An Innovative RAN Architecture for Emerging Heterogeneous Networks: "The Road to the 5G Era"

Shahab Hussain

Graduate Center, City University of New York

Recommended Citation

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Graduate Works by Year: Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
An Innovative RAN Architecture for Emerging Heterogeneous Networks: “The Road to the 5G Era”

By

Shahab Hussain

A dissertation submitted to the Graduate Faculty in Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2014
© 2014

Shahab Hussain

All Rights Reserved
This manuscript has been read and accepted for the Graduate Faculty in Engineering in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Prof. Mohamed A. Ali

Date Chair of Examining Committee

Dean Ardie D. Walser

Date Executive Officer

Prof. Roger Dorsinville, EE Dept, CCNY, CUNY

Prof. Samir Ahmed, EE Dept, CCNY, CUNY

Prof. Ahmed Mohamed, EE Dept, CCNY, CUNY

Dr. Andrew Wallace, AT&T

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
Abstract

An Innovative RAN Architecture for Emerging Heterogeneous Networks: “The Road to the 5G Era”

By

Shahab Hussain

Adviser: Professor Mohamed Ali

The global demand for mobile-broadband data services has experienced phenomenal growth over the last few years, driven by the rapid proliferation of smart devices such as smartphones and tablets. This growth is expected to continue unabated as mobile data traffic is predicted to grow anywhere from 20 to 50 times over the next 5 years. Exacerbating the problem is that such unprecedented surge in smartphones usage, which is characterized by frequent short on/off connections and mobility, generates heavy signaling traffic load in the network “signaling storms”. This consumes a disproportionate amount of network resources, compromising network throughput and efficiency, and in extreme cases can cause the Third-Generation (3G) or 4G (long-term evolution (LTE) and LTE-Advanced (LTE-A)) cellular networks to crash.

As the conventional approaches of improving the spectral efficiency and/or allocation additional spectrum are fast approaching their theoretical limits, there is a growing consensus that current 3G and 4G (LTE/LTE-A) cellular radio access technologies (RATs) won’t be able to meet the anticipated growth in mobile traffic demand. To address these challenges, the wireless industry and standardization bodies have initiated a roadmap for transition from 4G to 5G
cellular technology with a key objective to increase capacity by “1000× by 2020”. Even though the technology hasn't been invented yet, the hype around 5G networks has begun to bubble. The emerging consensus is that 5G is not a single technology, but rather a synergistic collection of interworking technical innovations and solutions that collectively address the challenge of traffic growth.

The core emerging ingredients that are widely considered the key enabling technologies to realize the envisioned 5G era, listed in the order of importance, are: 1) Heterogeneous networks (HetNets); 2) flexible backhauling; 3) efficient traffic offload techniques; and 4) Self Organizing Networks (SONs). The anticipated solutions delivered by efficient interworking/integration of these enabling technologies are not simply about throwing more resources and/or spectrum at the challenge. The envisioned solution, however, requires radically different cellular RAN and mobile core architectures that efficiently and cost-effectively deploy and manage radio resources as well as offload mobile traffic from the overloaded core network.

The main objective of this thesis is to address the key techno-economics challenges facing the transition from current Fourth-Generation (4G) cellular technology to the 5G era in the context of proposing a novel high-risk revolutionary direction to the design and implementation of the envisioned 5G cellular networks. The ultimate goal is to explore the potential and viability of cost-effectively implementing the 1000x capacity challenge while continuing to provide adequate mobile broadband experience to users. Specifically, this work proposes and devises a novel PON-based HetNet mobile backhaul RAN architecture that: 1) holistically addresses the key techno-economics hurdles facing the implementation of the envisioned 5G cellular technology, specifically, the backhauling and signaling challenges; and 2) enables, for the first
time to the best of our knowledge, the support of efficient ground-breaking mobile data and signaling offload techniques, which significantly enhance the performance of both the HetNet-based RAN and LTE-A’s core network (Evolved Packet Core (EPC) per 3GPP standard), ensure that core network equipment is used more productively, and moderate the evolving 5G’s signaling growth and optimize its impact.

To address the backhauling challenge, we propose a cost-effective fiber-based small cell backhaul infrastructure, which leverages existing fibered and powered facilities associated with a PON-based fiber-to-the-Node/Home (FTTN/FTTH) residential access network. Due to the sharing of existing valuable fiber assets, the proposed PON–based backhaul architecture, in which the small cells are collocated with existing FTTN remote terminals (optical network units (ONUs)), is much more economical than conventional point-to-point (PTP) fiber backhaul designs. A fully distributed ring-based EPON architecture is utilized here as the fiber-based HetNet backhaul. The techno-economics merits of utilizing the proposed PON-based FTTx access HetNet RAN architecture versus that of traditional 4G LTE-A’s RAN will be thoroughly examined and quantified. Specifically, we quantify the techno-economics merits of the proposed PON-based HetNet backhaul by comparing its performance versus that of a conventional fiber-based PTP backhaul architecture as a benchmark.

It is shown that the purposely selected ring-based PON architecture along with the supporting distributed control plane enable the proposed PON-based FTTx RAN architecture to support several key salient networking features that collectively significantly enhance the overall performance of both the HetNet-based RAN and 4G LTE-A’s core (EPC) compared to that of the typical fiber-based PTP backhaul architecture in terms of handoff capability, signaling
overhead, overall network throughput and latency, and QoS support. It will also been shown that the proposed HetNet-based RAN architecture is not only capable of providing the typical macro-cell offloading gain (RAN gain) but also can provide ground-breaking EPC offloading gain.

The simulation results indicate that the overall capacity of the proposed HetNet scales with the number of deployed small cells, thanks to LTE-A’s advanced interference management techniques. For example, if there are 10 deployed outdoor small cells for every macrocell in the network, then the overall capacity will be approximately 10-11x capacity gain over a macro-only network. To reach the 1000x capacity goal, numerous small cells including 3G, 4G, and WiFi (femtos, picos, metros, relays, remote radio heads, distributed antenna systems) need to be deployed indoors and outdoors, at all possible venues (residences and enterprises).
Acknowledgements

My thesis could not have been finished without the support and cooperation of many people.

I would first like to express my sincere appreciation and gratitude to Professor Mohammad Ali for his support, guidance, discussions and exhaustive editing of my publications and thesis. Professor Mohamed A. Ali has been a source of enthusiasm and encouragement over many years. His guidance, discussions and exhaustive editing of my publications and thesis make this ground breaking work possible. Without his support, I could not imagine the completion of my PhD at Graduate Center in CUNY.

I also like to extend my gratitude to Prof. Roger Dorsinville who is inspiration for me for various research publications. Moreover I am also thankful to Prof. Samir Ahmed, Prof. Ahmed Mohamed and Dr. Andrew Wallace for serving on my supervisory committee. Their insightful recommendations have improved the quality of this work. Much appreciation is given to my colleague Dr. S. Rashid Zaidi and Dr. Ajaz Sana for all the technical discussions and cooperation and for all the valuable and productive time we shared.
My dissertation is dedicated to my parents, sisters, my wife and my children who have been constantly supportive of me during my years of life, growth and education.
# TABLE OF CONTENTS

## 1 INTRODUCTION

1.1 Introduction .................................................................................................................. 2

1.2 Thesis Motivation ........................................................................................................... 5

1.3 Thesis Statement and Contribution ................................................................................ 7

## 2 OVERVIEW OF LTE AND LTE-A

2.1 LONG TERM EVOLUTION (LTE) .................................................................................. 13

2.1.1 LTE Basics ................................................................................................................. 13

2.1.2 Physical Parameters and Frame Structure................................................................. 17

2.1.3 MIMO Transmissions ............................................................................................... 19

2.1.4 Network simplification as compared to 3G UMTS Networks ................................... 20

2.1.5 Interface X2 (eNodeB-eNodeB) ............................................................................. 20

2.1.6 eNB, MME and SGW Pools .................................................................................... 22

2.1.7 Functional Mapping (from TR 25.813) .................................................................... 24

2.1.8 LTE functions in eNode-B....................................................................................... 24

2.1.9 MME Functions ....................................................................................................... 25

2.1.10 SGW Functions...................................................................................................... 26

2.1.11 PDN GW (PGW) Functions .................................................................................. 27

2.2 LONG TERM EVOLUTION ADVANCED (LTE-A) ...................................................... 28

2.2.1 Overview of LTE-Advanced .................................................................................... 28

2.2.2 LTE-Advanced Targets .......................................................................................... 29

2.2.3 LTE-Advanced main topics and issues for Release 12 ........................................ 30
LIST OF FIGURES

Figure 2-1: LTE Network Interfaces ................................................................. 14
Figure 2-2: LTE Network Architecture .............................................................. 14
Figure 2-3: Orthogonal sub-channels of OFDM system over bandwidth W .......... 16
Figure 2-4: LTE Frame ....................................................................................... 19
Figure 2-5: LTE Downlink Physical Resource Block .......................................... 19
Figure 2-6: EPC Network Simplification ............................................................. 20
Figure 2-7: X2 & S1 interfaces .......................................................................... 21
Figure 2-8: eNB, MME and SGW Pools .............................................................. 23
Figure 2-9: MME, SGW, eNB Pools ................................................................. 23
Figure 2-10: Functional Mapping E-UTRAN & EPC .......................................... 24
Figure 2-11: LTE-Advanced Targets ................................................................. 30
Figure 2-12: New Carrier Type (NCT) ............................................................... 31
Figure 2-13: Device to Device UE Communications .......................................... 32
Figure 2-14: HetNets layout ............................................................................. 33
Figure 2-15: HetNets Non Tower required locations .......................................... 34
Figure 2-16: Cell range expansion of low-power nodes under a macrocell .......... 36
Figure 2-17: Enhanced Inter-Cell Interference Coordination (eICIC) ................. 38
Figure 2-18: Combining two carriers yields 6.4 MHz of total useable bandwidth .. 39
Figure 2-19: Possible CC bandwidths 1.4, 3, 5, 10, 15, 20 MHz .......................... 40
Figure 2-20: CA allows more efficient use of expensive spectrum. .................................................. 41

Figure 2-21: CoMP scenarios................................................................. 43

Figure 3-1: EPON Downstream overview. ................................................................. 49

Figure 3-2: EPON Upstream overview................................................................. 49

Figure 3-3: Standalone EPON Architecture ................................................................. 52

Figure 3-4: ONU Architecture ........................................................................ 52

Figure 3-5: 10G EPON Protection State Architecture................................................... 54

Figure 3-6: 10 Nodes Network Architecture................................................................. 57

Figure 3-7: 7 Nodes Protected State Network Architecture................................................... 58

Figure 4-1: A three-pronged approach to capacity needs. [Source 19] ......................... 61

Figure 4-2: Factors affecting small-cell deployment. [Source 19] ...................................... 62

Figure 4-3: PON and LTE-A/HetNet Interconnection Models........................................... 66

Figure 4-4: (a) Architecture of the ONU-eNB, (b) Functional Modules hardware.............. 69

Figure 4-5: Proposed EPON-based HetNets Backhaul RAN Architecture....................... 70

Figure 4-6: Macro at Center of cell only................................................................. 71

Figure 4-7: Macro at Center Plus One Small Cell at the edge ........................................... 72

Figure 4-8: Macro at Center Plus 4 Small Cell at equal distance from center.................. 72

Figure 4-9: Example small cell covering problem [source 7] ........................................ 74

Figure 4-10: Normalized deployment cost of a typical PTP deployment scenario and the optimal PON-based solution under different split ratios [source 7]................................. 74
Figure 8-5: Average user US throughput gain for the HetNet with one macro BS and either 1 or 4 small cells over macro cell only network, for the hotspot (15/20) scenario............................... 126

Figure 8-6: Average user US throughput gain for the HetNet with one macro BS and either 1 or 4 WiFi APs over macro cell only network, for the hotspot (15/20) scenario. ........................................... 127

Figure 8-7: Overall average capacity gain of the proposed HetNet over Macro-only scales linearly with the number of deployed small cells................................................................. 128

Figure 8-8: Uplink Utilization Time Series for unevenly loaded Macro & HetNets ...................... 129

Figure 8-9: Uplink Utilization Time Series for Evenly loaded Macro & HetNets ......................... 129

Figure 8-10: DS throughput with 50% of local LAN Traffic ...................................................... 130

Figure 8-11: DS throughput with 100% of local LAN Traffic .................................................... 131

Figure 8-12: US Throughput with no local LAN Traffic ............................................................ 132

Figure 8-13: US Throughput with 0.5 local LAN Traffic ............................................................ 132

Figure 8-14: Average packet drop when one small cell transmits high data rate (> 200 Mbps) to the other small cell in the same ring.............................................................. 133

Figure 8-15: Traffic Throughput during Handoff ................................................................. 134
LIST OF TABLES

Table 2-1: Bandwidth assignments for LTE ................................................................. 18
Table 2-2: Example Requirements for LTE-A ............................................................ 30
Table 3-1: Parameters Used in the Model ................................................................. 56
Table 4-1: Several different sizes and versions of small cells ..................................... 63
Table 5-1: LTE-A Standardized QCI Characteristics ................................................. 78
Table 8-1: LTE Simulation Parameters ..................................................................... 119
Table 8-2: Wi-Fi (Unlicenced) System Parameters .................................................... 119
Table 8-3: Traffic Loading Scenarios ....................................................................... 120
Chapter 1

1 Introduction
1.1 Introduction

The global demand for mobile-broadband data services has experienced phenomenal growth over the last few years, driven by the rapid proliferation of smart devices such as smartphones and tablets. This growth is expected to continue unabated as mobile data traffic is predicted to grow anywhere from 20 to 50 times over the next 5 years. Most of this mobile data traffic (almost 80 percent) is being generated indoors, which requires increased link budget and coverage extension to provide satisfactory end-user experience. Indoor performance is significantly poorer than outdoor performance since the radio signals are seriously attenuated, distorted, and redirected by walls, ceilings, floors, etc.,. Thus, current cellular architectures that were originally tailored to serve large coverage areas and optimized for homogeneous traffic are no longer able to efficiently cope with such dominant indoor traffic patterns.

Exacerbating the problem is that such unprecedented surge in smartphones usage, which is characterized by frequent short on/off connections and mobility, generates heavy signaling traffic load in the network “signaling storms”. This consumes a disproportionate amount of network resources, compromising network throughput and efficiency, and in extreme cases can cause the Third-Generation (3G) or 4G (long-term evolution (LTE) and LTE-Advanced (LTE-A)) cellular networks to crash.

As the conventional approaches of improving the spectral efficiency and/or allocation additional spectrum are fast approaching their theoretical limits, there is a growing consensus that current 3G and 4G (LTE/LTE-A) cellular radio access technologies (RATs) won’t be able to meet the anticipated growth in mobile traffic demand. To address these challenges, the wireless industry and standardization bodies have initiated a roadmap for transition from 4G to 5G
cellular technology with a key objective to increase capacity by “1000× by 2020”. Even though the technology hasn't been invented yet, the hype around 5G networks has begun to bubble. The emerging consensus is that 5G is not a single technology, but rather a synergistic collection of interworking technical innovations and solutions that collectively address the challenge of traffic growth.

The envisioned 5G cellular network would allow people to be connected at all times – no matter where they are, who they connect to, and what their service needs are. The core emerging ingredients that are widely considered the key enabling technologies to realize the envisioned 5G era, listed in the order of importance, are: 1) Heterogeneous networks (HetNets); 2) flexible backhauling; 3) efficient traffic offload techniques; and 4) Self Organizing Networks (SONs). The anticipated solutions delivered by efficient interworking/integration of these enabling technologies are not simply about throwing more resources and/or spectrum at the challenge. The envisioned solution, however, requires radically different cellular RAN and mobile core architectures that efficiently and cost-effectively deploy and manage radio resources as well as offload mobile traffic from the overloaded core network.

Heterogeneous networks (HetNets), which comprise a combination of macro-cell base stations and low-cost low-power small cell base stations (BSs) operating over both licensed (e.g., femto and picocells) and unlicensed (e.g., WiFi access points) bands, have recently emerged as a viable solution to cope with the unprecedented mobile traffic growth [1-3]. Deployment of a large number of public access small cells (SCs) overlaying macro cells is expected to significantly increase the network capacity and expand the coverage while reducing the overall cost [4-7]. There are several different sizes and versions of small cells. They vary in the number
of users they can handle, their power, and their range. In virtually all cases, they include the essential 3/4G technologies of the carrier and Wi-Fi. They also have a power source and a backhaul connection to the cellular network. Extremely low-cost indoor SCs can be used at homes, offices, enterprises, shopping malls, etc., and can be installed by users themselves. SCs can also be deployed by operators as hotspots, cost-effectively serving highly concentrated indoor/outdoor traffic.

To handle the explosion of mobile data, offloading techniques have been proposed to improve the user experience for cellular services in overloaded areas. By offloading the cellular system, the network can handle more users with higher-speed data needs. In general, traffic offload can be classified into two types: “RAN offload” and “core network offload”. RAN offload is implemented through the use of WiFi, femtocells and SCs. Note that femtocells and SCs are typically deployed as a means to increase capacity and improve coverage, rather than as an offload solution. Typically, all IP traffic generated by/sent to a mobile device is routed to and through the mobile core network. However, because a majority of IP traffic is destined to best-effort Internet, it would be more cost-effective to divert this traffic away from the mobile core and offload it directly to the Internet. This is the definition of core network offload.

Core offload is implemented through the deployment of internet offload gateways, which splits out traffic bound for the internet from the traffic bound for the operator’s core network including signaling [12]. Selected IP Traffic Offload (SIPTO) and Local IP Access (LIPA) are two solutions that 3GPP is standardizing for core network offload. The major downside of core network offload is that by diverting traffic from the core, the network operator has no longer any control over this offloaded traffic (e.g., to meter usage, bill for traffic), since these functions all
reside in the core. As a consequence, mobility support for this offloaded traffic is rather limited. Note that core network offload is one form of “Internet offload” as Internet offload comes in several forms including, WiFi, femtocell, and core network.

1.2 Thesis Motivation

While it is a forgone conclusion that next-generation 5G cellular networks will be HetNets-based, however, HetNets also come with their own challenges, and there are significant techno-economics hurdles that still need to be addressed for successful widespread rollout and operations of these networks. A massive deployment of small access nodes introduces several challenges such as additional backhaul capacity, an adverse interference scenario, and mobility management requirements, which 5G needs to address. HetNets create a new challenge for the backhaul, which must provide connectivity at sufficient capacity and quality of service (QoS). The number of SC sites in certain macrocell coverage can rise up to several hundred (e.g. large city center) and every one of them needs to have a fast backhaul connection. Thus, implementing the connectivity between the mobile network and the SC BSs becomes problematic. The key challenge is how to provide cost-effective, scalable and flexible mobile backhaul solution to connect massive number of SCs to the mobile core network.

With small cells being deployed on sides of buildings, on street furniture and utility poles and even within large public areas such as airports and stadiums, a wide variety of backhaul access options, including microwave, copper and fiber as well as new wireless options can be used. But this presents a serious challenge since wired connectivity is cost-prohibitive, microwave backhaul requires line of sight, and low frequency that allow propagation in non-line-of-sight (NLOS) urban environments are simply not available. Fiber is considered to be an optimal access
technology offering the best characteristics in terms of capacity and QoS support. There are many fiber access options including Gigabit Passive Optical Network (GPON), Ethernet PON (EPON), Carrier Ethernet and dark fiber/wavelengths. However, fiber is not available to all sites, and the cost of deploying it strictly for small cell backhaul may be prohibitive.

Different levels of coordination/cooperation among small cells are key to enhance the network capacity and keep interference at an adequate level, to manage mobility and spectrum, and to improving the spectral efficiency. For instance, to improving the spectral efficiency of the system, the use of advanced Coordination/Cooperative schemes among BS/SC transmitters in order to combat the generated interference is required. This requires the exchange of enormous amount of signaling and control messages between a massive number of SCs and the macro BSs (mBSs) with very low latency. This is achieved via utilizing Coordinated Multipoint (CoMP) transmission and reception techniques between the participating cluster of mBS and small cells, in which the exchange of channel state information (CSI) and commands among the cluster must be implemented with very low latency (X2 delay should be in a range of 1 ms or lower).

Note that macro BSs and small cells exchange signaling and control messages via the standardized X2 interface, which is a logical interface (no direct physical connections between the BSs/SCs). Thus, to achieve intercommunication among the SCs cells, all exchanged signaling and control messages are transported first from the SC to the EPC over the mobile backhaul and then back from the EPC to the participating SCs. The typical X2 delay is then the sum of round trip propagation delay from the cells to the EPC and vice versa and the time taken to process these control messages at the EPC. Thus, to minimize the X2 latency, backhaul architecture and topology must be designed to facilitate rapid inter-BS/SC local communications.
Another example is Self-Organizing network (SON), which is one of the key enabling features on the road to 5G era, where a software solution is used to manage a HetNet. With SON, the HetNet will essentially manage itself. SON can automate configuration and coordinates between cells to maximize the performance of the entire network. This generates additional heavy signaling traffic load in the network. Since the overall signaling load is processed at the EPC, this will place a high signaling demand on the EPC gateway elements (serving gateway (S-GW) and packet gateway (P-GW)). Thus, to meet the anticipated 5G signaling demand, one needs to significantly scale the transaction rate performance of the EPC and related network elements (control-plane capability). This requirement is expected to be substantially more important than increases in raw throughput.

Overall, deployment of massive number of small cells including WiFi APs, their integration with the EPC, the dramatic surge in the number of short-lived connections (smartphones, emerging machine-to-machine (M2M) and “Internet of Things” services), will create new major control-plane and signaling challenges. The bottom line is that signaling is inherent to smartphones and M2M usage and will pose a major challenge in 5G networks. Addressing the above challenges, that is the focus of this thesis, specifically the evolving signaling challenge, which must be a core consideration in 5G network design, requires fundamentally different 5G RAN and mobile core design requirements.

1.3 Thesis Statement and Contribution

The main objective of this thesis is to address the key techno-economics challenges facing the transition from current Fourth-Generation (4G) cellular technology to the 5G era in the context of proposing a novel high-risk revolutionary direction to the design and implementation of the
envisioned 5G cellular networks. The ultimate goal is to explore the potential and viability of cost-effectively implementing the 1000x capacity challenge while continuing to provide adequate mobile broadband experience to users. Specifically, this work proposes and devises a novel PON-based HetNet mobile backhaul RAN architecture that: 1) holistically addresses the key techno-economics hurdles facing the implementation of the envisioned 5G cellular technology, specifically, the backhauling and signaling challenges; and 2) enables, for the first time to the best of our knowledge, the support of efficient ground-breaking mobile data and signaling offload techniques, which significantly enhance the performance of both the HetNet-based RAN and LTE-A’s core network (Evolved Packet Core (EPC) per 3GPP standard), ensure that core network equipment is used more productively, and moderate the evolving 5G’s signaling growth and optimize its impact.

To address the backhauling challenge, we propose a cost-effective fiber-based small cell backhaul infrastructure, which leverages existing fibered and powered facilities associated with a PON-based fiber-to-the-Node/Home (FTTN/FTTH) residential access network. Due to the sharing of existing valuable fiber assets, the proposed PON–based backhaul architecture, in which the small cells are collocated with existing FTTN remote terminals (optical network units (ONUs)), is much more economical than conventional point-to-point (PTP) fiber backhaul designs. Given the large investments many fixed-line carriers are making or have already made in PON-based FTTH/FTTC access infrastructure, the economic advantage of utilizing the existing fiber-based PON access infrastructure is quite compelling compared to the costly proposition of building up a new PTP fiber backhaul connection for each small cell.
In contrast to the typical star-based PON topology, a local access small ring-based PON topology is rather assumed here. Specifically, a fully distributed ring-based EPON architecture is utilized here as the fiber-based HetNet backhaul. The main characteristics of the proposed PON-based HetNet backhaul RAN architecture is that it supports a fully distributed control plane that enables direct intercommunication among the access nodes (SCs/mBSs) as well as signaling, scheduling algorithms, and handoff procedures that operate in a distributed manner. The techno-economics merits of utilizing the proposed PON-based FTTx access HetNet RAN architecture versus that of traditional 4G LTE-A’s RAN will be thoroughly examined and quantified. Specifically, we quantify the techno-economics merits of the proposed PON-based HetNet backhaul by comparing its performance versus that of a conventional fiber-based PTP backhaul architecture as a benchmark.

The significance of the purposely selected simple ring topology: 1) it enables direct intercommunication/connectivity among the SCs and among the macro BS (mBS) and SCs, allowing for the support of Efficient interference management and coordination, which requires SCs to be directly interconnected at lowest possible latency (via the direct physical connectivity among the SCs attached to the ring); 2) it facilitates highly accurate synchronization among the SC BS clocks. The faster real-time signaling information can be exchanged between SC BSs, the more accurately clocks are aligned and the less interference [9]; 3) minimizes the X2 interface latency, thus, allowing for harnessing the highest CoMP gains; and 4) the inherent self-healing mechanism of the ring architecture facilitates and guarantees the reliable delivery of mobile traffic.
It is shown that the purposely selected ring-based PON architecture along with the supporting distributed control plane enable the proposed PON-based FTTx RAN architecture to support several key salient networking features that collectively significantly enhance the overall performance of both the HetNet-based RAN and 4G LTE-A’s core (EPC) compared to that of the typical fiber-based PTP backhaul architecture in terms of handoff capability, signaling overhead, overall network throughput and latency, and QoS support. In addition, the proposed ground-breaking RAN as well as core network offload techniques are fully managed and controlled by the mobile core network, without resorting to typical deployment of Internet offload gateways.

Furthermore, the proposed HetNet-based RAN architecture enables redistributing some of the intelligence and network control and management (NCM) operations currently centralized in the EPC platform out into the RAN’s access nodes (SCs/mBSs). Specifically, as this work will show, it enables offloading sizable fraction of mobile data traffic and associated signaling overhead as well as the lengthy and complex processing of this traffic (e.g., LTE bearers/mobility tunnels switch/set-up, retain, and tear-down and associated signaling commands from the SCs to the EPC and vice-versa) from the typically overloaded EPC to the access nodes (SCs/mBS) of the RAN.

This has a significant impact on the performance of the envisioned 5G’s EPC. First, it frees up a sizable fraction of the badly needed network resources as well as processing on the overloaded EPC’s centralized serving nodes. Second, it frees up capacity and sessions on the typically congested mobile backhaul (from the small cells to the EPC and vice-versa). Third, the firmly held notion that the EPC’s control plane scalability might be a major stumbling block en-route to
the realization of the 5G will be shown to be no longer precise. This has a far-reaching implication as the small cells in the proposed HetNet RAN can now be deployed not only as typical means to increase capacity and improve coverage, but also as an effective EPC offload solution. This is significant as the proposed HetNet RAN is not only capable of providing the typical macro-cell offloading gain (RAN gain) but also can provide ground-breaking EPC offloading gain.

Overall, the proposed HetNet-based RAN architecture constitutes a complete cellular networking paradigm shift from the typically centralized RAN’s architecture and EPC-based NCM operations to a new disruptive fully distributed HetNet-based RAN’s architecture along with NCM operations in which substantial fraction of the typically centralized EPC-based NCM operations are migrated to and independently implemented by the HetNet access nodes (SCs/mBSs) in a distributed manner.

The simulation results indicate that the overall capacity of the proposed HetNet scales with the number of deployed small cells, thanks to LTE-A’s advanced interference management techniques. For example, if there are 10 deployed outdoor small cells for every macrocell in the network, then the overall capacity will be approximately 9x capacity gain over a macro-only network. To reach the 1000x capacity goal, numerous small cells including 3G, 4G, and WiFi (femtos, picos, metros, relays, remote radio heads, distributed antenna systems) need to be deployed indoors and outdoors, at all possible venues (residences and enterprises).
Chapter 2

2 Overview of LTE and LTE-A
2.1 LONG TERM EVOLUTION (LTE)

2.1.1 LTE Basics

The trend of ever increasing transmission bandwidths is challenging the limits of current 3G networks, hence it was decided by 3GPP (3rd generation partnership program) standardization body in 2005 to start work on next generation wireless network design that is only based on packet-switched data transmission. LTE is the latest standard in the mobile network technology tree that is being implemented within the Third Generation Partnership Project (3GPP) to ensure the competitiveness of 3G for the next 10 years and beyond. LTE supports both time-division duplex (TDD) and frequency division duplex (FDD). Moreover it supports a flexible and scalable bandwidth e.g., 1.25, 5, 10 and 20 MHz. Moreover LTE has a very flexible radio interface [1, 2].

LTE base station is referred to as enhanced NodeB (eNodeB) per 3GPP standard in order to differentiate it from UMTS (Universal mobile telecommunication system) base station which is known as NodeB. Enhanced NodeB (eNodeB) base stations are made more intelligent than NodeB by removing Radio Network Controller (RNC) and transferring the functionality to eNodeB and partly to the core network gateway. In LTE the base stations can also perform handovers as they can communicate directly over X2 interface. S1 interface connects eNodeB to the gateway nodes i.e., between radio network and core network. It is completely based on IP protocol. The gateway between radio access network and core network is divided into two entities Serving Gateway (Serving-GW) and the Mobility Management Entity (MME). MME is the control plane (c-plane) entity is mainly responsible for subscriber mobility, session management signaling, location tracking of mobile devices and selection of a gateway to the internet when mobile requests IP address from the network. On the other hand Serving-GW is
responsible for user plane (u-plane). Both components can be implemented on the same hardware or separated. If implemented separately, S11 interface is used to communicate between them. Basic LTE network interfaces [2] are shown in following figure 2-1. LTE Network architecture is shown in figure 2.2.

Figure 2-1: LTE Network Interfaces

Figure 2-2: LTE Network Architecture
S6 interface is between MME and database that stores subscription information, referred to as Home Subscriber Server (HSS). In LTE, the router at the edge of the wireless core network is known as Packet Data Network Gateway (PDN-GW) and the interface between PDN-GW and MME / Serving-GW is called S5. It uses GTP-U (user) protocol to tunnel user data from / to the Serving-GWs and the GTP-S (Signaling) protocol for the initial establishment of a user data tunnel and subsequent tunnel modifications when the user moves between cells that are managed by different Serving-GWs.

For air interface, LTE uses Orthogonal Frequency Division Multiple Access (OFDMA) in the downlink data transmission. In OFDMA, a big data stream is transmitted by using many narrow band sub-carriers simultaneously. The sub-carriers are spaced apart at fixed frequencies (15 KHz). This spacing provides orthogonality among carriers, as shown in Figure 2-3. Because many bits of data are transmitted in parallel, the transmission speed of each sub-carrier can be much lower than the overall data rate. This not only minimizes the multipath fading but also the effect of multipath fading and delay spread become independent of the channel bandwidth used. This is because the bandwidth of each sub-carrier remains same and only the number of sub-carriers is changed for different achievable overall bandwidth. Moreover OFDMA has more advantages like high spectral efficiency. The most common modulation techniques used are binary phase-shift keying (BPSK), quadrature phase-shift keying (QPSK) and quadrature amplitude modulation (QAM).
For OFDMA downlink transmission, a mathematical function Inverse Fast Fourier Transform (IFFT) transforms the signal from frequency domain to time domain. The resulting signal is then modulated and amplified and transmitted in the air. When the signal is received by the receiver, it first demodulates and amplifies the signal. After this the signal is converted back from time domain to frequency domain using Fast Fourier Transform (FFT). The multiple access (MA) in OFDMA refers to the data that is sent in the downlink is received by several users simultaneously. This is accomplished by the use of control messages to inform mobile devices, waiting for data, which part of data is addressed to them and which part they can ignore. On the physical layer it means the use of modulation schemes ranging from QPSK over 16QAM to 64QAM can be quickly changed for different sub-carriers to fulfill different reception conditions.

In LTE, for uplink transmission, a different transmission scheme is used as compared to in the downlink. This is known as Single Carrier Frequency Division Multiple Access (SC-FDMA). This is because OFDMA inherently suffers from high peak to average power ratio (PARP) which can drain the mobile device battery quickly. Since mobile device should consume as little energy as possible, a different transmission technique SC-FDMA is proposed for the uplink
transmission. In general this scheme is similar to OFDMA but has much lower PARP. This is the reason SC-FDMA is selected for uplink transmission. SC-FDMA also transmits data over the air interface in many subcarriers, but adds an additional processing step. A number of input bits are grouped and then passed through FFT first and then output of FFT is fed into IFFT block. Since not all the subcarriers are used by the mobile station, many of them set to zero. On the receiver side the signal is amplified, demodulated and then fed into FFT block. The resulting signal is fed into IFFT block to counter the effect of additional step in the transmission. The resulting time domain signal is fed into detector block which recreates the original signal bits.

2.1.2 Physical Parameters and Frame Structure

For LTE, physical parameters are chosen as follows:

• OFDM symbol duration, 66.667 µs
• Subcarrier spacing, 15 kHz
• Standard cyclic prefix (CP), 4.7 µs
• Extended cyclic prefix (CP), 16.67 µs

The cyclic prefix (CP) is transmitted before each OFDM symbol to prevent inter-symbol interference (ISI) which is evident because of different transmission paths of varying lengths.

Moreover in LTE different channel bandwidths ranging from 1.25 MHz to 20 MHz Table 2-1 shows the standardized transmission bandwidths, the number of subcarriers used and the FFT size used and physical Resource Block (PRB) for each bandwidth. Physical Resource Block (PRB) is the smallest element of resource allocation assigned by the base station scheduler.
Table 2-1: Bandwidth assignments for LTE

<table>
<thead>
<tr>
<th>Bandwidth (MHz)</th>
<th>Number of subcarriers</th>
<th>Subcarrier Bandwidth (kHz)</th>
<th>FFT size</th>
<th>Physical Resource Block (PRB)</th>
<th>PRB Bandwidth (kHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.25</td>
<td>76</td>
<td>128</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>151</td>
<td>256</td>
<td></td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>301</td>
<td>15</td>
<td>512</td>
<td>25</td>
<td>180</td>
</tr>
<tr>
<td>10</td>
<td>601</td>
<td>1024</td>
<td></td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>901</td>
<td>1536</td>
<td></td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1201</td>
<td>2048</td>
<td></td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

LTE generic frame structure is shown in figure 2-4. It is evident from the figure that LTE frame duration is 10 ms. It is then divided into 10 sub frames of 1 ms duration each. Each sub frame is further subdivided into two slots of 0.5 ms each. Each Slot of 0.5 ms consists of 12 subcarriers and 6 or 7 OFDM symbols depending upon either standard or extended cyclic prefix (CP) is used. When extended cyclic prefix is used then the number of OFDMA symbols reduced to 6. The grouping of 12 subcarriers results in PRB bandwidth of 180 kHz.

Two slots that grouped together to form a sub frame which is also known as Transmit Time Interval (TTI). In case of Time Division Duplex (TDD) operation, the sub frame can be used for downlink or uplink. This is decided by the network which frames are used for downlink or uplink. However in LTE most networks likely to use Frequency Division Duplex (FDD) in which separate bands are used for uplink and downlink.
Data is mapped to subcarriers and symbols, which are arranged in the time and frequency domain in a resource grid. LTE physical resource block is shown in Figure 2-5.

2.1.3 MIMO Transmissions

LTE standard defines the use of MIMO technology. Transmission of several independent signals over the same frequency band is also referred to as MIMO or multiple input multiple output. LTE standard defines two and four transmissions over the same band, which needs 2 or 4 antennas at both receiver and transmitter side respectively. A comprehensive mathematical treatment of MIMO is given in [4]. These transmissions are known as 2 x 2 MIMO and 4 x 4 MIMO. Since MIMO channels are separated from each other, 2 x 2 MIMO can increase overall
data rate by two and likewise 4 x 4 MIMO by four times. However LTE is only used in the
downlink transmissions since for uplink transmissions it is difficult to use MIMO for mobile
devices because of limited antenna size and power constraints.

2.1.4 Network simplification as compared to 3G UMTS Networks

User Plane: 3 functional entities: eNode B, Serving Gateway and PDN Gateway (the
gateways can be combined into a single physical entity). GGSN converges to S/P-GW

Control plane: SGSN converges to MME (Mobility Management Entity) and RNC
functionality moves to eNode B. No more RNC and RNC layers/functionalities moved to eNB.

X2 interface for inter-eNB mobility (i.e. data/context forwarding).

![EPC Network Simplification](image)

Figure 2-6: EPC Network Simplification

2.1.5 Interface X2 (eNodeB-eNodeB)

This interface is for eNodeB-to-eNodeB handover. During eNodeB handover another
eNodeB, the downlink data is forwarded from the source eNodeB target eNodeB over the X2
(direct forwarding). The X2 interface uses the Tunneling Protocol for the control plane (GTP-C).

When the UE receives the handover command it will remove any EPS bearers which it did
not receive and corresponding EPS radio bearers in the target part of handover execution,
downlink packets are forwarded from the source to the target eNodeB. When the UE has arrived to the target eNodeB, downlink forwarded from the source eNodeB can be sent to it. Uplink data from the delivered via the (source) SGW to the PGW.

The X2 user plane interface (X2-U) is defined between eNodeBs. The X2-interface provides non-guaranteed delivery of user plane PDUs.

The transport network layer is built on IP transport. GTP-U is used on top of UDP/IP to carry the user plane PDUs. The X2-UP interface protocol stack is identical to the S1-UP protocol stack. The X2 control plane interface (X2-CP) is defined between two neighbor eNodeBs. The transport network layer is built on SCTP on top of IP. The application signaling protocol is referred to as X2-AP (X2 Application Protocol).

![Figure 2-7: X2 & S1 interfaces](image)

S1 consists of S1-MME (control traffic) and S1-U (User Traffic). S1-MME is between eNB and MME. S1-U is between eNB and SGW. Flex Architecture for both interfaces S1-U and S1-
MME allows eNB to be connected to multiple MMEs and SGWs. It also allows creation of MME and SGW pools. The benefits of S1 architecture are:

- Network sharing
- Load balancing
- Network robustness

2.1.6 eNB, MME and SGW Pools

**Tracking Area (TA):** A group of base stations providing radio services for a wider area, each area is identified by a TA Identity (TAI). UE does not need to send a TA update as long as it is roaming in a TA.

**Pool Area:** Can be one or more TAs, served by one or more MME/SGW pools

**MME Pool:** One or more MMEs, can serve other (RAN) Pool areas

**SGW Pool:** One or more SGWs,

**MME Selection:** Performed by eNB, based on MME Load, UE state

**SGW Selection:** Performed by MME: Network topology/Service Area, SGW Load

All eNBs within the pool area (and overlapping areas) must have S1 (e.g., SCTP) connectivity to MME all eNBs within the pool area must have S1 (e.g., IP/UDP) connectivity to SGWs (and overlapping areas). The UE is served by any of the MME/SGWs within a pool. No MME/SGW relocation required within the MME/SGW pool. The eNBs must support S1-flex (which provides capability for eNB to perform MME selection function
Figure 2-8: eNB, MME and SGW Pools

Figure 2-9: MME, SGW, eNB Pools
2.1.7 Functional Mapping (from TR 25.813)

Following are salient functions of eNodeB:

- Selection of MME at UE attachment
- Routing towards SGW at UE initial access
- NAS messaging encapsulated by RRC for tx over radio
- Scheduling and transmission of paging messages
- Scheduling and transmission of System Information
- Dynamic allocation of resources to UEs in both UL and DL
- Configuration and provision of eNB measurements
• Radio Bearer Control
• Radio Admission Control
• Access restrictions in Active state
• Connection Mobility Control in LTE_ACTIVE state
• Active mode Handover handling
• RRC, header compression, encryption, RLC, MAC, PHY
• Security of User plane and RRC
• Encryption of both in PDCP, integrity check of RRC
• Scheduling and associated QoS handling

2.1.9 MME Functions

Following are key MME functionalities;

• NAS signalling
• NAS signalling security
• S101 – Interface between MME and eRNC for inter-RAT handoffs
• Inter CN node signalling for mobility between 3GPP access networks (terminating S3)
• UE reachability in ECM-IDLE state (including control and execution of paging retransmission)
• Tracking Area list management
• PDN GW and Serving GW selection
• MME selection for handovers with MME change
• SGSN selection for handovers to 2G or 3G 3GPP access networks
• Roaming (S6a towards home HSS)
• Authentication
• Bearer management functions including dedicated bearer establishment.
• Lawful Interception of signalling traffic.

2.1.10 SGW Functions

For each UE associated with the EPS, at a given point of time, there is a single Serving GW. The functions of the Serving GW, for both the GTP-based and the PMIP-based S5/S8, include:

• the local Mobility Anchor point for inter-eNodeB handover;
• assist the eNodeB reordering function during inter-eNodeB handover by sending one or more "end marker" packets to the source eNodeB immediately after switching the path.
• Mobility anchoring for inter-3GPP mobility (terminating S4 and relaying the traffic between 2G/3G system and PDN GW);
• ECM-IDLE mode downlink packet buffering and initiation of network triggered service request procedure;
• Lawful Interception;
• Packet routeing and forwarding;
• Transport level packet marking in the uplink and the downlink, e.g. setting the DiffServ Code Point, based on the QCI of the associated EPS bearer;
• Accounting on user and QCI granularity for inter-operator charging;
• UL and DL charging per UE, PDN, and QCI (e.g. for roaming with home routed traffic)
2.1.11 PDN GW (PGW) Functions

If a UE is accessing multiple PDNs, there may be more than one PDN GW for that UE, however a mix of S5/S8 connectivity and Gn/Gp connectivity is not supported for that UE simultaneously. PDN GW functions include for both the GTP-based and the PMIP-based S5/S8:

- Per-user based packet filtering (by e.g. deep packet inspection);
- Lawful Interception;
- UE IP address allocation;
- Transport level packet marking in the uplink and downlink, e.g. setting the DiffServ Code Point, based on the QCI of the associated EPS bearer;
- UL and DL service level charging as defined in TS 23.203 [6] (e.g. based on SDFs defined by the PCRF, or based on deep packet inspection defined by local policy);
- UL and DL service level gating control as defined in TS 23.203 [6];
- UL and DL service level rate enforcement as defined in TS 23.203 [6] (e.g. by rate policing/shaping per SDF);
- UL and DL rate enforcement based on APN-AMBR (e.g. by rate policing/shaping per aggregate of traffic of all SDFs of the same APN that are associated with Non-GBR QCI);s
- DL rate enforcement based on the accumulated MBRs of the aggregate of SDFs with the same GBR QCI(e.g. by rate policing/shaping);
• DHCPv4 (server and client) and DHCPv6 (client, relay and server) functions

2.2 LONG TERM EVOLUTION ADVANCED (LTE-A)

2.2.1 Overview of LTE-Advanced

LTE Advanced is an evolution of LTE. LTE Advanced is the next major milestone in the evolution of LTE and is a crucial solution for addressing the anticipated 1000x increase in mobile data. It incorporates multiple dimensions of enhancements including the aggregation of carriers, advanced antenna techniques. But most of the gain comes from optimizing HetNets, resulting in better performance from small cells.

The benefit of small cells in providing capacity where needed, is well understood. So are the challenges and solutions for managing the interference. Enhancements such as “Range Expansion,” introduced in LTE Advanced, increase the overall network capacity much more than what can be got by merely adding small cells. The interference management techniques of LTE Advanced make adding more small cells possible without affecting the overall network performance.

LTE-Advanced shall meet or exceed IMT-Advanced (IMT-A) requirement within ITU-R time plan. Extended LTE-Advanced targets are adopted in LTE Release 11 and Release 12. e.g. additional carrier aggregation band combinations. LTE-A also supports new frequency banks. LTE-A is backwards compatible with LTE Release 8. An LTE Rel. 8 UE can operate in an LTE-A network. Also an LTE-A UE (R10 or higher) can operate in an LTE Release 8 network. LTE-A deployment uses increased deployment of indoor eNB and HeNB. HeNB is home eNodeB, a type of femto cell with a very small coverage area, typically less than a 50 m radius.
2.2.2 LTE-Advanced Targets

In general LTE-Advanced needs to improve the capacity of LTE to meet the targets defined in IMT-Advanced. LTE-A increases the DL and UL peak data rates and peak spectral efficiency to exceed the targets defined in IMT-Advanced.

- Release-8 LTE numbers assume 4x4 MIMO in DL
- LTE-A numbers assume 8x8 MIMO in DL and 4x4 MIMO in UL

Following LTE Advanced features have been defined to meet LTE-A targets.

- Carrier Aggregation (CA)
- MIMO enhancements
- Heterogeneous network enhancements:
  - Enhanced Inter-Cell Interference coordination (eICIC)
  - further enhanced Inter-Cell Interference Coordination (feICIC)
- Coordinated Multi-Point (CoMP)
- eNodeB Relays
- Additional feature enhancements
  - New UE categories
  - New SON capabilities

Table 2.2 and Figure 2-11 show example target requirements for LTE-Advanced.
Table 2-2: Example Requirements for LTE-A

<table>
<thead>
<tr>
<th>Target</th>
<th>Rel. 8 LTE</th>
<th>LTE-Advanced</th>
<th>IMT-Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak data rate</td>
<td>DL</td>
<td>300 Mbps</td>
<td>&gt;1 Gbps</td>
</tr>
<tr>
<td></td>
<td>UL</td>
<td>75 Mbps</td>
<td>&gt;500 Mbps</td>
</tr>
<tr>
<td>Peak spectrum efficiency</td>
<td>DL</td>
<td>15</td>
<td>30.6</td>
</tr>
<tr>
<td>[bps/Hz/cell]</td>
<td>UL</td>
<td>3.75</td>
<td>16.8</td>
</tr>
</tbody>
</table>

Figure 2-11: LTE-Advanced Targets.

2.2.3 LTE-Advanced main topics and issues for Release 12

New carrier type (NCT): is sometimes called a “lean carrier”. The new lean carrier has reduced control channel and reference signal overhead. Because current cell reference signals are always on they create interference even though no data is being transmitted. The lean carrier’s
reduction of overhead reduces interference and energy consumption. Figure 2.13 shows a potential use scenario when Rel-12 dual carriers are being used. The carrier from a macro cell is used for signaling and the low power lean carrier from a small cell is used for high speed data. Macro/small cell split refers to using a macro cell for signaling and a small cell for data. Small cell enhancements also include X2 gateway.

Figure 2-12: New Carrier Type (NCT)

LTE-WiFi Integration: WiFi (i.e. non-3GPP) interworking is described in TS 23-402. In scenarios where 3GPP and non-3GPP access networks are available, UEs will need some help selecting the best network to use. A new function, Automatic Network Discovery and selection Function (ANDSF) is introduced to provide UEs with information about which network to use.

MIMO Enhancements: 3D MIMO is the concept of adjusting the beam in both the horizontal and vertical dimensions. Prior to Rel-12 and new antennas, the eNodeB transmitter was able to adjust the beam in the horizontal dimension only, and the down-tilt vertical dimension was fixed
for each user. This horizontal adjustment allows beams to be directed upward to floors in a building, or perhaps over a small cell to reduce interference.

**Device to Device (D2D):** allows the UEs to communicate directly using LTE spectrum rather than sending data through the eNodeB. Signaling is still sent to the eNodeB and UEs are still under the control of the eNodeB; the EPC must be enhanced to support this function. D2D is especially important to public safety where UE to UE communication may be required when the network is unavailable after a disaster. D2D can also be used for new proximity based social networking applications and services that allow the exchange of data because the devices are close to each other.

![Device to Device UE Communications](image)

**Figure 2-13: Device to Device UE Communications**

### 2.2.4 Heterogeneous Networks

Heterogeneous Networks (HetNets) include cells with different coverage areas (i.e. sizes) in the same geographic footprint. HetNets include;
- Macro (Inter Site Distance ~1Km)
- Micro (Inter Site Distance ~ 200 m)
- Small Cell (Pico) (Inter Site Distance ~ 100 m)
- Femto (Home eNodeB / HeNB)
- Relay Nodes

The term “small cells” refers to Micro, Pico and Femto cells. The most challenging aspect in the deployment of heterogeneous networks is the interference issues generated by sharing the carrier with the overlaid macro nodes, when operators have limited spectrum for LTE non-carrier aggregation based heterogeneous networks. Figure 2-14 shows HetNets layout and various cell coverage areas and range.

![HetNets layout](image)

Figure 2-14: HetNets layout

Heterogeneous networks have following benefits,
- Offload users and traffic from the macro eNB
- Increase capacity at traffic hot spots
- Improved coverage and performance at cell edges
- Fill coverage holes
- Provide coverage where real estate constraints do not allow macro
- Small cells have lower CAPEX and OPEX.

Following figures 2-15 shows the deployment of various non tower related HetNets locations,

![HetNets Non Tower required locations](image)

Figure 2-15: HetNets Non Tower required locations

The biggest challenges to metro small cell deployment include:

- Access to new types of sites (“Non-Towers”)
- Large scale installation workforces with the skill sets to perform carrier grade deployments
• Access to backhaul facilities
• The above assets are generally not all owned by any one company, requiring multiple partners

2.2.5 Cell Range Expansion (CRE)

Sharing the same carrier frequency between macrocells and small cells introduces new network design challenges. If the handoff boundary between cells is based on the received signal power at the UE, many UE devices that are very close to a picocell find themselves in the service area of a macrocell. This leads to severe uplink interference at the picocells. More important, high power transmission from the macrocells greatly shrinks the picocell coverage, leading to gross underutilization of low-power nodes. Even with optimized placement of small cells, they may become underutilized due to the temporal changes in data traffic demand. The technique of cell range expansion (CRE) is devised to address this problem.

Cell range Expansion (CRE) techniques allow improved performance. CRE used when a significant amount of traffic near the macro cell that has not been captured by the small cell due to its limits Tx power relative to the macro cell. CRE achieved through the use of UE-specific settings (cell association bias in idle mode and modification of handover parameters in active mode). Significant cell range expansion results in issues with PDCCH reception. eICIC is designed to handle cell edge interference problems.

PDCCH – Physical Downlink Control Channel, carries the layer one control. The PDCCH communicates who data is for, what data is sent, and how the data is sent over the air in the PDSCH. PDSCH is physical downlink shared channel, carries data and signaling messages.
RSRP- Reference signal received power, is a measurement of the signal strength of an LTE cell used to help rank cells as input for handover and cell reselection decisions. The RSRP is the average of the power of all resource elements which carry cell-specific reference signals over the entire bandwidth.

The cochannel deployment of low-power nodes in a macrocellular network does not necessarily reduce the number of users sharing the given base station. CRE overcomes this problem by biasing handoff boundaries in favor of small cells, causing most users to be served by the cell to which they are closest. This expands the service area of small cells, as illustrated in Figure 2-16.

![Figure 2-16: Cell range expansion of low-power nodes under a macrocell](image)

While CRE can significantly improve load balancing in the network and mitigate uplink interference from macro UE to picocells, it creates significant downlink interference for users in the CRE region, who are served by small cells but receive a much stronger signal from macrocells [7]. Downlink interference to CRE users can be overcome with resource partitioning techniques, where macrocells set aside certain restricted resources for the benefit of CRE users.
On these resources, macrocells only transmit the common control/paging/broadcast channels (CCCs) and common reference signals (CRSs). Pico users in a CRE region can achieve high enough signal-to noise-plus-interference ratio (SINR) \(= \frac{S}{I+N}\) on these resources by estimating/demodulating and cancelling the CCC and CRS from the macrocells. Although resource partitioning creates dimension loss at the macrocells, it results in a net system gain, because dimensions lost by each macrocell are exploited by many small cells under its footprint.

2.2.6 Enhanced Inter-Cell Interference Coordination (eICIC)

eICIC is used to mitigate interference in cell overlap in HetNets. It uses power, frequency and also time domain to mitigate intra-frequency interference. The most challenging aspect in the deployment of HetNets is the interference issues generated by sharing the carrier with the overlaid macro nodes, when operators have limited spectrum for LTE deployment. Enhanced Inter-Cell Interference Coordination (eICIC) has been defined in LTE Rel-10 to support non-carrier aggregation-based heterogeneous networks.

eICIC introduces “Almost blank subframes” (ABS). ABS subframes do not send any traffic channels and are mostly control channel frames with very low power. Macro cell configures ABS subframes allowing UEs connected to small cells to send control data during ABS subframes avoiding interference from macro cell. ABS configuration is shared via OAM or X2 interface.

Interference coordination between aggressor cell and victim cell is done by means of bitmap sent over X2 interface. Each bit is mapped to a single subframe and indicates an ABS subframe. Based on the data traffic demand, the pattern can change each 40 ms. Cell creating strong
interference controls which resources can be used by the victim cell to serve terminals in harsh interference conditions.

Figure 2-17 shows a time division scheme. There is also a frequency division scheme (not shown). When bandwidth is scarce use of the time division scheme is preferred over the frequency division scheme. UE4 does not experience interference from the macro because it is not at all small cell edge.

![Enhanced Inter-Cell Interference Coordination (eICIC)](image)

Figure 2-17: Enhanced Inter-Cell Interference Coordination (eICIC).

Rel-11/12 continues development of Further eICIC frequently called feICIC.

- Interference cancellation receiver in the terminal.
• Ensures that weak cells can be detected. Inter cell interference cancellation for control signals (pilot, synchronization signals).

• Ensures that remaining interference is removed. Inter cell interference cancellation for control and data channels (PDCCH/PDSCH).

• Interference cancellation done at the UE and Network.

2.2.7 Carrier Aggregation (CA)

Prior to 3GPP Release 10, an LTE UE could only perform Tx and Rx with a single DL and UL carrier from an eNodeB.

Release 10 introduces features that allow a UE to perform Tx and Rx with multiple carriers to increase the total available bandwidth and peak data rates. For example a 5MHz carrier could be aggregated with a 1.4 MHz carrier to create 6.4 MHz total available bandwidth. Figure 2-18 shows CA.

![Carrier Aggregation Diagram]

Figure 2-18: Combining two carriers yields 6.4 MHz of total useable bandwidth.

Carrier aggregation must be supported by the eNodeB and the UE. In Release 10 the carrier bandwidth remain the same as in Release-8 and Release 9. This remains backward compatibility
with existing Rel-9/9 UEs. Release 10 introduces the specification to allow up to 5 DL and UL carriers to be combined to allow up 100 MHz of total bandwidth.

Figure 2-19: Possible CC bandwidths 1.4, 3, 5, 10, 15, 20 MHz.

Figure 2-19 shows how 5 carriers (called component carriers) can be aggregated to allow up to 100 MHz total bandwidth. Aggregating carriers with lower bandwidth would result in a less than 100 MHz total aggregated bandwidth.

Carrier aggregation has following benefits:

1. Maximize the total peak data rate and throughput performance. Different frequencies have different propagation behavior.

2. Provide a higher quality of experience to end users by load-balancing traffic across carriers. A UE experiencing congestion in one band can access unused capacity available in another carrier as shown in figure 2-20.

3. Minimize inefficiencies inherent in wireless deployment in non-contiguous or narrow (5 MHz or less) channel bandwidths. One spectrum band may be fully utilized while another is under-utilized, aggregation allow use of the
under-utilized spectrum when needed. Carrier benefits from more cost effective use to licensed spectrum.

![Figure 2-20: CA allows more efficient use of expensive spectrum.](image)

**2.2.8 Coordinated Multipoint (CoMP)**

Carrier aggregation and CoMP are the two most important techniques that boost the data rate of the LTE-A to a new threshold.

In LTE, each UE will be served by a single cell and signals coming from cells on other eNBs can become interference to the UE. When the UE moved to the cell edge, the signal from the current cell becomes weaker and signals from other cells can become stronger. The UE will send measurements back to the current eNB to prepare for handover. This is also the time when the UE receives strong interference, and data rate will be very low. The situation will worsen quickly if the UE is moving at a high speed. Coordinated multipoint (CoMP) can coordinate transmissions from multiple eNBs to a single UW to reduce interference and improve performance at cell edges when interference is severe.
CoMP can be considered as a distributed MIMO system, where geographically distributed eNBs use multiple antennas and cooperate to transmit to and/or receive from UEs. There are some significant hurdles for CoMP to overcome. Feedback overhead, backhaul delay and interference channel estimation are examples.

CoMP can be applied to both the DL and UL. DL CoMP techniques can be classified according to the amount of information shared among cells. Joint processing is available when neighbouring cells share transmit data as well as the channel state information.

The joint processing can be realized in the form of joint transmission or dynamic cell selection. In joint transmission cooperating eNB’s jointly transmit data to one or more corresponding UE’s. Dynamic cell selection is a kind of fast cell selection; UE’s are handed over to the best cell in the interference situation. However, joint processing generally requires high-capacity X2 interface between eNB’s for sharing transmit data, and thus can cause excessive backhaul overhead and latency.

Coordinated scheduling/coordinated beam forming (CS/CB) can be realized only if the channel state information and scheduling information are shared between eNBs; data sharing is not required, only state and scheduling information. In the CS/CB, a UE receives data from only one eNB’s, its own serving node, while the precoding and scheduling are coordinated among related eNB’s in such a way to reduce interference and improve the throughput.

For the case of UL, joint detection and interference prediction are considered. Joint detection can be considered as a UL counterpart of the DL joint transmission. For joint detection, eNBs need to share received signal samples as well as channel state information and scheduling information. The basic principle of interference is to perform link adaption based on predicted
SINR values. Interference prediction is possibly by exchanging recourse allocation information among cells [8]. Figure 2-21 shows CoMP details.

![CoMP scenarios](image)

Figure 2-21: CoMP scenarios.

### 2.2.9 Self Organizing Networks (SON)

Self-organizing networks (SONs) are a software solution to managing a HetNet. While the interaction between macrocells is usually managed manually, with multiple small cells, such a manual task is overwhelming. With SON, the HetNet will essentially manage itself. SON can automate configuration and dynamically optimize the network based on the traffic loads [9].

SONs can be categorized by their three basic functions: self-configuration, self-optimization, and self-healing. Self-configuration adjusts the small-cell frequency, power level, and interfaces automatically as the device joins the system. It works with the automatic neighbor relations (ANR) software that builds and maintains a list of all cells in the network and the location and physical characteristics of each (see “Test ANR Functionality On Your LTE Devices” at
electronicdesign.com). If any new cell is added, the configuration is automatic and the list is updated. The same occurs if a cell is removed [9].

Self-optimization refers to the ability of the network to adapt itself to surrounding conditions and optimize its performance based on coverage, capacity, handover between cells, and interference. Two key functions are load balancing and interference mitigation. Load balancing is dividing the traffic between the cells so no one cell becomes too overloaded if adjacent cells are within range and have available capacity. Load balancing occurs automatically. This ability also helps balance the backhaul traffic load.

Interference management is essential in a HetNet since the small cells are generally closely spaced and could potentially interfere with one another. SON software uses the cells to measure the characteristics of nearby cells to determine if interference is a possibility. It then makes adjustments dynamically to change frequency or power level as necessary to minimize interference.

Self-healing refers to a SON’s ability to adjust to changing conditions such as cell failure. SON technology is a key part of HetNets, and the LTE standard supports it. Tests have shown that SON can monitor and update a network within milliseconds in some cases and dynamically adapt. Overall throughput can be improved by 10% to 45% in many cases [9].
Chapter 3

3 Fully Distributed Ring-Based EPON Architecture
3.1 Introduction

To date, mainstream Ethernet Passive Optical Network (EPON) bandwidth allocation schemes as well as the new IEEE 802.3ah Ethernet in the First Mile (EFM) Task Force specifications have been centralized, relying on a component in the central office (Optical Line Termination (OLT)) to provision upstream traffic. Hence, the OLT is the only device that can arbitrate time-division access to the shared channel. Since the OLT has global knowledge of the state of the entire network, this is a centralized control plane in which the OLT has centralized intelligence. One of the major problems associated with a centralized architecture is the “single-point of failure” problem that is the failure of the OLT software will bring down the whole access network. Another major problem is that the PON architecture is typically centralized but 4G RAN architecture is intrinsically distributed. Thus the PON architecture must support a distributed architecture as well as distributed radio network control and management (NCM) operations.

In this section we propose distributed solutions to this problem, and in the process to prove that these distributed networking architectures solutions and the associated bandwidth allocation algorithms and protocols have characteristics that make them far better suited for provisioning Quality of Service (QoS) schemes necessary for properly handling data, voice, video, and other real-time streaming advanced multimedia services over a single line.

3.1.1 Overview of Ethernet-PON (EPON) Technologies and Architectures

A PON is a point-to-multipoint fiber optical network with no active elements in the signal’s path [4–6]. It consists of a single, shared optical fiber connecting a service provider’s central office (head end) to a passive star coupler (SC)/optical splitter/combiner, which is located near
residential customers. The SC is intentionally positioned a substantial distance away from the central office (CO), but close enough to the customers in order to save fiber. Each customer receives a dedicated short optical fiber but shares the long distribution trunk fiber. All transmissions in a PON are performed between an Optical Line Terminal (OLT) and Optical Network Units (ONUs). Traffic from an OLT to an ONU is called ‘downstream’ (point-to-multipoint), and traffic from an ONU to the OLT is called ‘upstream’ (multipoint-to-point). Two wavelengths are used: typically 1310 nm (\(\lambda_{up}\)) for the upstream transmission and 1490 nm (\(\lambda_{d}\)) for the downstream transmission. The OLT resides in the central office, connecting the optical access network to the metro or backbone network, where the ONU is located at either the curb (Fiber To The Curb; FTTC solution) or the end-user location (Fiber To The Building and Fiber To The Home; FTTB and FTTH respectively). A single PON typically serves from 16-64 customers. PONs can be deployed in a 1:N tree, tree-and-branch, ring, or bus topology.

In the downstream direction, figure 3-1 [7] shows an overview simplified illustration of EPON downstream, an EPON operates as a broadcast and select network. The OLT has the entire bandwidth of the channel to broadcast standard formatted 802.3 Ethernet frames to all ONUs. Each ONU extracts those packets that contain the ONU’s unique Media Access Control (MAC) address. In the upstream direction, figure 3-2 [7] shows an overview simplified illustration of EPON upstream, multiple ONUs share the transmission channel. Thus, the ONUs need to employ some arbitration mechanism to avoid collisions. In that case, each ONU transmits within a dedicated time slot and the OLT receives a continuous stream of collision-free frames from multiple ONUs.
The IEEE 802.3ah task force is actively standardizing the control and management messages used to control the data exchange between the OLT and the ONUs as well as the processing of these messages through the development of Multi-Point Control Protocol (MPCP). Note that MPCP is not concerned with any particular bandwidth allocation; it is merely a supporting protocol that facilitates the implementation of various bandwidth allocation algorithms in EPON. The protocol relies on two Ethernet control messages (GATE and REPORT) in its regular operation. The OLT assigns the Transmission Windows (TWs) via the GATE messages.

In general, the OLT arbitrates the upstream transmissions by allocating an appropriate timeslot/transmission window to each ONU. An ONU is only allowed to transmit during the TW allocated to it by the OLT. Each ONU uses a set of queues to store its Ethernet frames and starts transmitting them as soon as its TW starts. An ONU can support up to 8 priority queues as defined in 802.1Q [3]. Within each cycle, in order to inform the OLT about its bandwidth requirements, ONUs use REPORT Messages that are also transmitted along with the data in the TW. The ONU should also account for additional overhead when requesting the next time slot; this includes 8 bytes frame preamble and 12 bytes Inter-Frame Gap (IFG) between two consecutive frames. Between the TW of two ONUs there is a certain guard time “g” needed to account for the laser on and off times, receiver recovery times, round trip delay (which relates to the physical distance between communicating ONUs) and other optic related issues. Upon receiving a REPORT, the OLT passes the message to a Dynamic Bandwidth Allocation (DBA) module, which performs the bandwidth allocation computation.
3.2 Standalone Ring-Based EPON Architecture

3.2.1 Normal State Operation

The standalone architecture refers here to just the wire line segment of the hybrid architecture without incorporating the wireless segment the small cells. Fig. 3-3 illustrates the standalone ring-based EPON architecture. An OLT is connected to N ONUs via a 20 km trunk feeder fiber, a passive 3-port optical circulator, and a short distribution fiber ring. To cover the same local access area as in the typical tree-based architecture, the small ring at the end of the trunk is assumed to have a 1-2 km diameter. The ONUs are joined by point-to-point links in a closed loop around the access ring. The links are unidirectional: both downstream (DS) and upstream (US) signals (combined signal) are transmitted in one direction only. The US signal is
transmitted sequentially, bit by bit, around the ring from one node to the next where it is terminated, processed, regenerated, and retransmitted at each node (ONU). Since US transmission is based on a TDMA scheme, inter-ONU traffic (LAN data and control messages) is transmitted along with upstream traffic destined to the OLT (MAN/WAN data) within the same pre-assigned time slot. Thus, in addition to the conventional transceiver maintained at each ONU (a $\lambda_{\text{up}}$ US transmitter (Tx) and a $\lambda_{\text{d}}$ DS receiver), this approach requires an extra receiver (Rx) tuned at $\lambda_{\text{up}}$ to process the received US/LAN signal.

DS signal is coupled to the ring at port 2 of the optical circulator. After recombining it with the re-circulated US signal via the 2x1 CWDM combiner placed on the ring directly after the optical circulator, the combined signal then circulates around the ring (ONU$_1$ through ONU$_N$) in a Drop-and-Go fashion, where the DS signal is finally terminated at the last ONU. The US signal emerging from the last ONU is split into two replicas via the 20:80 1x2 passive splitter (Fig. 3-3) placed on the ring directly after the last ONU. The first replica (80 %) is directed towards the OLT via circulator ports 1 and 3, where it is then received and processed by the US Rx (housed at the OLT), which accepts only MAN/WAN traffic, discards LAN traffic, and process the control messages, while the second replica (20 %) is allowed to recirculate around the ring after recombining with the DS signal via the 2x1 CWDM combiner.

The detailed ONU architecture is shown in Fig. 3-4. Each ONU attaches to the ring via the input port of a 1x2 CWDM DMUX housed at each ONU (incoming signal at point A in Fig.3-2) and can transmit data onto the ring through the output port of a 2x1 CWDM combiner (outgoing signal at point E in Fig. 3-4). At each ONU, the incoming combined signal is first separated into its two constituent: DS and US signals via the 1x2 CWDM DMUX housed at the ONU. As can
be seen from Figure 3-4, the separated US signal is then received and processed via the US Rx housed at the ONU, where it is regenerated and retransmitted along with the ONU’s own local control and data traffic. Note that DS signal is terminated at the last ONU via removing the 2x1 CWDM combiner and 1x2 passive splitter.

As can also be seen from Figure 3-4, the separated DS signal is coupled to the input port of the (10: 90) 1x2 passive splitter, which splits the DS signal into a small (10%) “Drop-signal-portion” and a large (90%) “Express-signal-portion”. The small portion (Drop-Signal) is then received and processed by the DS Rx housed at the ONU. The remaining large portion emerging from the 90% output splitter’s port (Express-Signal) is further transmitted through the ring to the next ONU, where it is, once again, partially split and detected at the corresponding DS Rx and partially transmitted towards the rest of the ring. Note that the Express-Signal recombines again with the retransmitted US signal (all previous ONU’s regenerated US signals plus its own US signal) via the 2x1 CWDM combiner to form the outgoing combined signal (incoming signal for next ONU) that circulates around the ring.

Since the ring is a closed loop, US traffic will circulate indefinitely unless removed. The process of removing, regenerating, and retransmitting the second replica of the US signal at each node (ONU) is implemented as follows: first, the US Rx (housed at each ONU) terminates all US traffic, examines the destination MAC address of each detected Ethernet frame, and then performs one or more of the following functions: (1) the source node removes its own transmitted frames that complete one trip around the ring through re-circulation; (2) once the destination address of the LAN traffic matches the node’s MAC address, it is copied and delivered to the end users; (3) all US traffic (including LAN and control frames), excluding those
that match items 1 and 2 above, is processed, regenerated, and then retransmitted to the next node.

Figure 3-3: Standalone EPON Architecture

Figure 3-4: ONU Architecture
3.2.2 Protected State Architecture

The protected architecture as shown in figure 3-5 is identical to that of the normal working architecture except for the following additional components: i) a redundant trunk fiber and distribution fiber ring; ii) a redundant transceiver pair located at the OLT; iii), Automatic Protection Switching (APS) module located at each ONU. The APS module attached to each ONU monitors the state of its adjacent distribution fiber paths and the state of the ONU and performs both fault detection and the APS functions. Each APS module houses a commercially available low loss 4x4 bidirectional Optical Switch (OS) that is capable of switching from any port to any port used for switching between working and protection fibers. It also includes two detection circuits comprised of a 1×2 CWDM filter (to separate the combined DS/US signal), a control circuit to configure the OS, and a p-i-n detector (except the first ONU (ONU₁), which has two p-i-n detectors at the first detection circuit). The first detection circuit of each ONU (except the first ONU) is used to detect only the US signal via taping a small portion (about 1%) of the incoming combined (DS/US) signal and passing it through the CWDM filter. On the other hand, the first detection circuit of the first ONU is used to detect both US and DS signals. Likewise, the second detection circuit of each ONU is used to detect the outgoing US signal via taping a small portion (about 1%) of the outgoing combined signal.
Figure 3-5: 10G EPON Protection State Architecture
3.2.3 Recovery Time

Recovery time is defined here as the time elapsed from when a failure occurs to when service is fully restored and a new cycle resumes. The total recovery time is the sum of several delay components including timeout, fault detection time, REPORT/GATE transmission time/propagation delays/processing times, and OS switching time. In general, the switching time is much longer than all other delay components combined and, therefore, the total recovery time is mainly dominated by the switching time (about 13 ms) [1].

3.2.4 Power Budget and scalability Analysis

The scalability of the proposed working state architecture is mainly limited by the concatenated splitter losses encountered by the DS signal at each node. Since the US signal is regenerated at every node, typical limited US power budget problems as well as the utilization of the 10 Gbps US burst-mode Tx/Rx and associated design challenges at the ONU/OLT are totally eliminated. To examine the performance impact of the DS power budget under the assumption of a fixed (10:90) tap ratio at each ONU, we consider the worst-case scenario by calculating the total ODN loss (passive optical elements (e.g., splitters, combiners, fibers, connectors, switches and splices forming an optical path), incurred by the DS signal on its optical path from the OLT to the second to last ONU (ONU_{N-1}).

There are two types of losses encountered by the DS signal at each node. The first type is along the path I-A-B in Fig. 3-3 (Drop-component, $IL_{\text{Drop}}$) and the second type is along the path I-A-E-O (Express-component, $IL_{\text{Express}}$). Table I quantifies both types of losses assuming typical commercially available CWDM components.
Table 3-1: Parameters Used in the Model

<table>
<thead>
<tr>
<th>Type of Loss</th>
<th>Path I-A-B (DROP)</th>
<th>Path I-A-E-O (EXPRESS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Splitter-10/90 (A)</td>
<td>10.0</td>
<td>0.45</td>
</tr>
<tr>
<td>CWDM</td>
<td>0.5</td>
<td>2×0.5</td>
</tr>
<tr>
<td>Access Ring Fiber Loss</td>
<td>0.0</td>
<td>0.125</td>
</tr>
<tr>
<td>Switch (I-A)/(E-O)</td>
<td>0.5</td>
<td>2×0.5</td>
</tr>
</tbody>
</table>

TOTAL IL (dB)

<table>
<thead>
<tr>
<th></th>
<th>Working</th>
<th>Protected</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.5</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>1.60</td>
<td>2.60</td>
<td></td>
</tr>
</tbody>
</table>

The total ODN loss incurred by the downstream signal on its path to ONU\(_{N-1}\) is:

\[ IL_{Total\ Loss}^{ONU_{N-1}} = IL_{fiber}^{trunk} + 2 \ IL_{CWDM} + (N-2) \ IL_{Express}^{ONU} + IL_{Drop}^{ONU} + IL_{fiber}^{Ring} \] \hspace{1cm} (1)

Assuming a 20 km trunk feeder fiber (0.25 dB/km loss), the first ONU is 20 km away from the OLT, and the last ONU is 23.2 km away from the OLT (ring circumference is about 3.2 km; 1 km diameter), and the IEEE 802.3av 10G-EPON highest power budget class (PR/PRX30) parameters [2] with a DS Rx (APD w/FEC) sensitivity of – 28.5 dBm and OLT Tx optical power of + 2 dBm, the total number of ONUs that can be adequately supported is equal to 10 ONUs, (see Fig. 3-6). As for the protected state architecture, the signals encounter the additional OS and tap loss at each node. Assuming a 0.5 dB insertion loss per OS, the total number of ONUs that can be adequately supported by the protected architecture is reduced to 7 ONUs shown in Fig. 3-7.
1. **Express Path**
-1.6 dB per node for express path to next node

2. **Drop Path**
-10.5 dB loss

3. **Fiber Ring**
-0.8 dB loss

---

**10 Nodes Network Architecture**

Figure 3-6: 10 Nodes Network Architecture
Figure 3-7: 7 Nodes Protected State Network Architecture
Chapter 4

4 A Novel Intelligent Mobile Backhaul RAN Architecture for Emerging Heterogeneous Networks
4.1 Introduction

Heterogeneous networks (HetNets), which comprise a combination of macro-cell base stations and low-cost low-power small cell base stations (BSs) operating over both licensed (e.g., femto and picocells) and unlicensed (e. g., WiFi access points) bands, have recently emerged as a viable solution to cope with the unprecedented mobile traffic growth [1-3]. Deployment of a large number of public access small cells (SCs) overlaying macro cells is expected to significantly increase the network capacity and expand the coverage while reducing the overall cost [4-7]. While deploying large number of SCs close to users will certainly help to solve the RAN’s capacity and coverage problem, however, there is a significant price to pay --- HetNets/SCs create a new challenge for the backhaul, which must provide connectivity at sufficient capacity and quality of service (QoS). The key challenge is how to provide cost-effective, scalable and flexible mobile backhaul solution to connect SCs to the mobile core network.

HetNet backhauling leads to new challenges compared to Macro backhauling. In contrast to the typically centralized 2G/3G RAN infrastructure, the more distributed architecture associated with LTE-A/SCs-based HetNet necessitates fundamentally different RAN design requirements. Specifically, the applicability of Coordinated Multipoint (CoMP) transmission and reception techniques between neighbors macro BSs/SCs depends to a great extent on the backhaul characteristics (latency and capacity), which is driven by the transport technology (e.g., optical fiber, microwave or copper-based technologies) and the RAN topology. It is critical that HetNet backhaul RAN architecture, topology, capacity, and latency, be taken into account for efficient offloading strategy that ensures a better and seamless user experience. Deployment will
depend on several factors, such as, existing infrastructure, spectrum and license costs, availability of equipment, operator business situation, etc [8].

Operators are shifting their focus to a three-pronged approach to squeezing out more capacity and coverage. Moving the base station closer to the user equipment results in a higher-quality air interface which provides better spatial efficiency. Spectrum increase: more spectrum is being freed up in an attempt to meet demand. Spectrum efficiency: moving to LTE delivers better spectrum efficiency

Figure 4-1: A three-pronged approach to capacity needs. [Source 19]

With higher signal quality using small cells, more bits can be transmitted at the same time, which leads to better throughput. When you combine this with new spectrum it has a multiplier effect. Couple that with the spatial efficiency of small cells and you get the force-multiplier effect of a theoretical 1000x capacity increase as highlighted in figure 4-1. Note that for
completeness, other methods from various vendors get to the 1000x by 10x more performance and 10x more spectrum with 10x more cells. Apart from the capacity increase, small cells enable

- Better latency: users will experience faster download and upload times
- In-building coverage: small cells invariably provide better in-building coverage and this can represent a significant source of revenue for network operators
- Better cell-edge coverage: small cells provide better cell-edge performance than macro cells, resulting in better quality of experience.

Informa [19] published a report recently that highlighted the industry’s views on what the important factors are concerning small cells, summarized in figure 4-2.

![Factors affecting small-cell deployment. [Source 19]](image-url)
With metro (small) cells being deployed on sides of buildings, on street furniture and utility poles and even within large public areas such as airports and stadiums, a wide variety of backhaul access options, including microwave, copper and fiber as well as new wireless options [6-7], will be used to meet service requirements at the lowest possible cost. Fiber is considered to be an optimal access technology offering the best characteristics in terms of capacity and QoS support. There are many fiber access options including Gigabit Passive Optical Network (GPON), Ethernet PON (EPON), Carrier Ethernet and dark fiber/wavelengths. However, fiber is not available to all sites, and the cost of deploying it strictly for metro cell backhaul may be prohibitive.

There are several different sizes and versions of small cells. They vary in the number of users they can handle, their power, and their range. In virtually all cases, they include the essential 3G technologies of the carrier, LTE and Wi-Fi. They also have a power source and a backhaul connection to the cellular network. Several different sizes and versions of small cell [20] are shown in Table 4-1.

Table 4-1: Several different sizes and versions of small cells. [Source 20]

<table>
<thead>
<tr>
<th>LICENSED SMALL CELLS</th>
<th>Femto</th>
<th>Pico</th>
<th>Micro/metro</th>
<th>Macro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indoor/outdoor</td>
<td>Indoor</td>
<td>Indoor or Outdoor</td>
<td>Outdoor</td>
<td>Outdoor</td>
</tr>
<tr>
<td>Number of users</td>
<td>4 to 16</td>
<td>32 to 100</td>
<td>200</td>
<td>200 to 1000+</td>
</tr>
<tr>
<td>Maximum output power</td>
<td>20 to 100 mW</td>
<td>250 mW</td>
<td>2 to 10 W</td>
<td>40 to 100W</td>
</tr>
<tr>
<td>Maximum Cell radius</td>
<td>10 to 50 m</td>
<td>200 m</td>
<td>2km</td>
<td>10 to 40 km</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>10 MHz</td>
<td>20 MHz</td>
<td>20, 40 MHz</td>
<td>60 to 75 MHz</td>
</tr>
<tr>
<td>Technology</td>
<td>3G/3G/WiFi</td>
<td>3G/4G/WiFi</td>
<td>3G/4G/WiFi</td>
<td>3G/4G</td>
</tr>
<tr>
<td>MIMO</td>
<td>2x2</td>
<td>2x2</td>
<td>4x4</td>
<td>4x4</td>
</tr>
<tr>
<td>Backhaul</td>
<td>DSL, Cable, fiber</td>
<td>Microwave, mm</td>
<td>Fiber, microwave</td>
<td>Fiber, microwave</td>
</tr>
</tbody>
</table>
The smallest is the femtocell, which is a single-box BS used by the consumer to improve local cellular service. Femtos have been around for years, and millions have been installed by most of the larger carriers. Backhaul is by way of the customer’s high-speed Internet connection via a cable TV or DSL telecom provider. There are also enterprise femtos that handle more users and provide a significant boost in indoor accessibility. There are progressively larger small cells such as the picocell, microcell, and metrocell, each with increasing capacity, power, and range. Virtually all handle legacy 3G, LTE, and Wi-Fi. Many future small cells will also feature LTE-Advanced.

In this section we propose and devise a novel PON-based HetNet mobile backhaul RAN architecture that enables, for the first time to the best of our knowledge, the support of efficient ground-breaking radio access network (RAN) as well as core network offload techniques, which are fully managed and controlled by the mobile core network, without resorting to typical deployment of Internet offload gateways. We quantify the performance impact of utilizing PON-based FTTx access network architecture to backhaul a large number of small cells. In contrast to the PON–based small-cell backhaul architecture reported in [7], which utilizes the typical star-based PON topology, a local access small ring-based PON topology is rather assumed here. Specifically, a fully distributed ring-based EPON architecture is utilized here as the fiber-based HetNet/SCs backhaul.

The significance of the purposely selected simple ring topology is: 1) it enables direct intercommunication /connectivity among the SCs and among the macro BS (mBS) and SCs, allowing for the support of Efficient interference management and coordination, which requires SCs to be directly interconnected at lowest possible latency (via the direct physical connectivity among the SCs attached to the ring); 2) it facilitates highly accurate synchronization among the
SC BS clocks. The faster real-time signaling information can be exchanged between SC BSs, the more accurately clocks are aligned and the less interference [9]; 3) minimizes the X2 (logical connectivity between neighboring SCs) interface latency; thus, allowing for harnessing the highest CoMP gains; 4) meets the stringent requirement to fully meshing the SCs, in conformity with the LTE standards; and 5) the inherent self-healing mechanism of the ring architecture facilitates and guarantees the reliable delivery of mobile traffic.

The main characteristics of the proposed PON-based HetNet backhaul RAN architecture is that it supports a fully distributed control plane that enables direct intercommunication among the access nodes (ONUs/SCs/ mBS) as well as signaling, scheduling algorithms, and handoff procedures that operate in a distributed manner. We quantify the technical merits of the proposed PON-based HetNet backhaul by comparing its performance versus that of a conventional PTP backhaul architecture as a benchmark. The purposely selected ring-based RAN architecture along with the supporting distributed control plane enable the proposed EPON-based backhaul RAN architecture to support several key salient networking features that collectively significantly enhance the performance of both the RAN and LTE’s core network (Evolved Packet Core (EPC) per 3GPP standard) compared to that of the typical PTP backhaul architecture in terms of handoff capability, signaling overhead, overall network throughput and latency, and QoS support.

In addition to supporting the typical macro-cell offloading gain (RAN gain) that mainly corresponds to the saving in macro-cell resources, the proposed backhaul RAN architecture also supports an innovative EPC offloading gain, which ensures that core network equipment is used more productively. The EPC offloading gain is defined here as offloading a significant volume of IP traffic (including both real-time IP traffic (VOIP, video) and best effort traffic), which is
typically routed to and through the mobile core network (EPC), directly to the RAN (not to the
Internet) and away from EPC. This traffic is routed and processed at the RAN but still under the
full control and management of the EPC. Note that this in radical contrast with the typical core
network offload (Internet offload), which requires the deployment of Internet offload gateways
to offload/divert only best effort traffic from the mobile core directly to the Internet.

4.2 PON and LTE-A/HetNet Interconnection Models

Figure 4-3: PON and LTE-A/HetNet Interconnection Models

As shown in Figure 4.3, there are two interconnection models (depending on how the
ONUs are interconnected to the BS/SC, namely, the overlay (independent) model and the
integrated model [18-19]. Under this simple overlay (independent) model, the PON and HetNet
systems are operated independently where the RAN system is assumed to have its own NCM
operations, independent of those for the PON. The mBS/SC is assumed to be collocated with an
ONU or treated as a generic user attached to it. The ONU and mBS/SC can be interconnected as long as they support a common standard interface. Thus, the OLT, AGW, ONUs, and mBSs/SCs, are all assumed to support a common standard interface (e.g., 802.3ah Ethernet interface). Each ONU is assumed to have two different Ethernet port ranges; the first port range will support wired users, while the second port range will support mobile users. The port ranges will be used by the ONUs to identify and differentiate between mobile users versus fixed users. Since EPC aggregate traffic from thousands of mBS/SCs, numerous OLTs can be attached to it (only two are shown in Figure 4-3 for simplicity).

Under the integrated model, an ONU and LTE’s mBS/SC can be functionally integrated into a single module either in terms of software or both software and hardware functionalities. The following are the main technical requirements needed to support the functional integration of the PON and HetNet access infrastructure: 1) the OLT, S-GW, ONUs, and mBS/SCs, are all assumed to support a common standard interface (e.g., 802.3ah Ethernet interface); 2) each ONU is assumed to have two different Ethernet port ranges, the first port range will support wired users, while the second port range will support mobile users. The port ranges will be used by the ONUs to identify and differentiate between mobile versus fixed users; 3) Depending on the selection of either Layer-3 or Layer-2 connectivity at the transport layer, all the intermediate nodes (e.g., OLT, ONU, mBS/SC) in Figure 4-3 are assumed to be equipped with either an IP access router to forward IP packets or GE Ethernet switch to forward the traffic using either PBB, PBT, or VPLS. This is a critical issue that will be thoroughly investigated during the course of this work.

Figure 4-4b illustrates the three main control modules of the functionally integrated ONU-eNB access node, namely, ONU’s control module, eNB’s control module, and the common
control module, where each module can be a single CPU in hardware [18]. ONU’s module interfaces with the PON section and runs the PON protocols; eNB’s module interfaces with LTE section and runs the LTE protocols. The common module interfaces to both the PON and eNB sections, manages and coordinates joint optical-radio resources, and executes the integrated DBA and packet scheduling algorithms. ONU and eNB modules report their queue statuses and bandwidth request details to the common module; the latter utilizes this information to make decisions, and to optimally allocate upstream/LAN resources to the ONU and eNBs.

The functional modules for provisioning upstream traffic corresponding to the three modules in Figure 4-4b are shown in Figure 4-4a. Specifically, the ONU’s control module that interfaces with the PON section includes the functional components of PON packet scheduler, priority queues management, and PON packet classifier. Similarly, the LTE’s module that interfaces to the LTE section includes the functional components of two LTE mapping modules (one to map UE’s radio bearers to mobility tunnels), eNB packet classifier, and LTE upstream scheduler. Finally, the third at the bottom of figure 4.4 (a) corresponds to the ONU-eNB common coordinator controller, which comprises the functional components required to map QoS between PON and LTE and performs global admission and congestion control as well as integrated DBA and resource allocation and sharing protocols and algorithms.
Figure 4-4: (a) Architecture of the ONU-eNB, (b) Functional Modules hardware

4.3 Proposed PON-Based HetNets Backhaul RAN Architecture

As shown in Figure 4-5, the standalone ring-based EPON architecture can be evolved to a HetNet backhaul RAN architecture by simply collocating (overlying) the SCs and the macro BS (mBS) with the ONU, while capitalizing on existing fibered (available fiber backhaul over dark fibers) and powered ONU associated with the PON-based FTTx residential access network. The SCs can be deployed using a low-height (2-4 m) antenna mounted on or near the ONU (e.g., on an adjacent light post) [7]. The coverage radius of a small-cell is typically assumed in the 100-300 m range and the small-to-small inter-site distance (ISD) is assumed to be in the 400-500 m range. It is further assumed that the SCs are placed around the periphery of existing macro-cells serving area (> 700 m from the nearest macro-cell site), thus improving the poor coverage near the macro-cell perimeter [7]. The Central Office (CO) houses the OLT, which connects with metro/EPC via the metro terminal equipment collocated at the CO.

Because EPC is designed to be access-independent, it can support the integration of both the LTE-A SCs and WiFi APs. However, the integration of WiFi APs, according to the EPC
standards for 3GPP and Non-3GPP interworking, depends on whether these APs are classified as “Trusted or Un-Trusted Non-3GPP Access Networks”. Trusted Wi-Fi Networks mean that the WiFi APs are deployed and managed by the Operator, so that UE can connect to the WiFi network directly using the radio interface without requiring any additional security measures. In contrast, Un-trusted WiFi networks do not have any trust relationship to the operators, so that the operators require that the UE establish a secure tunnel (i.e. IPSec tunnel) to a trusted node in the operator core network. Typically, such a node is termed “Evolved Packet Data Gateway” (ePDG) in EPC networks. Because the proposed PON-based architecture, which is used to backhaul both the LTE-A SCs and WiFi APs, is likely to be considered untrusted IP/Ethernet backhaul, IPSec termination will be needed. As shown in Figure 3, the ePDG is likely to be installed at the edge of the EPC to terminate and aggregate the high number of incoming tunnels/connections.

Figure 4-5: Proposed EPON-based HetNets Backhaul RAN Architecture
The significance of the purposely selected simple ring topology: 1) it enables direct intercommunication/connectivity among the SCs and among the macro BS (mBS) and SCs, allowing for the support of Efficient interference management and coordination, which requires SCs to be directly interconnected at lowest possible latency (via the direct physical connectivity among the SCs attached to the ring); 2) it facilitates highly accurate synchronization among the SC BS clocks. The faster real-time signaling information can be exchanged between SC BSs, the more accurately clocks are aligned and the less interference; 3) minimizes the X2 interface latency, thus, allowing for harnessing the highest CoMP gains; and 4) the inherent self-healing mechanism of the ring architecture facilitates and guarantees the reliable delivery of mobile traffic.

4.4 Optimal Small Cell Location Problem
The following scenarios for small cell location will be considered:

1. Macro at Center of cell radius=1 km
2. Macro at Center Plus One Small Cell at the edge
3. Macro at Center Plus 4 Small Cell at equal distance from Macro

Figure 4-6: Macro at Center of cell only
Finding the optimal deployment configuration by enumeration is only feasible for very small instances. As the number of potential sites and city blocks increases, one needs to resort to an optimization algorithm. The small cell covering problem is a maximum covering problem, and integer programming formulations for it exist. However, the problem is NP-hard, implying that as the problem size increases, integer programming will eventually not be able to find the optimal solution to the problem. In such situations, one often resorts to heuristics. Several heuristics for maximum covering have been proposed in the literature (e.g., [13,14]). A very efficient software for maximum covering is POPSTAR [15], which formulates the maximum
covering problem as the NP-hard p-median problem and applies GRASP with the evolutionary path-relinking heuristic for p-median described in [16]. It can quickly find optimal and near-optimal solutions to small cell covering problems having thousands of potential cell locations and tens of thousands of city blocks.

Consider the example in Fig. 4-9 where there are nine city blocks and four potential small-cell sites: $a$, $b$, $c$, and $d$. Small-cell coverage is indicated by the shaded blocks [7]. If only one small cell will be deployed (i.e., SC = 1), the optimal choice is site $b$, since it alone covers 41 people, while the other sites each cover fewer people. If two small cells are deployed (i.e., SC = 2), the optimal choice is $a$ and $d$, since together these cells would cover 70 people, while $\{a, b\}$, $\{a, c\}$, $\{b, c\}$, $\{b, d\}$, and $\{c, d\}$ each cover fewer. If three cells are to be deployed (i.e., SC = 3), the optimal choice is $b$, $c$, and $d$ which covers 79 people, while $\{a, b, c\}$, $\{a, b, d\}$, and $\{a, c, d\}$ each cover fewer. Finally, if SC = 4 (i.e., all cells are chosen), 84 people are covered. Note that the incremental coverage decreases as more cells are deployed, going from a 71 percent increase from SC = 1 to SC = 2, to a 13 percent increase from SC = 2 to SC = 3, to only a 6 percent increase going from SC=3 to SC=4 [7, Copyright © 2013, IEEE].
Fig. 4-10 [7] shows the total deployment cost of a typical PTP deployment scenario and the optimal PON-based solution under different split ratios for one CO serving area of AT&T’s existing FTTN network [7]. In addition to the total deployment cost, Fig. 4-10 also shows the
cost contribution of each major cost component involved in the deployments. Note that the values shown in Fig.4-10 are normalized with respect to the total deployment cost of the PTP solution such that the total deployment cost of the PTP solution is 100. In contrast to green field deployments, where labor is typically the dominant cost, Fig. 4-10 shows that the main cost contributor in the optimal PON-based deployment is the equipment cost. Conversely, the main cost contributor in the PTP deployment is the fiber. Moreover, the deployment costs of the PON-based solution increase when the split ratio decreases. In particular, the equipment and labor costs increase while the fiber cost decreases as the split ratio decreases. This occurs because the number of splitter locations and the number of PONs that are required for such a deployment are increased, resulting in higher costs for equipment and labor. Overall, for this test case, the cost of the optimal PON-based solution saves more than 50 percent of deployment cost in comparison to that of the PTP case [7].
Chapter 5

5 Core Innovative Building Blocks To Realize the Proposed PON-Based HetNet RAN Architecture
In this section we present and devise the key building blocks, which enable the realization of the proposed PON-based HetNet RAN architecture including: 1) QoS support and mapping; 2) A fully distributed Control Plane; 3) Fully Distributed Dynamic Bandwidth Allocation schemes at the ONU/SCs; and 4) Layer-2 versus Layer-3 Connectivity at the Transport Layer.

5.1 Overview of QoS in LTE-A

The 3GPP specifications define eight standardized QCI as indicated in Table 5-1, each with its corresponding standardized characteristics including bearer type (GBR versus non-GBR), priority, packet delay, and packet-error-loss rate. To allow for traffic separation in the transport network (IP cloud connecting the eNBs to the EPC), P-GW and eNB map each QCI onto a corresponding diffserv code point (DSCP) in order to translate a bearer-based QoS (QCI) to transport-based QoS (DSCP) [1-2]. Using this mapping function, packets on a bearer associated with a specific QCI are marked with a specific DSCP for forwarding in the transport network. The QCI to DSCP mapping is performed based on operator policies, which are configured into the network nodes. P-GW performs the mapping for DL packets while SC performs it for UL packets.
Table 5-1: LTE-A Standardized QCI Characteristics

<table>
<thead>
<tr>
<th>QCI</th>
<th>Resource Type</th>
<th>Priority (ARP)</th>
<th>Packet Delay Budget (PDB)</th>
<th>Packet Error Loss Rate (PELR)</th>
<th>Example Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GBR</td>
<td>2</td>
<td>100 ms</td>
<td>$10^{-2}$</td>
<td>Conversational voice</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>4</td>
<td>150 ms</td>
<td>$10^{-3}$</td>
<td>Conversational video (live streaming)</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>3</td>
<td>50 ms</td>
<td>$10^{-3}$</td>
<td>Real time gaming</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>5</td>
<td>300 ms</td>
<td>$10^{-6}$</td>
<td>Non-conversational video (buffered streaming)</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1</td>
<td>100 ms</td>
<td>$10^{-6}$</td>
<td>IMS Signaling</td>
</tr>
<tr>
<td>6</td>
<td>Non-GBR</td>
<td>6</td>
<td>300 ms</td>
<td>$10^{-6}$</td>
<td>Video (Buffered streaming)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TCP-based (e.g., www, email etc)</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>7</td>
<td>100 ms</td>
<td>$10^{-3}$</td>
<td>Voice</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Video (live streaming)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Interactive gaming</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>8</td>
<td>300</td>
<td>$10^{-5}$</td>
<td>Video (Buffered streaming)</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>9</td>
<td></td>
<td></td>
<td>TCP-based (e.g., www, email etc)</td>
</tr>
</tbody>
</table>

5.2 QoS Mapping

The QoS model of EPS, which was standardized in 3GPP release 8, is based on the logical concept of an “EPS bearer” [1-5]. The term “bearer” refers to a logical IP transmission path between the UE and the EPC with specific QoS parameters (capacity, delay, packet loss error rate, etc.). Each bearer is assigned one and only one QoS class identifier (QCI) by the network and is composed of a radio bearer and a mobility tunnel. The QCI is a scalar that is used within the access network to identify the QoS characteristics that the EPC is expected to provide for the IP SDFs. This scalar (bearer ID) is used by routers to access node-specific parameters that control packet forwarding treatment (e.g., scheduling policy, admission thresholds, link layer configurations,
queue management policy, etc.), which are specified and preconfigured by the operator. An EPS bearer uniquely identifies packet flows that receive the same packet forwarding treatment between the UE and EPC. Thus, the aggregated IP flows constituting a bearer are carried from the UE over the radio interface to the eNB, from the eNB to the S-GW, and then onwards to the P-GW as a single logical bearer with the same level of QoS (or packet forwarding treatment). Services with IP flows requiring a different packet forwarding treatment would therefore require more than one EPS bearer.

An IP flow is defined by a five-tuple (the source and destination IP addresses, source and destination port numbers, and protocol ID, typically are referred to as the IP five-tuple), which is used by the packet filter to identify different IP flows. Downlink (DL) IP flows are identified by downlink packet filters located at the P-GW, while uplink (UL) IP flows are identified via uplink packet filters located at the UE. Thus, the UE/P-GW performs UL/DL packet filtering to map the outgoing/incoming IP flows onto the appropriate bearer (bearer binding). There are two types of bearers: guaranteed bit-rate (GBR) and non-guaranteed bit-rate (non-GBR) bearers. A GBR bearer has a guaranteed bit-rate (GBR) and maximum bit-rate (MBR) while more than one non-GBR bearer belonging to the same UE shares an Aggregate Maximum Bit Rate (AMBR). Non-GBR bearers can suffer packet loss under congestion, while GBR bearers are immune to such losses (via admission control functions that reside at the eNB and P-GW). A bearer can also be classified as either a default or a dedicated bearer. The default bearer is set up when the UE attaches to the network to provide the basic connectivity. The 3GPP specifications mandate that the default bearer is a non-GBR bearer. The dedicated bearer can be either a GBR or a non-GBR bearer.
For a given bearer, QoS characteristic is completely defined by two parameters: QCI (bearer ID) and Allocation and Retention Priority (ARP) that specifies the control plane treatment that the bearers receive. ARP does not have any impact on packet forwarding behavior but is used to decide whether a bearer establishment/modification request can be accepted or rejected. The 3GPP specifications define eight standardized QCIs, each with its corresponding standardized characteristics including bearer type (GBR versus non-GBR), priority, packet delay, and packet-error–loss rate. To allow for traffic separation in the transport network (IP cloud connecting the eNBs to the EPC), P-GW and eNB map each QCI onto a corresponding diffserv code point (DSCP) in order to translate a bearer-based QoS (QCI) to transport-based QoS (DSCP) [2]. Using this mapping function, packets on a bearer associated with a specific QCI are marked with a specific DSCP for forwarding in the transport network. The QCI to DSCP mapping is performed based on operator policies, which are configured into the network nodes. P-GW performs the mapping for DL packets while eNB performs it for UL packets.

As can be seen from the eNB/SC module shown on Figure 5-1, the UE uses the packet filters to classify IP packets to authorized IP SDFs. This process is referred to as SDF detection. The UE then performs the binding of the detected uplink IP SDFs to the appropriate bearers. Once the UE’s radio bearers are terminated at the eNB/SC, they are mapped into the appropriate mobility tunnels based on their bearer-IDs. The eNB’s packet classifier then maps their constituent IP flows into their appropriate priority queues based on the bearer-IDs attached to the IP packets, which is the basic enabler for traffic separation. Finally, to allow for traffic separation in the transport network, the eNB maps each OCI (bearer-ID) onto the corresponding DSCP value.
Figure 5-1: (a) Architecture of the ONU-eNB, (b) Functional Modules hardware layout

On the other hand, EPON technology does not support this type of bearer-based connection. Rather, bandwidth requests are queue-oriented; an aggregate bandwidth is allocated to each ONU, and then the latter makes a local decision to allocate the granted bandwidth and schedules packets transmission for up to eight different priority queues in the ONU. In a typical centralized EPON, QoS support is implemented via two independent scheduling mechanisms [7]: 1) inter-ONU scheduling: an aggregate bandwidth is allocated to each ONU by the OLT. 2) intra-ONU scheduling: each ONU makes a local decision to allocate the granted bandwidth and schedules packets transmission for up to eight different priority queues in the ONU. Under the proposed integrated architecture, however, instances of the same DBA algorithm are executed simultaneously at each ONU. Thus, both scheduling mechanisms (inter and intra-ONU scheduling) are performed at each ONU in a fully distributed approach, leading to the notion of
integrating both scheduling mechanisms at the ONU. This enables the proposed distributed architecture to provide better QoS support and guarantees.

Both EPON and LTE classify data traffic in a differentiated services mode. However, EPON supports only enhanced QoS through prioritization where packets are classified, stored in different priority queues and, then, scheduled for service according to their priority. On the other hand, LTE supports guaranteed QoS through logical bearer reservation where each router/node on the RAN/EPC is configured to forward the packets of different IP flows based on their bearer-IDs (QCIs) in which resources are reserved (queue space, queuing management strategy, scheduling strategy) accordingly.

To achieve a truly integrated model, an effective mapping mechanism is required between EPON priority queues and QCI/bearer-based LTE IP flows. Specifically, the mapping has to identify which LTE IP flow should be stored in which EPON priority queue for equivalent QoS. EPON has up to eight different priority queues in each ONU, while LTE defines eight standardized QCIs that classify data traffic into eight different classes of service, ranging from real-time gaming to the lowest priority best effort TCP bulk data. This theoretically facilitates a one-to-one mapping from eNB’s eight priority queues to ONU’s eight priority queues (e.g., packets of highest/lowest eNB’s priority queue are mapped onto highest/lowest ONU’s priority queue) and vice versa in both upstream and downstream directions. However, devising an efficient viable mapping strategy that enables a unified QoS model for both wired and wireless services requires the implementation of the following critical functions:

1) Since the bearers are not visible to the ONUs/OLTs, each and every ONU/OLT must be directly configured (semi-statically) with all eight LTE’s standardized QCIs (QoS characteristics) or more precisely with the corresponding DSCP values (QCI to DSCP mapping
is performed based on operator policies). This configuration enables each ONU/OLT to forward the packets of different UL/DL IP flows based on their DSCP values such that the packets-forwarding treatment received by these flows at the ONU/OLT is identical to that received at the eNB/P-GW. This is achieved by ensuring that the queue management schemes and scheduling algorithms implemented at the ONU/OLT are identical to those implemented at the eNB/P-GW.

2) The PON’s packet scheduler at the ONU/OLT must apply the same packet forwarding treatment for both wired and wireless upstream/downstream traffic for each and every configured QCI/DSCP value that is associated with a given IP flow. This further enhances the typical PON’s prioritization-based QoS support for wired users as well as simplifies the implementations of queue management schemes and scheduling algorithms at the ONUs and OLTs.

3) The typical PON’s cycle-based approach for DBA and QoS support must be drastically modified at both the ONUs and OLTs. None of EPON scheduling mechanisms can guarantee bandwidth for real-time IP flows because the bandwidth allocated by the OLT to one ONU can only be guaranteed for a significantly short time (e.g., a fraction of one cycle) and may vary from one cycle to another cycle according to the load at other ONUs. Thus, each ONU is required to reserve bandwidth for its real-time IP flows for the whole duration of the flow (and not on a per cycle basis) in order to satisfy their QoS requirements as specified by the attached DSCP value.

4) In addition to bandwidth allocation and service differentiation, a global admission and congestion control (AC) mechanism for both wired and wireless traffic that makes decisions on whether or not to admit/block a new wired/wireless real-time IP flow based on its requirements and the upstream channel usage condition. Ideally, this AC module should be housed at the common control module (Figure 5-1) since the critical information needed by the AC module to make appropriate admission/denial decisions (e.g., available fixed optical and mobile radio
resources as well as both available wired (ONUS-OLT) and wireless (UE-eNBs) uplink channel capacities) is always dynamically available to the common control module. For instance, when the congestion bottleneck is at the backhaul and not at the radio interface, the common control module can block the admission of any new mobile user’s traffic until congestion subsides.

The combination of a distributed PON-RAN architecture along with a fully distributed/unified control plane with global information about the entire fixed-mobile network status collectively enable the implementation of a simple and efficient QoS-aware DBA scheme, in which resources are reserved (e.g., queue space and bandwidth) via signaling. Note that the overall process of QoS mapping and support can be further simplified by reducing the number of standardized QoS levels for both PON and LTE from eight to the typical three DiffServ’s classes of services (Expedited Forwarding (EF), Assured Forwarding (AF), and Best effort (BE)), which are commonly and widely used by operators.

5.3 Fully Distributed Control Plane

This work utilizes the control and management messages defined by the IEEE 802.3ah multi-point control protocol (MPCP) standard [9] that facilitate the exchange of control and management information between the ONUs/SCs/macro BS and OLT. The protocol relies on two Ethernet control messages, GATE (form OLT to ONUs) and REPORT (from ONUs to OLT and between ONUs/SCs/mBS) messages in its regular operation. Direct communication among ONUs/SCs/ mBS is achieved via the US wavelength channel {control messages along with both LAN and US data share the same US channel bandwidth (in-band signaling)}, which is terminated, processed, regenerated, and retransmitted at each ONU.

Since control messages are processed and retransmitted at each node, the ONUs can directly communicate their US/LAN queue status and exchange signaling and control information with
one another in a fully distributed fashion. Likewise, SCs/mBS can also directly communicate the status of their queues and radio resources and exchange signaling and control messages with one another. The control plane utilized among the ONUs/SCs/mBS can thus support a distributed HetNet RAN architecture, where each access node (ONU/SC/mBS) deployed around the ring has now a truly direct physical connectivity and is, thus, capable of directly communicating with all other access nodes, in conformity with LTE standards.

Each access node maintains a database about the states of its own queue and every other ONU/mBS/SC’s queue on the ring. This information is updated each cycle whenever the ONU receives new REPORT messages from all other ONUs. During each cycle, the access nodes sequentially transmit their REPORT messages along with both US and LAN data in an ascending order within their granted timeslots around the ring from one node to the next, where each REPORT message is finally removed by the source ONU after making one trip around the ring. The REPORT message typically contains the desired size of the next timeslot based on the current ONU’s buffer occupancy. Note that the REPORT message contains the aggregate bandwidth of mobile data buffered at each SC’s queue (requested size of next timeslot).

An identical dynamic bandwidth allocation (DBA) module, which resides at each access node (ONU/SC/mBS), uses the REPORT messages during each cycle to calculate a new US timeslot assignment for each ONU. ONUs sequentially and independently run instances of the same DBA algorithm outputting identical bandwidth allocation results each cycle. The execution of the algorithm at each ONU starts immediately following the collection of all REPORT messages. Thus, all ONUs must execute the DBA algorithm prior to the expiration of the current cycle so that bandwidth allocations scheduled for the next cycle are guaranteed to be ready by the end of the current cycle. Once the algorithm is executed, the ONUs sequentially and orderly transmit
their data without any collisions, eliminating the OLT’s centralized task of processing requests and generating grants for bandwidth allocations. Thus, supported by the distributed control plane, most of the typical radio control functions including radio resource management, handover control, admission control, etc, can be independently implemented at each SC/mBS in a distributed approach without resorting to a central control entity (e.g., EPC’s AGW).

5.4 Fully Distributed Dynamic Bandwidth Allocation at the ONUs/SCs

5.4.1 Overview of Typical PON Scheduling Schemes

In order for mainstream centralized EPON architectures to support differentiated QoS, two independent scheduling mechanisms are required:

a) Scheduling at the OLT (inter-ONU scheduling): The OLT is the only device that can arbitrate the upstream transmissions by allocating an appropriate TW to each ONU. In this case, the OLT passes the request messages to a dynamic bandwidth allocation module (co-located with the OLT) that performs the bandwidth allocation computation and generates grant messages.

b) Scheduling at the ONU (intra-ONU scheduling): In this case, queue management and priority queuing are used to divide the bandwidth allocated by the OLT to a given ONU among the different class of services (based on their priorities) supported by that ONU.

Since the two scheduling schemes are independent of each other, the final bandwidth allocated to a particular class of service for a given ONU may not be the optimum choice.

Several centralized tree-based DBA schemes have been reported in the literature [10-13]. An OLT-based polling scheme, called Interleaved Polling with Adaptive Cycle Time (IPACT) based on Grant and Request messages, has been presented in [10]. Using IPACT, several DBA
schemes were studied in [10]; namely fixed, limited, gated, constant credit, and linear credit. Amongst these algorithms, the limited was shown to exhibit the best performance. The limited DBA scheme is cycle-based, where a cycle \( T_{CYC} \) is defined as the time that elapses between two executions of the scheduling algorithm. A cycle has a variable length size confined within certain lower and upper bounds, which we denote as \( T_{MIN} \) and \( T_{MAX} \) (sec) respectively. Thus, the algorithm schedules between \( B_{MIN} \) and \( B_{MAX} \) (bytes) at a time, where \( B_i \) is determined by multiplying \( T_i \) with the line rate. In this scheme, the ONU will be granted the requested number of bytes, but no more than a given predetermined maximum \( B_{MAX} \). If \( R_i \) is the requested bandwidth of ONU\(_i\), then the granted bandwidth \( (B_{Granted}) \) is equal to:

\[
B_{Granted} = \begin{cases} 
R_i & \text{If } R_i \leq B_{MAX} \\
B_{MAX} & \text{If } R_i > B_{MAX}
\end{cases}
\]  

(1)

\( B_{MAX} \) is determined by the maximum cycle time \( T_{MAX} \) [9]:

\[
B_{max} = \frac{1}{N} [R_{EPON} \cdot (T_{MAX} - (N \cdot T_G))] 
\]

(2)

where \( N \) is the number of ONUs, \( T_G \) is the guard band slot, and \( R_{EPON} \) is EPON line rate.

5.4.2 Decentralized Dynamic Bandwidth Allocation Scheme

All of the above referenced DBA schemes are OLT-based, that is the OLT has centralized intelligence. The performance of most of these centralized schemes, including the limited scheme, suffers from several limitations, including: (1) the bandwidth granted by the OLT, during cycle \( n \), to ONU\(_i\) is only determined by the content of a single REPORT message...
transmitted in the previous cycle n-1 by ONU\textsubscript{i} (i.e., the bandwidth computation module does not take into account the remaining requests of other ONUs). Thus, the process of bandwidth allocation is not globally optimized; (2) due to the bursty nature of Ethernet traffic, some ONUs might have less traffic to transmit while other ONUs may require more bandwidth than \(B_{\text{max}}\). For instance, assume that ONU\textsubscript{i} requests an amount of bandwidth \(R_i < B_{\text{max}}\), while ONU\textsubscript{j} requests an amount of bandwidth \(R_j > B_{\text{max}}\). Although there is an excess amount of bandwidth \((B_{\text{max}} - R_i)\) that can be granted to ONU\textsubscript{j}, however, due to limitation # 1 cited above, the maximum bandwidth that may be granted to ONU\textsubscript{j} is only \(B_{\text{max}}\).

The proposed distributed ring-based EPON architecture, however, enables instances of the same DBA algorithm (inter-ONU scheduling) to be executed simultaneously at each ONU. Thus, both scheduling tasks (inter and intra-ONU scheduling) schemes are performed at the ONU leading to the notion of integrating both scheduling mechanisms at the ONU.

In this work, the centralized limited service scheme reported in [10], along with the appropriate changes needed to accommodate the distributed architecture, is used here as the basis for the decentralized DBA scheme presented here. As mentioned above, to globally optimize the bandwidth allocation process, the proposed DBA algorithm execution is performed only after each ONU receives and processes all other ONUs requests.

Based on bandwidth demands, ONUs can be classified into two groups, namely: lightly loaded ONUs that has bandwidth demands less than \(B_{\text{MAX}}\); and heavily loaded ONUs that have bandwidth demands more than \(B_{\text{MAX}}\). During each cycle, the DBA module must keep track of the unclaimed bandwidth from the set of lightly loaded ONUs. It then must redistribute this excess bandwidth to other heavily loaded ONUs based on their requested bandwidth, i.e. two
ONUs requesting bandwidths B1 and B2 more than $B_{MAX}$ will be assigned excess bandwidths proportional to B1 and B2.

During each cycle, the lightly loaded ONUs with $R_i < B_{MAX}$ will contribute a total remainder cycle bandwidth:

$$B_{Cycle\_Remainder} = \sum_{i=1}^{L} (B_{MAX} - R_i)$$

$L$: Number of lightly loaded ONUs

The heavily loaded ONUs with $R_i > B_{MAX}$ will require a total over the limit cycle bandwidth:

$$B_{Cycle\_OverLimit} = \sum_{i=1}^{H} (R_i - B_{MAX})$$

$H$: Number of heavily loaded ONUs

The total remainder cycle bandwidth can be fairly distributed amongst the heavily loaded ONUs to expand their maximum transmission window as follows [11]:

$$\Delta B_{extra} = B_{Cycle\_Remainder} \left[ \frac{R_i - B_{MAX}}{B_{Cycle\_OverLimit}} \right]$$

$$\text{(3)}$$

where $\Delta B_i$ is the extra bandwidth allocated to ONU$_i$. The granted bandwidth, $B_{GH}$, for a heavily loaded ONU$_i$ is given by:

$$B_{GH} = \Delta B_{extra} + B_{MAX}$$

$$\text{(4)}$$

If $R_i$ is the requested bandwidth of ONU$_i$, $B_{Granted}$ is the bandwidth granted using the proposed limited service-based distributed DBA scheme (Eqs. 1 and 4), then $B_{Granted}$ can be expressed as:

$$B_{Granted} = \begin{cases} 
R_i & \text{if } R_i \leq B_{MAX} \\
B_{GH} & \text{if } R_i > B_{MAX} \text{ and } B_{Cycle\_Remainder} \geq B_{Cycle\_OverLimit} \\
B_{GH} & \text{if } R_i > B_{MAX} \text{ and } B_{Cycle\_Remainder} < B_{Cycle\_OverLimit}
\end{cases}$$

$$\text{(5)}$$
Note that the lightly loaded ONUs ($R_i < B_{MAX}$) can be scheduled instantaneously “on-the-fly” without waiting for DBA module to perform its end of cycle computations. Whereas, the heavily loaded ONUs ($R_i > B_{MAX}$) will have to wait until all REPORT messages have been received and the DBA algorithm has computed their bandwidth allocations. Thus, lightly loaded ONUs can be scheduled ahead of heavily loaded ones.

Thus, the proposed decentralized EPON architecture addresses some of the limitations of the centralized DBA schemes cited above and can further provide several advantages as follows:

- Since the bandwidth allocation computation is performed after receiving and processing all ONUs requests (processing period) (i.e., the computation takes into account the entire network status), the bandwidth allocation process now reflects the entire network information collectively, leading to a globally optimized decision.

- In contrast to the centralized architectures where the order of ONUs transmission is fixed in each cycle (sequential), the decentralized architecture has the added flexibility of varying the order of ONUs transmission according to the ONUs traffic demands and priority. Thus, the order of ONUs transmission may be different in each cycle and need not be fixed.

- Since the DBA computation is based on the global network information, the heavily loaded ONUs may be allocated the remaining excessive bandwidth that is not utilized by the lightly loaded ONUs.

- Given that DBA and priority queuing scheduling tasks are both executed at the ONU, the DBA module can integrate both scheduling information to yield a globally optimized bandwidth allocation to a particular class of service in a given ONU.
Because the centralized limited DBA scheme was shown to exhibit the best performance in [13], we will consider this scheme as a reference model for comparing the performance of our distributed architecture versus that of the centralized scheme reported therein.

5.5 Layer-2 versus Layer-3 Connectivity at the Transport Layer

Determining the most effective and efficient mix of layer-2 and layer-3 in the backhaul is a major issue worldwide. There are myriad approaches to support LTE backhaul. First, the transport network choice and architecture could have a significant impact on EPC – for instance, where layer-2 carrier Ethernet is utilized, a more centralized EPC gateway (S-GW and P-GW) is preferred. Alternatively, if layer-3 dominates, the gateway can be distributed and perhaps integrate that capability with edge routers. The EPC initial deployment is expected to be one of dedicated mobility nodes (S-GW and P-GW) installed on top of IP/MPLS core networks. However, over the longer term, there is a potential for eliminating the boundaries between the IP network and EPC “mobility layer” [8]. Under this scenario, we can assume that EPC applications can be implemented on a router where a dedicated module or blade is added to the router to provide EPC functions, resulting in a “Carrier-grade edge-router” that performs both typical routers and EPC functionality.

Thus, under the assumption of layer-3 connectivity, the EPC is modeled as a distributed architecture by pushing the S-GW and P-GE nodes to the edge and assume that these nodes are multiservice carrier-grade edge-routers incorporated with the typical IP/MPLS core network. On the contrary, the second assumption is that rather than pushing layer-3 routing and S-GW and P-GW out towards the edge of the network, the focus instead should be on low-cost layer-2 Ethernet transport, backhauling traffic to a more centralized S-GW and P-GW that are implemented in a blade server-platform or other non-router platforms [8]. Under this assumption
(layer-2 Ethernet connectivity), several hierarchical carrier grade Ethernet transport solutions can be utilized including: a) the IEEE 802.1ad (Qin Q or “double tagging”); b) the IEEE 802.1ah (MAC-in-MAC or Provider Backbone Bridges (PBB)); c) Provider Backbone Transport (PBT), or PBT-TE (PBT with traffic engineering).

To avoid the complexity of IP/MPLS control planes, along with the fact that Ethernet is considered as the most effective method to transport IP packets, the LTE backhaul can, for example, use MEF-compliant interfaces on the eNB/SC and on the S-GW and MME. The mobile operator can send VLAN-tagged frames toward the EPC. The backhaul can now identify the VLAN tag, then maps these frames to the EVCs (Ethernet virtual circuits). A multipoint EVC can be used to support X2 among a cluster of eNBs that need to exchange protocols. Initially, we lean more towards the approach of leaving the IP functionality to the mobile endpoints that actually need it (e.g., the eNB and the EPC), and avoiding it in the backhaul network by utilizing carrier Ethernet. A detailed technical and economic study is needed to weigh the pros and cons of each transport technology to determine the optimum solution.
Chapter 6

6 Key Salient Innovations Enabled By the Proposed HetNet RAN Architecture
The distributed ring-based PON architecture along with the supporting control plane enable the proposed EPON-based HetNet RAN architecture to support several key salient networking features that collectively significantly enhance the performance of both the HetNet RAN and EPC compared to that of the typical PTP backhaul architecture in terms of handoff capability, signaling overhead, overall network throughput and latency, and QoS support. These include:

### 6.1 Ground Breaking EPC Offload Techniques

To appreciate the significance of the proposed offload techniques, it is important to first review the current offload mechanisms’ status. In general, traffic offload can be classified into two types: “RAN offload” and “core network offload”. RAN offload is implemented through the use of WiFi, femtocells and SCs. Note that femtocells and SCs are typically deployed as a means to increase capacity and improve coverage, rather than as an offload solution [3]. Generally, all IP traffic generated by/sent to a mobile device is routed to and through the mobile core network. However, because a majority of IP traffic is destined to best-effort Internet, it would be more cost-effective to divert this traffic away from the mobile core and offload it directly to the internet. This is the definition of core network offload.

The benefit of small cells in providing capacity where needed, is well understood. So are the challenges and solutions for managing the interference. Enhancements such as “Range Expansion,” introduced in LTE Advanced, increase the overall network capacity much more than what can be got by merely adding small cells. The interference management techniques of LTE Advanced make adding more small cells possible without affecting the overall network performance.
As shown in figure 6-1, Core offload is implemented through the deployment of internet offload gateways, which splits out traffic bound for the internet from the traffic bound for the operator’s core network including signaling [3]. Selected IP Traffic Offload (SIPTO) and Local IP Access (LIPA) are two solutions that 3GPP is standardizing for core network offload. The major downside of core network offload is that by diverting traffic from the core, the network operator has no longer any control over this offloaded traffic (e. g., to meter usage, bill for traffic), since these functions all reside in the core. As a consequence, mobility support for this offloaded traffic is rather limited. Note that core network offload is one form of “Internet offload” as Internet offload comes in several forms including, WiFi, femtocell, and core network.

Because EPC is designed to be access-independent, it can support the integration of both the LTE-A SCs and WiFi APs. However, the integration of WiFi APs, according to the EPC
standards for 3GPP and Non-3GPP interworking, depends on whether these APs are classified as “Trusted or Un-Trusted Non-3GPP Access Networks”. Trusted Wi-Fi Networks (see Fig. 6.3) mean that the WiFi APs are deployed and managed by the Operator, so that UE can connect to the WiFi network directly using the radio interface without requiring any additional security measures [4]. In contrast, as shown in Fig. 6.2, Un-trusted WiFi networks do not have any trust relationship to the operators, so that the operators require that the UE establish a secure tunnel (i.e. IPSec tunnel) to a trusted node in the operator core network. Typically, such a node is termed “Evolved Packet Data Gateway” (ePDG) in EPC networks [4]. Because the proposed PON-based architecture, which is used to backhaul both the LTE-A SCs and WiFi APs, is likely to be considered untrusted IP/Ethernet backhaul, IPSec termination will be needed. The ePDG is likely to be installed at the edge of the EPC to terminate and aggregate the high number of incoming tunnels/connections.

Figure 6-2: EPC Architecture for Access via Untrusted WLAN.
6.1.1 Significance of Local Mobile LAN Traffic

Local mobile LAN traffic is defined here as bidirectional multimedia traffic exchange (including VOIP, video, and best-effort traffic) between two mobile users served by two SCs or by a SC and the mBS that are attached to two different ONUs on the same ring (same PON domain). In the proposed backhaul RAN architecture, this traffic is directly routed on the ring from the source SC/mBS directly to the target SC/mBS and vice-versa as local LAN traffic, without the typical lengthy bidirectional re-routing from/to the SCs/mBS to/from the EPC. This is significant as the volume of VOIP calls and multimedia data exchange between local mobile users that are served by the same PON domain is substantial. Note that this traffic is still under the full control and management of the EPC. In a typical PTP LTE fiber backhaul, however, this traffic represents bidirectional US/DS data exchange between the two mobile users, which must be routed first from the source SC/mBS to the EPC (US traffic) and then from the EPC to the target SC/mBS (DS traffic), and vice-versa.
Thus, a substantial volume of mobile data traffic and associated signaling overhead as well as the lengthy and complex processing of this traffic (e.g., LTE bearers/mobility tunnels switch/set-up, retain, and tear-down and associated signaling commands from the SCs to the EPC and vice-versa) have been offloaded from the typically overloaded EPC to the access nodes (SCs/mBS) of the RAN. This has a significant impact on the performance of the EPC. First, it frees up a sizable fraction of the badly needed network resources as well as processing on the typically overloaded centralized serving nodes (AGW) in the EPC. Second, it frees up capacity on the typically congested mobile backhaul (from the SCs to the EPC and vice-versa). Third, the firmly held notion that the EPC’s control plane scalability might be a major stumbling block en-route to the realization of the 5G will be shown to be no longer precise.

This has a far-reaching implication, as it is clear that the SCs in the proposed RAN can now be deployed not only as a typical means to increase capacity and improve coverage, but also as an effective offload solution. This is significant as the proposed HetNet RAN is not only capable of providing the typical macro-cell offloading gain (RAN gain) but also can provide EPC offloading gain.

While both the proposed core network offload (EPC offload) and the typical core network offload (Internet offload) techniques ultimately provide EPC offloading gains, however, there are three significant advantages that distinguish the proposed EPC offload technique from that of the typical core offload: 1) since the offloaded IP traffic in the case of a typical core offload is only best-effort traffic, it would have then required almost no or slight processing at the EPC. However, since the offloaded IP traffic in the case of the proposed core offload is a mix of real-time and best-effort IP traffic, it would have then required much more processing at the EPC. Thus, for the
same amount of offloaded traffic, the EPC offloading gain provided by the proposed offload technique is significantly higher than that provided by a typical core offload. 2) The offloaded traffic in the case of the proposed core offload is still under the full control and management of the EPC. Thus, an efficient mobility control and traffic management can be supported. However, as explained above, this is in drastic contrast with a typical core network offload. 3) Implementing a typical core network offload requires the additional deployment of Internet offload gateways, which incurs additional cost and management complexity.

6.2 Mobility Management and Inter-Macro BS Handoff Capabilities

Seamless mobility that enables the support of VoIP and other real-time IP applications is one of the most important functionalities of the proposed converged architecture. The converged architecture must support seamless distributed handoff (HO) procedures that conform to the distributed nature of the LTE architecture. In LTE there is no soft handover support and at each HO the user context (defines the radio-bearer configurations) and the coupling between mobility tunnels and radio bearers need to be relocated from one eNB to the other. LTE defines three mobility-states of the UE, LTE-DETACHED, LTE-IDLE, and LTE-ACTIVE. In LTE-ACTIVE, when a UE roams between two LTE eNB cells, “backward” handover is carried out. Based on measurement reports from the UE, the source cell determines the target cell and queries the target cell if it has enough resources to accommodate the UE [5-8]. The target cell prepares radio resources before the source cell commands the UE to handover to the target cell.

Because data buffering in the downlink (DL) occurs at the eNB, mechanisms to avoid data loss during inter-eNB handoffs are more critical compared to the 3G architecture where data buffering occurs at the centralized RNC and inter-RNC handoffs are less frequent. The proposed architecture efficiently addresses this issue as described below. In this work, HO is classified into
two different scenarios, namely, intra-OLT HO and inter-OLT HO. The former is a HO between two neighboring eNBs (cells) that are located on the same ring and managed by the same OLT (same PON domain), while the latter is between two eNBs located on different adjacent rings, where each eNB is managed by a different OLT (each belongs to a different PON domain) but still managed by the same EPC.

6.2.1 Registration & Handoff

When a UE enters a domain served by a new PON-RAN, it needs to register itself to the new domain OLT’s access router and updates the new location in its home subscriber server (HSS). The new OLT initiates a location update request to the HSS indicating the change of location to a new OLT. As long as the UE is roaming within the same PON-RAN domain, it needs not to reregister again. The remaining procedures follow the typical LTE registration process.

6.2.2 Intra-OLT Handoff

The message sequence diagram of the intra-OLT handoff (HO) procedure between the source eNB1 and the target eNB2 is shown in Figure 6.4. The figure shows both the control plane signaling messages (solid arrows) and the flow of the user (data) plane packets (dashed arrows). The UE sends measurement reports to the source eNB (eNB1), which may decide on the execution of a HO based on these reports. The source eNB1 sends the coupling information and the UE context to the target eNB2 requesting the preparation of a HO (HO request context transfer). The target eNB2 performs admission control to check whether the established QoS bearers of the UE can be accommodated in the target cell.

Once eNB2 signals that it is ready to perform the HO (HO accept), eNB1 commands the UE to change the radio bearer to eNB2 (HO command). At the same time, to ensure seamless HO,
eNB1 suspends the RLC/MAC protocols and may start to forward the buffered service data units (SDUs) that have not yet been successfully sent to the UE along with all the incoming SDUs from OLT1, if there is any, toward the target eNB2. According to typical LTE standards, whether SDU forwarding is employed at all by the eNB is left as a vendor specific implementation detail. However, in the proposed converged architecture, it is a simple and straightforward procedure for the source eNB1 to forward the SDUs directly to the target eNB2 as a local LAN traffic on the ring, where the needed direct physical connectivity between them exists. However, in LTE, creating a logical connectivity between eNB1 and eNB2 requires the lengthy process of signaling to the MME/S-GW to coordinate the mobility-tunnel switch from eNB1 to eNB2.

Next, the UE sends the HO Complete message to the target eNB2, which is used by the target eNB2 to verify that it is the right UE that is accessing the target cell. At that point the target eNB2 can start sending DL data to the UE. For the HO to complete, eNB2 then signals OLT1 to inform it that the HO is complete (HO complete) and to update its records with the new eNB, i.e., to add ONU2/eNB2 to the forwarding list for the UE. This means that the scheduler at the OLT will just redirect the traffic destined to the UE from downstream Q1 (connected to dedicated downstream wavelength $\lambda_1$ serving ONU1/eNB1) to downstream Q2 (connected to dedicated downstream wavelength $\lambda_2$ serving ONU2/eNB2). After receiving the HO complete message, OLT1 first redirects UE’s traffic from Q1 to Q2 and then removes ONU1/eNB1 from the forwarding list of the UE. Then, OLT1 sends redirect traffic acknowledgement (ACK) to eNB2. Upon receiving the ACK, eNB2 triggers the release of resources at the source eNB1. Finally, OLT1 signals MME to update the UE’s new location.
Clearly, the proposed distributed ring-based unified PON-RAN architecture enables the support of a seamless distributed intra-OLT HO scheme (inter-eNBs) that has several additional significant features compared to the typical LTE’s inter-eNB HO scheme, including: 1) no path switch/setup command is needed since the path (mobility tunnels) from EPC to the UE remains unchanged; 2) the EPC is not involved at all except for the simple signaling from OLT1 to the MME to report the location update of the UE; 3) re-registration procedures to the HSS when the UE moves from eNB1 to eNB2 is avoided. It is also avoided as long as the UE roams within the coverage area served by the cells (eNBs) attached to the ring.

Overall, the proposed architecture significantly reduces the signaling overhead and handoff latency. Furthermore, the proposed HO scheme eliminates the lengthy process of the frequent registration and forwarding path setup, when the UE repeats crossing the boundary of two adjacent eNBs. Thus, with very small signaling overheads, the proposed architecture supports seamless and speedy handoff service for the mobile nodes when they roam in any PON-RAN domain attached to the EPC. In addition to directly routing on the ring the buffered SDUs that have not yet been successfully sent toward the target eNB from the source eNB during intra-OLT HO, all bidirectional upstream data exchange (including VOIP, video, and data sessions) between any two mobile users served by two different eNBs that are attached to the same ring is also directly routed on the ring from the source eNB directly to the destination eNB and vice-versa as local LAN traffic, without the direct participation of either the OLT or the EPC. This is significant as the volume of voice calls and/or multimedia data exchange between local mobile users is substantial. Consequently, a sizable fraction of the mobile path switch/setup signaling commands as well as actual local upstream traffic transport and processing are offloaded from the EPC to the access nodes (ONUs/eNBs).
6.2.3 Inter-OLT Handoff

The first 7 steps of the inter-OLT HO are almost identical with those of the intra-OLT HO shown in Figure 6-4. Starting from step # 8, as shown in Figure 6-4, the message sequence diagram of the inter-OLT HO procedure between the source eNBR1 located on the first ring and the target eNBR2 located on the second ring are different. Figure 6.5 shows both the control plane signaling messages (dashed arrows) and the mobility tunnels (solid arrows). First, the UE sends a registration request to the new OLT (OLT2) once it enters the new domain of the second ring. Next, OLT2 signals MME to coordinate the mobility tunnel switch from eNBR1 to eNBR2 and to initiate a location update request to the HSS indicating the change of location to a new OLT.

MME then triggers the update at the S-GW to switch the mobility tunnel, based on the signaling received from OLT2 via eNBR2 indicating that radio bearer was successfully transferred. Once the UE completes the registration to the HSS and new OLT (OLT2), S-GW will begin to forward packets for it through the new domain access root router at OLT2. At the
same time, the HSS notifies the old OLT (OLT1) to cancel the location process for this UE. As a result, the old OLT removes UE from its visitor list and releases its resources. Finally, eNB2 triggers the release of resources at the source eNB1.

6.2.4 Paging & Efficient Idle Mobility

In idle mode, according to LTE standards, the UE is in power-conservation mode and does not inform the network of each cell change. In this state, the location of the UE is only known at the MME and only at the granularity of a few cells, called the Tracking Area (TA). When there is a UE-terminated call, the MME knows the TA in which the UE last registered and paging is necessary to locate the UE to a cell. This approach, which registers to MME/HSS for idle nodes at every few handoffs, introduces significant signaling overheads and reduces the efficiency of EPC, especially when the idle node moves quickly. To further reduce the registration signaling overhead, the TA is redefined here to include all the cells (eNBs) attached to the ring (minimum of 16 cells versus 3-5 cells according to LTE standards). Thus, the idle UE sends a re-registration request to the new OLT when it only crosses a PON domain boundary. The new OLT records the idle UE in its paging list and reports the location update to the MME/HSS, but it does not allocate resources and does not set up a mobility tunnel for the idle UE.

To eliminate the paging signaling overhead, for every PON-RAN domain, the paging information is broadcasted periodically via the downstream Ethernet control frame associated with each wavelength channel. When the idle UE moves within the same paging domain, it only need to monitor the current paging information in the control frame and need not send any message to the OLT. If the new OLT receives data destined for the idle UE, it buffers the data in its cache and broadcasts a paging request message for the UE within its domain. Upon receiving the paging message, the UE reports its current location to the OLT, which then forward the data
to the UE. With the application of this paging scheme, the unnecessary signaling overheads and power waste, which are associated with the frequent re-registration for idle MN, can be significantly reduced.

Figure 6-5: Sequence of the inter-OLT handoff procedure between the source eNB and the target eNB

6.3 Enhanced Inter-Small Cell Handoff Capabilities

In LTE-A standards hard handoff (HHO) is mandatory. The HHO is a break-before-make procedure, in which LTE user equipment (UE) breaks its connections with the serving SC (SSC) before setting up new connections with the target SC (TSC) and this is when traffic interruption and packet loss take place. By exploiting both the distributed nature of the ring-based RAN architecture and the supporting control plane, the proposed architecture enables the support of seamless and speedy inter-SC HOs in which, as the simulation results will show, packet loss is almost totally avoided and VoIP and other real-time IP applications can be adequately supported during HO. This is accomplished as follows:

1) When a UE enters a domain served by the PON-RAN, it needs to register itself to the domain OLT’s access router and updates the new location in its home subscriber server (HSS).
As long as the UE is roaming within the same PON-RAN domain, it needs not to reregister again.

2) The physical connectivity among the both the SSC and TSC attached to the ring allows direct data exchange and intercommunications among them during HO (compare the simplicity and reduced latency and signaling overhead of this direct approach versus that of the typical 4G indirect bidirectional lengthy intercommunications and logical connectivity among the SSC and TSC via the EPC). Thus, once the TSC accepts the HO command, the SSC may immediately start to forward the buffered data (which have not yet been successfully sent to the UE), to the TSC directly on the ring as local LAN traffic. This is significant as creating the typical 4G logical connectivity among the SSC and TSC, which requires the lengthy process of signaling to the AGW to coordinate the mobility-tunnel set up/switch from the SSC to TSC (and vice-versa) via the EPC, is totally avoided as well as the direct participation of the AGW/OLT.

3) For the HO to complete, the TSC signals the OLT/EPC to inform it that the HO is complete and to update its records with the new TSC, i.e., to add TSC (and corresponding target ONU (TONU) that is collocated or attached with/to the TSC) to the forwarding list for the UE. Then, under the typical 4G RAN scenario, to resume normal operation and forward DS traffic to the TSC, the typical lengthy process of setting up a mobility tunnel form the EPC to the TSC is essential. Under the proposed PON-based RAN architecture, however, the scheduler at the OLT just simply redirects the UE’s DS traffic from the DS queue that was serving the SONU/SSC before the HO (the OLT houses N dedicated DS queues, each serving one of the N ONUs-SCs attached to the ring) to the new DS queue that is now serving the TONU/TSC. To further reduce the signaling latency and packet loss during the HO, the OLT may concurrently broadcast DS traffic destined to the UE to both the SSC and TSC.
Overall, the proposed EPON-based RAN architecture introduces several significant advantages versus that of a typical LTE/LTE-A RAN, including: 1) significantly reduces the signaling overhead and handoff latency; 2) offloading a sizable fraction of the local mobile sessions switch/set-up and tear-down and associated lengthy and complex signaling processing from the overloaded EPC to the RAN’s access nodes; 3) re-registration procedures to the HSS when the UE moves from a SC to another is avoided as long as the UE roams within the coverage area served by the SCs attached to the ring; 4) during inter-SCs HOs, no path switch/setup command is needed since the path (mobility tunnels) from EPC to the UE remains unchanged.
Chapter 7

7 Distributed Antenna System (DAS)
7.1 **Introduction**

Small cell techniques can remarkably improve frequency reuse factor and have been recognized as the best way to deliver high capacity in cellular communications. Reducing cell size implies increasing the number of cells, which typically leads to significant increase in hardware, operation, maintenance and installation costs. Currently, two innovative systems, femtocell and distributed antenna system (DAS, also known as remote radio head, RRH), have been developed and deployed, and enable cellular systems to efficiently reduce cell size.

A DAS is deployed by cellular operators. In DAS, the radio frequency (RF) components and antennas are located far away from BSs, and connected to the BSs typically by fibers using the radio over fiber (RoF) technique. One BS can have multiple such extended RRHs, and the signal processing is done centrally in the BS. A single large macrocell is equivalently divided to multiple smaller picocells, which can cooperate efficiently under centralized processing, and the network capacity can be increased significantly [2]. However, this requires dedicated deployment of the optical network and RRHs, and the cost significantly increases with the density of RRHs.

7.2 **Centralized Baseband Processing and Backhaul Network**

A radio base station can be functionally separated into

- Baseband Unit (BBU, sometimes also referred to as Digital Unit DU), which generates and processes a digitized baseband Radio Frequency (RF) signal
- Radio Unit (RU), which creates the analog transmit RF signal from the baseband signal and sources it to the antenna, and respectively digitizes the RF receive signal
With today’s Radio Base Stations, both units are integrated into a single network element. Figure 7-1 shows a scenario with overlapping cells in which the radio inter-cell communication is handled through the X2 interface [3].

![Diagram of radio base station architecture](image)

Figure 7-1: Small Cells and eNBs use X2 interface to communicate with each other.

Separating both units creates opportunities for network optimization. Figure 7-2 shows how the architecture is impacted by introducing a split radio base station [3]. The active radio frequency unit, which is called Remote Radio Head (RRH) is connected to the pooled digital units by means of a CPRI (Common Public Radio Interface) interface. This interface was specified by an industry cooperation with participation from Ericsson AB, Huawei Technologies Co. Ltd, NEC Corporation, Alcatel Lucent and Nokia Siemens Networks GmbH & Co. KG. It transports the digitized radio frequency signal as well as management and control data. The transmission network connecting RRH with BBU is called Backhaul network.
difference with the backhaul network, which connect the DUs with the edge of the evolved Packet Core (ePC) [3].

Small form factor Remote Radio Heads (RRH) simplify installation and reduce power consumption of active equipment at the antenna site. As the characteristic of the RF signal is generated at the collocated, pooled Baseband Units, a tight coordination of the radio signals is achieved. Besides the cost advantages, the improved interference management translates into a higher cell utilization as well as improved quality of service [3].

Figure 7-2: Connecting Remote Radio Heads with a pool of Baseband Units.

Optical Backhaul networks form basis for the next step of innovation towards software defined radio access networks, which can be upgraded from one radio technology to another simply by management command. As the CPRI interface does not depend on the radio technology, a upgrade from 3G to LTE or LTE-A only increases data rate in the Backhaul
transmission network. Bit rate transparent transmission allows a network upgrade without any impact on the transmission network.

Transmission between BBUs and the Remote Radio Heads will in most cases be done with fiber systems as data rates of several Gbit/s need to be transported and distances of up to 40km need to be bridged with low latency and low jitter in the range of 10ns. Copper and Microwave transmission systems might be an alternative in certain cases, however, both technologies come with some limitations which make a wider application quite unlikely.

Although the latest microwave transmission systems are capable of transporting data at multiple Gbit/s speed, restrictions on availability of spectrum and distance limitation at high frequencies, e.g., in the E-Band at 60/80 GHz, need to be considered. In addition, cost of scaling capacity is significantly less favorable with microwave transmission, making fiber-based solutions ideal. Copper is a theoretical option as well, however, it requires highly sophisticated vectoring and bonding technologies for achieving the required data rates. Distance limitations further reduce the relevance of this technology.

Although CPRI interfaces can be connected by grey interfaces and dedicated fiber, CWDM/DWDM will improve fiber utilization. As fewer fibers are used, cost for fiber provisioning is lower. Active C/DWDM technology can monitor the transmission network for fast and efficient fault isolation. Resilient optical transmission improves availability while optical switching allows implementing 1:N protection of BBUs [4].
7.3 Distributed Antenna System (DAS) Approach

In current LTE implementations the BS is co-located with Antenna tower containing the MIMO antennas [5] connected with electrical cables (Figure 7-4).

The distributed antenna System (DAS) is shown is Figure 7-5 [5] and connected to the same base station is more efficient to enhance the range and rate of LTE. Power amplifier and Radio
can be moved to each tower ONU making the base station simpler so that more towers can be connected with the same Base Station.

Figure 7-5: Base Station with Remote Radio Heads at ONU [Source 5]

The idea to connect BBU and RRH using PON based ring and Star configurations has been presented in [1] as shown in figure 7.6 & 7.7 respectively.

A DAS is deployed by cellular operators. In DAS, the radio frequency (RF) components and antennas are located far away from BSs, and connected to the BSs typically by fibers using the radio over fiber (RoF) technique. One BS can have multiple such extended RRHs, and the signal processing is done centrally in the BS. A single large macrocell is equivalently divided to multiple smaller picocells, which can cooperate efficiently under centralized processing, and the network capacity can be increased significantly. However, this requires dedicated deployment of the optical network and RRHs, and the cost significantly increases with the density of RRHs.
We can make additional significant usage of the proposed PON-based HetNet backhaul architecture by simply replicating the small cell with Remote Radio Head (RRH).

Overall, as shown in Fig. 7.8 [4], to reach the envisioned 1000x capacity goal, WDM transmission can easily be used scale to higher bandwidth by increasing the data rate of an
optical channel or by adding additional wavelengths. This allows expanding the capacity of a network without significant investment. Low fiber attenuation allows larger distances, which make it possible to further, centralize BBU pools and reduce the number of active sites in a network [3].

Figure 7-8: The envisioned PON-Based HetNet RAN Architecture to reach the 1000x capacity goal. [Source 4]
Chapter 8

8 Performance Evaluation & Simulation Results
In this section, we compare the performance of the proposed PON-based HetNet backhaul with that of the conventional fiber-based PTP HetNet backhaul. Two simulation programs were developed using event driven C++ along with simulation development environment using LTESim and OMNET [1,3]: one for the typical PTP HetNet backhaul and the other one for the PON-based HetNet backhaul. We consider the practical case of non-uniform traffic load in which, during a given period, some SCs/mBS might be lightly loaded/idle, while other SCs might be heavily loaded. At a given total network load, different SCs/mBS have different average traffic loads. Under this non-uniform traffic load scenario, the significance of utilizing PON-based HetNet RAN architecture is clearly established (Table 8.3). LTE-A and WiFi systems parameters used in the simulation are shown in Tables 8.1 and 8.2, respectively [2].
### Table 8-1: LTE Simulation Parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Radius</td>
<td>1 km</td>
</tr>
<tr>
<td>Scenario of HetNets</td>
<td>Macro only</td>
</tr>
<tr>
<td></td>
<td>Macro + 1 Small Cell</td>
</tr>
<tr>
<td></td>
<td>Macro + 4 Small Cell</td>
</tr>
<tr>
<td></td>
<td>Macro + 1 WiFi AP</td>
</tr>
<tr>
<td></td>
<td>Macro + 4 WiFi AP</td>
</tr>
<tr>
<td>Number of Ues</td>
<td>20</td>
</tr>
<tr>
<td>UE Distribution</td>
<td>(a) Uniform: User Equipments(UE) are randomly and uniformly distributed in the 1 km radius of Macro Cell.</td>
</tr>
<tr>
<td></td>
<td>(b) Hotspot: 15/20 Distribution. 15 Ues are associated to a hotspot cell and remaining are randomly and uniformly distributed within the macro cell.</td>
</tr>
<tr>
<td>Traffic Flows / UE</td>
<td>VoIP Flow, Video Flow and Best Effort Flow</td>
</tr>
<tr>
<td>Scheduler Type</td>
<td>PF= Proportional Fair Algorithm</td>
</tr>
<tr>
<td>FRAME STRUCT</td>
<td>FDD</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>20 Mhz</td>
</tr>
<tr>
<td>Channel Interference</td>
<td>No Interference</td>
</tr>
<tr>
<td>Number of RBs</td>
<td>50</td>
</tr>
<tr>
<td>Number of Subcarriers/RB</td>
<td>600</td>
</tr>
<tr>
<td>Antenna Configuration</td>
<td>2Tx / 2 Rx</td>
</tr>
<tr>
<td>Macro Tx Power</td>
<td>46 dBm</td>
</tr>
<tr>
<td>Pico Tx Power</td>
<td>30 dBm</td>
</tr>
<tr>
<td>UE Tx Power</td>
<td>23 dBm</td>
</tr>
<tr>
<td>Subcarrier spacing</td>
<td>15 kHz</td>
</tr>
<tr>
<td>Propagation Model</td>
<td>Path-loss: $128.1+37.6\log_{10}(d)$, where $d$ is the distance between the user and the two nodes in km</td>
</tr>
<tr>
<td></td>
<td>Penetration loss: 10 dB</td>
</tr>
<tr>
<td></td>
<td>Multi-path loss: Jakes model</td>
</tr>
<tr>
<td></td>
<td>Shadow fading: log-normal distribution with a mean value and standard deviation of 0 dB and 10 dB, respectively</td>
</tr>
</tbody>
</table>

### Table 8-2: Wi-Fi (Unlicenced) System Parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier Frequency</td>
<td>5.5 GHz</td>
</tr>
<tr>
<td>Antenna Gain</td>
<td>3 dbi</td>
</tr>
<tr>
<td>Antenna Configuration</td>
<td>AP = 2Tx / 2Rx</td>
</tr>
<tr>
<td></td>
<td>Client = 1 Tx / 1 Rx</td>
</tr>
<tr>
<td>Shadowing Model</td>
<td>Lognormal Sdev = 10 db for Wi-Fi AP-UE link</td>
</tr>
<tr>
<td>Penetration Loss</td>
<td>Fixed 20 db</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>20 MHz</td>
</tr>
<tr>
<td>AP Tx Power</td>
<td>24 dBm</td>
</tr>
<tr>
<td>Client Tx Power</td>
<td>18 dBm</td>
</tr>
<tr>
<td>Packet Size</td>
<td>1500 Bytes</td>
</tr>
<tr>
<td>RTS/CTS</td>
<td>None</td>
</tr>
<tr>
<td>Scheduler</td>
<td>Round-Robin</td>
</tr>
</tbody>
</table>
Table 8-3: Traffic Loading Scenarios

<table>
<thead>
<tr>
<th></th>
<th>1 Macro 4 Small Cell Scenario</th>
<th></th>
<th>Total Network Load</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.07</td>
<td>0.02</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>0.12</td>
<td>0.07</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>0.3</td>
<td>0.05</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>0.46</td>
<td>0.06</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1 Macro 1 Small Cell Scenario</th>
<th></th>
<th>Total Network Load</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.15</td>
<td>0.05</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.4</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.65</td>
<td>0.15</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.8</td>
<td>0.2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1.2</td>
<td>0.3</td>
<td>1</td>
</tr>
</tbody>
</table>

In this simulation, LTE-A evaluation methodology specified in 3GPP (Table 8.1) for a co-channel macro/small cell HetNet deployment is used. We use advanced receivers at the UE with interference cancellation to complement network based enhanced inter-cell interference coordination (eICIC) [2]. As part of the eICIC scheme, within the coverage of each macro cell, some subframes are exclusively used by small cells to serve UEs in each small cell's extended range, while other subframes are used by both macro cells and small cells.

In the subframes exclusively used by small cells, the macro cell does not transmit any traffic but still transmits the common signals (sync, broadcast and reference) and the UEs cancel this interference emanating from the macro cell using their advanced receiver capability. In the subframes used by the macro cell, the embedded small cells can still schedule UEs in each small cell's non-extended range. The partitioning is orchestrated by the macro cell by negotiating with the embedded small cells and the exact partitioning ratio adapts to the traffic pattern across macro and small cells in the network.
Figure 8-1 shows the actual physical layout of the simulated PON-based HetNet RAN architecture. As can be seen from Figure 8-1, a single macro cell covering a geographical area of 1 km radius is modeled with LTE-A macro BS (mBS) located at the center of the ring along with 4 Small SCs located at the horizontal and vertical axis of the ring at equi-distance from each other. Each SC is located 1 km from the center. The good coverage range of the mBS is assumed to cover a 750 m radius area. Figure 8-2 shows the actual physical layout of the simulated PTP typical RAN architecture.
The following are the system parameters used for simulating the PON-based HetNet RAN architecture: (1) a PON with 4 ONUs/SCs and one ONU/mBS; (2) aggregate access link data rate from the UEs to a given ONU/SC/mBS is 320 Mb/s; (3) the RAN DS line rate (from the OLT to the SCs) is assumed to be same as the US line rate (from the SCs to the OLT) and is equal to 1 Gb/s; (4) the average distance between the OLT and SCs is 20 km; (5) the buffer size in each SC/mBS is 1 Mbyte; (6) the IEEE 802.3ah MPCP REPORT/GATE message is 64 bytes; (7) the total mobile traffic is divided equally among US and local LAN traffic; (8) the DBA scheme reported in [10] is used here to provision the PON DS and US/LAN traffic, whereas the proportional fairness algorithm is used to provision the HetNet US traffic; (9) the mobile traffic modeled here uses the typical LTE CoSs (GBR, non-GBR) and are mapped into the EPON CoSs (EF, AF, BE); (10) the maximum EPON cycle time is assumed to 2 ms for US transmission, while a typical fixed periodic cycle of 10 ms is assumed for LTE US transmission (from the UEs to the SCs/mBS).
To have a fair comparison, all the PON-based RAN parameters listed above are also used for simulating the typical PTP HetNet backhaul except for the following: each and every dedicated fiber link data rate of the typical PTP backhaul in either US (5 dedicated point-to-point links between the SCs/mBS and the OLT) or DS (5 dedicated point-to-point links between the OLT and the SCs/mBS) direction is set to 200 Mbps. Thus, the aggregated link data rate in either direction is: $200 \text{ Mbps} \times 5 = 1 \text{ Gbps}$, which is equal to that of the PON-based RAN. The performance metrics used here are network utilization and average user throughput gains comparing HetNet with one macro BS along with either one or 4 low power small cells network over macro cell only network.

**WiFi Association Method**

In this simulation scenario, each UE stays in one location and is associated with either a macro cell or a Wi-Fi AP. A UE is offloaded from WAN and associated with a Wi-Fi AP whenever it can be served by the Wi-Fi AP with at least the lowest modulation and coding scheme (MCS) of Wi-Fi (6.5 Mbps for 802.11n) [2]. Once the UE joins the Wi-Fi network, it becomes a Wi-Fi client. Since the client has lower power (18 dBm) than the AP (21 dBm/antenna), the coverage range of Wi-Fi is typically limited by the uplink.

**Small Cell Association Method**

The association rule in the presence of small cells is based on the maximum downlink received power with a bias adjustable between 0 dB and 18 dB towards small cells [2]. This implies that the common signal C/I of a UE being served by a weak pico cell can be as low as -18 dB. If no such small cells are available, the UE will usually be served by the macro cell.

Figure 8.3 shows the simulation results for the average upstream (US) user throughput gain for the macro cell deployment, with either one or four small cells per macro cell, for the uniform
user distribution scenario. As can be seen from Figure 8-3, PTP HetNet offers little improvement for the average throughput gain over macro-only scenario. For instance, the average US throughput gain with one and four SCs is only about 20% and 80%, respectively, over the macro-only scenario. However, in the case of the proposed HetNet RAN architecture, the average US throughput gain with one and four SCs is almost doubled compared to that of the PTP scenario; about 40% and 160%, respectively, over the macro-only scenario.

Figure 8-3: Average user US throughput gain for the HetNet with one macro BS and either 1 or 4 small cells over macro cell only network, for the uniform user distribution scenario.

Figure 8.4 shows the simulation results for the average upstream user throughput gain for the macro cell deployment, with either one or four WiFi APs per macro cell, for the uniform user distribution scenario. In this case, as can be seen form Fig. 8.4, Wi-Fi APs provide little or no throughput improvement for both the typical PTP and proposed HetNets. For example, the gain with four Wi-Fi APs is only 9% mainly due to limited association range of Wi-Fi APs within the
large macro cell coverage. The limited range is due to 18 dBm transmit power of typical Wi-Fi clients and minimum MCS of 6.5 Mbps. However, as shown in figure 8-3, with four small cells, one can achieve 160% throughput gain over macro-only network, because LTE Advanced techniques (eICIC and IC) lead to expanded range of small cells.

Figure 8-4: Average user throughput gain for the HetNet with one macro BS and either 1 or 4 WiFi APs over macro cell only network, for the uniform user distribution scenario.

Figures 8.5 and 8.6 present the same simulation results obtained above in Figures 8.3 and 8.4, however, for a hotspot scenario. In the hotspot scenario, many UEs (15) are located in the vicinity of low power cells. A small cell can therefore offload a large number of UEs from the macro cell compared to the uniform user distribution scenario. LTE-A small cells provide even higher gains in this hotspot scenario. As shown in Fig. 8.5, in the case of the proposed HetNet RAN architecture, the average user throughput gain with one and four SCs is increased
significantly over the macro-only scenario. For instance, as shown in Fig. 8.5, four small cells deliver a gain of about 900% for average user throughput; this is a significant improvement over what can be realized using Wi-Fi APs. Adaptive resource partitioning between macro and small cells allows more resource and capacity allocated to small cells to serve large number of UEs in the hotspots.

Figure 8-5: Average user US throughput gain for the HetNet with one macro BS and either 1 or 4 small cells over macro cell only network, for the hotspot (15/20) scenario.

On the other hand, as can be seen from Figure 8.6, Wi-Fi APs can also offer significant throughput improvement in the case of a hotspot scenario. For example; four Wi-Fi APs can offer 230% improvement in terms of average user throughput gain.
Overall, the above simulation results indicate that the overall capacity of the proposed HetNet almost scales linearly with the number of deployed small cells, thanks to LTE-A’s advanced interference management techniques. For example, if there are 10 deployed outdoor small cells for every macrocell in the network, then the overall capacity will be approximately 11x capacity gain over a macro-only network as shown in Figure 8.7.

Figure 8-6: Average user US throughput gain for the HetNet with one macro BS and either 1 or 4 WiFi APs over macro cell only network, for the hotspot (15/20) scenario.
Figure 8-7: Overall average capacity gain of the proposed HetNet over Macro-only scales linearly with the number of deployed small cells.

Figures 8.8 shows the uplink utilization versus time at a given single network load of 0.8 for unevenly loaded Macro & HetNets, for both the typical PTP and proposed HetNet RAN architectures. The results demonstrate that the proposed HetNet RAN has higher utilization as well as stability with less variation with time compared to typical PTP HetNet. This enhances the network’s stability and predictability.
Figure 8-8: Uplink Utilization Time Series for unevenly loaded Macro & HetNets

Figure 8.9 shows the uplink utilization versus time at a given single network load of 0.8 for evenly loaded Macro & HetNets, for both the typical PTP and the proposed HetNet RAN architectures. The results demonstrate that the proposed HetNet RAN has higher stability with less variation with time compared to typical PTP LTE. This enhances the network’s stability and predictability.

Figure 8-9: Uplink Utilization Time Series for Evenly loaded Macro & HetNets
Figures 8.10 and 8.11 show DS throughput for two different local LAN/ upstream traffic loads of 50% and 100%, respectively. In the typical PTP LTE backhaul, DS throughput is badly impacted as the local upstream traffic is increased. This is because the local LAN/upstream traffic is typically re-routed back to the local UEs through the OLT/EPC as DS traffic and, thus, shares the network downlink capacity with native DS traffic originated from the EPC. On the other hand, DS throughput of the proposed PON-based backhaul is independent of local traffic. This indicates that the proposed HetNet backhaul architecture can also enhance the native network downlink capacity.

![DS Throughput Comparison with local LAN Traffic load upto 50%](image)

Figure 8-10: DS throughput with 50% of local LAN Traffic
Figure 8-11: DS throughput with 100% of local LAN Traffic

Figure 8-12 shows actual US throughput with no local LAN traffic. Figure 8-13 shows US throughput as the local LAN traffic is increased to 0.5. As can be seen from Figure 8.13, as the local LAN traffic increases, the US throughput is adversely impacted in case of typical PTP HetNet while in case of the proposed HetNet, US throughput remains independent of increasing local LAN traffic.

Figure 8-14 shows average packet drop, for both the typical PTP and the proposed HetNet RAN architectures, vs DS Load when one SC transmits high data rate (> 200 Mbps) to the other SC in the same ring. As shown in Figure 8.14, packet drop rate increases as the DS load increases in typical PTP HetNet, while in the case of the proposed HetNet packet drop remains independent of increasing DS load.
Figure 8-12: US Throughput with no local LAN Traffic

Figure 8-13: US Throughput with 0.5 local LAN Traffic
Figure 8-14: Average packet drop when one small cell transmits high data rate (> 200 Mbps) to the other small cell in the same ring.

Figure 8.15 shows the throughput versus time for a UE during HO when moving away from the source SC (SSC) that is attached to ONU1 and approaching the Target SC (TSC) that is attached to a neighboring ONU2 for both the typical PTP and the proposed HetNet RAN architectures. A unidirectional BE application traffic is configured between UE and the server at the rate of 64 Kbps. The UE has trajectory that starts moving around 120 seconds and converges to the TSC between 120 to 125 seconds. Same scenario is set up for both traditional PTP and the proposed HetNets RAN architectures. Parameters collected for comparison are the traffic received/dropped and HO latency. HO latency is computed from the time the SSC sends a Handover Request (HO_REQ) message to initiate the HO process until initial ranging with the TSC is successfully completed. As expected, the proposed HetNet RAN (X2 HO) shows lower
HO latency (15 ms versus 20 ms) and almost no packets drop as compared to typical PTP HetNet (S1 HO).

Figure 8-15: Traffic Throughput during Handoff
Chapter 9

9 Conclusion
This thesis has addressed the key techno-economics challenges facing the transition from current Fourth-Generation (4G) cellular technology to the 5G era to explore the potential and viability of cost-effectively implementing the 1000x capacity challenge. Specifically, this work has proposed and devised a novel PON-based HetNet mobile backhaul RAN architecture that: 1) holistically addresses the key techno-economics hurdles facing the implementation of the envisioned 5G cellular technology, specifically, the backhauling and signaling challenges; and 2) enables, for the first time to the best of our knowledge, the support of efficient ground-breaking mobile data and signaling offload techniques, which significantly enhance the performance of both the HetNet-based RAN and LTE-A’s core network (Evolved Packet Core (EPC) per 3GPP standard), ensure that core network equipment is used more productively, and moderate the evolving 5G’s signaling growth and optimize its impact.

To address the backhauling challenge, we have proposed a cost-effective fiber-based small cell backhaul infrastructure, which leverages existing fibered and powered facilities associated with a PON-based fiber-to-the-Node/Home (FTTN/FTTH) residential access network. Due to the sharing of existing valuable fiber assets, the proposed PON–based backhaul architecture, in which the small cells are collocated with existing FTTN remote terminals (optical network units (ONUs)), is much more economical than conventional point-to-point (PTP) fiber backhaul designs. A fully distributed ring-based EPON architecture is utilized here as the fiber-based HetNet backhaul.

It is shown that the purposely selected ring-based PON architecture along with the supporting distributed control plane enable the proposed PON-based FTTx RAN architecture to support several key salient networking features that collectively significantly enhance the overall performance of both the HetNet-based RAN and 4G LTE-A’s core (EPC) compared to that of
the typical fiber-based PTP backhaul architecture in terms of handoff capability, signaling overhead, overall network throughput and latency, and QoS support. It has also been shown that the proposed HetNet-based RAN architecture is not only capable of providing the typical macro-cell offloading gain (RAN gain) but also can provide ground-breaking EPC offloading gain.

The simulation results have indicated that the overall capacity of the proposed HetNet scales with the number of deployed small cells, thanks to LTE-A’s advanced interference management techniques. For example, if there are 10 deployed outdoor small cells for every macrocell in the network, then the overall capacity will be approximately 9x capacity gain over a macro-only network. To reach the 1000x capacity goal, numerous small cells including 3G, 4G, and WiFi (femtos, picos, metros, relays, remote radio heads, distributed antenna systems) need to be deployed indoors and outdoors, at all possible venues (residences and enterprises).

Overall, the proposed HetNet-based RAN architecture constitutes a complete cellular networking paradigm shift from the typically centralized RAN’s architecture and EPC-based NCM operations to a new disruptive fully distributed HetNet-based RAN’s architecture along with NCM operations in which substantial fraction of the typically centralized EPC-based NCM operations are migrated to and independently implemented by the HetNet access nodes (SCs/mBSs) in a distributed manner.
10 References

10.1 Chapter 1


Chapter 2


10.3 Chapter 3


10.4 Chapter 4


10.5 Chapter 5


10.6 Chapter 6


10.7 Chapter 7


10.8 Chapter 8

