virtualized RAN setup by the means of introducing a hypervisor in each eNB to ensure isolation among service slices and tenants.

SDN allows adjustment of size/capacity of the slices and associated TDD frame configuration reflecting evolving load conditions, while providing policies towards eNBs for scheduling resource blocks with respect to the QoS of particular services. An overview of the SDN-based architecture for TDD network slicing is illustrated in Fig. 5-10.

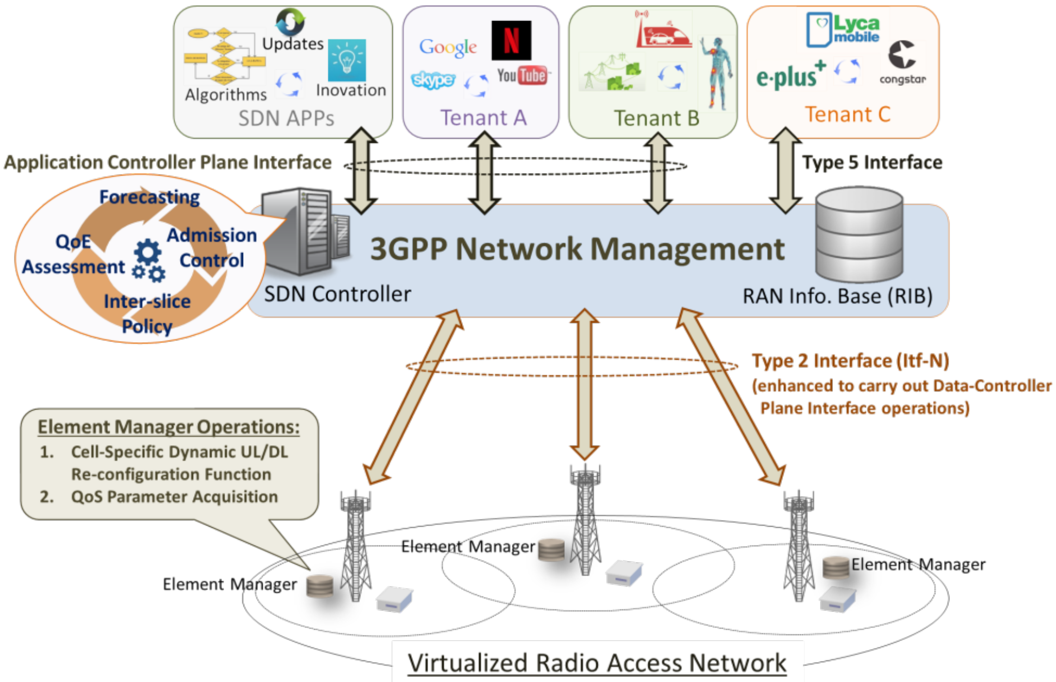


Figure 5-10: SDN-based network architecture for TDD network slicing

The SDN controller can be co-located at the 3GPP network management system in order to take advantage of the conventional network monitoring information and the existing interfaces. In particular, the 3GPP network management system can be retrieved via the Type 2 or northbound interface (Itf-N), such as interference, UL/DL load and Key Performance Indicators (KPIs), which can be stored in a RAN Information Base (RIB) creating a global network view. Such global information can then assist the SDN controller to forecast the capacity availability, providing admission control and slice allocation for incoming requests according to the specified SLA.

Slice requests arrive from MVNOs via the Type 5 interface, which connects the management systems of two mobile network providers directly, or through the SDN Application-Controller Plane Interface (A-CPI), which facilitates connectivity for application providers and vertical segments. Slice requests need to specify the SLA in terms of the amount of resources, e.g. resource blocks, the type of service, the size of the file if the service concerns a best-effort download, the desired starting time and the time duration. Hence, Type 5 interface and the SDN A-CPI need to support a new type of signalling that negotiates such parameter extensions, while the Itf-N interface should also be extended with new configuration signalling to carry out the allocation and adjustment of network slices. Besides the configuration of the slice size/capacity and timing information with respect to allocated TDD slices, Itf-N should also support

(i) the re-configuration of the cell-specific dynamic UL/DL function, responsible for the TDD frame allocation and (ii) QoS parameter acquisition collects performance parameters such delay, loss, throughput, etc. and feedback the QoE assessment function at the SDN controller making the application performance visible to the SDN controller. In addition, Itf-N should accommodate conventional SDN operations allowing SDN APPs to program the network, e.g. scheduling, virtual function algorithms, etc., but this is beyond the scope of this thesis.

The SDN controller acts as a mediator mapping SLA requests into physical resources, considering also the UL/DL service ratio and consistently monitoring the traffic demands to continuously adjust the allocated network slice resources. In particular, the SDN controller performs:

- Traffic forecasting considering UL/DL traffic demands at regular time intervals based-on the global network monitoring information collected at the RIB.
- Admission control for network slice requests considering the desired SLA and the resource availability based-on traffic forecasting. The goal is to assure adequate resources for particular service-oriented network slices within the specified time duration.
- The inter-slice policy maps the SLA of an incoming request to a slice specific policy introducing traffic prioritization, the TDD UL/DL ratio and frame reconfiguration time scales, ensuring no service conflicts or SLA violation. It also helps in maintaining isolation among slices in accordance with the RRM polices.
- QoE assessment to periodically optimize the slice resource allocation and the selection of UL/DL frame re-configuration considering the inter-slice policy and traffic forecasting.

### 5.3.3 Adaptive Network Slicing and TDD Frame Re-configuration Mechanisms

In this section, three new mechanisms are introduced namely Joint Dynamic Slice-Explicit (DSE) CSDR-10ms (Joint DSE-CSDR-10ms), Joint Dynamic Slice-Explicit (DSE) CSDR-640ms (Joint DSE-CSDR-640ms) and Distributed Dynamic Slice-Explicit (DSE)

CSDR (Distributed DSE-CSDR). The proposed mechanisms are compared with the state-of-the art CSDR and SE-CSDR schemes explained in sections 5.2.2 and 5.2.3 respectively. The results of the state-of-the art CSDR have been reproduced in a slightly different deployment scenario, however, it follows the similar pattern as in [33].

Therefore in this section, four different TDD frame re-configuration mechanisms are discussed that are used for studying the network resource allocation strategies considering UL/DL ratio adaptation and network slicing.

- Cell Specific Dynamic Reconfiguration (CSDR)- CSDR was proposed in [33] with the aim to improve resource utilization. It is discussed in more detail in section 5.2.2.

- Service-oriented Slice-explicit Dynamic TDD (SE-CSDR)- In this scheme, the concept of network slicing is applied in evolving 5G TDD networks. The network slices are created based on periodic Holt-winters exponential smoothing measurements for individual application types which includes Video, VoIP and Best effort (FTP) respectively [84]. SE-CSDR is explained in detail in section 5.2.3.

- Joint Dynamic Slice-Explicit (DSE) CSDR or Joint DSE-CSDR- In this scheme, slice size or capacity is dynamically scaled at regular time intervals, using traffic forecasting mechanisms such as Holt-winter's exponential smoothing. This is performed together with UL/DL frame reconfiguration within each slice independently considering UL/DL traffic demands in each slice. As the application traffic load and number of users may vary across different eNBs, the slice size of a particular application can be different within each eNB. The slice dimension and its corresponding frame reconfiguration is periodically monitored and can be adapted as the traffic conditions change. Thus, to the OTT application provider or the tenants the resource allocation to the application specific slices and frame reconfiguration within individual slices across the RAN is an abstract process. The overall slice size/capacity or dimension assigned to a particular tenant should satisfy the SLAs between the network operator and the tenant. However, at each eNB across the RAN the slice size/capacity or dimension may vary depending on traffic load and periodically monitored and adapted as the traffic conditions change. A simple example is illustrated in Fig. 5-11, providing an overview of the proposed concept. The slice dimension module takes the application/service load, eNB, and application type as an input and defines an optimized adaptive slice for an individual application. The slice size of each application and the users' UL/DL application load are then used to independently and dynamically re-configure the frames within each slice. The slice is monitored and application KPIs such as UL/DL buffer statistics are collected. After a period of time, the slice dimension module re-adapts the slice to match the performance demand of the application and avoid any over or under provisioning of resources. This process is repeated until an efficient slice dimension is achieved. It is worth noting that dynamic provisioning of slices on per application or service basis as well as scaling slice size/capacity on demand may require adaptation of other slice dimensions within the network therefore management of virtual resources among slices and execution of policies/decisions to solve conflicting requirements among slices for sharing and allocation of (virtual) resources is an important and complex task. The analytical logic for slice capacity dimensioning and UL/DL frame reconfiguration within each slice is discussed in detail in section 5.3.3. The isolation can be maintained via the RRM policies and SLAs between the tenant and the network operator with the help of SDN controller. However, in reality, how RAN supports resource isolation is implementation dependent. Dynamic slice size/capacity adaptation, monitoring the slice at regular intervals and Adaptive UL/DL frame reconfiguration within each slice acts as protection mechanisms against any potential breach in SLAs. Such mechanisms helps in avoiding breach in SLAs with another slice, especially in the case when a shortage of shared resources in one slice occurs. i.e. negative impacts that may occur due to the shortage of shared resources in one slice on the SLAs of another slice. There are two variants of Joint DSE-CSDR. First the Joint DSE-CSDR-10ms where the network traffic demands, slice and UL/DL frame configurations are monitored at 10ms timescales and changes in the network are only performed if any severe QoE or QoS degradation is observed. Second, Joint DSE-CSDR-640ms which is similar the Joint DSE-CSDR-10ms, with the only difference that the network is monitored every 640 ms instead of 10ms.

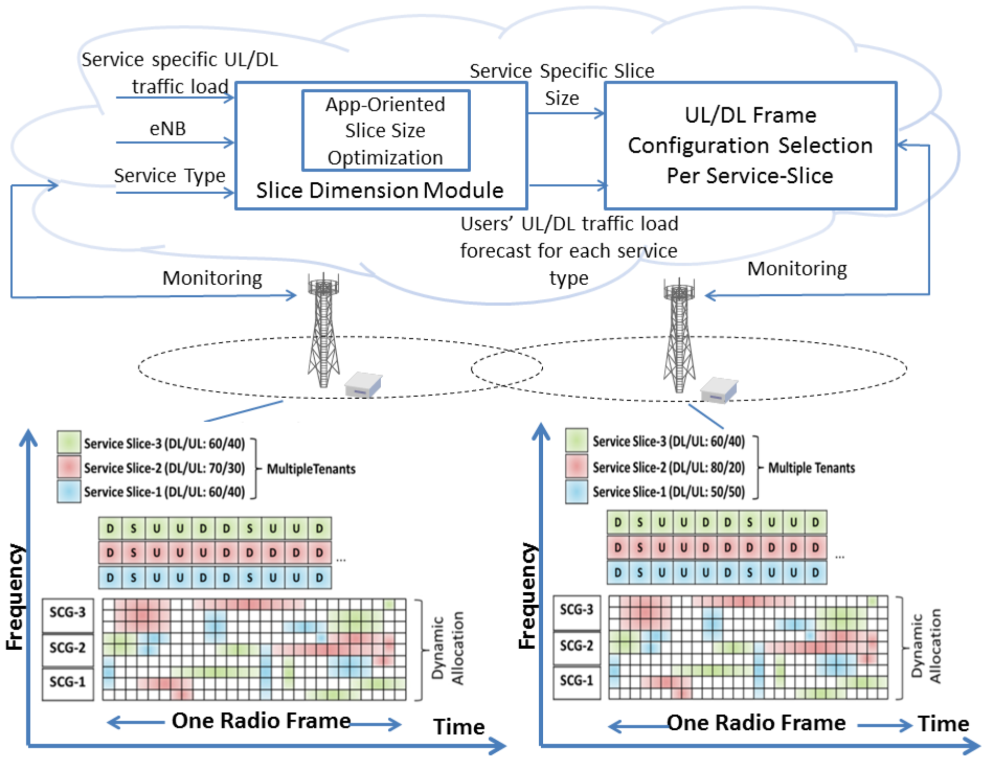


Figure 5-11: Joint DSE-CSDR

- Distributed Dynamic Slice-Explicit (DSE) CSDR or Distributed DSE-CSDR- In this scheme, slice capacity or slice size is dynamically scaled at regular time intervals using traffic forecast mechanisms such as Holt-winter's exponential smoothing in the same way as explained in Joint DSE-CSDR scheme. However, UL/DL frame reconfiguration is performed in a distributed and independent manner for each application slice within each eNB, considering UL/DL traffic demand of the corresponding slice. Therefore, the DSE-CSDR can have two different periodicities or monitoring timescales, one for individual application slice dimensioning and the other for frame reconfiguration within each slice across different eNBs within the RAN. The slice capacity adaption is managed in a similar way as in Joint DSE-CSDR and is discussed in section 5.3.3. In contrast to the Joint DSE-CSDR, in Distributed DSE-CSDR, the slice adaptation and the UL/DL frame reconfiguration can be performed as separate processes at a different timescale. The frame reconfiguration is performed independently at each eNB and for each slice. Each slice may also have a different UL/DL frame configuration that suits best its respective traffic demand. This helps in avoiding over provisioning of resources by periodically adapting the slice size considering its corresponding application load within each eNB in the deployment scenario. Fig. 5-12 shows the concept and overview of the proposed Application-Oriented Adaptive Slice Dimension Adaptation and Distributed UL/DL frame reconfiguration scheme.

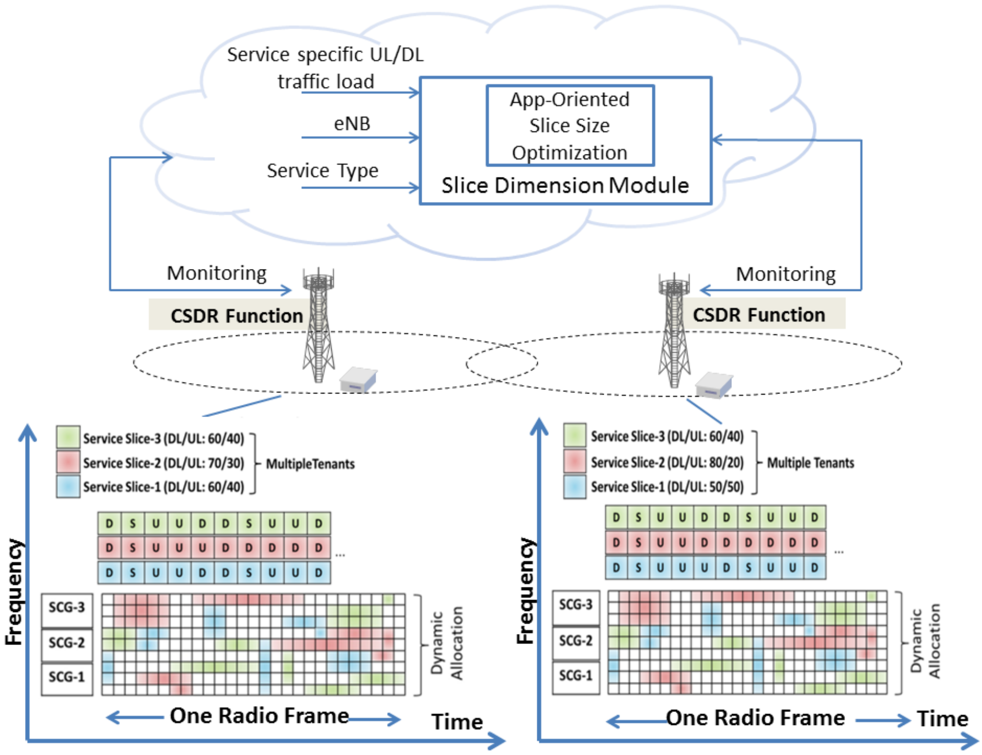


Figure 5-12: Distributed DSE-CSDR

### Analytical Model Description

We introduce a method to dynamically adjust the application-oriented slice capacity or slice size, on-demand according to the varying traffic conditions in UL/DL directions respectively. The slicer needs to decide the slice bandwidth dimensions considering uplink and downlink traffic demands. This results in a weighted maximization problem with the objective of maximizing the allocated slice bandwidth to each corresponding load. The rationale behind is to prevent over-(or under-) provisioning of resources aiming at delivering the desired performance for the respective application or service.

Let us consider the following optimization problem:

**Problem** `Slicer`:

$$\begin{aligned} \text{maximize} \quad & \sum_{s=1}^{|S|} w_s \log X_s \\ \text{subject to} \quad & \sum_{s=1}^{|S|} w_s \leq 1, \\ & w_s \in \{0; 1\}; \end{aligned}$$

where the output weight $w_s$ describes the normalized slice size/capacity tailored to service $s$. In addition, $X_s$ is the aggregate traffic load related to service $s$ and can be expressed as the follows

$$X_s = \sum_{d=1}^{2} X_s^d = \sum_{d=1}^{2} \sum_{i=1}^{|U^d|} b_i^d \tag{5.11}$$

where $b_i^d$ is the amount of data to be served for user $i$ in downlink ($d = 1$) or uplink ($d = 2$) transmissions and $\mathscr{U}_d$ is the set of users asking for traffic in a particular direction $d$. The number of required DL and UL subframe for an individual slice $S_(i, j)$ belonging to application $i$ and eNB $j$ may be determined according to the following equation (5.12):

$$N^{\text{DL}} = \frac{\sum_{i=1}^{|U^{\text{DL}}|} u_i^{\text{DL}} \frac{b_i^{\text{DL}}(t)}{R_i^{\text{DL}}}}{\sum_{i=1}^{U^{|\text{UL}|}} u_i^{\text{UL}} \frac{b_i^{\text{UL}}(t)}{R_i^{\text{UL}}} + \sum_{i=1}^{N^{|\text{DL}|}} u_i^{\text{DL}} \frac{b_i^{\text{DL}}(t)}{R_i^{\text{DL}}}} \tag{5.12}$$

where $u_i^{\text{DL}}$ and $u_i^{\text{UL}}$ are the UE application-based weights for DL and UL, expressed as a value between 0 and 1. Such weights help in determining the application type, e.g., Best effort, FTP, VoIP or Video, and the total traffic generated within the slice by an application. $b_i^{\text{DL}}(t)$ and $b_i^{\text{UL}}(t)$ are the buffer sizes for application $i$ at time $t$, while $R_i^{\text{DL}}$ and $R_i^{\text{UL}}$ is the past average data rate for the application $i$ at time $t$ in DL and UL directions respectively. The frame configuration matching closest to the calculated *DL:UL subframe ratio* is selected. Once the dimension of the network slice has been adjusted, a dynamic TDD configuration is done independently per slice with the help of equation (5.12), where $b_i^d, R_i^d$ and $N^d$ are replaced with $b_{i,s}^d, R_{i,s}^d$ and $N_s^d$ to account for a specific slice $s$. A slice may also have independent schedulers and UL/DL reconfiguration time scales depending on the degree of traffic asymmetry within the slice resulting in improved resource utilization efficiency and reduced loss of multiplexing gains. Specifically, customization of the network slice can be done by mapping corresponding SLA requirements onto the inter-slice policy.

Due to the complexity while solving Problem `Slicer` with big instances, we might run the instance of such a problem within a shorter time window. However, this incurs in much more traffic overhead, as the slicer needs to gather all user traffic requests ($b_i$) with a higher frequency. Therefore, forecasting schemes might be applied to boil down the complexity while predicting user traffic requests ($b_i^{\prime d}$).

**Joint Optimization Process**

While the previous approach is practical and lightweight, it provides suboptimal solutions which might be not sufficient for the overall system spectral efficiency enhancement. Therefore, we blend together the slice adjustment process and the TDD UL/DL pattern configuration in order to introduce robustness and accuracy in our scheme.The equation (5.1) and equation (5.2) (and even more accurate, equation (5.12)) provides an easy-to-use technique to identify the number of timeslots dedicated for transmitting requested traffic based on the amount of data still to be served as well as the average serving rate related to any specific user. The obtained ratio will guide the system to properly select a predefined TDD pattern. In our proposal, we select the appropriate ratio between number of admitted uplink slots and downlink slots by means of an advanced version of Problem `Slicer` as follows:

**Problem** `TDD Slicer Optimizer`:

$$\begin{aligned}
\text{maximize} \quad & \sum_{s}^{|S|} \sum_{d=1}^{2} w_s^d \log\left( \frac{\sum_{i=1}^{|U_s^d|} b_i^{\prime d}}{\mu_s^d} \right) \\
\text{subject to} \quad & \sum_{s=1}^{|S|} \sum_{d=1}^{2} w_s^d \leq 1, \\
& w_s^d \in \{0; 1\};
\end{aligned}$$

where $w_s^d$ identifies the bandwidth portion assigned to slice $s$ for downlink ($d = 1$) or uplink ($d = 2$) traffic. The slice size or capacity assigned for application $s$ is calculated as $w_s = \sum_d w_s^d$. The number of timeslots $N^{d,s}$ assigned to the $d$ transmissions within a slice $s$ are obtained as $N^{d,s} = \frac{w_s^d}{\sum_d w_s^d}$. This directly provides the ratio used for properly selecting the TDD pattern configuration within a slice $s$. Last, $\mu_s^d$ provides the average spatial serving rate within slice $s$ for all the traffic sent through direction

$d$. In particular, we can express such a value as follows:

$$\mu_s^d = \int_A \eta(x,y)\, L_s^d(x,y)\,\mathrm{d}(x,y) \tag{5.13}$$

where $\eta(x,y)$ is the punctual throughput value obtained by a reference user placed into position $(x,y)$ experiencing a specific SINR given the scenario's topology, whereas $L_s^d(x,y)$ is the probability of encountering a user attached to slice $s$ performing traffic through direction $d$ in position $(x,y)$. The latter value strictly depends on the mobility model we are assuming, although in realistic scenario could be empirically obtained after a long observation period. Please note that we first assume the same network slicing operations along the whole network. This is the reason why we consider the overall system area $A$ in equation (5.13).

### Cell Diversity and User Mobility

When network slicing operations are considered along the whole network, general user statistics may be collected to drive the system towards sub-optimal operational points. However, in other cases we may need to apply the network slicing concept per slice independently. User mobility information help in modelling the probability of finding a user within a given location. Different mobility models may lead to different system behaviours as the average spatial user throughput might change based on user movements. Let us start considering a set of users uniformly distributed between different services (or slices $s$) over the spatial domain. This may be expressed by $L_s^d(x,y) = \frac{1}{A}, \forall (x,y) \in \mathscr{A}$ as a single user may be attached to multiple slices as well as could transmit and receive in both directions resulting in independent statistical distributions. We can rewrite Eq. equation (5.13) as follows:

$$\begin{aligned} \mu_s^d &= \int_A \eta(x,y)\,\mathrm{d}(x,y) &&= \sum_{m=1}^{|\mathscr{M}|} \int_{A_m} \bar{\eta}_m\,\mathrm{d}(x,y) \\ &= \sum_{m=1}^{|\mathscr{M}|} \bar{\eta}_m A_m &&= \sum_{m=1}^{|\mathscr{M}|} \bar{\eta}_m \pi (l_m^2 - l_{m-1}^2) \end{aligned} \tag{5.14}$$

where we assume a uniform channel condition distribution [1] which incurs in dif-

[1] Our assumption relies on a wrap-around scenario wherein base stations are uniformly distributed at the same inter-site distance. The user channel condition regularly degrades moving towards the edge

ferent modulation and coding schemes (MCSs). Each MCS determines the spectral efficiency each user may experience within a certain geographical area $A_m$ (annulus) around the base station. Such an area is determined between two distances $l_m$ and $l_{m-1}$ whereas we consider the first annulus as a regular circle with a specific ray $l_1$ and $l_0 = 0$.

For advanced analysis, we stress our solution within a realistic scenario by taking into account a well-known mobility model, namely SLAW model [85].

### 5.3.4 Simulation Results and Analysis (Phase-2)

This section presents the simulation set-up and performance evaluation of the proposed framework using Matlab as the system-level simulation tool. A standard 3GPP scenario with seven TD-LTE macro eNBs was considered, with each eNB having 3 sectors forming 21 macro cells in total. Users are distributed uniformly across the entire area, with the users' mobility modeled following the Self-Similar Least Action Walk (SLAW) [85]. A detailed simulation parameter list, based-on 3GPP specification [74], is summarized in table 5.2.

Table 5.2: Simulation Parameters-II

| | |
|---|---|
| Number of eNBs | 7 |
| System Bandwidth | 10 MHz |
| Duplexing Scheme | TDD |
| eNB inter site distance | 500m |
| eNB Max Tx Power | 46 dBM |
| eNB Antenna Gain | 15 dBi |
| UE Total Tx Power | 23 dBM |
| UE Antenna Gain | 0 dBi |
| Path loss Model | 128+37.6 log10(R), R in Km |
| Application Traffic Model | G.729 VoIP, H.264 Video, FTP Best Effort |
| Traffic Prediction Model | Holt-Winter's Exponential Smoothing |
| Mobility Model | SLAW |

The performance evaluation results of the two proposed variations of the DSE-

of the cell.

CSDR namely the Joint DSE-CSDR and the Distributed DSE-CSDR were compared with respect to the SE-CSDR [91] and CSDR [33] [32], which comprise the current state of the art. Three different traffic models were considered in the simulations, similar to the description in section 5.2.4, each representing a different service provider including: (i) Voice over IP (VoIP), (ii) video [78] and (iii) File Transfer Protocol (FTP) best effort traffic following the model in [74]. In the simulations, it is assumed that 30% of the users are engaged in VoIP calls, 30% stream video and 40% download files via an FTP best effort service, while for network slicing proposals, each service provider is associated with a separate slice. The video performance is evaluated considering the throughput and video MOS. Fig. 5-13 and Fig. 5-14 illustrate the video throughput CDF in the DL and UL direction respectively. In the DL direction, in Fig. 5-13, the performance of Joint DSE-CSDR-10ms shows a 36% improvement compared to SE-CSDR and approximately 45% improvement compared to CSDR while Distributed DSE-CSDR shows mean overall improvement of around 33% and 38% with respect to SE-CSDR and baseline CSDR respectively.The Joint DCE-CSDR-640ms performs the worst with 12% gains compared to SE-CSDR and 17% gains compared to CSDR in the DL. Video throughput performance in the UL can be observed in Fig. 5-14. The Joint DSE-CSDR-10ms performs the best with an overall mean improvement in performance of 30% compared to the SE-CSDR and 38% improvement over CSDR. The Distributed DSE-CSDR on the other hand shows mean overall improvement of 28% compared to SE-CSDR while upto 34% improvement over CSDR scheme where no slicing is employed. Joint DCE-CSDR-640ms shows in least performance gains in the UL direction as well with 10% improvement over SE-CSDR and 15% improvement over CSDR.

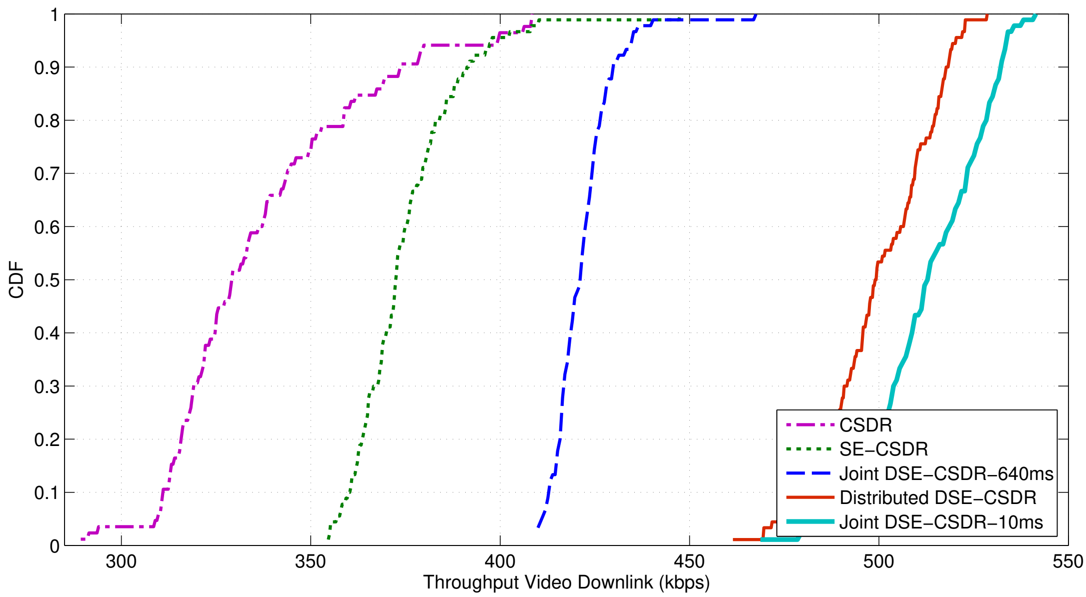


Figure 5-13: Video Application Throughput CDF DL (kbps)

Thus the performance evaluation results for the Video throughput in UL and DL show that the proposed Joint DSE-CSDR-10ms scheme outperforms the other schemes followed by the Distributed DSE-CSDR TDD that performs the second best. The reason for the better performance of Joint DSE-CSDR-10ms is that it monitors the slice every 10ms and assess the amount of resources allocated to the slice. Any variations in the traffic that may affect the QoE of the services that a particular slice is serving is addressed by adjusting the allocation of resources to the slice i.e. the size of the slice together with the UL/DL frame reconfiguration. It is worth noting that any changes are done only when it is foreseen that the services will suffer performance degradation due to traffic variations or if the resources are over provisioned to a particular slice. No changes are made within the slice if the slice has been allocated sufficient resources to maintain the desired QoE for a specific service.

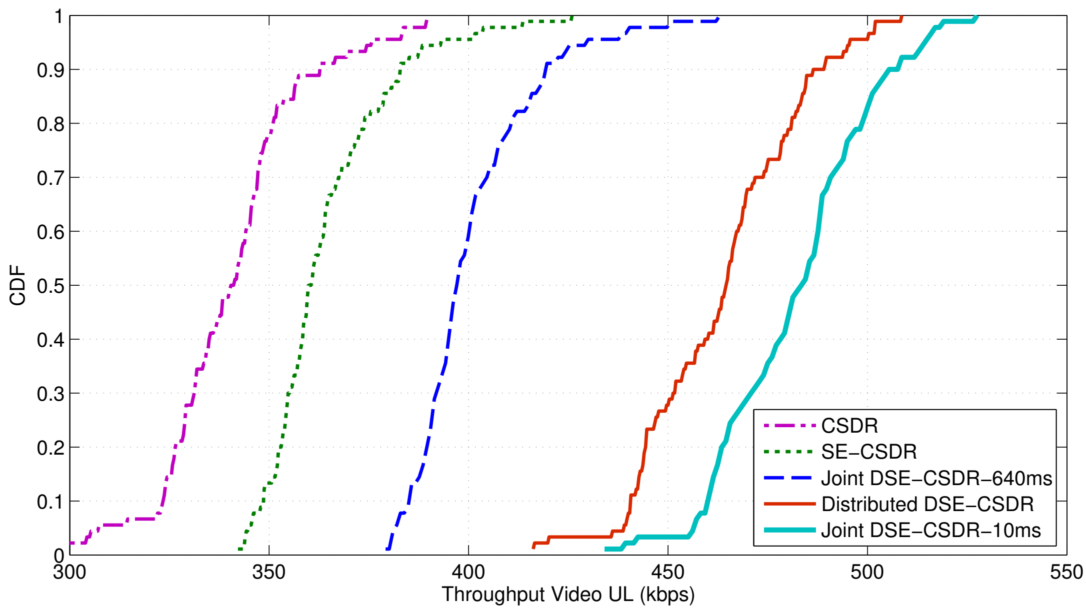

Figure 5-14: Video Application Throughput CDF UL (kbps)

The isolation and flexible resource allocation offered by the slicing mechanism, helps in minimizing the delay of video applications thereby ensuring that the end users' experience the desired quality. The concept of slicing enhances the performance of specific applications/services in addition to that the optimization of slice size/capacity regularly, ensures that the resources are neither over provisioned nor under provisioned while maintaining the desired performance of the target application or service. Dynamic frame reconfiguration ensures that resources in the UL/DL transmission direction are allocated in an efficient and flexible way.

The Joint DSE-CSDR-10ms and Distributed DSE-CSDR perform better than the Joint DSE-CSDR-640ms, SE-CSDR and CSDR, this is because in Joint DSE-CSDR-10ms and Distributed DSE-CSDR schemes, as the load varies among different traffic types and across different eNBs, the slice size/capacity is periodically adjusted according to the traffic variations as explained in section 5.3.3. These schemes also use traffic prediction to periodically reconfigure UL/DL frame ratios and customize the slice size which enables them to have a better assessment of the future traffic demand. Distributed DSE-CSDR offers the flexibility that allows each slice within an eNB to adaptively select an UL/DL frame configuration that best match their traffic demand considering UL and DL traffic separately. This further enables efficient use of the system resources. The isolation offered by slicing also allows different UL/DL reconfiguration timescales that is suitable for a particular slice and depends on the degree of traffic asymmetry. This implies that how often UL/DL sub-frames needs to be adapted within a slice may also be customized on-demand based on the traffic pattern within the slice. It is also worth noting that in high interference scenarios adaptive Joint DSE-CSDR-10ms and Joint DSE-CSDR-640ms schemes may be more suitable as it jointly optimizes slice capacity at regular intervals and selects most suitable UL/DL frame configuration for each application slice independently within every eNB. The selected configuration for a specific application is applied to all the slices of that particular application across all the eNBs. In Joint DSE-CSDR-10ms and Joint DSE-CSDR-640ms schemes, different application slices may have different frame configuration but in this case, the frame configuration for a particular application slice remains the same across all the eNBs. This can help limit the cross slot interference that may occur due to transmission in the opposite directions.

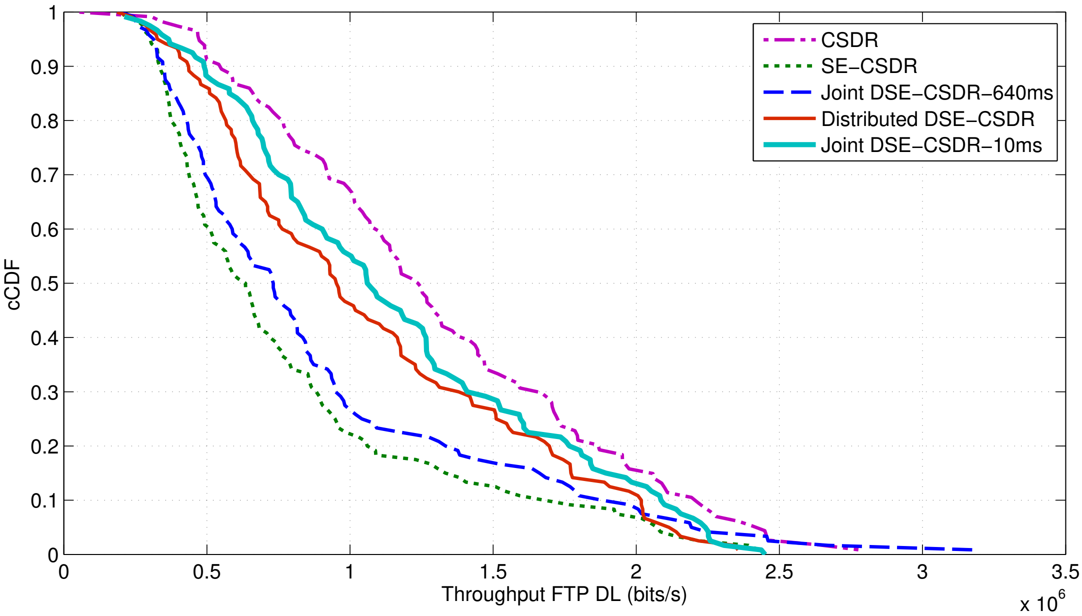

Figure 5-15: DL Application Throughput cCDF BE Traffic (bits/s)

The loss in the throughput performance of FTP best effort traffic can be observed in Fig. 5-15 and Fig. 5-16 which illustrates the FTP Throughput complementary CDF (cCDF) in the DL and UL direction respectively. Overall, the proposed Joint DSE-CSDR-10ms, Distributed DSE-CSDR and Joint DSE-CSDR-640ms schemes provide significantly low throughput loss compared to SE-CSDR, performing close to the Baseline CSDR. Here, the Baseline CSDR is considered as the target scheme as in this case there is no slicing and hence it offers the best performance for the best effort traffic with least probability of having performance degradation due to losses in multiplexing gain.

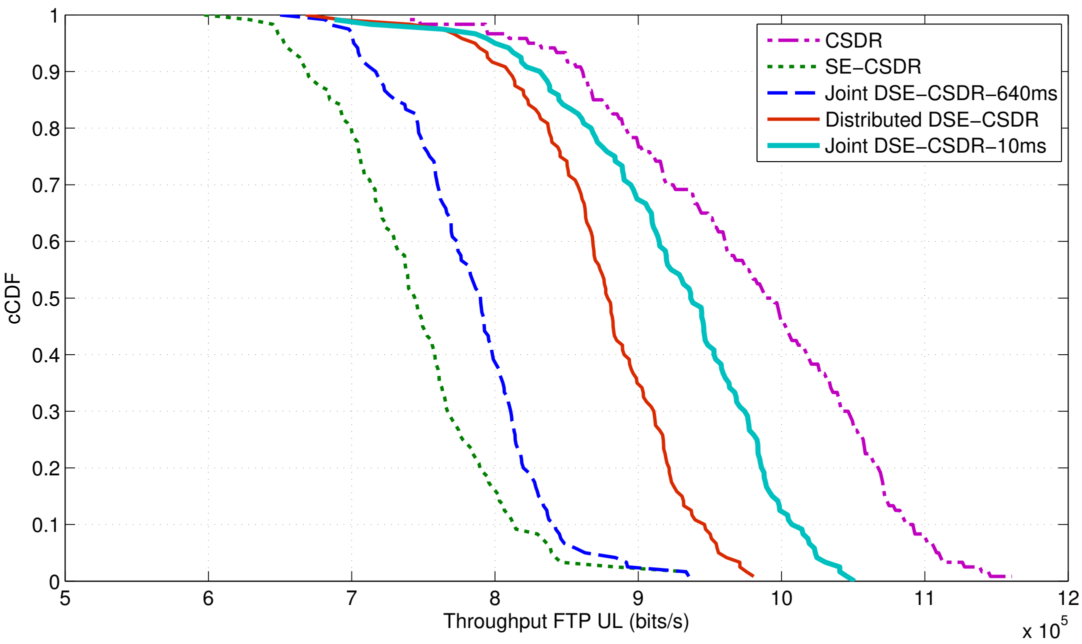


Figure 5-16: UL Application Throughput cCDF BE Traffic (bits/s)

In particular, it can be seen in Fig. 5-15, that the Joint DSE-CSDR-10ms performs the best with approximately 12% less throughput than CSDR which is the target Scheme for measuring loss in multiplexing gain. Following the Joint DSE-CSDR-10ms, the Distributed DSE-CSDR provides an overall 15% less throughput than the CSDR while Joint DSE-CSDR-640ms have 25% less throughput compared to the CSDR. SE-CSDR performs worst with 35% less throughput than the target Baseline CSDR scheme.

In the UL direction, in Fig.5-16, the Joint DSE-CSDR-10ms scheme performs close to the CSDR scheme which is the performance target scheme, with an overall 10% less throughput. Following the Joint DSE-CSDR-10ms, the Distributed DSE-CSDR provides overall 12% less throughput than the CSDR while the Joint DSE-CSDR-640ms have 20% less Throughput than the Baseline CSDR scheme. SE-CSDR on the other hand has the highest loss in throughput performance due to loss in multiplexing gain providing 30% less throughput than the Baseline CSDR scheme therefore the loss in the throughput performance for the best effort traffic in SE-CSDR is higher compared to Joint DSE-CSDR-10ms, Distributed DSE-CSDR and Joint DSE-CSDR-640ms have respectively.

As previously mentioned, the slicing provides isolation to specific applications/services which often have incompatible, conflicting and diverse QoS requirements. Although, slicing helps in enhancing application/service specific performance, it comes at the cost of loss in multiplexing gain due to reduced resource diversity affecting the low priority best effort traffic. Thus, the throughput performance presented in Fig. 5-15 and Fig. 5-16 shows the impact of slicing on the BE traffic, which in this case is the FTP traffic. Here, the aim is to be able to match the CSDR scheme which is considered as the target performance scheme in the BE case. The CSDR scheme has no slicing and provides maximum throughput as well as delay performance for the BE traffic.

The slicing schemes (or the aforementioned schemes that employ slicing) that perform closest to the Baseline CSDR performance is considered to be the best as it shows that even though it accommodates and maximizes the performance of certain applications within the slice, it is able to minimize the loss in performance of the BE traffic due to loss in multiplexing gain that arises due to reduced resource diversity as a consequence of network slicing. In this case Joint DSE-CSDR-10ms performs the best followed by Distributed DSE-CSDR and Joint DSE-CSDR-640ms. The SE-CSDR scheme improves the performance of the applications within the slice compared to the CSDR, however, the loss in performance of BE traffic due to loss in multiplexing gain is highest in case of the SE-CSDR.

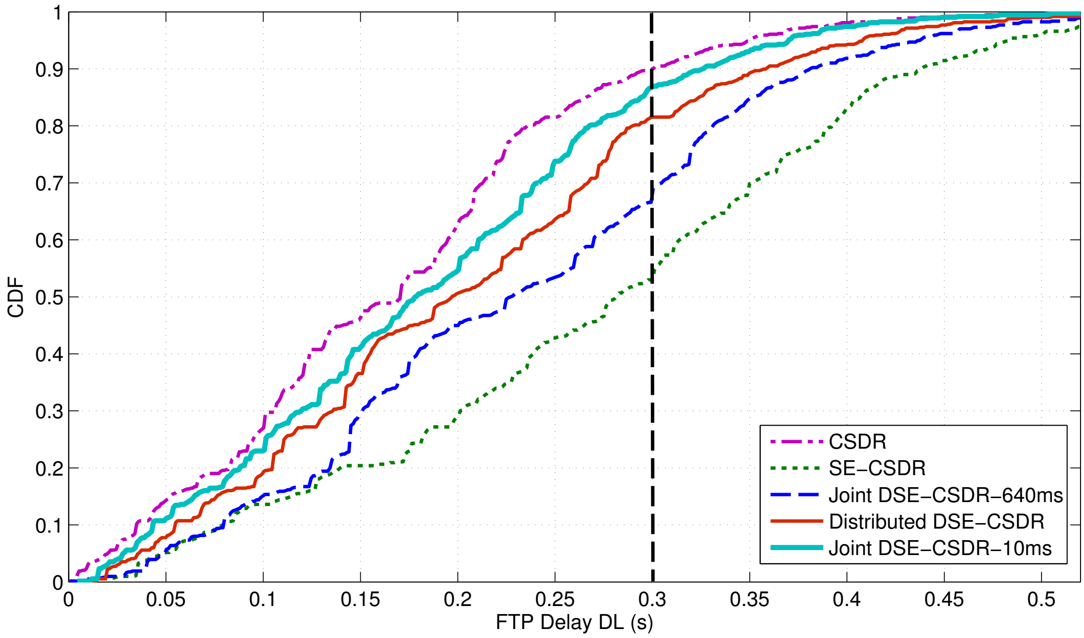


Figure 5-17: DL Application Delay CDF BE Traffic (s)

The delay performance comparison of the proposed schemes with the target CSDR for the FTP traffic in DL and UL directions is captured in Fig.5-17 and Fig. 5-18. A target delay of 300ms is considered for the FTP best effort traffic, which is in accordance with [73], [40]. From Fig. 5-17 and Fig. 5-18, it can be observed that in the

DL direction, in Fig. 5-17, the Joint DSE-CSDR-10ms scheme performs better than the other three schemes having 7% less users below the target delay of 300ms [73], while the Distributed DSE-CSDR have 10% less users below the target delay compared to the Baseline CSDR, which offers the best delay performance. Joint DSE-CSDR-640ms have 20% less users below the BE target delay compared to the Baseline CSDR. While in the case of the SE-CSDR, 35% less users are below the BE target delay compared to the Baseline CSDR. This shows that the negative impact of network slicing on the BE traffic is maximum in the SE-CSDR. Here, the delay performance of the Baseline CSDR which has no slicing involved is selected as the reference scheme as it has the potential to provide maximum performance gain and the scheme performing closest to the Baseline CSDR is considered to be the best.

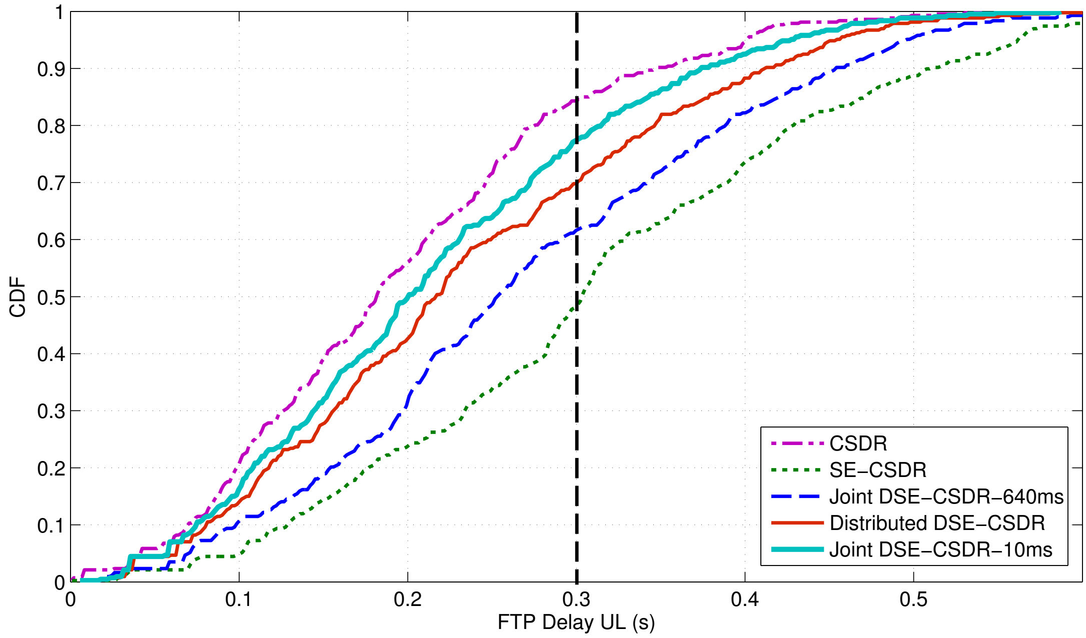


Figure 5-18: UL Application Delay CDF BE Traffic (s)

In the UL direction, in Fig. 5-18, the Joint DSE-CSDR-10ms scheme outperforms the other schemes having 10% less users below the target BE delay followed by Distributed DSE-CSDR with 15% fewer users below the BE target delay. Joint DSE-CSDR-640ms has 25% less users below the BE target delay compared to the Baseline CSDR which is slightly better than the SE-CSDR. SE-CSDR in this case shows delay performance loss of 36% compared to the Baseline CSDR which exhibits the maximum loss.

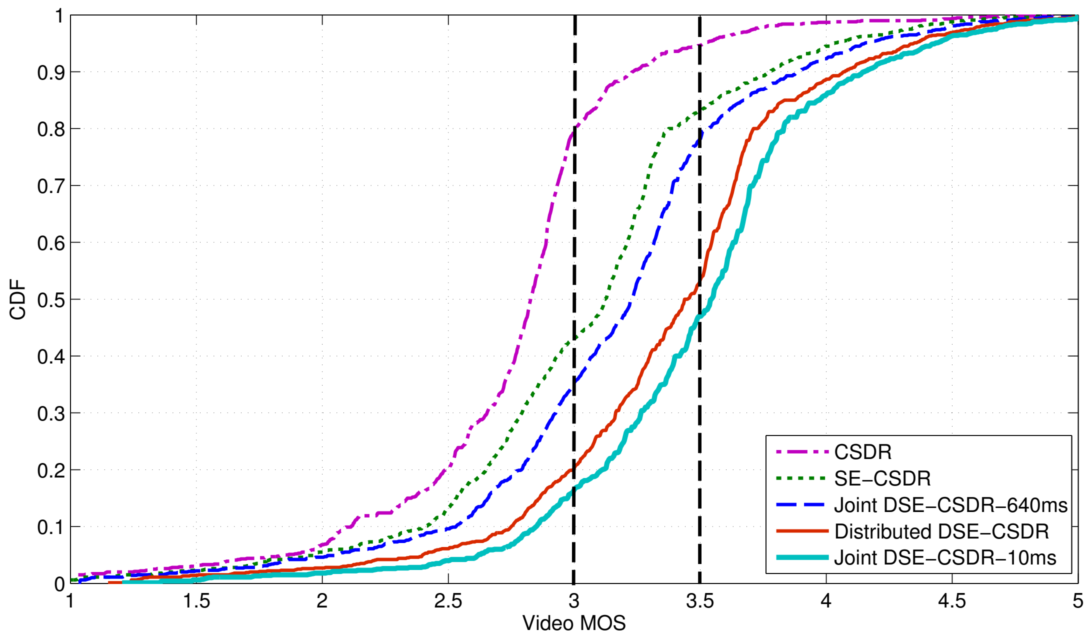

Figure 5-19: Video MOS CDF

The video MOS CDF is shown in Fig. 5-19. It can be observed from Fig. 5-19 that the Joint DSE-CSDR-10ms outperforms the other solutions, with Distributed DSE-CSDR performs better than the DSE-CSDR-640ms but exhibits slightly higher video MOS performance losses than the DSE-CSDR-10ms. For video applications an acceptable MOS is between 3 to 3.5, while a MOS above 3.5 is considered as excellent [90]. For the VoIP applications, a minimum target MOS of 3.5 is considered, which is sufficient to provide an acceptable QoE to the end users [90].

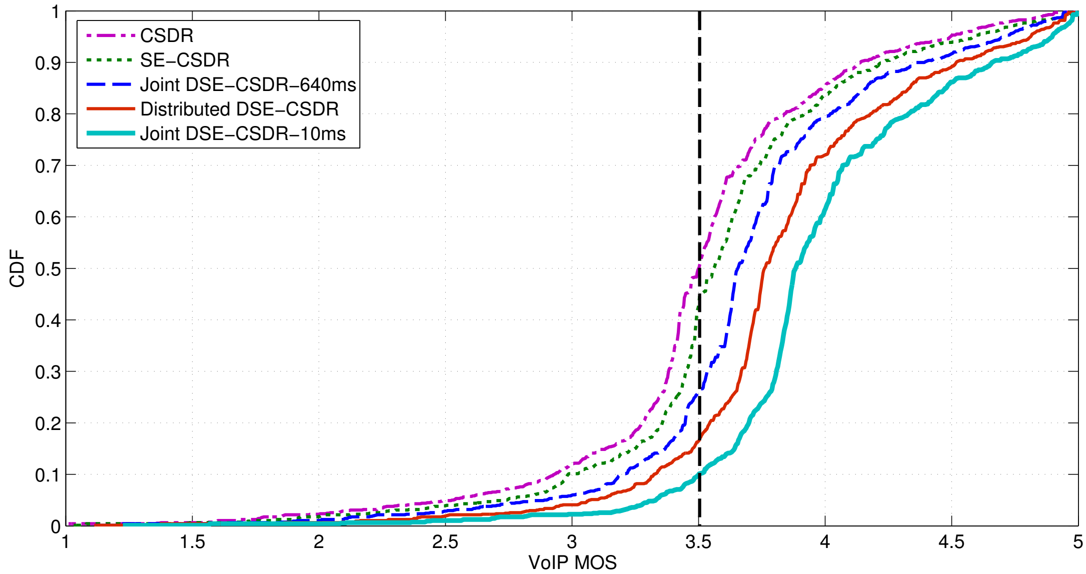

Figure 5-20: VoIP MOS CDF

Fig. 5-20 shows the MOS CDF for the VoIP traffic. From Fig. 5-20, it can be observed that the proposed Joint DSE-CSDR-10ms solution outperforms the SE-CSDR and CSDR showing 30% gain compared to the SE-CSDR and 40% gain compared to the CSDR, considering the VoIP target MOS. Distributed DSE-CSDR offers 20% while Joint DSE-CSDR-640ms offers 12% gain compared to SE-CSDR while 22% gain compared to the baseline CSDR respectively. Therefore the Joint DSE-CSDR-10ms offers the maximum gain in the MOS performance for both Video and VoIP applications followed by Distributed DSE-CSDR and Joint DSE-CSDR-640ms. The reason behind these gains are that the Joint DSE-CSDR-10ms scheme is able to adapt more often to match the traffic demands in UL and DL directions respectively and the more frequent traffic prediction mechanism helps it to obtain a more accurate view of the traffic situation thereby allowing it to efficiently program the network to provide the desired QoE for the respective applications.

The delay performance comparison of the proposed schemes with the target CSDR and SE-CSDR for the Video traffic in DL direction is captured in Fig. 5-21. A target delay of 100ms is considered for the video traffic, which is in accordance with [73]. From Fig. 5-21 it can be observed that in the DL direction, the Joint DSE-CSDR-10ms scheme performs better than the other four schemes having 90% users below the target delay of 100ms, while Distributed DSE-CSDR has 85% users below the target delay. Joint DSE-CSDR-640ms have 72% users below the Video target delay. While in the case of SE-CSDR, 62% users are below the video target delay. The CSDR performs poorly with only 55% users below the target video delay. The delay performance of the video traffic in the UL direction follows similar trend as of the DL direction.

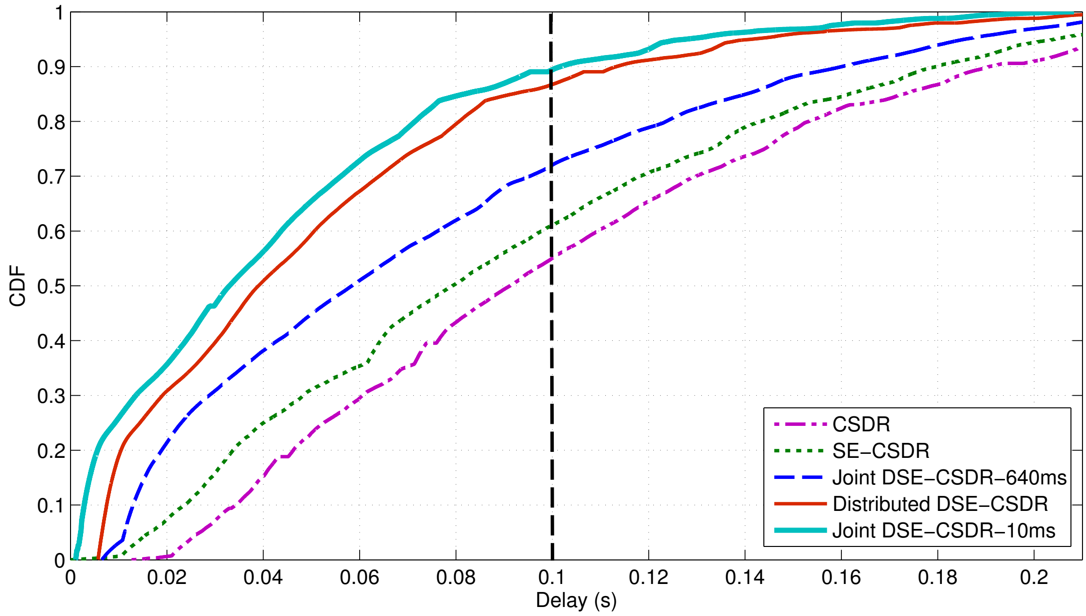


Figure 5-21: Video Delay CDF DL (s)

The delay performance comparison of the proposed schemes with the target CSDR and SE-CSDR for the VoIP traffic in DL is captured in Fig. 5-22. A target delay of 100ms is considered for the VoIP traffic, which is in accordance with [73]. From Fig. 5-22, it can be observed that in the DL direction, the Joint DSE-CSDR-10ms scheme performs better than the other three schemes having approximately 86% to users below the target delay of 100ms, while the Distributed DSE-CSDR have 82% users below the target delay. Joint DSE-CSDR-640ms have 65% users below the Video target delay. While in the case of the SE-CSDR, 60% users are below the video target delay The CSDR performs poorly with only 52% users below the target VoIP delay. In the UL direction, the delay performance of the video traffic exhibit similar trend as of the DL direction.

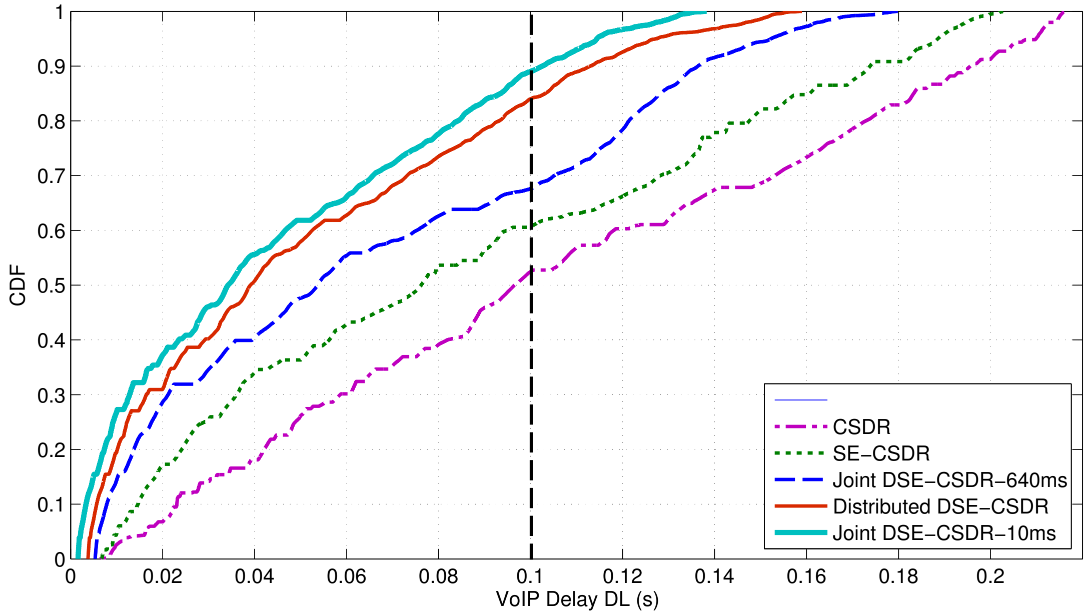

Figure 5-22: VoIP Delay CDF DL (s)

The aggregate throughput in the UL and DL comparing all the 5 schemes is shown in Fig.5-23. In the DL, Joint DSE-CSDR-10ms shows an overall 36% gain compared to the SE-CSDR while 48% compared to the CSDR respectively. Distributed DSE-CSDR demonstrates 24% improvement in performance compared to SE-CSDR and 36% compared to Baseline CSDR respectively. In case of Joint DSE-CSDR-640ms, around 8% improvement is observed compared to SE-CSDR while 17% improvement in performance compared to the CSDR which is the target solution. In the UL direction, the Joint DSE-CSDR-10ms exhibit an overall improvement in performance of 27% compared to the SE-CSDR and 38% compared to the CSDR while the Distributed DSE-CSDR scheme shows an improvement in performance of 18% compared to SE-CSDR and 27% compared to Baseline CSDR respectively. The improvement in UL, in the case of Joint DSE-CSDR-640ms was observed to be 6% compared to the SE-CSDR and 12% compared to the CSDR. In general, the proposed Joint DSE-CSDR-10ms provide finer granularity and allows custom dimensioning of slice size/capacity in line with the traffic variations in UL/DL directions respectively. The ability to adapt UL/DL frames within a slice and periodic slice size/capacity dimensioning with the help of traffic prediction algorithms to match the application specific UL/DL traffic requirements. It also provides the flexibility to program the network in ways that leverage efficient resource usage, enhance the overall system performance and satisfy the application related QoE demands.

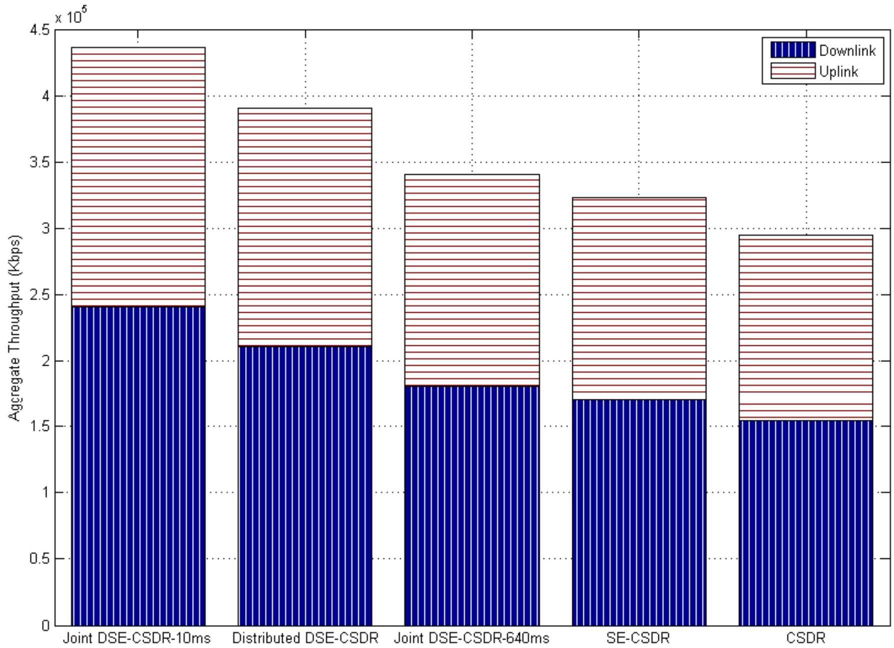


Figure 5-23: Aggregate Throughput (kbps)

## 5.4 Summary and Conclusion

This chapter has presented a novel framework to enhance the benefits offered by application-oriented network resource slicing in evolving TDD networks towards 5G. The proposed framework aims at improving the performance of high priority traffic and at the same time minimize the loss in multiplexing gain that is offered by existing rigid and consolidated network architecture. First, an innovative Slice-Explicit CSDR (SE-CSDR) concept is introduced. Among the several benefits offered by the proposed

SE-CSDR scheme, the most significant is the flexibility improvement by allowing the creation of dynamic customized network slices for a specified duration of time considering individual application requirements. Application-driven network slices enable independent scheduling and CSDR operation that better matches instantaneous traffic demands in both UL and DL. The isolation offered by slicing helps in reducing congestion and eliminates the negative effects of traffic variation between slices. Furthermore, an advanced 5G TDD adaptive network slicing concept is also introduced that further enhances the benefits offered by the SE-CSDR concept. Three variants of adaptive service-oriented adaptive network slicing namely Joint DSE-CSDR-10ms, Joint DSE-CSDR-640ms and Distributed DSE-CSDR are introduced and evaluated against the state-of-the-art CSDR and SE-SCDR solutions. The SDN based network management architecture proposed in this thesis allows 3rd party applications or services to interact with the network and create on-demand service-oriented network slices for specific time durations. The proposed solutions periodically adapt service-oriented network slices considering parameters such as UL/DL traffic demands, application type, user mobility, traffic forecast algorithms etc. The slices are monitored at specified timescales that is different for every scheme. An advanced mechanism to re-configure UL/DL ratio within each slice is also introduced that may be employed jointly with adaptive slice configuration or independently in a distributed way. The system level simulations have shown that the adoption of the proposed SE-CSDR, with independent TDD frame reconfiguration per slice, can significantly improve the system resource utilization as well as the QoE/QoS of high priority traffic. However, the resource utilization efficiency, flexibility and overall network performance can be further enhanced with an adaptive approach with the help of SDN control and traffic forecast algorithms. The three variants of the proposed solution also provide options to select the most appropriate solution depending on the deployment scenario.

# Chapter 6

# Summary and Conclusion

## 6.1 Summary and Conslusion

This thesis has investigated and introduced novel concepts related to service-oriented flexible 5G networks. Various aspects related to 5G enabling technologies such as network virtualization, SDN based network management, multi-connectivity and network slicing etc. have been discussed in detail. First, an innovative multi-connectivity concept called the virtual cell concept for TDD networks has been introduced. Then, a SDN based QoE enhancement mechanism for multiple applications that employ the virtual cell concept is proposed. Ideas related to Multi-tenant resource sharing in FDD/TDD HetNets and network slicing for evolving 5G TDD have been presented to perform efficient resource management and enhance overall system performance. This work has attracted a lot of interest and contributed to a number of research works on multiple topics collaboratively with National and Kapodistrian University of Athens (Greece), Radio Systems Research, Nokia Networks (Finland), Eurecom, Sophia Antipolis (France), within NEC Europe labs (Germany) as well as University of York (U.K.). There were also fruitful discussions and collaboration with various CROSSFIRE project partners. This work has also directly contributed to the European Commission ITN FP7 Marie Curie project called CROSSFIRE (MITN 317126) and partially to the H2020-ICT-2014-2 project 5G NORMA. A brief summary and conclusions for the thesis are given below:

The first chapter provides a brief introduction to the thesis and the purpose of this work. Chapter 2 presents a comprehensive literature review on the research related to this work. This focuses mainly on mobile network virtualization, overview of TD-LTE and 5G enabling technologies such as SDN , Network Slicing etc. Chapter 3 introduces the concept of virtual cells in Time-Division Long Term Evolution (TD-LTE) systems, which enables users residing in overlapping cell coverage regions to utilize resources from multiple base stations. The benefits of virtual cells are realized by the efficient resource utilization, by adapting the network resource availability with the traffic demand additionally taking into account the impact of cross-slot interference. The virtual cell concept also helps in avoiding and resolving congestion issues. The virtual cell concept is further developed with the Software Defined Network (SDN) paradigm, which monitors network resource utilization and allows applications or services to request resources. The resources requested by the applications or services can be allocated on-demand by adjusting the Time Division Duplex (TDD) frames in different regions of the geographical area being considered. This is accomplished by creating virtual cells in the overlapping regions that can customize services for the residing users. An extensive simulation study has been carried out to elaborate the benefits of this approach and the performance enhancements in comparison with the state-of-the art and conventional TD-LTE configurations are presented.

Chapter 4 discusses SDN based Elastic Resource Management and QoE Enhancements in 5G Networks. First, an SDN-based framework for enhanced Quality of Experience (QoE) in the presence of TD-LTE pico cell hotspots is proposed. The proposed framework enables elastic radio resource management and allows users to utilize radio-resources from multiple base stations and enhance their QoE, by using the virtual cell concept. TD-LTE UL/DL subframes are customized as per the UL/DL traffic demands at the pico cell base stations. Following this, a novel SDN-based framework that enables efficient and elastic spectrum utilization among multiple operators (i.e. in a multi-tenant scenario) in a 3GPP LTE-A HetNet scenario is investigated. It was observed that the sharing of resources with the FDD macro eNB network may degrade the performance of the TDD pico eNB system, however, the application of the proposed SDN-based framework with the dynamic re-configuration of the UL/DL TDD frames at each pico eNBs at a specific timescale, is shown to minimize the losses and improve the performance of the overall TDD pico eNB system.

A framework to enhance the benefits offered by the application-oriented network resource slicing in evolving TDD networks towards 5G is proposed in chapter 5. The proposed framework improves the performance of the applications and services and at the same time reduces the loss in multiplexing gain that the conventional mobile network with no network slicing employed offers. In this chapter, multiple adaptive and static slicing mechanisms are discussed and simulations are carried out to evaluate the performance enhancements as compared to the state-of-the-art schemes. Out of the several benefits offered by the proposed mechanism, some of the most significant include improved flexibility achieved via on-demand creation of dynamic customized network slices for a specified duration of time considering individual application requirements. The isolation property of the network slicing concept helps in countering congestion problems and ensures desired performance for the respective service that the slice serves. Application-driven network slices enable independent scheduling and customized UL/DL frame re-configuration operations in each slice that better match instantaneous traffic demands in both UL/DL directions and improves resource utilization efficiency. In this thesis, novel mechanisms to adaptively form service-oriented slices and configure UL/DL TDD subframes within each slice, in either, a combined way or distributed manner are developed and analyzed. It was observed that the proposed solutions improve the overall system performance and enhance the QoE of the applications and services.

Some new ideas and directions to further evolve the work presented in this thesis are discussed in chapter 7. Finally, chapter 6 highlights novel contributions of this work and concludes this thesis.

## 6.2 Summary of Novel Contributions

This thesis focuses on various aspects of network virtualization and SDN-based flexible network management including concepts such as multi-connectivity and network slicing that are essential for supporting the demands of the future applications and services. The performance of the proposed solutions and concepts in this thesis have been evaluated in comparison to the state-of-the art work done in this area. Most of the work has been published in a number of peer-reviewed journals and conferences. This thesis work was carried out as a part of European Marie Curie Innovative Training Network (ITN) project called CROSSFIRE. Also, some of the work done in this thesis, partially contributed in European H2020 project-5G NORMA. The thesis work was carried out at NEC Europe Labs located in Heidelberg, Germany.

### 6.2.1 TD-LTE Virtual Cells: An SDN Architecture for User-centric Multi-eNB Elastic Resource Management

In chapter 3, the concept of virtual cell is proposed which enables users residing in overlapping cells' coverage regions to utilize resources from multiple base stations. To date the state of the art has focused on configuring specific UL/DL ratio that matches the best long term traffic demands. In this case, the neighboring cells typically follow the same UL/DL configuration in order to avoid cross-slot interference, i.e. interference among neighboring eNBs or among UEs residing within a close proximity as described in [35]. Such synchronization introduces limitations in resource allocation when neighbor cells have different traffic demands. Methods that aim to enhance resource flexibility for TD-LTE systems concentrate on relaxing such synchronization [92], and on introducing dynamic resource management via channel allocation and scheduling [93]. The idea of Cell Specific UL/DL frame reconfiguration is discussed in [33] wherein, the benefits of adaptive UL/DL reconfiguration are verified via simulations.

Different from the state of the art, in this thesis, the concept of virtual cells has been developed. The benefits offered by the virtual cells are realized by the efficient resource utilization via adapting the network resource availability with the traffic demand taking also into account the impact of cross-slot interference. Besides the increased resource flexibility, virtual cells also resolve pseudo congestion, and introduce a customized, user specific, resource utilization. Such a feature enables mobile users to utilize sub-frames from different base stations creating an on demand virtual frame.

The notion of allowing users to utilize resources from multiple eNBs is also supported by the 3GPP Coordinated Multi-Point (CoMP) [94] to enhance the cell edge throughput via joint processing or coordinated scheduling and beamforming methods. By contrast to the Virtual Cell Concept proposed in this thesis, CoMP requires coherent transmission and detection at physically separated eNBs, which jointly process the transmit/receive signal in order to gain from array and diversity gains. CoMP relies on the fundamental requirement to align the UL and DL resources of multiple cells without providing any means to counteract pseudo congestion as the applied UL/DL sub-frame pattern is assumed to be fixed for all cells, i.e. a UE utilizes resources from different eNBs but in the same transmission direction. The virtual cell proposal concentrates on customizing the TDD frame according to the user traffic demand, which is beyond the use of conventional CoMP advancing the current efforts towards more efficient resource allocation. The detailed explanation and performance evaluation analysis of the virtual cell concept have been published in IEEE Vehicular Technology Conference (VTC) [45] and Elsevier Computer Communications Journal [95].

Furthermore, this thesis introduces the mechanisms and network management architecture that can facilitate a broader adoption of virtual cells in TDD 5G networks. It encounters also a closer coordination with cell specific adaptive UL/DL frame re-configuration, i.e. enforcing a selected frame re-configuration on particular cells that may enable an improved virtual cell formation, without compromising the performance of other existing users. Considering the network management architecture, this thesis adopts the SDN paradigm to facilitate a fine-grained network resource control based-on a global network view, which also enables application and service providers to acquire QoS and resources via the Application-Controller Plane Interfaces (A-CPI) [68] . Although this later feature is not explicitly explored into this work, the proposed mechanisms can accommodate on-demand resource allocation, which reflects the generic case for OTT resource acquisition via the use of the A-CPI. Based-on this information, and on knowledge of incoming requests from application and service providers, the SDN controller: (i) feeds an algorithm, which identifies the TD-LTE frame configuration for particular eNBs inside the radio access network and (ii) enforces such a new TD-LTE frame configuration at the selected eNBs allowing in this way a more efficient resource management with the option of enabling virtual cells at desired locations. In particular, the contributions of this thesis are:

- A framework for adopting the SDN paradigm for virtual cells considering a macro cellular deployment scenario, where SDN is used to enforce a re-configuration of the UL/DL TDD ratio at particular eNBs, in order to secure resources for virtual cells. Such resources can serve the required traffic demands as closely as possible resolving pseudo congestion, without increasing the interference for all other users beyond a certain limit, i.e. making sure that the overall impact of virtual cells is positive.
- The introduction of an SDN network management architecture that supports

the monitoring and dynamic control of virtual cells based-on observed traffic demands.

- The analytical logic for allocating resources on virtual cells considering the amount of resource blocks that enhance UEs' resource allocation and network utilization without compromising the performance of other existing users.

- The algorithm for forming virtual cells and enforcing TD-LTE frame re-configuration on selected cells in order to overcome pseudo-congestion.

- The simulation study that provided a comparative evaluation of the proposed virtual cell solution with the cell specific adaptive re-configuration and with the simple static configuration.

### 6.2.2 SDN-based Elastic Resource Management and QoE Enhancements in 5G Networks

An SDN framework is proposed in chapter 4, which aims at enhancing the QoE perceived by the users running OTT applications with the help of a multi-connectivity virtual cell concept discussed in chapter 3.

So far the state-of-the art has focused on various techniques with and without SDN for improving QoS and QoE for the end users. The state-of-the art in this domain mainly aims to meet the individual QoS requirements at the end users side (i.e. in terms of throughput and packet delay). For example, the authors in [41] have proposed a QoS aware dynamic uplink-downlink reconfiguration algorithm in TD-LTE HetNet, however, they mainly focus on the QoS aspects and did not consider SDN based network management or multi-connectivity concepts to improve the QoE of the end user applications. In contrast to the state-of-the art, this thesis focus on the scenario where the LTE network is composed by TD-LTE picocells that can better handle asymmetric traffic and effectively match instantaneous traffic demands by adapting the UL/DL frame ratio. Further extending the state-of-the art, the framework proposed in this thesis enables customized TD-LTE frame ratio in line with the QoE demands at the end users. Besides, in this work we utilize the SDN paradigm [17] to achieve efficient synchronization of the TD-LTE picocells in order to enable an SDN-enhanced formation of virtual cells. In this manner, the novel method to enable interaction between the OTT application providers and network operators is introduced. Moreover, an SDN controller is considered which obtains the global view of the TD-LTE picocell and the QoE of the users with the help of 3GPP OAM system. The SDN controller then can use the available information to create instances of virtual cells and adapt the UL/DL frames of then serving and the neighboring picocells on-demand. Thus in this way, the proposed framework helps in resolving and avoiding congestion problems including pseudo-congestion. Besides, the proposed framework enable cell-edge users utilize resources from multiple picocells via virtual cells. Hence with the help of SDN based network resource management, UL/DL frame adaptation and application of virtual cells can improve service quality and network resource utilization. The analysis and discussion related to this proposed concept has been published in the IEEE Conference on Quality of Multimedia Experience (QoMEX) [48].

### 6.2.3 An SDN Framework for Elastic Resource Sharing in Integrated FDD/TDD LTE-A HetNets

In chapter 4.3, an LTE-A HetNet scenario is considered where a common infrastructure provider owns a set of macro cell eNBs, termed as macro eNBs, and a set of picocell eNBs, termed as pico eNBs. All the macro eNBs are assumed to use FDD, whereas all pico eNBs are assumed to use TDD. Existing literature reviews focus on enabling on-demand sharing of base stations that belongs to multiple operators in an efficient manner. In this direction, an SDN-based framework is proposed in [61] where the main idea is to allow the cellular users to attach to the nearest base station. The OpeRAN concept proposed in [62] aims to enhance the convergence benefits of HetNets and leverage network customization. To achieve higher granularity, network virtualization concept is applied at the application, spectrum and network level. In, [63], SDN concepts to manage cellular network are analyzed.

The idea proposed within this thesis considers a scenario where the infrastructure provider leases its infrastructure to multiple macrocell operators and multiple picocell operators. Under this model, we address the problem of elastic resource sharing between the FDD macro eNB operators and the TDD pico eNB operators in a dynamic manner. The proposed concept provides flexibility by introducing an innovative SDN-based architecture that enables dynamic resource sharing among the different operators and employs a dynamic re-configuration mechanism that allows the TDD pico eNBs to adapt their UL/DL ratio in order to release or embody a set of radio resources to/from the FDD macro eNBs operators.

Different from the state-of-the art ([61]- [63]), this thesis focus on the resource sharing problem in a multi-operator integrated FDD macrocell and TDD picocell LTE-A system, employing network programmability through re-configuration of the TDD frames at the pico eNBs. In addition, the work in this thesis considers the amount of resources shared in the FDD system and the requirements of the users associated with the TDD picocells. In [76], a joint TDD and FDD scenario is considered where in the TDD operate in the guard band spectrum between the FDD UL and DL. In contrast to this, the work in this thesis investigates elastic resource sharing in a multi-tenant, multi-operator FDD/TDD LTE-A HetNet environment. System level simulation results demonstrate that the combination of these solutions in a resource sharing scenario between a TDD pico eNB operator to a FDD macro eNB operator achieves notable performance gains in terms of communication delay in the application layer. This work has been published in the IEEE CloudNet [66].

### 6.2.4 Service-Oriented Resource Virtualization for Evolving TDD Networks Towards 5G

In chapter 5, this thesis proposes a slicing concept for evolving 5G TDD networks, considering a network deployment with heterogeneous and asymmetric traffic conditions. In this case TDD is adopted for macro base stations. Effectively, distinct network slices are provided to each tenant that employs a different type of service. In particular, a framework is proposed for network virtualization, wherein a novel service-oriented network resource slicing scheme is adopted. Different aspects of network slicing and network programmability in 5G networks considering also the impact on services and applications have been discussed in [96] [47]. While an SDN based network slicing architecture that aims to accommodate heterogeneous RAN requirements is presented in [5]. The main idea behind the proposed architecture in [5] is to satisfy QoE needs of different services and support resource sharing in a multi-tenant environment.

In contrast to the above mentioned state-of-the art, the proposed concept introduces a new notion of network slicing in TDD networks that has never been discussed before and aims to enhance the flexibility and performance of application aware network slices, while limiting the loss of multiplexing gains. The proposed SDN based network slicing and resource management concept in this thesis has been filed as a patent [97]. The performance evaluation and detailed analysis related to this work has been published in IEEE Wireless Communications and Networking Conference (WCNC) conference [91].

The proposed slicing concept aims at enhancing system network performance, applications' QoE and network resource utilization efficiency by introducing additional flexibility in resource allocation to different tenants. To achieve this, the proposed scheme deploys a Cell Specific Dynamic UL/DL Re-configuration (CSDR) per slice in line with the real time traffic demands. Such dynamic TDD operation per slice is shown to reduce the loss of multiplexing gain caused by the isolation property of the resource slicing. Thus, the proposed mechanism advances the current state of the art towards more flexible application oriented network management considering the heterogeneous requirements of diverse applications in 5G.

The contributions include:

- A service specific network slicing concept in TDD networks that permits efficient resource provision and allows each slice to support a different UL/DL ratio in order to effectively follow the service-oriented traffic dynamics.
- An SDN-based architecture that provides network slicing considering different tenants and the mechanism for performing UL/DL re-configuration per slice.
- The performance evaluation of the proposed scheme using a QoE-based methodology considering a set of different variations of the CSDR scheme that takes into account either the aggregate traffic load or the load of a particular application.

### 6.2.5 Service-tailored Network Slicing for emerging 5G-TDD Networks

The proposed static 5G TDD network slicing in this thesis, in chapter 5, is further developed towards an on-demand, adaptive 5G TDD network slicing approach. The proposed adaptive on-demand 5G TDD network slicing concept allows adaptation of the amount of allocated resources to the each slice, including the corresponding UL/DL frame ratio of each slice, based on a traffic forecast, user mobility and real time UL/DL traffic demand of the respective slice applications/services. This approach leads to overall improvement in resource utilization efficiency. The proposed solution provides a flexible allocation of resource blocks considering the entire spectrum rather than the a-priori static spectrum reservation, adopting the notion of Network Virtualization Substrate (NVS) [98] that defines a two-step scheduling process, one controlled by the slice tenant, while the other is controlled by the virtualization layer that allocates network slices.

The contributions include:

- Service-oriented network slicing operations for 5G TDD networks that allocates and adjusts network resources dynamically considering UL/DL forecasted traffic demands and user mobility for each service type, instead of a fixed slice size with a reactive UL/DL adjustment based-on aggregate or service-specific traffic conditions.

- A flexible spectrum allocation mechanism per slice. This involves flexible allocation of resource blocks to particular slices and periodically adapting the slice capacity as the traffic load varies within the available spectrum band. Hence, allowing in this way, variable application-oriented slice dimensions across different eNBs in the RAN, rather than the a-priori static spectrum reservation, wherein the allocation of resource blocks considers only a particular sub-set of the spectrum band.

- A novel metric for selecting a suitable TDD frame configuration of each slice based-on the predicted traffic load in the UL and DL direction respectively and spatial throughput across a specified RAN region, which reflects the probability that a user visits a particular location at a certain known point in time. Such a metric is used to adjust the TDD frame configuration, instead of the conventional UL/DL buffer status and past average throughput.

- An analytical model, formulating the network slice resource allocation as a weighted optimization problem considering the newly introduced resource allocation metric in order to optimize the capacity offered for each slice and the corresponding UL/DL ratio

- Two TDD slice (re-)configuration mechanisms that regularly adjust the resources allocated per slice considering the desired UL/DL service ratio that avoids over-provisioning of network resources without compromising the desired SLAs.

  (i) Joint slice resource allocation and TDD frame configuration (Joint DSE-CSDR-640ms and Joint DSE-CSDR-10ms), wherein the slice size is optimized jointly, i.e. in a single step, with the desired UL/DL service ratio. The DSE-CSDR-640ms and Joint DSE-CSDR-10ms configures the same UL/DL ratio across the entire RAN slice, updating each slice independently at regular intervals for example 640ms or 10ms.

  (ii) Slice allocation with distributed UL/DL ratio adjustment (Distributed DSE-CSDR), where the slice size is allocated in a centralized manner, lasting relatively long time durations, while the UL/DL ratio is adjusted independently at each eNB in a distributed manner.

- Simulation study that compares the proposed algorithms with the state of the art considering

  (i) Fixed slicing for the duration of the service with distributed cell-specific UL/DL ratio adjustments (SE-CSDR) and (ii) Cell specific dynamic UL/DL re-configuration with no slicing (CSDR).

In this way, with the help of proposed SDN architecture and the adaptive 5G TDD network slicing concept in this thesis, resource flexibility is improved. Thus, in this thesis, a novel idea of real time adaptive service-oriented network slicing in 5G TDD RAN is introduced.

### 6.2.6 Evolving LTE/LTE-A/5G System Level Simulator

A novel 3GPP compliant system level simulator is developed in this work that accurately evaluates the effects and performance gains of the proposed concepts and network architectures. The simulator integrates the network slicing concepts introduced within the framework of this thesis. It also evaluates various algorithms and concepts proposed in this work. The 3GPP RAN with different scheduling techniques such as round robin, EXP rule, video and VoIP oriented scheduling have been modelled. A modular approach is adopted to maximize the flexibility of the simulator and provide compatibility with future developments. A number of modules are developed to model different aspects of the system. These modules include a location module, propagation module, RRM module and traffic module with different types of traffic models such as video, VoIP, FTP and best effort. Traffic models are implemented following the standard state of the art and 3GPP defined methodologies. The mobility module uses the self-similar least action walk (SLAW) model to emulate user mobility. In some cases, fading is modeled using Jakes model. Functionalities such as traffic prediction are performed following the Holt-Winter's Exponential Smoothing Concept. Application MOS calculation is done employing the MOS models of the respective applications. Detailed description of the simulation, analysis and evaluation methodology for every proposed solution within this thesis is provided in each chapter separately.

# Chapter 7

# Future Work

## 7.1 Virtual Cell Concept for Evolving TDD networks-5G and Beyond

The virtual cell concept may be explored in the context of TDD Full Duplex Communication (FDC). More specifically, it will be interesting to investigate how the virtual cell concept can be employed and how much performance enhancement it can provide when the FDC enabled eNBs or UEs are available. FDC has the potential to increase the spectral efficiency to upto 2 times the current limit [99] in UL and DL directions respectively. The FDC has the potential to double the capacity of the Half Duplex system at the link level. However, the improvement in capacity is restricted due to the increase in overall aggregate interference in each communication link. FDC systems introduce an additional form of interference called self-interference which needs to be taken into account when considering an FDC enabled system. This interference problem gets even more serious when a dynamic TDD operation is employed where in, the UL/DL frames are adapted at a specific timescale to match the real time traffic demands in UL and DL directions respectively. Within such a dynamic TDD system, in addition to the UE-UE, BS-BS and self-interference (in case of FDC), there is also a possibility for cross-slot interference to occur, in addition to the UE-UE, BS-BS and self-interference (in the case of FDC). Hence such an arrangement makes the scenario even more complex. Therefore it would interesting to explore how much capacity improvement can be achieved in such a dynamic environment considering also multi-connectivity concepts such as the virtual cell concept introduced and investigated in this thesis.

## 7.2 Considering Different Frequency Ranges and mmWave Technology

Further research is envisioned to analyze more complex scenarios considering different frequency ranges for example, below 6GHz and above 6GHz or mmWave. In case of higher frequency bands, mechanisms to efficiently manage high bandwidth under inteference limitations considering different application or service requirements may be investigated. On-demand multi-connectivity across different or same frequency bands would also be interesting to explore. More advanced and complex mechanisms may be investigated to perform UL/DL reconfiguration (with or without network slicing) under alternating traffic conditions considering Enhanced Mobile Broadband (eMBB), as well as Low Latency and High Reliability to enable some Ultra-Reliable and Low Latency Communications (URLCC) use cases. How the concepts and mechanisms introduced in this thesis, will work in a 5G New Radio (NR) scenario would also be interesting to explore.

## 7.3 Multi-Access Edge Computing (MEC) or Cloud RAN (C-RAN) Scenarios

Further research is also recommended in studying the use of virtual cells in a MEC or cloud RAN scenario similar to [100] where, a centralized entity or controller controls a set of radio access points. Such a scenario helps in reducing infrastructure costs especially in the case of dense deployments. The centralized approach allows optimization of certain operations, that may be computation intensive, related to control and data plane of the associated cells. MEC is designed to offer high performance computing capabilities and IT service environment at the edge of the mobile network (MEC doc). It provides cloud computing platform to application developers and content providers. MEC provides an environment that is essential to support ultra-low latency and high bandwidth and provide real time access to the RAN information that can be exploited by applications and services to meet their performance targets. The operators can open their MEC platform at the RAN to authorized third parties and allow them to flexibly and rapidly deploy innovative applications and services to serve the end users, enterprises and vertical markets. The SDN based mechanisms proposed in this thesis may also be extended considering split bearers and towards dynamic adjustments in the mobile backhaul provisioning resources for lower layer transport mechanisms. Novel SDN-based frameworks that enable QoE-aware network management for OTT mobile services have also been proposed in this thesis therefore it will be interesting to have the option of dynamically placing certain core network and RAN functions at selected locations inside the network or at the network edge according to the QoS/QoE requirements of individual applications. Flexible Joint RAN and Backhaul management mechanisms considering different backhaul technologies may be developed. Such mechanisms may also help in exploiting path diversity gains in the backhaul. Fig. 7-1 illustrates some of the main functional blocks of the ideas discussed in this section, where B-Type indicates the type of backhaul technology used.

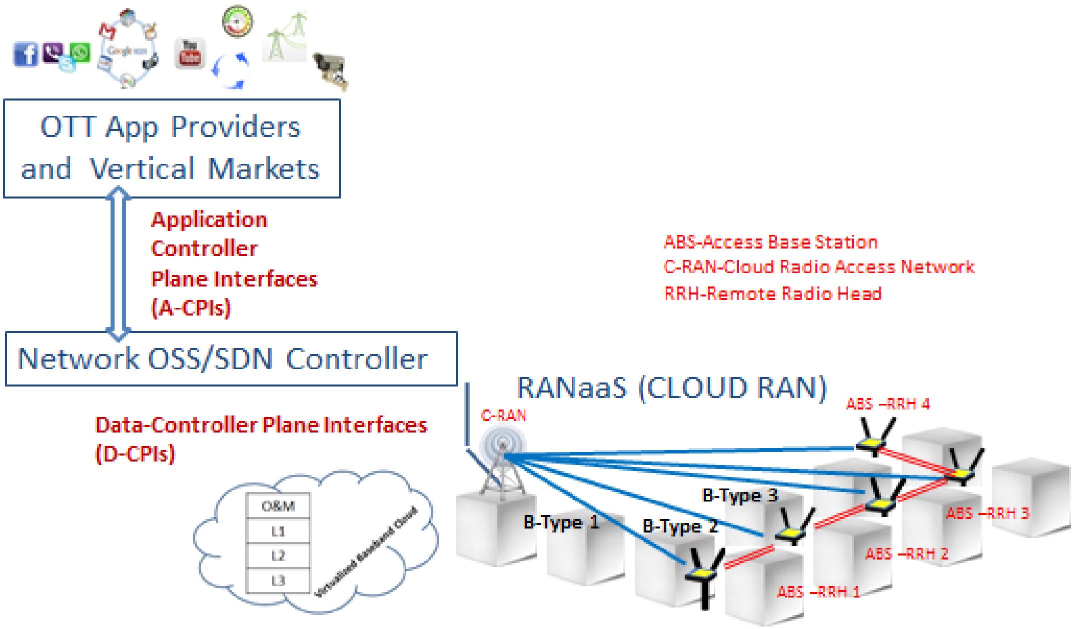


Figure 7-1: Flexible C-RAN Scenarios

## 7.4 Network Slicing for 5G and Beyond Networks

The concept of network slicing refers to the notion of deploying multiple dedicated logical mobile networks on top of same physical infrastructure with varying levels of mutual isolation. A network slice maybe composed of mobile network functions (groups) and a specific set of radio access technologies (RATs) or specific RAT configurations required to operate an end-to-end independent logical network [24]. The set of network functions and configurations may be grouped in a way that slice specific data and control plane functionality is customized as per the requirements of various use cases, users or clients and business models. Network slicing acts as an enabler for multi-tenancy and allows service-oriented mobile network operations. The work in this thesis introduces new concepts involving network slicing and service-tailored network management for 5G networks and beyond. The work can be further extended to explore network slicing concept considering heterogeneous applications in different deployment scenarios. The configuration of network slices considering 5G new radio (NR) and multiple RATs may also be explored. The 5G NR is still under development in 3GPP. With different RATs and new 5G technologies, efficient and dynamic management of network resources considering 5G heterogeneous applications requirements could be a challenge. In this direction it would be interesting to investigate machanisms to adaptively and efficiently manage the network in real-time with the aim to support demands of the 5G and beyond applications and services.

# Abbreviations and Acronyms

Table 7.1: List of Abbreviations and Acronyms

| **Notation** | **Description** |
|---|---|
| $3GPP$ | Third Generation Partnership Project |
| $A-CPI$ | Application-Controller Plane Interface |
| $AMR$ | Adaptive Multi-Rate |
| $AoA$ | Angle of Arrival |
| $API$ | Application Program Interface |
| $AS$ | Access Stratum |
| $AWGN$ | Additive White Gaussian Noise |
| $BER$ | Bit Error Rate |
| $BGP$ | Border Gateway Protocol |
| $BS$ | Base Station |
| $CA$ | Carrier Aggregation |
| $CBR$ | Constant Bit Rate |
| $CDF$ | Cumulative Distribution Function |
| $CQI$ | Channel Quality Indicator |
| $CSDR$ | Cell-Specific Dynamic Re-configuration |

Table 7.1 continued..

| **Notation** | **Description** |
|---|---|
| $D2D$ | Device-to-Device |
| $DCI$ | Downlink Control Information |
| $D-CPI$ | Data-Controller Plane Interface |
| $DL$ | Downlink |
| $DPI$ | Deep Packet Inspection |
| $eNB$ | evolved NodeB |
| $EPC$ | Evolved Packet Core |
| $E-UTRA$ | Evolved Universal Terrestrial Radio Access |
| $DL$ | Downlink |
| $FDD$ | Frequency Division Duplex |
| $FFR$ | Fractional Frequency Reuse |
| $FRF$ | Frequency Reuse Factor |
| $FTP$ | File Transfer Protocol |
| $GBR$ | Guaranteed Bit Rate |
| $GW$ | Gateway |
| $GWCN$ | Gateway Core Network |
| $HAS$ | HTTP Adaptive Streaming |
| $HeNB$ | Home evolved NodeB |
| $HetNets$ | Heterogeneous Networks |
| $ICIC$ | Inter-Cell Interference Coordination |
| $IMS$ | IP Multimedia Subsystem |
| $ISP$ | Internet Service Provider |

Table 7.1 continued..

| Notation | Description |
|---|---|
| $ITU$ | International Telecommunication Union |
| $KPI$ | Key Performance Indicator |
| $LTE$ | Long Term Evolution |
| $LTE-A$ | Long Term Evolution-Advanced |
| $MAC$ | Medium Access Control |
| $MCS$ | Modulation and Coding Scheme |
| $MIMO$ | Multiple Input Multiple Output |
| $MME$ | Mobility Management Entity |
| $MNO$ | Mobile Network Operator |
| $MOCN$ | Multi-Operator Core Network |
| $MO-NM$ | Master Operator-Network Manager |
| $MOS$ | Mean Opinion Score |
| $MO-SR-DM$ | Master Operator-Shared RAN-Domain Manager |
| $MPLS$ | Multi-Protocol Label Switching |
| $MVNO$ | Mobile Virtual Network Operator |
| $NFV$ | Network Functions Virtualization |
| $NGN$ | Next Generation Network |
| $NVS$ | Network Virtualization Substrate |
| $OAM$ | Operations Administration and Maintenance |
| $OFDM$ | Orthogonal Frequency Division |
| $OFDMA$ | Orthogonal Frequency Division Access |
| $OTT$ | Over-The-Top |

Table 7.1 continued..

| **Notation** | **Description** |
|---|---|
| $PCRF$ | Policy and Charging Rules Function |
| $PEP$ | Packet Error Probability |
| $PF$ | Proportional Fair |
| $PSNR$ | Peak Signal-to-Noise Ratio |
| $QoE$ | Quality of Experience |
| $QoS$ | Quality of Service |
| $RAC$ | Radio Admission Control |
| $RAN$ | Radio Access Network |
| $RB$ | Resource Block |
| $RIB$ | RAN Information Base |
| $RRC$ | Radio Resource Control |
| $RRM$ | Radio Resource Management |
| $RSRP$ | Reference Signal Received Power |
| $RTT$ | Round Trip Time |
| $SC$ | Small Cell |
| $SDN$ | Software-Defined Networking |
| $SINR$ | Signal to Interference plus Noise Ratio |
| $SLA$ | Service Level Agreement |
| $SO-NM$ | Sharing Operator-Network Manager |
| $TBS$ | Transport Block Size |
| $TDD$ | Time Division Duplex |
| $TTI$ | Transmission Time Interval |

Table 7.1 continued..

| **Notation** | **Description** |
|---|---|
| $UDP$ | User Datagram Protocol |
| $UE$ | User Equipment |
| $UL$ | Uplink |
| $VPN$ | Virtual Private Network |

# Bibliography

[1] Cisco Networks. "*Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018*".

[2] GSMA. *The Mobile Economy*. GSMA, 2014.

[3] *3GPP TR 22.891, "Feasibility Study on New Services and Markets Technology Enablers", Stage 1, Rel. 14*.

[4] 5G Infrastructure Association. *5G Vision - The 5G Infrastructure Public Private Partnership: The Next Generation of Communication Network and Services, White paper, 5G Infrastructure Association*. February 2015.

[5] Carlos J Bernardos, Antonio De La Oliva, Pablo Serrano, et al. "An architecture for software defined wireless networking". In: *IEEE Wireless Communications* 21.3 (2014), pp. 52–61.

[6] ONF. *Open Networking Foundation (ONF), TR-521, SDN Architecture (2016)*. Tech. rep. ONF, 2016.

[7] 3GPP. *3GPP TS 36.300, Technical Specification Group Radio Access Network; E-UTRA and E-UTRAN; Overall description, Rel 12*. Tech. rep. 3GPP.

[8] 4G Americas. *Executive Summary Inside 3GPP Release-13: Understanding the Standards for HSPA+ and LTE-Advanced Enhancements*. September 2015.

[9] Nick McKeown, Tom Anderson, Hari Balakrishnan, et al. "OpenFlow: Enabling Innovation in Campus Networks". In: *SIGCOMM Comput. Commun. Rev.* 38.2 (Mar. 2008), pp. 69–74. ISSN: 0146-4833. DOI: 10.1145/1355734.1355746. URL: http://doi.acm.org/10.1145/1355734.1355746.

[10] ONF. *Open Networking Foundation, "OpenFlow-Enabled Mobile and Wireless Networks", ONF Solution Brief*. September 2013.

[11] Marco Hoffmann and Markus Staufer. "Network virtualization for future mobile networks: General architecture and applications". In: *2011 IEEE International Conference on Communications Workshops (ICC)*. IEEE. 2011, pp. 1–5.

[12] Yasir Zaki, Liang Zhao, Carmelita Goerg, et al. "LTE mobile network virtualization". In: *Mobile Networks and Applications* 16.4 (2011), pp. 424–432.

[13] Jorge Carapinha and Javier Jiménez. "Network virtualization: a view from the bottom". In: *Proceedings of the 1st ACM workshop on Virtualized infrastructure systems and architectures*. ACM. 2009, pp. 73–80.

[14] *RAN sharing: NEC's Approach towards Active Radio Access Network Sharing, White Paper, 2013.*

[15] Ravi Kokku, Rajesh Mahindra, Honghai Zhang, et al. "CellSlice: Cellular wireless resource slicing for active RAN sharing". In: *2013 Fifth International Conference on Communication Systems and Networks (COMSNETS)*. IEEE. 2013, pp. 1–10.

[16] ONF. *Open Networking Foundation (ONF), TR-518, Relationship of SDN and NFV*. Tech. rep. October 2015.

[17] ONF. *Open Networking Foundation (ONF), TR-502, SDN Architecture (1)*. Tech. rep. ONF, June 2014.

[18] ETSI GS NFV-MAN. *Network Function Virtualization (NFV); Management and Orchestration, v1.1.1*. Dec. 2014.

[19] T. Nadeau P. Quinn. *Problem Statement for Service Function Chaining*. IETF RFC 7498. Apr. 2015.

[20] NFV. *Network Functions Virtualization, Introductory White Paper.*

[21] NGMN Alliance. *5G White Paper*. Tech. rep. Feb. 2015.

[22] ONF. *TR-526, Applying SDN Architecture to 5G Slicing*. Tech. rep. Apr. 2016.

[23] *3GPP TR 23.799, Study on Architecture for Next Generation System, Rel.14*.

[24] P. Rost, A. Banchs, I. Berberana, et al. "Mobile network architecture evolution toward 5G". In: *IEEE Communications Magazine* 54.5 (May 2016), pp. 84–91. ISSN: 0163-6804. DOI: 10.1109/MCOM.2016.7470940.

[25] P. Marsch, I. Da Silva, O. Bulakci, et al. "5G Radio Access Network Architecture: Design Guidelines and Key Considerations". In: *IEEE Communications Magazine* 54.11 (Nov. 2016), pp. 24–32. ISSN: 0163-6804. DOI: 10.1109/MCOM.2016.1600147CM.

[26] A. Aijaz. "Hap-SliceR: A Radio Resource Slicing Framework for 5G Networks With Haptic Communications". In: *IEEE Systems Journal , vol.PP, no.99, pp.1-12* (2017).

[27] Oriol Sallent, Jordi Perez-Romero, Ramon Ferrus, et al. "On Radio Access Network Slicing from a Radio Resource Management Perspective". In: *IEEE Wireless Communications* (2017).

[28] K. Samdanis, X. Costa-Perez, and V. Sciancalepore. "From network sharing to multi-tenancy: The 5G network slice broker". In: *IEEE Communications Magazine* 54.7 (July 2016), pp. 32–39. ISSN: 0163-6804. DOI: 10.1109/MCOM.2016.7514161.

[29] Matías Richart, Javier Baliosian, Joan Serrat, et al. "Resource slicing in virtual wireless networks: A survey". In: *IEEE Transactions on Network and Service Management* 13.3 (2016), pp. 462–476.

[30] Shanzhi Chen, Shaohui Sun, Yingmin Wang, et al. "A comprehensive survey of TDD-based mobile communication systems from TD-SCDMA 3G to TD-LTE (A) 4G and 5G directions". In: *China Communications* 12.2 (2015), pp. 40–60.

[31] 3GPP. *3GPP TS 36.133 v10.3.0, 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (EUTRA); Requirements for support of radio resource management (Release 10)*. Research rep. 3GPP.

[32] *3GPP TR 36.828, "Further Enhancements to LTE TDD for DL-UL Interference Management and Traffic Adaptation", Rel 11, Jun. 2012*.

[33] Zukang Shen, Alexey Khoryaev, Erik Eriksson, et al. "Dynamic uplink-downlink configuration and interference management in TD-LTE". In: *IEEE Communications Magazine* 50.11 (2012), pp. 51–59.

[34] 3GPP. *3GPP TS 36.211 v10.2.0, Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (EUTRA); Physical Channels and Modulation (Release 10)*.

[35] Howon Lee and Dong-Ho Cho. "Combination of dynamic-TDD and static-TDD based on adaptive power control". In: *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*. IEEE. 2008, pp. 1–5.

[36] Peter Rost and Konstantinos Samdanis. “The two-way interference channel: Robust cooperation to exploit limited backhaul efficiently”. In: *Systems, Communication and Coding (SCC), Proceedings of 2013 9th International ITG Conference on*. VDE. 2013, pp. 1–6.

[37] Yufeng Wang, Kimmo Valkealahti, Kodo Shu, et al. “Performance evaluation of flexible TDD switching in 3gpp LTE system”. In: *Sarnoff Symposium (SARNOFF), 2012 35th IEEE*. IEEE. 2012, pp. 1–4.

[38] 3GPP. *3GPP TS 36.331, v10.3.0, Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (EUTRA); Radio Resource Control (RRC); Protocol specification (Release10)*.

[39] Alexis A Dowhuszko, Olav Tirkkonen, Juha Karjalainen, et al. “A decentralized cooperative Uplink/Downlink adaptation scheme for TDD Small Cell Networks”. In: *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE. 2013, pp. 1682–1687.

[40] Mohammed S ElBamby, Mehdi Bennis, Walid Saad, et al. “Dynamic uplink-downlink optimization in TDD-based small cell networks”. In: *2014 11th International Symposium on Wireless Communications Systems (ISWCS)*. IEEE. 2014, pp. 939–944.

[41] Yanchao Lin, Yuehong Gao, Yuancao Li, et al. “QoS aware dynamic uplink-downlink reconfiguration algorithm in TD-LTE HetNet”. In: *2013 IEEE Globecom Workshops (GC Wkshps)*. IEEE. 2013, pp. 708–713.

[42] Davide Catania, Marta Gatnau Sarret, Andrea F Cattoni, et al. “The potential of flexible UL/DL slot assignment in 5G systems”. In: *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*. IEEE. 2014, pp. 1–6.

[43] Li-Chun Wang and Yi-Cheng Chen. “A novel link proportional dynamic channel assignment for TDD-CDMA systems with directional antennas”. In: *Proc. IEEE Int. Conf. Networking, Sensing and Control*. Vol. 1. Mar. 2004, 164–169 Vol.1. DOI: 10.1109/ICNSC.2004.1297427.

[44] R. Susitaival, H. Wiemann, J. Ostergaard, et al. “Internet Access Performance in LTE TDD”. In: *Proc. IEEE 71st Vehicular Technology Conf*. May 2010, pp. 1–5. DOI: 10.1109/VETECS.2010.5493986.

[45] Konstantinos Samdanis, Rudraksh Shrivastava, Athul Prasad, et al. "Virtual cells: enhancing the resource allocation efficiency for TD-LTE". In: *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*. IEEE. 2014, pp. 1–5.

[46] Roberto Riggio, Karina Gomez, Leonardo Goratti, et al. "V-Cell: Going beyond the cell abstraction in 5G mobile networks". In: *2014 IEEE Network Operations and Management Symposium (NOMS)*. IEEE. 2014, pp. 1–5.

[47] Aditya Gudipati, Daniel Perry, Li Erran Li, et al. "SoftRAN: Software defined radio access network". In: *Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking*. ACM. 2013, pp. 25–30.

[48] Salvatore Costanzo, Rudraksh Shrivastava, Dionysis Xenakis, et al. "An SDN-based virtual cell framework for enhancing the QoE in TD-LTE pico cells". In: *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*. IEEE. 2015, pp. 1–6.

[49] Rafal Stankiewicz, Piotr Cholda, and Andrzej Jajszczyk. "QoX: what is it really?" In: *IEEE Communications Magazine* 49.4 (2011), pp. 148–158.

[50] Janne Seppänen, Martín Varela, and Aggeliki Sgora. "An autonomous QoE-driven network management framework". In: *Journal of Visual Communication and Image Representation* 25.3 (2014), pp. 565–577.

[51] M Sajid Mushtaq, Brice Augustin, and Abdelhamid Mellouk. "QoE-based LTE downlink scheduler for VoIP". In: *2014 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE. 2014, pp. 2190–2195.

[52] Amanpreet Singh, Xi Li, Indika Abeywickrama, et al. "QoE-based access network dimensioning". In: *Telecommunications Network Strategy and Planning Symposium (Networks), 2014 16th International*. IEEE. 2014, pp. 1–6.

[53] Gerardo Gómez, Javier Lorca, Raquel García, et al. "Towards a QoE-driven resource control in LTE and LTE-A networks". In: *Journal of Computer Networks and Communications* 2013 (2013).

[54] Tarik Taleb, Yassine Hadjadj-Aoul, and Konstantinos Samdanis. "Efficient solutions for enhancing data traffic management in 3GPP networks". In: *IEEE Systems Journal* 9.2 (2015), pp. 519–528.

[55] Fabrizio Granelli and Claudio Sacchi. “An OFDMA RRM strategy based on QoE maximization and radio resource redistribution”. In: *Aerospace Conference, 2012 IEEE*. IEEE. 2012, pp. 1–9.

[56] 3GPP. *3GPP TS 22.951, "Service Aspects and Requirements for Network Sharing", Rel.11*. September 2012.

[57] 3GPP. *3GPP TS 23.251, "Network sharing; Architecture and functional description", Rel. 12*. December 2013.

[58] Jignesh S Panchal, Roy D Yates, and Milind M Buddhikot. “Mobile network resource sharing options: Performance comparisons”. In: *IEEE Transactions on Wireless Communications* 12.9 (2013), pp. 4470–4482.

[59] Wolfgang Kiess, Petra Weitkemper, and Ashiq Khan. “Base station virtualization for OFDM air interfaces with strict isolation”. In: *2013 IEEE International Conference on Communications Workshops (ICC)*. IEEE. 2013, pp. 756–760.

[60] S. Chen and J. Zhao. “The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication”. In: *IEEE Communications Magazine* 52.5 (May 2014), pp. 36–43. ISSN: 0163-6804. DOI: 10.1109/MCOM.2014.6815891.

[61] S. Costanzo, D. Xenakis, N. Passas, et al. “OpeNB: A framework for virtualizing base stations in LTE networks”. In: *Proc. IEEE Int. Conf. Communications (ICC)*. June 2014, pp. 3148–3153. DOI: 10.1109/ICC.2014.6883805.

[62] Mao Yang, Yong Li, Depeng Jin, et al. “OpenRAN: a software-defined ran architecture via virtualization”. In: *ACM SIGCOMM computer communication review*. Vol. 43. 4. ACM. 2013, pp. 549–550.

[63] Kostas Pentikousis, Yan Wang, and Weihua Hu. “Mobileflow: toward software-defined mobile networks”. In: *IEEE Communications magazine* 51.7 (2013), pp. 44–53.

[64] Alexey Khoryaev, Andrey Chervyakov, Mikhail Shilov, et al. “Performance analysis of dynamic adjustment of TDD uplink-downlink configurations in outdoor picocell LTE networks”. In: *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2012 4th International Congress on*. IEEE. 2012, pp. 914–921.

[65] P. W. C. Chan, E. S. Lo, R. R. Wang, et al. “The evolution path of 4G networks: FDD or TDD?” In: *IEEE Communications Magazine* 44.12 (Dec. 2006), pp. 42–50. ISSN: 0163-6804. DOI: 10.1109/MCOM.2006.273098.

[66] Rudraksh Shrivastava, Salvatore Costanzo, Konstantinos Samdanis, et al. “An SDN-based framework for elastic resource sharing in integrated FDD/TDD LTE-A HetNets”. In: *Cloud Networking (CloudNet), 2014 IEEE 3rd International Conference on*. IEEE. 2014, pp. 126–131.

[67] Wang Weidong, Zhang Yinghai, Zhang Nan, et al. “A call admission control algorithm based on inter-link load balance in heterogeneous networks”. In: *Broadband Network & Multimedia Technology, 2009. IC-BNMT’09. 2nd IEEE International Conference on*. IEEE. 2009, pp. 356–360.

[68] ONF. *Open Flow Foundation, SDN Architecture, Issue 1*. Tech. rep. June 2014.

[69] *3GPP TS 32.425, "Performance measurements Evolved Universal Terrestrial Radio Access Network*".

[70] Andrea Goldsmith. *Wireless communications*. Cambridge university press, 2005.

[71] P. Mogensen, W. Na, I. Z. Kovacs, et al. “LTE Capacity Compared to the Shannon Bound”. In: *Proc. IEEE 65th Vehicular Technology Conf. - VTC2007-Spring*. Apr. 2007, pp. 1234–1238. DOI: 10.1109/VETECS.2007.260.

[72] A. Prasad, K. Doppler, M. Moisio, et al. “Distributed Capacity Based Channel Allocation for Dense Local Area Deployments”. In: *Proc. IEEE Vehicular Technology Conf. (VTC Fall)*. Sept. 2011, pp. 1–5. DOI: 10.1109/VETECF.2011.6093152.

[73] 3GPP. *3GPP TS 23.203, "Policy and Charging Control Architecture"', (E-UTRAN)”, Rel.13*. December 2014.

[74] 3GPP. *3GPP TR 36.814, "Further Advancements of E-UTRA: Physical Layer Aspects", Rel.9*. March 2010.

[75] A. Prasad, A. Maeder, and C. Ng. “Energy efficient small cell activation mechanism for heterogeneous networks”. In: *Proc. IEEE Globecom Workshops (GC Wkshps)*. Dec. 2013, pp. 754–759. DOI: 10.1109/GLOCOMW.2013.6825079.

[76] S. Obeso, J. Luo, R. Halfmann, et al. “Intra-Operator Inter-Mode Spectrum Sharing”. In: *Proc. IEEE 69th Vehicular Technology Conf. VTC Spring 2009*. Apr. 2009, pp. 1–5. DOI: 10.1109/VETECS.2009.5073616.

[77] 3GPP. *3GPP TS 32.101, "Telecommunication management; Principles and high level requirements", (V 12.0) Rel-12*. September 2014.

[78] G. Piro, L. A. Grieco, G. Boggia, et al. “Simulating LTE Cellular Systems: An Open-Source Framework”. In: *IEEE Transactions on Vehicular Technology* 60.2 (Feb. 2011), pp. 498–513. ISSN: 0018-9545. DOI: 10.1109/TVT.2010.2091660.

[79] D. Xenakis, N. Passas, L. Merakos, et al. “Mobility Management for Femtocells in LTE-Advanced: Key Aspects and Survey of Handover Decision Algorithms”. In: *IEEE Communications Surveys Tutorials* 16.1 (2014), pp. 64–91. ISSN: 1553-877X. DOI: 10.1109/SURV.2013.060313.00152.

[80] ONF. *O. M. E. Committee, "Software-defined Networking: The New Norm for Networks", Open Networking Foundation,* 2012.

[81] X. Xu, H. Zhang, X. Dai, et al. “SDN based next generation Mobile Network with Service Slicing and trials”. In: *China Communications* 11.2 (Feb. 2014), pp. 65–77. ISSN: 1673-5447. DOI: 10.1109/CC.2014.6821738.

[82] T. Shimojo, Y. Takano, A. Khan, et al. “Future mobile core network for efficient service operation”. In: *Proc. 1st IEEE Conf. Network Softwarization (NetSoft)*. Apr. 2015, pp. 1–6. DOI: 10.1109/NETSOFT.2015.7116190.

[83] D. Tikunov and Toshikazu Nishimura. “Traffic prediction for mobile network using Holt-Winter ’s exponential smoothing”. In: *Proc. Telecommunications and Computer Networks 2007 15th Int. Conf. Software*. Sept. 2007, pp. 1–5. DOI: 10.1109/SOFTCOM.2007.4446113.

[84] Anne B Koehler, Ralph D Snyder, and J Keith Ord. “Forecasting models and prediction intervals for the multiplicative Holt–Winters method”. In: *International Journal of Forecasting* 17.2 (2001), pp. 269–286.

[85] K. Lee, S. Hong, S. J. Kim, et al. “SLAW: A New Mobility Model for Human Walks”. In: *Proc. IEEE INFOCOM 2009*. Apr. 2009, pp. 855–863. DOI: 10.1109/INFCOM.2009.5061995.

[86] Sofiene Jelassi, Gerardo Rubino, Hugh Melvin, et al. “Quality of experience of VoIP service: A survey of assessment approaches and open issues”. In: *IEEE Communications surveys & tutorials* 14.2 (2012), pp. 491–513.

[87] Christos Bampis, Zhi Li, and Alan Bovik. “Continuous Prediction of Streaming Video QoE using Dynamic Networks”. In: *IEEE Signal Processing Letters* (2017).

[88] *International Telecommunication Union (ITU-T) G.107: "The E-model: a computational model for use in transmission planning."*.

[89] S. Khan, M. Sgroi, E. Steinbach, et al. “Cross-layer optimization for wireless video streaming-performance and cost”. In: *Proc. IEEE Int. Conf. Multimedia and Expo*. July 2005, 4 pp.–. DOI: 10.1109/ICME.2005.1521575.

[90] N. Khan, M. G. Martini, and Z. Bharucha. “Quality-aware fair downlink scheduling for scalable video transmission over LTE systems”. In: *Proc. IEEE 13th Int. Workshop Signal Processing Advances in Wireless Communications (SPAWC)*. June 2012, pp. 334–338. DOI: 10.1109/SPAWC.2012.6292922.

[91] S. Costanzo, R. Shrivastava, K. Samdanis, et al. “Service-oriented resource virtualization for evolving TDD networks towards 5G”. In: *Proc. IEEE Wireless Communications and Networking Conf*. Apr. 2016, pp. 1–6. DOI: 10.1109/WCNC.2016.7565171.

[92] Yuanye Wang, Simone Frattasi, Troels B Sorensen, et al. “Network time-synchronization in TDD based LTE-Advanced systems”. In: *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*. IEEE. 2009, pp. 1–5.

[93] Ellina Foutekova, Patrick Agyapong, and Harald Haas. “Channel Asymmetry in Cellular OFDMA-TDD Networks”. In: *EURASIP Journal on Wireless Communications and Networking* 2008.1 (2008), pp. 1–14.

[94] *3GPP TR 36.819, Coordinated Multi-Point Operation for LTE Physical Layer aspects, Rel 11*. 2011.

[95] Konstantinos Samdanis, Rudraksh Shrivastava, Athul Prasad, et al. “TD-LTE virtual cells: An SDN architecture for user-centric multi-eNB elastic resource management”. In: *Computer Communications* 83 (2016), pp. 1–15.

[96] Mikio Iwamura. “NGMN View on 5G Architecture”. In: *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*. IEEE. 2015, pp. 1–5.

[97] Konstantinos Samdanis, Rudraksh Shrivastava, and Xavier Costa-Perez. “Network Slicing in 5G TDD networks”. Pat. PCT/GB2016/050914. 2016.

[98] R. Kokku, R. Mahindra, H. Zhang, et al. “NVS: A Substrate for Virtualizing Wireless Resources in Cellular Networks”. In: *IEEE/ACM Transactions on Networking* 20.5 (Oct. 2012), pp. 1333–1346. ISSN: 1063-6692. DOI: 10.1109/TNET.2011.2179063.

[99] Sanjay Goyal, Pei Liu, Shivendra Panwar, et al. “Full duplex operation for small cells”. In: *arXiv preprint arXiv:1412.8708* (2014).

[100] Vincenzo Sciancalepore, Konstantinos Samdanis, Rudraksh Shrivastava, et al. “A service-tailored TDD cell-less architecture”. In: *PIMRC 2016, 27th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, September 4-7, 2016, Valencia, Spain*. 2016.