5-2018

Performance Evaluation of v-eNodeB using Virtualized Radio Resource Management

Sai Keerti Teja Boddepalli
bskteju@gmail.com

PERFORMANCE EVALUATION OF V-ENODEB USING VIRTUALIZED RADIO RESOURCE MANAGEMENT

by

Sai Keerti Teja Boddepalli

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfilment of Requirements

For the Degree of Master of Science

Major: Computer Science

Under the Supervision of Professor Jitender S. Deogun

Lincoln, Nebraska

May, 2018

PERFORMANCE EVALUATION OF V-ENODEB USING VIRTUALIZED RADIO RESOURCE MANAGEMENT

Sai Keerti Teja Boddepalli, M.S.

University of Nebraska, 2018

Advisor: Jitender S. Deogun

With the demand upsurge for high bandwidth services, continuous increase in the number of cellular subscriptions, adoption of Internet of Things (IoT), and marked growth in Machine-to-Machine (M2M) traffic, there is great stress exerted on cellular network infrastructure. The present wireline and wireless networking technologies are rigid in nature and heavily hardware-dependent, as a result of which the process of infrastructure upgrade to keep up with future demand is cumbersome and expensive.

Software-defined networks (SDN) hold the promise to decrease network rigidity by providing central control and flow abstraction, which in current network setups are hardware-based. The embrace of SDN in traditional cellular networks has led to the implementation of vital network functions in the form of software that are deployed in virtualized environments. This approach to move crucial and hardware intensive network functions to virtual environments is collectively referred to as network function virtualization (NFV). Our work evaluates the cost reduction and energy savings that can be achieved by the application of SDN and NFV technologies in cellular networks.

In this thesis, we implement a virtualized eNodeB component (Radio Resource Management) to add agility to the network setup and improve performance, which we compare with a traditional resource manager. When combined with dynamic network resource allocation techniques proposed in Elastic Handoff, our hardware

agnostic approach can achieve a greater reduction in capital and operational expenses through optimal use of network resources and efficient energy utilization.

Our simulation results of the experimental prototype show that the proposed model can deliver up to 33% reduction in energy consumption and revenue increase of as much as 27% by taking advantage of dynamic pricing. In addition, we study the ability of the model to honor service level agreements and propose plans to better handle these agreements under peak network load.

ACKNOWLEDGMENTS

I am grateful to my advisor Dr. Jitender S. Deogun for his guidance and support throughout my graduate program. His wholehearted feedback and encouragement have always been the motivation for completing my research. I thank Dr. Vinodchandran N. Variyam and Dr. Ashok Samal for serving on my master's committee.

Many thanks to my friends at the University of Nebraska-Lincoln for their timely feedback and to Lakshmi Prabhala for her help in bringing the illustrations to life. I would like to take this opportunity to thank Saravanan Raju for his suggestions, help in this thesis and throughout my masters journey. I appreciate my team at Rehrig Pacific Company for their support during the course of completion of my thesis.

Lastly, I want to thank my family for their love and continuous support.

Contents

List of Abbreviations

RSSI	Received Signal Strength Indicator
IoT	Internet of Things
M2M	Machine to Machine
SDN	Software-defined Networks
NFV	Network Function Virtualization
BS	Base Station
eNodeB	Evolved Node B
RRM	Radio Resource Management
UE	User Equipment
OPEX	Operation Expenditure
CAPEX	Capital Expenditure
CRAN	Cloud based Radio Access Network
MIMO	Multiple Input, Multiple Output
LTE	Long Term Evolution
PSTN	Public Switched Telephone Network
IMTS	Improved Mobile Telephone Service
FCC	Federal Communications Commission
RCC	Radio Common Carriers
FDMA	Frequency Division Multiple Access
GSM	Global System for Mobile
CDMA	Code Division Multiple Access
ISDN	Integrated Services Digital Network
UMTS	Universal Mobile Telecommunications System
AMPS	Advanced Mobile Phone Service

DCS	Digital Cellular System
EDGE	Enhanced Data rates for GSM Evolution
OWA	Open Wireless Architecture
CR	Cognitive Radio
SoC	System on Chip
API	Application Programming Interface
VNF	Virtual Network Function
IP	Internet Protocol
WiMAX	Worldwide Interoperability for Microwave Access
VoIP	Voice over Internet Protocol
InP	Infrastructure Provider
SIM	Subscriber Identity Module
ARPU	Average Revenue Per User
SNR	Signal to Noise Ratio
CDP	Call Dropping Probability
CBP	Call Blocking Probability
SCBE	Spectrum Configuration and Bandwidth Estimation
MNO	Mobile Network Operator
MVNO	Mobile Virtual Network Operator
PDF	Probability Distribution Function

List of Figures

List of Tables

Chapter 1

Introduction

In the last two decades, cellular networks have revolutionized telecommunications and are responsible for connecting more people and devices than ever. The rapid advances in cellular network technologies have solved most connectivity challenges, have handled increase in the number of subscribers and large amounts of data traffic. At the end of the year 2017, the total number of mobile subscriptions reached 7.8 billion by adding almost 100 million new subscriptions in a span of three months. With the number of subscribers increasing and consumption of multimedia content, in particular video, data traffic rose by as much as 65% [2]. This trend is predicted to continue due to the penetration of cellular technologies in the developing world, demand for bandwidth intensive services such as video streaming, and the surge in machine-to-machine (M2M) traffic. There are also studies that predict the total number of devices connected to the Internet will reach 50 billion by 2020 from just 12.5 billion in 2010. The explosive growth in the number of cellular devices results in significant green house emissions due to substantial electricity usage of which up to 90% is consumed by the network infrastructure [3].

The cellular networks that are in existence today needs considerable design modifications to effectively handle the forecasted increase in user base or resulting high bandwidth demands and at the same time reduce energy utilization levels. Also, current networks are highly hardware-dependent making the process of upgrading to newer wireless access technologies both expensive and painstakingly slow. To address these challenges in cellular networks, researchers and carriers have been looking into virtualization of networks. As the adoption of virtualization technologies in addressing similar challenges in storage and computing in the information technology industry has had remarkable success, the research community is exploring cloudification of internet-based services with the goal of creating hardware independent, flexible, and efficient design of network services.

The recent advances in networking technologies are strengthening the case for virtualization of cellular networks through Software Defined Networks (SDN), Network Function Virtualization (NFV), and Cloud-based Radio Access Networks (CRAN). Apart from providing a flexible network architecture, we believe these virtualization technologies can decrease capital and operational expenses considerably. In this thesis, we delve into the idea of designing efficient cellular network components using SDN and NFV with the goal of increasing energy efficiency.

1.1 Motivation

For all the advantages they offer, current cellular networks are very rigid in nature and not energy efficient. With the increase in the adoption of Internet of things (IoT) and the addition of new users, the struggle for network resource is increasing. Cellular service providers have depended on procuring more frequency bands and efficient communication techniques such as multiple input, multiple outputs (MIMO), etc. to improve network performance till now. These Efforts alone cannot completely solve the issues of ever increasing cellular users and bandwidth requirements of modern applications. Apart from these challenges, a carrier should also invest considerable financial resources in the form of capital expenditure (CAPEX) for procuring new spectrum and operational expenses involved in upgrading specialized hardware in the face of decreasing revenue per user (RPU). The adoption of virtualization in technology infrastructure which include traditionally hardware-based devices such as firewalls, load balancers etc., has led to efforts to move crucial and hardware intensive network functions to virtual environments, collectively under the SDN and NFV initiatives.

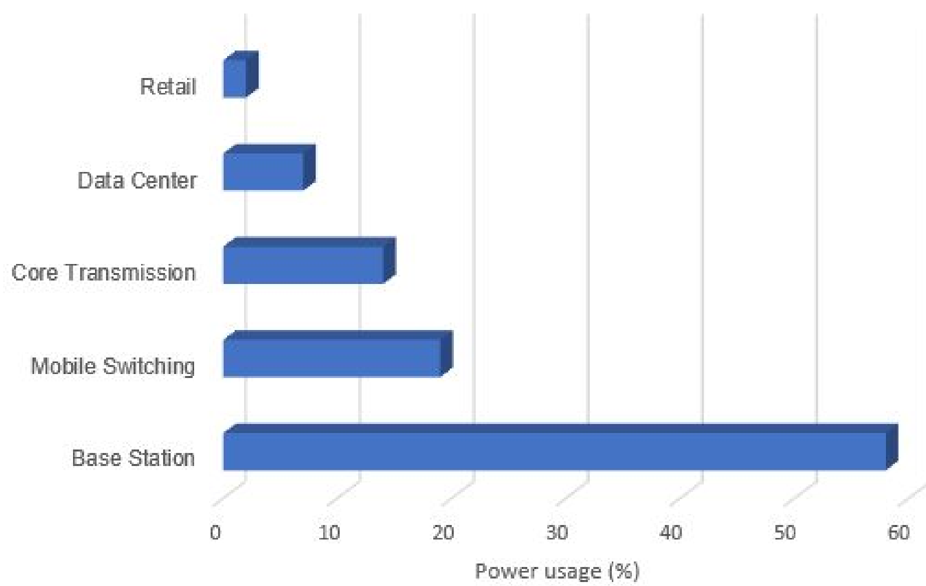


Figure 1.1: Energy utilization in current cellular network components

Another major concern related to current networks are not energy efficient and are responsible for approximately 3% - 5% of worldwide carbon emissions. In the developing world, where electricity grids are not very dependable, carriers use diesel-powered generator to power their networks. By studying the power usage statistics of various network components, it is evident that the base stations and access points are some of the most power hungry components accounting for as much as 60% of all energy consumed. The telecom towers, which includes the base station alone may consumes an estimated 2.5 billion liters of fossil fuel annually, leading to an approximate addition of 6.5 million metric tonnes of CO_2 emissions.

While a number of studies look at improving the power efficiency of various hardware components at the base stations ranging from efficient antenna to transreceivers, the question "*What if we can redesign the base station in cellular networks by applying the principles of SDN and NFV?*", is the major motivator of this research. In the current cellular network model, displacing some or all of these functions into modular hardware units is a very complex task. Hence, the motivation for this study is the application of virtualization to provide these services as software-based application that utilizes SDN and NFV will aid in adding agility to the cellular network infrastructure while also reducing carbon footprint.

1.2 Objective and Scope

The eNodeB which is the base station in the cutting-edge Long Term Evolution (LTE) cellular networking technology handles several key functions such as supporting flexible bandwidth allocation, radio-related functions, location management, handover functionality, etc. The main objective of this research is to realize a redesigned eNodeB by applying Elastic Handoff, SDN, NFV, and to quantify the overall efficiency that can be achieved by such a redesign. The following sub-objectives and their scope are derived from the main objective:

- To identify related work and research in the field of networking, cloud-based service model, SDN, and NFV technologies and their implementations in the area of cellular networks;

- To identify the relevant network technologies and architectures along with their major benefits, issues, and concerns;
- To identify the essential network functions of eNodeB to virtualize for the purpose of this research;
- To propose an implementation approach and its strategy to realize the main research objective;
- To demonstrate a proof-of-concept implementation of the proposed architecture;
- Finally, to validate the proposed implementation approach by an experimental performance evaluation.

1.3 Contributions

The contributions of this thesis include:

- *Spectrum Sensing Data Collection*: We capture, curate and study spectrum utilization data in the 2.4 GHz ISM band for a period of 30 days. This dataset with 5 million+ data points can be used for future research work.
- *Analytical Model to Perform Call Admission Control*: We propose a model and use proven mathematical approaches to study the state of the network based on three parameters: i) data rate the network can support, ii) a carrier's cost to service a subscriber, and iii) quality of service.
- *Implement Virtual RRM*: We implement and evaluate our proposed call admission process in a virtualized eNodeB component called Radio Resource Management (RRM) by applying SDN and NVF technologies.

- *Performance Analysis*: We study and compare total network utilization, cost of service and quality of service of our virtual eNodeB implementation against a traditional base station.

1.4 Structure

Chapter 2 presents the literature review of telecommunications in order to provide the necessary background information. This is required to understand the problems with current networks and fully appreciate the need for Software defined network and network function virtualization. Chapter 3 introduces the concept of virtualized Radio Resource management and takes a detail look the various sub functions of RRM. We also identify the sub functions that will be studied as part of this research. This chapter also provides the various design decisions for the implementation and simulations. This chapter is the most informative part of the thesis since it provides most of the information required to understand the simulations and analysis of results presented later in the thesis. Chapter 4 presents the various simulation setups, use-cases and discusses the results of various simulations. Finally, Chapter 5 provides a conclusion for the entire thesis with pointers for future work.

Chapter 2

Background

There are very few inventions that can boast to have driven the history of humans similar to what the telephone has. Having its origins in Greek and meaning 'communications at a distance', telecommunications has played a huge role in the development of human civilization and all the scientific advancements. Starting from fire (light) and horns (sounds) in ancient times to the ultra sophisticated wireless technologies of today, telecommunications has evolved rapidly in a very short span compared to the human history.

2.1 Wireline vs Wireless

From the days of smoke and horn signals, telecommunications technology had a jump start with the invention of electricity. Electric signals, with their literal lightening speed fueled experimentation of advanced telecommunication systems. While there were several proposals and competing technologies during this time, it was in the year 1843, Samuel Morse proposed an efficient and affordable way to communicate using sounds and what is today commonly termed as Morse's code. Evolving into telegraph, these inventions could only be used by trained personnel and were confined to offices and governmental institutions.

The big steps from telegraph to creating the technology to convert sounds to an electrical signal and vice-versa were accomplished in 1850. These technological breakthroughs lead to a race to actualize a machine that can transmit and receive long distance sounds (telephony). It was in the year 1857 Graham Bell independently built a prototype and patented his invention. With the invention of the telephone and consequent rapid adoption, the need for routes and circuitry also went up. Early telephone networks had exchanges where a person was responsible for connecting the two parties that were communicating. This solution for the challenge of routing was later dubbed as Public Switched Telephone Network (PSTN). This early solution to routing had the inherent issue of privacy, which lead to the development of automatic switching by as early as 1891. Years 1955 and 1956 saw the laying of a transatlantic telephone cable connecting New York and London. By 1969, almost 90% of US households had a telephone. This mighty technology still had a problem, the wire.

The initial proposal of total wire-free telephone by Bell System was made in 1947 with hexagonal cells for mobile phones in vehicles. These early examples required that the mobile phone stays within the coverage area of one base station for the duration of the entire call. There was no concept of handoff. The concepts of frequency reuse and handoff, as well as a number of other concepts that formed the basis of modern cell phone technology, were described in the late 1960s, in papers by Frenkiel and Porter.

Wireless networks offered users the mobility that wireline networks could not provide. Early cellular networks struggled with several shortcomings like limited range, bulky user equipment, etc. But with time, there were several technical breakthroughs such as, hexagonal cells, invention of integrated circuits, microprocessors and allocation of spectrum by governments towards cellular technology which made wireless cellular networks of today possible.

A cellular network today is distributed over a geographic area divided into multiple cells. Each cell is served by at least one fixed transceiver known as a base station (BS) that provides coverage to a specific cell, as a result, the operating frequencies can be re-used to serve a number of user equipment (UE) in each cell. In order to deal with the issue of mobility across cells, UEs are handed off from one base station to another seamlessly.

2.2 Evolution of Wireless

The concept of wireless cellular telephone system was first discussed in 1947 by Bell Labs' D.H.Ring and W.R.Young. In the US, Bell Labs and other Radio Common Carriers (RCC) later procured channels of the spectrum from the Federal Communications Commission (FCC). In 1948, these RCCs operated the first fully automated radiotelephone service, which eliminated the need for an operator in most cases. Bell system went on to build its first mobile phone technology called Mobile Telephone System (MTS).

The next breakthrough came in the form of the invention of the transistor which revolutionized not just the computing industry but telecommunications as a whole. The vacuum tubes used till then were replaced by transistors. The telecommunications industry and underlying technologies kept improving gradually and in the year 1958 Jack Kilby invented the integrated circuits at Texas Instruments, Dallas, which showed how different electronic components can co-exist together.

In 1964, the Bell Labs introduced the Improved Mobile Telephone Service (IMTS) to replace the aging MTS. IMTS did not require that users push to talk and it also allowed for direct dialing, automatic channel selection, and reduced bandwidth. In January 1969 the Bell System made commercial cellular radio operational for the first time by employing frequency reuse in a small zone system. This was realized by using six channels in 450 MHz bands again and again in nine zones between New York City and Washington, DC. Worldwide commercial cellular deployments blossomed in the 70s and early 80s. The subsections in this chapter trace the evolution of telecommunications technology from early telephone networks to the next-generation telecommunications technologies.

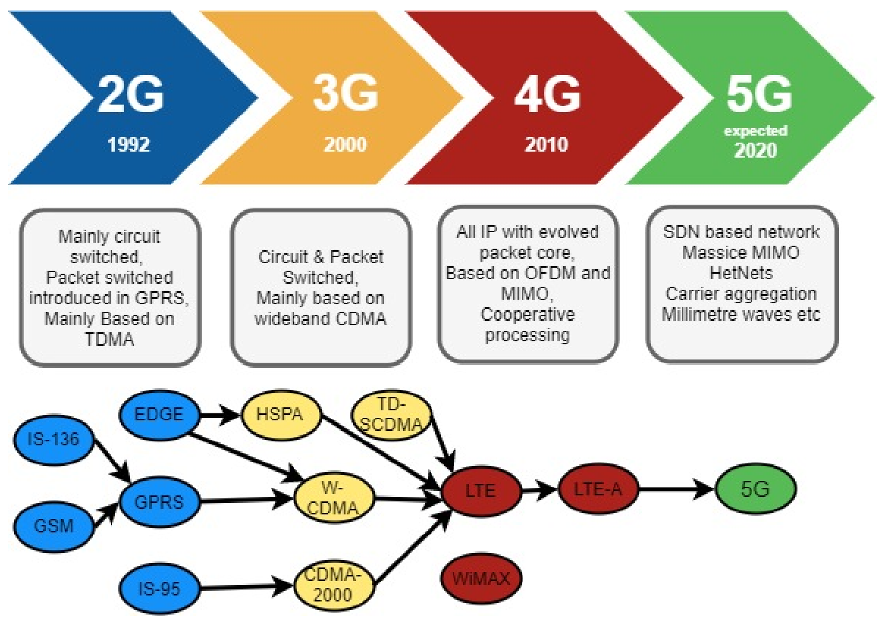


Figure 2.1: Evolution of Cellular Networks

The various cellular technologies were developed and rolled out in waves or generations.

2.2.1 1G - Analog Cellular

The first automatic analog cellular systems were deployed by Nippon Telegraph and Telephone (NTT) and first used in Tokyo in 1979 before gaining widespread adoption to the whole of Japan, and in the Nordic countries in 1981.

The first analog cellular system widely deployed in North America was the Advanced Mobile Phone System (AMPS). It was commercially introduced in the Americas in October 1983, Israel in 1986, and Australia in 1987. AMPS was a pioneering technology that helped drive mass market usage of cellular technology, but it had multiple issues. The conversations were not encrypted and easily vulnerable to eavesdropping via a scanner, it was susceptible to cell phone "cloning" and it used a Frequency-Division Multiple Access (FDMA) scheme and required significant amounts of wireless spectrum to support.

Many of the early commercial cell phones using AMPS were eventually superseded by Digital AMPS (D-AMPS) in 1990, and AMPS service was shut down by most North American carriers in 2008.

2.2.2 2G - Digital Cellular

In the 1990s, the 'second generation' mobile phone systems emerged. Two systems competed for supremacy in the global market: the European developed GSM standard and the U.S. developed CDMA standard. These differed from the previous generation by using digital instead of analog transmission, and also fast out-of-band phone-to-network signaling. The rise in mobile phone usage as a result of 2G was tremendous and this period also saw the introduction of prepaid mobile phones.

GSM first stood for Groupe Spciale Mobile, after the study group that created the standard. Its now known as Global System for Mobile Communications. It is the most successful mobile digital communications system in today's world, with networks in over 130 countries and more than 100 million users worldwide [4]. At the start of the 1980's, there were many competing standards amongst analog mobile networks in

Europe all based on similar standards, for example, NMT 450, however, running on slightly different carrier frequencies resulting in interoperability issues. To avoid this problem in the second generation mobile phone system, the Groupe Spciale Mobile (GSM) was created in 1982.

The primary goal of GSM was to provide a mobile phone system that would allow its subscribers to use their mobile phone in any European country i.e. make roaming possible. This system would have to provide voice services comparable with ISDN and other PSTN systems. While GSM was an important replacement for old analog first-generation system, it did not offer high data rates. GSM was initially deployed in Europe using the 890 - 915 MHz band for the uplinks and the 935 - 960 MHz band for the downlinks. This version is commonly known as GSM 900, and later versions as GSM 1800 (1710 - 1785MHz uplink, 1805 - 1880MHz downlink) which commonly known as DCS (Digital Cellular System). In 1991, the first commercial GSM network was launched in Finland. In 1989 Qualcomm successfully demonstrated a prototype Code Division Multiple Access(CDMA) cellular system to a group of 250 network operators and suppliers from around the world.

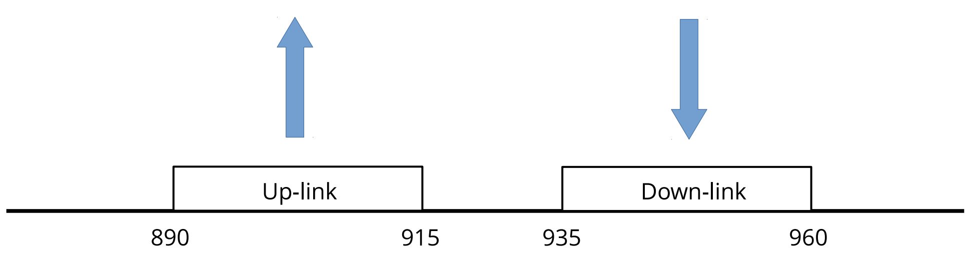


Figure 2.2: Up-link and Down-link frequencies for GSM

The second generation cellular technologies introduced a new variant of communication called Short Messaging Service(SMS) or commonly termed as text messaging. It was initially available only on GSM networks but spread eventually to all digital

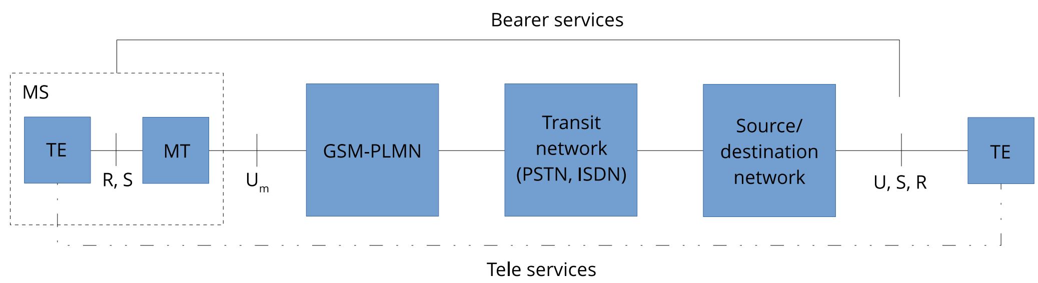


Figure 2.3: Mobile services in GSM

networks. The first machine-generated SMS message was sent in the UK in 1992 to followed in 1993 by the first person-to-person SMS sent in Finland.

In 1993, IBM Simon was introduced, which was possibly the world's first smart phone. It was a mobile phone, pager, fax machine, and a personal digital assistant all rolled into one. It included a calendar, address book, clock, calculator, notepad, email, and a touchscreen with a QWERTY keyboard. The IBM Simon had a stylus which was used to tap the touch screen with. Coinciding with the introduction of 2G systems was a trend away from the larger "brick" phones toward smaller form-factor hand-held devices. This change was possible not only through technological improvements such as more advanced batteries and more energy-efficient electronics but also because of the higher density of cell sites to accommodate increasing usage. The latter meant that the average distance transmission from phone to the base station shortened, leading to increased battery life while on the move.

By the mid-1990s,voice quality with every cellular radio scheme reached enough maturity and the interest data-centric services started gaining traction with system designers as a result of which while voice remained the essential service for the large majority of mobile phones, developing better and faster data networks over cellular radio became the priority.

2G introduced the ability to access media content on mobile phones. In 1998, the first downloadable content sold to mobile phones was the ring-tone in Finland. Mobile payments were experimented in 1998 in Finland and Sweden, where a mobile phone was used to pay for a Coca-Cola vending machine and car parking. Commercial launches followed in 1999 in Norway and the first commercial payment system to mimic banks and credit cards was launched in the Philippines in 1999 simultaneously by mobile operators Globe and Smart.

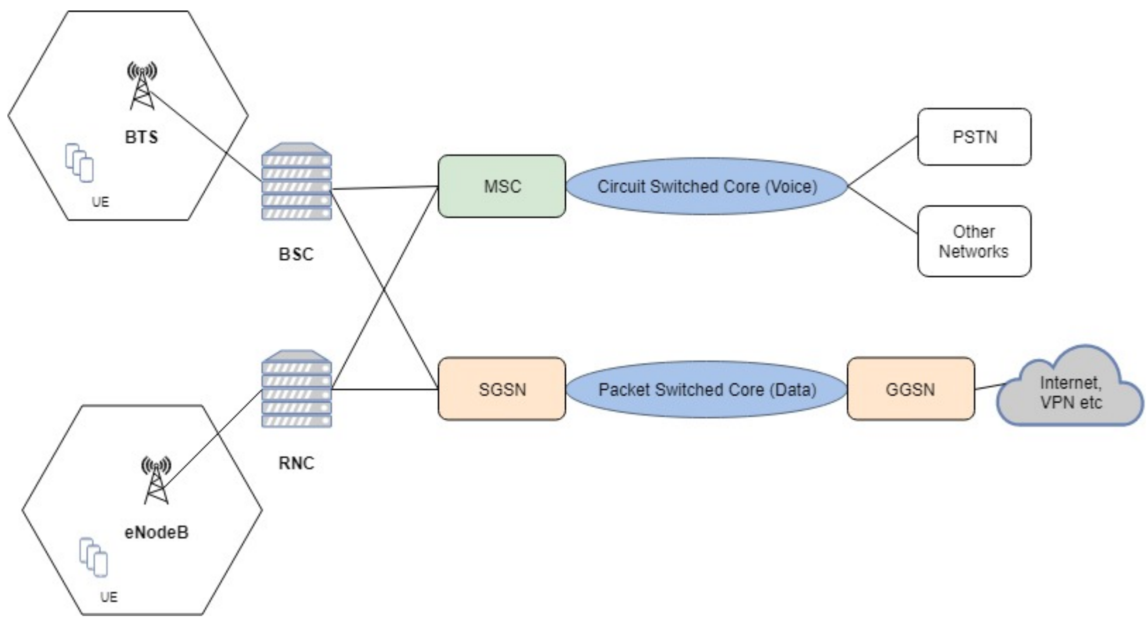


Figure 2.4: 2G/3G Network Architecture

2.2.3 3G - Mobile broadband

As the use of 2G phones became more widespread and people began to use mobile phones in their daily lives, it became clear that demand for data (such as access to browse the internet) was growing. Further, experience from fixed broadband services showed there would also be an ever-increasing demand for greater data speeds. As data is not efficiently handled by circuit switching, the need to transition from circuit to packet switching was deemed essential.

The internet became commercial in the mid-1990s with the creation of graphical browsers Mosaic and then Netscape. Internet user growth rivaled cellular telephony between 1995 and 2000. The internet runs on the aptly titled Internet Protocol or IP, a packet switching technique cellular data network operators quickly chose to adopt. All 3G systems use IP to provide high data speeds by utilizing packet switching rather than circuit switching for data transmission.

During the development of 3G systems, 2.5G systems such as CDMA2000 1x and GPRS were developed as extensions to existing 2G networks. These provide some of the features of 3G without fulfilling the promise of high data rates or full range of multimedia services. In Japan, the national carrier NTT DoCoMo launched the first commercial 3G network in 2001, using the Wideband Code Division Multiple Access (WCDMA) technology. In 2002, the first 3G networks on the rival CDMA2000 1xEV-DO technology were launched by SK Telecom and KTF in South Korea, and Monet in the US. The year 2003 saw eight commercial launches of 3G, six based on WCDMA and two more based on the EV-DO standard.

In the mid-2000s, an evolution of 3G technology began to take shape, named High-Speed Downlink Packet Access (HSDPA). It is an enhanced 3G mobile telephony communications protocol in the High-Speed Packet Access (HSPA) family, also coined 3.5G, 3G+ or turbo 3G, which allows networks based on Universal Mobile Telecommunications System (UMTS) to offer higher data transfer speeds and capacity.

The high connection speeds of 3G technology enabled a transformation in the industry: for the first time, media streaming of radio (and even television) content to 3G handsets became possible. By the end of 2007, there were 295 million subscribers on 3G networks worldwide, which reflected 9% of the total worldwide subscriber base. Although mobile phones had long had the ability to access data networks such as the Internet, it was not until the widespread availability of reliable 3G coverage in the mid-2000s that devices started accessing the mobile web.

2.2.4 4G - Native IP networks

The 4G systems were originally envisioned by the Defense Advanced Research Projects Agency (DARPA). DARPA selected the distributed architecture and end-to-end Internet Protocol (IP), and believed at an early stage in peer-to-peer networking in which every mobile device would be both a transceiver and a router for other devices in the network, eliminating the spoke-and-hub weakness of 2G and 3G cellular systems. In 4G systems, the circuit-switched infrastructure is abandoned and only a packet-switched network is provided, while 2.5G and 3G systems require both packet-switched and circuit-switched network nodes, i.e. two infrastructures in parallel. This implies that in 4G, traditional voice calls are replaced by IP telephony.

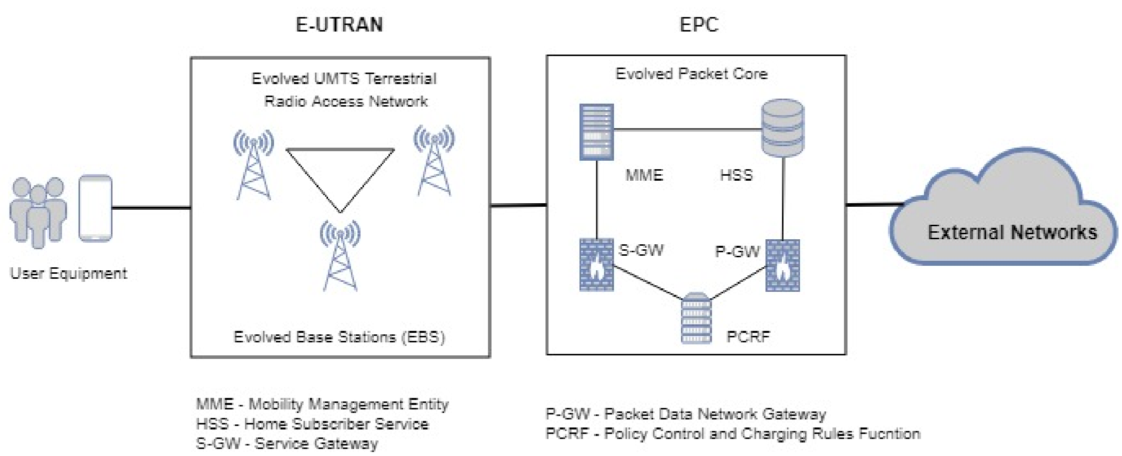


Figure 2.5: 4G (LTE) cellular network architecture

By 2009, it had become clear that, at some point, 3G networks would be overwhelmed by the growth of bandwidth-intensive applications like streaming media. Consequently, the industry began looking at data-optimized 4th-generation technologies, with the promise of speed improvements of up to 10-fold over existing 3G technologies. The first two commercially available technologies billed as 4G were the WiMAX standard (offered in the U.S. by Sprint) and the LTE standard, first offered in Scandinavia by TeliaSonera.

These 4G technologies eliminated circuit switching, instead employed an all-IP network. Thus, 4G ushered in a treatment of voice calls just like any other type of streaming audio media, utilizing packet switching over Internet, LAN or WAN networks via VoIP.

One of the key technologies for 4G and beyond is the Open Wireless Architecture (OWA) which supports multiple wireless air interfaces in an open architecture platform, one of which is Software Defined Radio (SDR) that promises radio convergence.

2.2.5 Beyond 4G

Pervasive networks are an amorphous concept where the user can be simultaneously connected to several wireless access technologies and can seamlessly move between them. These access technologies can be Wi-Fi, UMTS, EDGE, or any one of the many future access technologies. Included in this concept are SDRs and its next iteration, Cognitive Radio(CRs) technology, to efficiently manage spectrum use and transmission power with the goal of creating a pervasive network, which enables on-demand network access.

2.3 Hardware-Centric Radio vs SDR

A radio is an electronic component to send or receive data over a set of radio frequency. This simple definition of radio classifies almost all modern electrical appliances which are common form today as some form of radio.

Figure 2.6: Traditional network vs Software-defined network

In order to utilize a radio, there was a need for complex and expensive transmitter and receiver, a known frequency to operate on, and other intricate pieces of hardware like modulator, mixer, filter etc. Unlike in the past, most of these hardware components are abstracted thanks to the invention of modern electronic technology like ICs, microprocessor, etc. and large scale industrial design and hardware fabrication. Inspite of all these advancements, the dependence of radio on custom hardware for each application has not been eliminated and this hardware dependency is one of the reasons for slow and evolutionary nature of cellular technology. As traditional hardware based radio devices limit cross-functionality and can only be modified through physical intervention, it results in higher production costs and minimal flexibility in supporting multiple and/or new access technology standards.

SDRs addresses some of these problems by providing control over how the radio hardware functions can be softwarized. The advent of SDRs and CRs has also been boosted by the availability of inexpensive hardware which can be produced at scale leading to wider adoption but is yet to reach the volumes where major network equipment vendors would adopt.

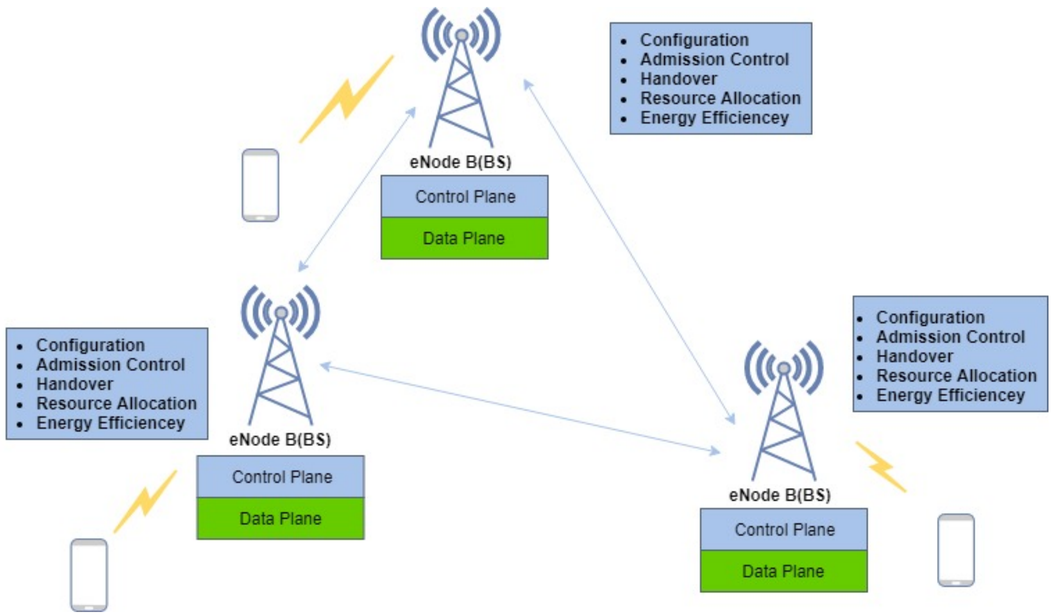


Figure 2.7: Control plane and data plane in current cellular networks

2.3.1 Benefits of SDR

SDR offers several compelling benefits to all the stakeholders, some of which are listed below [5]:

2.3.1.1 For Radio Equipment Manufactures and Systems Integrators:

- Implement a common platform architecture, allowing for the introduction of new radio technologies as radio "products" in the market.
- Promote software reusability across radio "products", reducing development costs dramatically.

- Over-the-air reprogramming, allowing "bug fixes" to occur while a radio is in service, thus reducing the time and costs associated with operations and maintenance.

2.3.1.2 For Radio Service Providers:

- New features and capabilities can be added to existing infrastructure without requiring major capital expenditures, allowing service providers to quasi-future proof their networks.
- Use of a common radio platform for multiple markets, significantly reducing logistical support and operating expenditures.

2.3.1.3 For End Users:

- Reduce costs in providing end-users with access to ubiquitous wireless communications enabling them to communicate with whomever they need, whenever they need and in whatever manner is appropriate.

2.3.2 Early adopters of SDR

There are several mainstream adoptions of SDR, some of which are:

- Thousands of software defined radios have been successfully deployed in defense applications.
- Cellular infrastructure systems are increasingly using programmable processing devices to create "common platform" or "multiband-multiprotocol" base stations supporting multiple cellular infrastructure standards.

- Cellular handsets are increasingly utilizing System on Chip (SoC) devices that incorporate programmable "DSP Cores" to support the baseband signal/modem processing.
- Satellite "modems" in the commercial and defense markets make pervasive use of programmable processing devices for intermediate frequency and baseband signal processing.

While the aforementioned types of systems are often not marketed as "SDRs, they utilize and benefit from SDR technologies to solve market specific problems such as; cost of development, cost of production, cost of upgrades and maintenance, time to market in supporting new and evolving air interface standards, or problems associated with network interoperability.

2.4 Need for NFV and CRAN

2.4.1 Network Function Virtualization

Network functions virtualization (NFV) is an initiative to virtualize network services traditionally run on proprietary, dedicated hardware. With NFV, functions like routing, load balancing and firewalls are packaged as virtual machines (VMs) on commodity hardware. Individual virtual network functions, or VNFs, are an essential component of NFV architecture.

Multiple VNFs can be hosted on a standard x64 server which can be monitored and controlled by a hypervisor. NFV's mission to use commodity hardware is important as it relinquishes network managers to no longer purchase and manually configure dedicated hardware devices in order to build a service chain that links specific functions in order to achieve a desired result. As NFV virtualizes network functions and eliminates function-specific hardware, network managers can add, move or change network functions at the server level in a simplified provisioning process. If a VNF running on a virtual machine requires more bandwidth, for example, the manager can move the VM to another physical server or provision another virtual machine on the original server to handle part of the load. Having this flexibility allows the networking staff to respond in a more agile manner to changing business goals and network service demands.

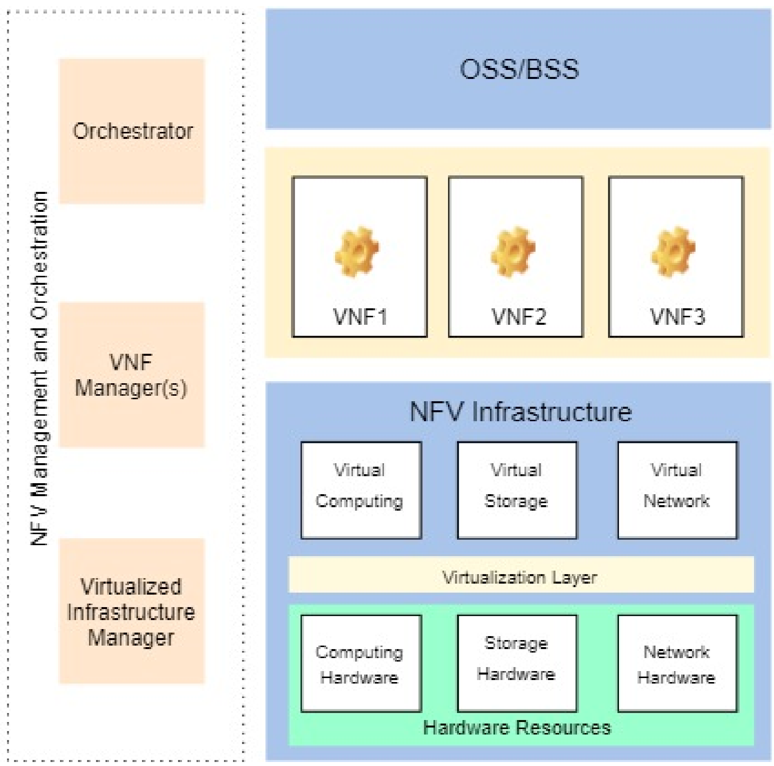


Figure 2.8: A simplified NFV architecture.

2.4.1.1 Benefits of NFV

The NFV concept was originally presented by a group of network service providers at the SDN and OpenFlow World Congress in October 2012. These service providers wanted to simplify and speed up the process of adding new network functions or applications. The European Telecommunications Standards Institute (ETSI) Industry Specification Group for Network Functions Virtualization proceeded to spearhead NFV development and standards.

While NFV can benefit enterprises, service providers have a more immediate use-case for it as it can improve scalability and better utilize network resources. If a service provider's customer requests a new function, for example, NFV enables the service provider to more easily add that service in the form of a virtual machine without upgrading or buying new hardware. In addition, NFV can reduce power consumption and increase physical space, since NFV eliminates most traditional hardware appliances reducing both CAPEX and OPEX.

2.4.1.2 NFV Challenges

NFV deployment has seen relatively slow progress due to a lack of standards in NFV management, automation and orchestration (MANO). MANO provides the framework for provisioning VNFs and managing NFV infrastructure. It also helps components within NFV infrastructure communicate with existing operational and billing support systems (OSS/BSS).

Part of the challenge for NFV adoption is the number of standards and open source projects being implemented to promote NFV development. For example, a short list includes ETSI, Open Platform for NFV, Open Network Automation Platform, Open Source MANO and MEF—formerly the Metro Ethernet Forum. With so many competing approaches—all backed by various service providers and operators—settling on an approach that offers usable capabilities for the whole industry is problematic. As a result, some service providers are uncertain which standards will be adopted and are more hesitant to invest.

2.4.1.3 Relation between NFV and SDN

Often confused as being the same, NFV and SDN are complementary technology initiatives. In essence, SDN is an approach to build data networking equipment and software that separates and abstracts elements of these systems. It does this by decoupling the control plane and data plane from each other, such that the control plane resides centrally and the forwarding components remain distributed. The control plane interacts both northbound and southbound. In the northbound direction the control plane provides a common abstracted view of the network to higher-level applications and programs using Application Programming Interfaces(APIs). In the southbound direction the control plane programs the forwarding behavior of the data plane, using device level APIs of the physical network equipment distributed around the network.

Thus, NFV is not dependent on SDN or SDN concepts. It is entirely possible to implement a virtualized network function (VNF) as a standalone entity using existing networking and orchestration paradigms. However, there are inherent benefits in leveraging SDN concepts to implement and manage an NFV infrastructure, particularly when looking at the management and orchestration of VNFs, and that's why multivendor platforms are being defined that incorporate SDN and NFV in concerted ecosystems.

An NFV infrastructure needs a central orchestration and management system that takes operator requests associated with a VNF, translates them into the appropriate processing, storage and network configuration needed to bring the VNF into operation. Once in operation, the VNF potentially must be monitored for capacity and utilization, and adapted if necessary.

All these functions can be accomplished using SDN concepts and NFV could be

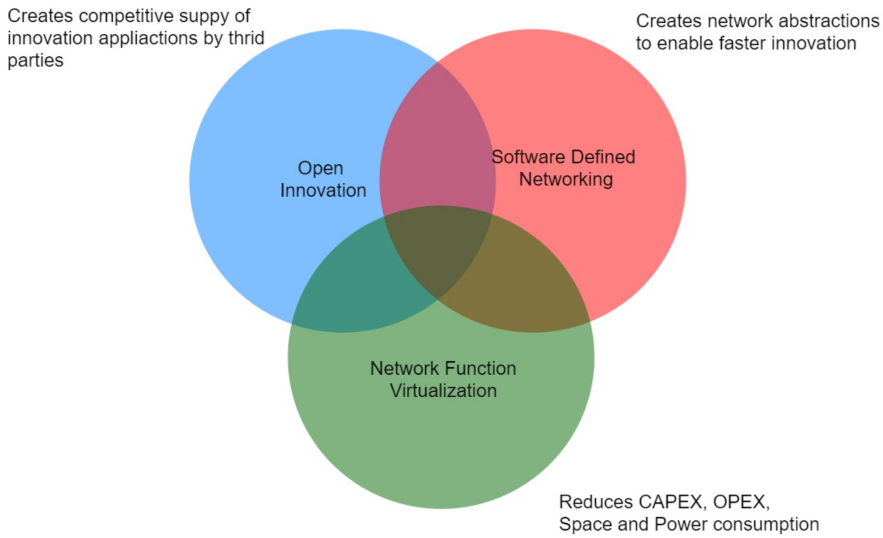


Figure 2.9: Relationship between SDN, NFV and Open Innovation (From [1])

considered one of the primary SDN use-cases in service provider environments. It is also apparent that many SDN use-cases can incorporate concepts introduced in the NFV initiative. Examples include where the centralized controller is managing a distributed forwarding function that could in fact be also virtualized on existing processing or routing equipment.

2.4.2 C-RAN

C-RAN (Cloud-RAN), sometimes referred to as Centralized-RAN, is a proposed architecture for future cellular networks. It was first introduced by China Mobile Research Institute in April 2010 in Beijing, China, 9 years after it was disclosed in patent applications filed by U.S. companies. Simply speaking, C-RAN is a centralized, cloud computing-based architecture for radio access networks that supports 2G, 3G, 4G and future wireless communication standards. Its name comes from the four C's in the main characteristics of C-RAN system, "Clean, Centralized processing, Collaborative radio, and a real-time Cloud Radio Access Network".

2.4.2.1 Evolution of Base Station Architecture

C-RAN may be viewed as an architectural evolution of traditional base station system. It takes advantage of many technological advances in wireless, optical and information technology communications systems. For example, it uses the latest CPRI standard, low cost Coarse or Dense Wavelength Division Multiplexing (CWDM/ DWDM) technology, and mmWave to allow transmission of baseband signal over long distance thus achieving large scale centralized base station deployment. It applies recent data centre network technology to allow a low-cost, high reliability, low-latency and high-bandwidth interconnect network in the baseband unit(BBU) pool. It utilizes open platforms and real-time virtualization technology rooted in cloud computing to achieve dynamic shared resource allocation and support of multi-vendor, multi-technology environments.

C-RAN architecture has the following characteristics that are distinct from other cellular architectures:

- **Large scale centralized deployment**: Allows hundreds of thousands of remote radio heads (RRHs) to connect to a centralized BBU pool. The maximum distance can be 20 km in fiber link for 4G (LTE/LTE-A) system, even longer distance (40 km 80 km) for 3G (WCDMA/TD-SCDMA) and 2G (GSM/CDMA) systems. There are reports that some operators in Asia have deployments of C-RAN systems with 1,200 RRHs centralized in one central office.

- **Native support of collaborative radio technologies**: Any BBU can talk with any other BBU within the BBU pool with very high bandwidth (10Gbit/s and above) and low latency (10μs level). This is enabled by the interconnection of BBUs in the pool. This is one major difference from BBU Hotelling, or base station hotelling; in the latter case, the BBUs of different base stations are simply stacked together and have no direct link between them to allow physical layer coordination.

- **Real-time virtualization capability based on open platform**: This is different from the traditional base station built on proprietary hardware, where the software and hardware are proprietary and provided by one single vendor. C-RAN BBU pool is built on commodity hardware such as, x86/ARM CPU based servers, with interface cards that connects fiber links to RRHs and inter-connection in the pool. Real-time virtualization ensures the resources in the pool can be allocated dynamically to the base station software stacks, say 4G/3G/2G function modules from different vendors according to network load. However, to satisfy the stringent timing requirements of wireless communications systems, the real-time performance for C-RAN is in the order of 10s of microseconds, which is two orders of magnitude higher than the millisecond level "real-time" performance usually seen in a cloud-computing environment.

2.4.2.2 Relation with NFV

C-RAN is considered as one of the applications of NFV. ASOCS Ltd has recently announced development of a full-virtual base station. C-RAN can also enable cell virtualization, a technique for virtualizing wireless spectrum resources. As one of the promising evolution paths for future cellular network architecture, C-RAN has attracted high academic research interest. Some of these research topics are discussed in Section 2.5.

2.5 Research on SDN, NFV and CRAN

The distributed nature of control plane in current cellular networks is not optimal for coherently managing spectrum, radio resources, call handoff, and other essential network functions in a way to manage future network demands. SDN-based approaches present opportunities in making the process of deploying and managing cellular networks that are traditionally not turnkey. These features includes seamless mobility through efficient handover procedures [6], [7], load balancing [8], [9], creation of on-demand virtual access points (VAPs) [8], [10], downlink scheduling (e.g., an OpenFlow switch can do a rate shaping or time division) [10], dynamic spectrum usage [10], enhanced intercell interference coordination [10], [7], shared wireless infrastructures [11], seamless subscriber mobility and cellular networks [7], QoS and access control policies made feasible and easier [7], [12], and easy deployment of new applications [8], [9], [11]. Some of the recent initiatives in the area of network virtualization, include XBone and Tempes focusing on networking technology; UCLP, VNET, AGAVE and VIOLIN focusing on layers of virtualization; VNRMS, NetScript, Genesis and FEDERICA focusing on architectural domain and management; and PlanetLab, GENI, VINI, CABO, 4WARD and NouVeau focusing on the granularity of virtualization; and VITRO focusing on virtualization of wireless sensor networks.

2.5.1 Virtualization in Cellular Networks

With respect to cellular networks, the concept of virtualization can be traced back to [13], where the basic framework of virtual node and virtual radio was proposed. The decoupling of *data plane* and *control plane* is defined such that multiple protocols and management strategies are able to run at different virtual nodes and links. Due to the popularity of LTE, most approaches on wireless network virtualization discussed in cellular networks are based on 3GPP LTE systems. The authors of [14] investigate virtualizing eNodeB in 3GPP LTE and point out that virtualizing the eNodeB is similar to node virtualization that has a number of solutions. However, the issue of resource allocation, isolation and energy efficiency in cellular networks were not explored in [13] which this research as a first step attempts to address.

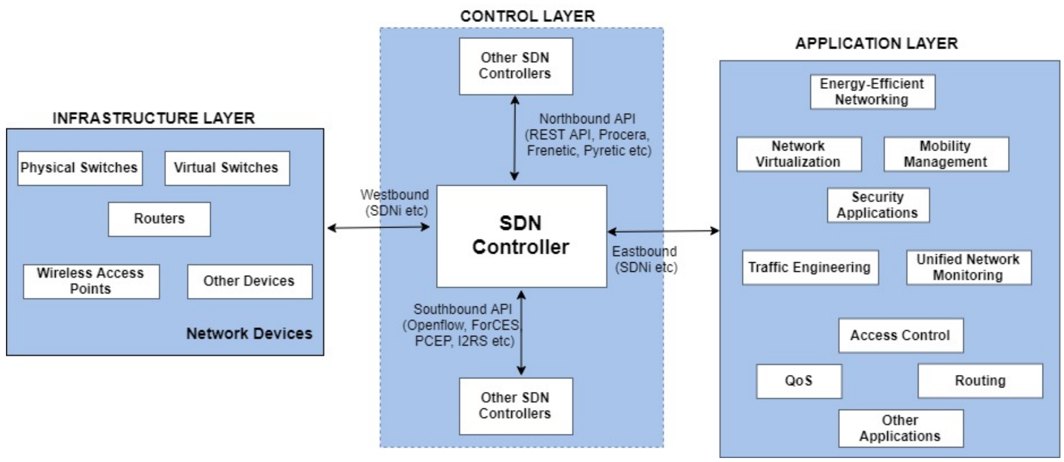


Figure 2.10: Instance of LTE Virtualization

As depicted in Fig. 2.10, a hypervisor [15] is physically added to the LTE eNodeB and logically placed between physical resource and virtual eNodeB. The LTE hypervisor takes the responsibility of virtualizing the eNodeB into a number of virtual eNodeBs, such as virtual machines (e.g., CPU, memory, I/O devices, etc.), and spectrum, which can be used by different service providers (SPs) or mobile virtual network operators (MVNOs). Moreover, the LTE hypervisor is also responsible for scheduling the air interface between eNodeBs and user equipment resources. There are two proposed entities in the hypervisor, one of which is the Spectrum Configuration and Bandwidth Estimation (SCBE) module that logically is stored in each virtual eNodeB, and the other one is Spectrum Allocation Unit (SAU) that is logically located in the hypervisor. To estimate the requirement of spectrum at virtual eNodeB is one of the main functions performed by SCBE. From each virtual operator, at frequent time intervals, spectrum bandwidth estimation is calculated by SCBE based on Exponential Moving Average and sent back to the SAU of the hypervisor which is then used for PRBs scheduling. Another main function of SCBE is to configure the spectrum where each virtual eNodeB operates.

The air interface resources that the hypervisor schedules are actually the Physical Radio Resource Blocks (PRB). Since the PRB is the smallest unit that the LTE MAC scheduler can allocate to a user, scheduling PRBs among virtual eNodeBs implies splitting the spectrum among various virtual eNodeBs. To split the spectrum for multiple virtual eNodeBs, the hypervisor has to schedule a number of PRBs based on a set of criteria which includes bandwidth, data rates, power, interference, predefined contracts, channel conditions, traffic load. The SAU is used to schedule air interface through a contract-based hypervisor algorithm to divide the spectrum among virtual eNodeBs based on these four pre-defined contracts: 1) Fixed guarantees where assured spectrum and bandwidth will be allocated 2) Dynamic guarantees where PRBs are allocated according to the requirements of virtual eNodeB limited by an upper bound 3) Best effort with minimum guarantees where essential bandwidth will be allocated and additional bandwidth may be added in a best effort manner 4) Best effort with no guarantees where bandwidth is allocated only on a best effort manner. The approach presented in [16] is a practical and integrated mechanism to realize virtualization in LTE-based RANs. Nevertheless, there are still some aspects that need to be improved, including control signaling, isolation among virtual eNodeBs, and upper layers virtualization.

A mechanism that is similar to [14] is used in [17], [18], and [19] to address multiple issues: The multiplexing gain by eNodeB virtualization is investigated in [17] from an analytical perspective and a generalized multiparty model is proposed to enable

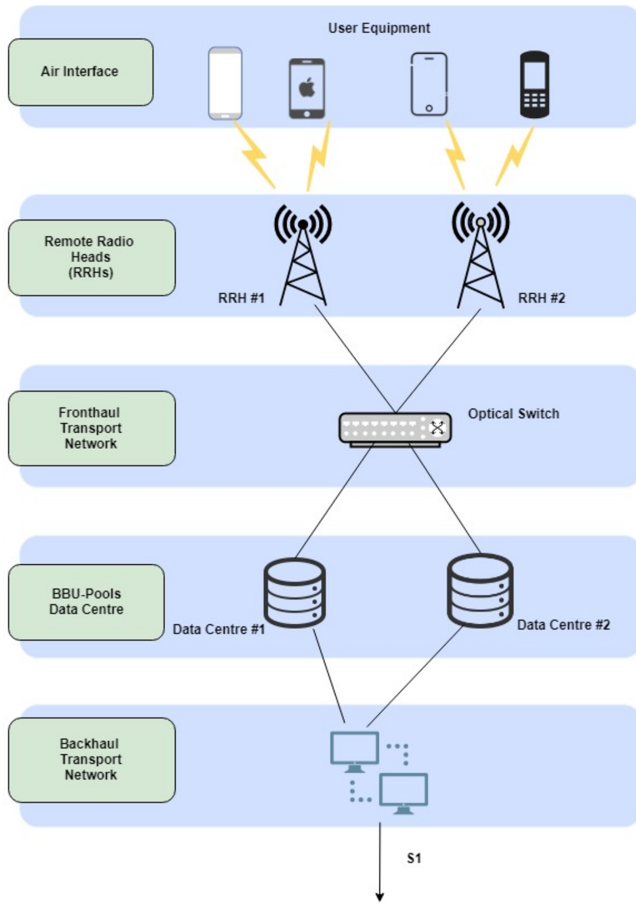


Figure 2.11: C-RAN Architecture.

centralized spectrum sharing and spectrum budget estimation for real-time services. In [18], load balancing techniques are proposed and embedded to the framework proposed in [14]. Using the dynamic load balance mechanism, heavy-load virtual eNodeBs can offload the excessive traffic to a low-load virtual eNodeB, which brings a significant gain to user performance. Unlike others approaches, [19] introduces a bankruptcy game into the resource allocation of LTE virtualization. Assuming that both big (higher traffic load and rate requirements) and small (lower traffic load and rate) MVNOs coexist, the PRBs owned by the infrastructure providers are limited and scarce such that the required PRBs are less than the available PRBs. Thus, the authors model the Infrastructure providers (InP), who own the PRBs, and MVNOs as the bankrupt company and players in the game, respectively. By solving the game, the InPs guarantee relative fairness among operators when allocating the PRBs.

OpenFlow, as a technology to realize SDN, is also introduced in LTE virtualization by [20]. [20] provides architecture level solution on LTE virtualization of eNodeB. The eNodeB is sliced into a number of virtual eNodeBs by FlowVisor [21] policy, and the same number of controllers are created to assign to the corresponding SPs or MVNOs. When one SP sends information to their virtual eNodeB, FlowVisor stops the traffic and maps it to the allowed resource based on the policies set at the eNodeB. Similarly, the FlowVisor only forwards the traffic originating from eNodeB to the respective controller, which is operated by the SP whose flowspace matches this traffic. Thus, the SP does not realize that the eNodeB has been sliced to multiple virtual eNodeBs.

The concept of SDN is applied to C-RAN [22] and contains three main parts: wireless spectrum resource pool (WSRP), cloud computing resource pool (CCRP) and SDN controller. In this architecture, WSRP consists of multiple physical remote radio units (pRRUs) distributed at various locations and it virtualizes one pRRU into several virtual RRUs (vRRUs) with different access protocols (GSM, UMTS, or LTE) coexisting in one shared pRRU. CCRP is comprised of a large amount of physical processors constructing a high speed cloud computing network and virtualized to virtual BBUs and virtual BSCs. The WSRP and CCRP create several complete virtual RANs and the SDN controller takes the responsibility of the control plane of this heterogeneous RANs.

Chapter 3

Design of Energy Efficient eNodeB

The issue of energy consumption in cellular networks has led to numerous research efforts whose goal is to reduce eNodeB energy usage by looking into power management as a primary means for energy saving. Limiting of power transmission to reduce both the amount of interference and energy consumption is discussed in [23]. However, the applicability of this approach by reducing transmission power of eNodeB will impact the quality of service a carrier can offer. In [24], dynamic power control during a period of low-load, such as, night time is recommended while ensuring full-load at all times. Another dominant energy saving technique for eNodeB is energy-aware cooperation, where significant fluctuations of traffic load in cellular networks in both space and time due to changing user behavior is considered.

In this thesis, we propose virtualized eNodeB as a solution to achieve energy savings in cellular networks. We implement and study a virtualized radio resource management (RRM) component in an NFV and C-RAN-based network infrastructure. The following section describes RRM in a traditional eNodeB and highlights the sub-functions we choose to virtualize for the purpose of this research.

3.1 Review of Radio Resource Management

RRM is is concerned with multi-user and multi-cell network capacity issues and it crucial for network operators to meet their subscribers QoS requirements. It can increase system capacity by an order of magnitude by utilizing processes and algorithms to efficiently use radio resources, while serving subscribers according to configured QoS parameters.

RRM can adjust network parameters adaptively based on traffic load, users location, QoS requirements, among other factors, to ensure radio resources are efficiently used. The functions that RRM provides can be broadly classified into:

- **Admission control**, which decides whether to serve or deny a new network access connection. As servicing new connections should not affect the guaranteed QoS of any active sessions, this control mechanism investigates the prevailing resource situation, QoS requirements for the new bearer, as well as the priority levels and the currently provided QoS of any active sessions in order to make a determination of whether to accept any new network access request. Any new request is admitted up until such time the packet scheduler in the cell can arrive at a feasible solution. Admission control may also trigger handover procedures, transport channel type switching, etc.; the procedure may differ based on the function triggering it. In addition, admission control algorithms has to take user and/or service priorities into account. In case, there are no more resources available to admit a new call or connection, algorithms may decide to assign more weight on some users or services.

- **Packet scheduling and link adaptation** is used to maximize cell capacity, ensure minimum QoS requirements for bearers is fulfilled and that there are adequate resources for best effort bearers. Besides, link adaptation provides information to the packet scheduler on the supported modulation and coding scheme for a user depending on the selected bit rate. These techniques can be used for resource scheduling, where each have their own characteristics, advantages and drawbacks. Some of the common scheduling algorithms include Maximal-Rate Scheduling, Round-Robin Scheduling, and Proportional Fairness Scheduling.

- **Power control** manages transmit power to provide a better Signal-to Interference-plus-Noise Ratio (SINR) at the receiver. A high transmit power causes more interference to other cells or users and also increases energy consumption. Taking dynamic channel properties into account, including channel attenuation, noise and interference level at the receiver side, power control is crucial in radio networks.

- **Handover** strategies in RRM maximizes the number of active call sessions over a set of cells. Although handover (hard handover) is an inherent procedure in cellular networks, CDMA introduced new handover types which enables UEs to be connected to more than one BS simultaneously i.e. soft handover.

- **Radio Access Technology (RAT) selection** is needed in order to realize RRM over heterogeneous access networks and it includes initial network selection followed by vertical handover (VHO). In case two or more RATs are co-located in the same coverage area, it is expected that the RAT selection algorithm maximizes system performance and QoS by allocating users to the most suitable RAT.

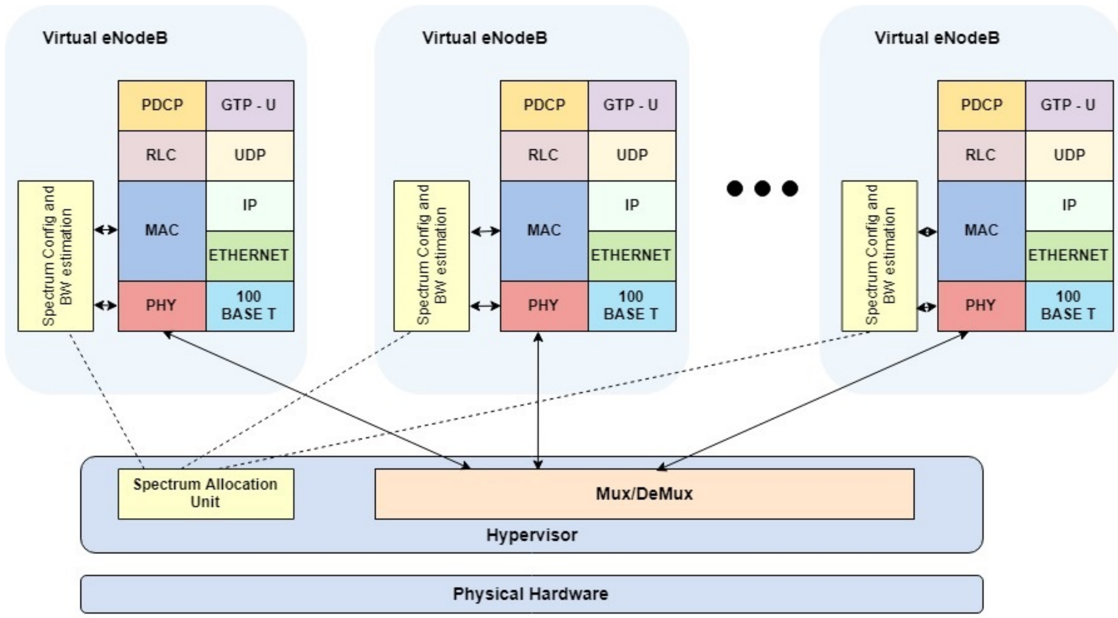


Figure 3.1: Architecture of a virtual eNodeB—Multiple virtualized instances of eNodeBs are concurrently hosted on commodity hardware and are logically separated through a hypervisor.

3.2 Design Choices for Implementation

Of the various functions of a RRM listed in the previous section, we choose to implement and study Admission Control. The various design choices for the implementation are listed and explained below:

3.2.1 Choice of Access Technology

Our proposed model is based on an all-Internet Protocol (IP) LTE access technology. LTE offers these basic advantages:

- LTE represents a significant shift from legacy mobile systems as the first all-IP network technology that will impact the way networks are designed, deployed and managed.

- With the long-term evolutionary access technology that is LTE, the 3GPP (Third Generation Partnership Project) systems are well positioned to remain competitive for decades to come.
- LTE also meets current market needs for improved performance at reduced cost particularly in the face of anticipated increase in M2M subscriptions.

Apart from these advantages offered by LTE, our design decision is also based on the quantum of previous and ongoing research being carried out in the area of SDN, NFV and C-RAN towards their adoption in cellular networks.

3.2.2 Network Infrastructure Sharing Solution

The case for infrastructure sharing in cellular networks was made several times in the last few chapters of this thesis. With adoption of virtual-RAN, a carrier can choose to be either a hardware operator and service provider or just a service provider. The terminology hardware operator refers to an independent hardware operator who is not in the business of providing network access service to subscribers. On the other hand, a Mobile Network Operator (MNO), provides service to both normal subscribers and other network operators who do not own any hardware. In such a model not all network operators own and manage physical hardware.

For the purpose of this thesis, our model treats network sharing as a combination of spectrum and infrastructure sharing. As a result, both radio resource and network infrastructure can be shared among multiple network operators based on mutual access agreements.

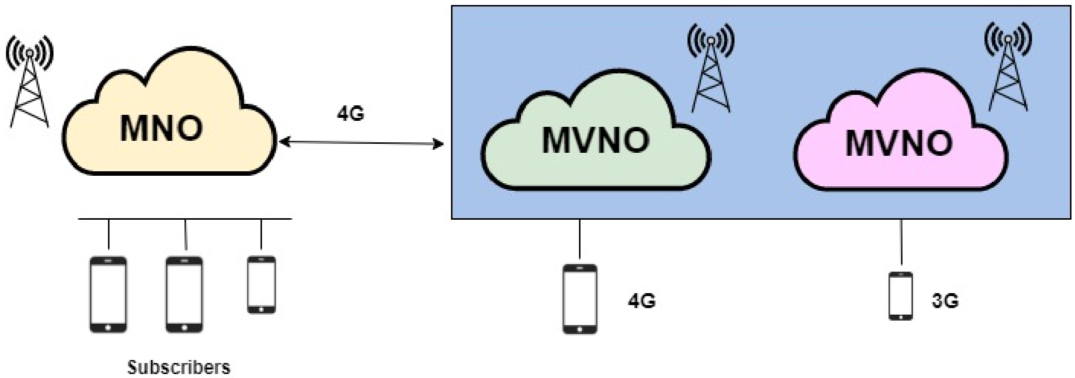


Figure 3.2: Network sharing model (MNO & MVNO)

3.2.3 Type of Network Users

The various types of network users considered for the purpose of our thesis are as follows:

- **Mobile Network Operators (MNO)** are carriers who own and manage the physical hardware required for operating a cellular network. In a SDN-based cellular network, this hardware is generic and most network functions are virtualized. This allows the MNO to offer network-as-a-service to other carriers who ar interested in providing services to consumers without having to manage their own hardware.

- **Mobile Virtual Network Operators (MVNO):** MVNOs are carriers who procure RAN-as-a-service from the hardware operators. MVNOs operate at the carrier level and may have their own billing systems and customer support but not own and manage any of the underlying network infrastructure. Such an abstraction helps the operators scale their operations based on consumer demand. MVNOs in current cellular arena are famous for their innovation in network subscription plans and pricing. SDN-based MVNOs are empowered to

Traffic class	Fundamental characteristics	Example
Conversational	Preserves time relation (variation) between information entities of the stream with stringent and low network delay.	Voice over IP
Streaming	Preserves time relation (variation) between information entities of the stream where some delay is acceptable.	Video streaming
Interactive	Request-response pattern, where pay-load content is to be preserved on a best-effor basis.	Web browsing
Background	Destination is not expecting the data within a certain time.	Background downloads.

Table 3.1: Classification of traffic based on service type

a greater degree to build on such innovations. Similar to individual subscribers, MVNOs have a SLA with the MNO or hardware operator.

- **Subscribers:** Subscribers are individual consumers with a UE, such as, a smartphone which consumes radio resources from a carrier at a pre-determined price. The price is based on voice and/or data usage as mentioned in the SLA between the carrier and consumer. The various SLA classification are discussed in the following subsections.

3.2.4 Service Requirement for Traffic Classes

The end users of cellular networks have various QoS requirements, meeting these requirements impose various constraints on the operators and in turn on the hardware providers. Hence, the classification of various services based on their traffic type is pivotal for resource allocation problem in RRM. The various services offered can be classified [25] as listed in Table 3.1. The data and time-slot QoS requirements for these traffic classes can be defined as follows:

$$\text{Conversational services } \rightarrow \begin{cases} R_{ji}^{Srv} = R_{ji}^{Nom}, & \text{for Data Rate} \\ t_{Rsp_{ji}} \leq t_{Rsp_{ji}}^{Nom}, & \text{for Response Time} \end{cases} \tag{3.1}$$

Where:

- R_{ji}^{Nom}: Nominal data rate for user i serviced by carrier j .
- R_{ji}^{Srv}: Served data rate for user i serviced by carrier j.
- $t_{Rsp_{ji}}$: response time for user i serviced by carrier j.
- $t_{Rsp_{ji}}^{Nom}$: Nominal response time for user i serviced by carrier j.

Equation 3.1 can be modified for *Streaming services* based on the number of time intervals allocated is as follows:

$$\text{Streaming services } \rightarrow \begin{cases} R_{ji}^{Srv}(t_i)\Delta t \geq R_{ji}^{Nom} n_{\Delta t}\Delta t \\ t_{Rsp_{ji}} \leq t_{Rsp_{ji}}^{Nom} \end{cases} \tag{3.2}$$

Where:

- $n_{\Delta t}$: Number of time intervals.
- $R_{ji}^{Srv}(t_i)$: Served data rate for timeslot t_i.
- Δt: Size of the time interval.

QoS requirements for *Interactive services* are as follows:

$$\text{Interactive services } \rightarrow \begin{cases} R_{ji}^{Srv} \geq R_{ji}^{Nom} \\ t_{Srv_{ji}} \leq t_{Srv_{ji}}^{Nom} \\ t_{Rsp_{ji}} \leq t_{Rsp_{ji}}^{Nom} \end{cases} \tag{3.3}$$

Where:

- $t_{Srv_{ji}}$: Time for servicing user i serviced by carrier j.
- $t_{Srv_{ji}}^{Nom}$: Nominal serving time for user i serviced by carrier j.

Finally, QoS constraints for *Background services* is as follows:

$$\text{Background services } \rightarrow \begin{cases} t_{Srv_{ji}} \leq t_{Srv_{ji}}^{Nom} \end{cases} \quad (3.4)$$

3.2.5 Spectrum Usage and Network Traffic Patterns

In this subsection, we present and explain the spectrum usage and network traffic pattern based on a real-world dataset that we use for our evaluating the performance of the virtual eNodeB.

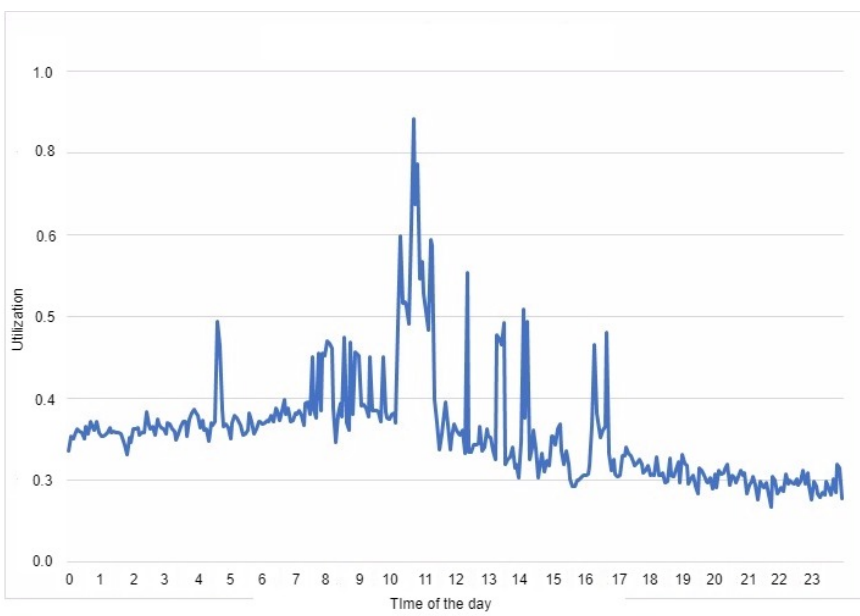


Figure 3.3: Observed spectrum utilization of the unlicensed band 2.4 GHz over a period of 24 hours.

A. **Dataset**

The dataset we use in this thesis was collected over a twenty-four hour cycle, spanning thirty days, using HackRF, a low cost, open source SDR peripheral. We capture spectrum utilization of the unlicensed industrial, scientific, and medical (ISM) radio band for a cell located in the densely-populated area of Lincoln, Nebraska. In particular, our dataset captures spectrum utilization levels at a rate of 512 data points per millisecond using the spectrum analyzer from the HackRF toolbox. For the purpose of this thesis, this huge dataset is curated and filtered to represent the spectrum utilization for a twenty-four hour period spaced at time intervals of four minutes as represented in Fig. 3.3.

B. **Limitations**

Our dataset has limitations when compared to a real world cellular network, which we elaborate in this section ahead of model building and simulations to ensure readers understand what our simulations represent. The experiments we conduct rely on existing mathematical studies performed on various traffic models to predict the composition of different traffic types. As the entire dataset captured is voluminous to study and arrive at a specific usage pattern, we opt to partition the data on per-day basis and process it to arrive at meaningful spectrum utilization pattern.

C. **Traffic Model**

In [26], a single two-state activity model for VoIP is considered, as presented in Fig. 3.4. The two states of this model are *silence* or *inactive* state (state 0) and *talking* or *active* state (state 1). The transition probability from state 0 to state 1 is α, while the probability of staying in state 0 is $(1 - \alpha)$. Moreover, transition from state 1 back to state 0 happens with a probability of β, while (1

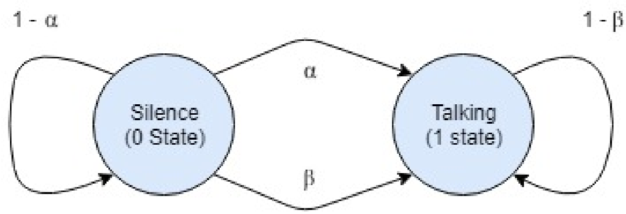


Figure 3.4: Two-state voice activity model

$-\ \beta$) is the probability of staying in state 1. The probability of being in state 0 and state 1, P_0 and P_1, respectively, are given as follows:

$$P_0 = \frac{\beta}{\alpha + \beta} \tag{3.5}$$

$$P_1 = \frac{\alpha}{\alpha + \beta} \tag{3.6}$$

The probability of being in the talking state, state 1, is referred to as Voice Activity Factor (VAF):

$$VAF = P_1 = \frac{\alpha}{\alpha + \beta} \tag{3.7}$$

The mean silence duration and mean talking duration in terms of number of voice frames can be written as:

$$\text{The mean silence duration} \rightarrow E[\tau_s] = \frac{1}{\alpha} \tag{3.8}$$

$$\text{The mean talking duration} \rightarrow E[\tau_t] = \frac{1}{\beta} \tag{3.9}$$

The probability of n-frame long silence or talking is given by:

$$n\text{-frame silence duration} \rightarrow p_{\tau_s=n} = \alpha(1-\alpha)^{(n-1)} \tag{3.10}$$

$$n\text{-frame talking duration} \rightarrow p_{\tau_t=n} = \beta(1-\beta)^{(n-1)} \tag{3.11}$$

The distribution of time period τ_{AE} between two talking states is the combination of the distribution of τ_s and τ_t. The probability that this duration is n-voice frame long is:

$$p_{\tau_{AE}=n} = \frac{\alpha}{\alpha-\beta}\beta(\alpha-\beta)^{(n-1)} + \frac{\beta}{\beta-\alpha}\alpha(1-\alpha)^{(n-1)} \tag{3.12}$$

Since the state transitions from state 0 to state 1 and vice-versa are independent, the mean time between successive transitions into the talking state is the sum of the mean time in each state:

$$E_{\tau_{AE}} = E_{\tau_{AE}} = E_{\tau_{AE}} = \frac{1}{\alpha} + \frac{1}{\beta} \tag{3.13}$$

The rate of arrivals into the active state can serve as a guide on the number of resource requests needed for persistent allocation of resources for VoIP traffic. A single resource request and/or scheduling grant will be required for persistent allocation of resources when a user moves from inactive to talking state.

3.2.6 Classification of Service Level Agreements

Apart from QoS requirements, resource allocation manager must take into consideration various service level agreements which define the type and priority of the service. Based on these conditions, there can be several service level agreements (SLA) that can be broadly categorized into the following types:

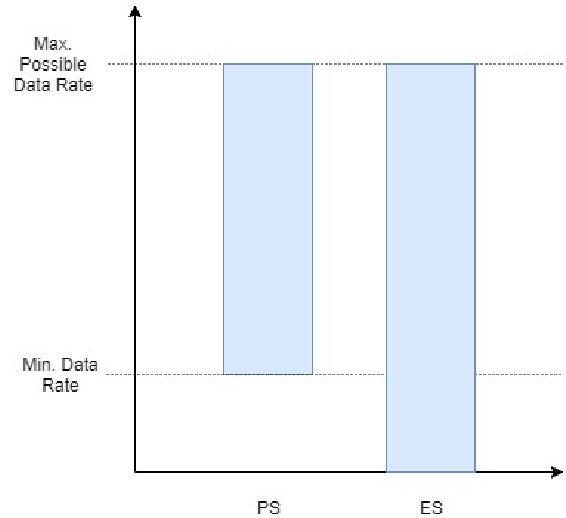


Figure 3.5: Comparison of SLAs

A. **Promised Service** (PS) is one in which a user is guaranteed a minimum level of data rate, regardless of network status. In this service level, users are guaranteed a good QoE at greater expense. This tier is the equivalent of high availability service.

B. **Essential Service** (ES) is one in which a user is served based on the network traffic and subscriber demand for resources at a specific time. For instance, operators may have resources at higher than normal utilization levels during business hours.

3.2.7 Distribution of Users and their SLAs

For the purpose of our implementation, we consider 10,240 users or subscribers for the MNO. The International Telecommunications Union estimates the number of cellular connections meant for humans to be around 30 billion and that of machines to be 70 billion in 2030 [27]. Hence, we distribute the SLAs of our sample size into two: 30% for PS and 70% for ES. We also consider two MVNOs which operate with one of the two defined SLAs.

3.2.8 Cost of Service

Dynamic pricing policies allow network operators to charge varying cost per-unit time depending on the availability of network resources, which may have an impact on call arrival rates to the network. This implies network service requirements such as availability, reliability, security, bandwidth, congestion, routing, stability, delays, etc., are maintained at an optimum level. Traditionally, dynamic pricing strategies have been used in wireline networks that offer Internet-based services. We use a dynamic pricing model as described in [28], which is a key value-add of this thesis. Our dynamic pricing function has pricing values from 0.1 to 1.0 based on the demand for network resources at any time t.

3.3 Analytical Model for Admission Control

The primary task of a RRM is to observe the demand for network resources and allocate these resources to users. This resource allocation problem is complicated due to the dynamic nature of networks. The major steps in solving this problem are tracking the available radio resources and allocating these resources based on the factors defined in the previous sections of this chapter.

We assume that the total network resources in a given Radio Resource Unit (RRU) is $\mathcal{R}$, which is the net capacity of the RRU when there is no user traffic. This $\mathcal{R}$ is the amount of resource RRM has to allocate to the users as and when they appear on the network, which translates to $\mathcal{R}_{Net}$, the total data rate from the RRU.

The total connections on the network can be calculated as the sum of subscribers serviced by the MNO at any specific instance and the total number of subscribers in each of the two MVNOs. This can be represented as:

$$\mathcal{S}_{Total}(t) = \sum_{\mathcal{S}_{i=1}}^{n} \mathcal{S}_i + \sum_{\mathcal{V}_{i=1}}^{m} \left[\sum_{\mathcal{S}_{j=1}}^{\mathcal{V}_i} \mathcal{S}_j \right] \tag{3.14}$$

where:

- $\mathcal{S}_i$: Subscribers present in MNO at time t
- $\sum_{\mathcal{S}_{j=1}}^{\mathcal{V}_i} \mathcal{S}_j$: Total number of subscribers in MVNO j at time t.

The objective function to solve the resource allocation and decision making problems in our experiment is based on three components as described below:

3.3.1 Data Rate Component

The effective data rate of a connection in the network is a dynamic function based on SINR, channel quality, modulation, etc. Different users using different terminals require different average data rates. As an evidence to this claim, consider that the M2M terminals need the lowest average data rates (i.e., 1 Mbps). On the other hand a terminal such as a Laptop or smartphone may need higher data rates. The MVNOs based on the diversity of their subscribers terminals have to contract different data rates. Hence, the allocated virtual radio resources to different MVNOs change. Also, the number of subscribers to be served in the same physical infrastructure may change in relation to the terminals used by subscribers. For the purposes of this experiment, we consider SINR as the only factor responsible for varying the effective data rate.

$$\mathcal{R}_i(\rho_i) \in \left[0, \mathcal{R}_i^{MAX}\right] \tag{3.15}$$

Where:

- $\mathcal{R}_i$ is the data rate as observed by user i.

- ρ_i is the SINR value of user i's device.

- $\mathcal{R}_i^{MAX}$ is the upper limit of data rate observed by user i.

From [29], a probability distribution function of SINR $\mathcal{P}_\rho$ can be deduced and represented as:

$$\mathcal{P}_\rho = \frac{0.2}{\alpha_p} ln(10) e^{-\frac{0.2}{\alpha_p}.ln(10).\rho_{in}} \tag{3.16}$$

where:

- α_p is the path loss exponent ≥ 2.

The effective data rate (from Equations 3.15, 3.16) of each network connection can be extrapolated to all the current users being serviced in order to find the network's effective data rate. This is a key function used to track available radio resources in the network, which is key in identifying a solution to the decision making and resource allocation problem. The function for tracking the resource consumption of the network is as follows:

$$\mathcal{R}_{Usage}(t) = \sum_{\mathcal{R}_{i=1}}^{n} \mathcal{R}_i(\rho_i) + \sum_{\mathcal{V}_{i=1}}^{m} \Big[\sum_{\mathcal{R}_{j=1}}^{\mathcal{V}_i} \mathcal{R}_j(\rho_j) \Big] \tag{3.17}$$

From $\mathcal{R}_{Net}$ and Equation 3.17, we can estimate the total radio resource available for virtual RRM to allocate to new users. This available data rate $\mathcal{R}_{Avail}$ can be computed as follows:

$$\mathcal{R}_{Avail}(t) = \mathcal{R}_{Net} - \left[\sum_{\mathcal{R}_{i=1}}^{n} \mathcal{R}_i(\rho_i) + \sum_{\mathcal{V}_{i=1}}^{m} \left[\sum_{\mathcal{R}_{j=1}}^{\mathcal{V}_i} \mathcal{R}_j(\rho_j) \right] \right] \tag{3.18}$$

3.3.2 Cost Component

The cost component of the decision problem is a dynamic function that depends on the observed traffic pattern and real-time demand. The cost function achieves social-welfare maximization for both users and operators. The concept of dynamic pricing and advantages are studied in detail in [30].

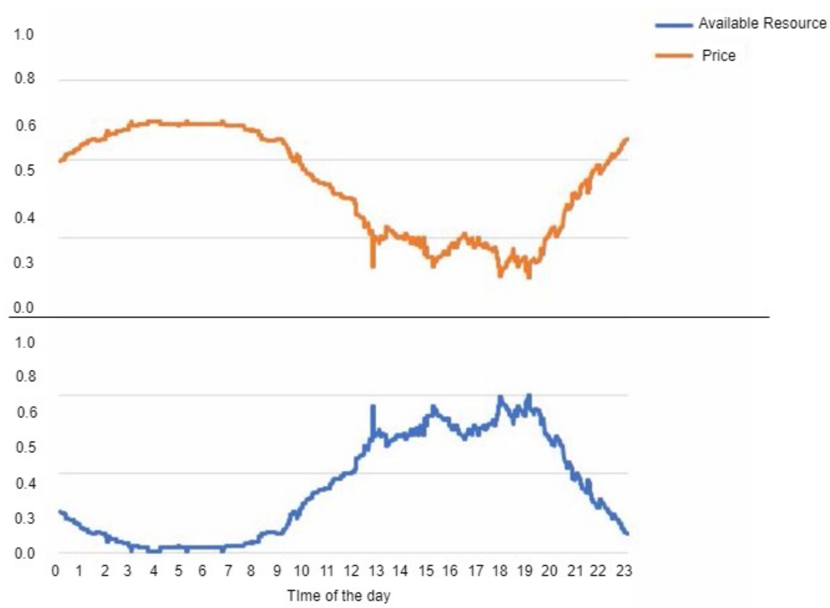


Figure 3.6: Fractional network availability vs. relative change in subscription pricing.

In [31], the demand function for a single service class was represented as shown in Equation 3.19, where p_0 was the normal price and $p(t)$ was the dynamic price computed based on user arrival rate at time t. This pricing function was derived based on arrival rates for new calls (λ_n), retry calls (λ_T), and the rate of arrival of substitute users $\lambda_s(t)$.

$$D\big[p(t)\big] = e^{-\left(\frac{p(t)}{p_0}-1\right)^2},\ p(t) \geq p_0 \tag{3.19}$$

The pricing function based on demand function defined in Equation 3.19 is:

$$p(t) = D_{-1}\left(min\left(\frac{\lambda_n^*}{\lambda_n(t) + \lambda_T(t) + \lambda_s(t)}, 1\right)\right) \tag{3.20}$$

3.3.3 QoS Component

The QoS component of the decision problem is used to observe the current state of network using parameters such as packet loss, number of calls dropped, etc. Our implementation considers call dropping and call blocking as the QoS variables to evaluate network state and allocate resources accordingly. In case of an increase in call drop rate, our model blocks new users in-order to satisfy the SLAs of existing users in the network. This preemption control function of the RRM uses subscribers' SLA values to identify ES subscribers that must be offloaded in order to free-up resources for PS subscribers. During excessive load conditions, the QoS component of objective function in the RRM can also initiate load shedding by identifying ES subscribers based on their SLA value.

The call dropping probability function used in this implementation is based on [32] and is represented as:

$$P_d(t) = 1 - \frac{1}{e^\rho - 1}\sum_{k=1}^{\infty}\frac{\rho^k}{k!}\int_0^\infty f_T(t)e^{-(\frac{\gamma_d t}{k})}dt \tag{3.21}$$

Where:

- $\gamma_d t$ is the drop-call rate

- $f_T(t)$ is the probability distribution function of call duration for user terminated calls

- ρ is the network utilization factor that depends on the number of users serviced and resources used at time t

The combination of the aforementioned three components is used to arrive at a objective function to solve the resource allocation and decision making problem which is represented as follows:

$$\mathcal{F}_t = f\Big(\mathcal{R}_{Avail}(t), p(t), P_d(t)\Big) \tag{3.22}$$

Once the VRRM is initialized, it gathers information about the subscribers, MVNOs, and their SLAs. This stage is referred to as "Network Initiation" and the next step involves the estimation of total network capacity. VRRM uses the information about the number of available RRUs and estimates the probability function of network capacity. The allocation of available capacity to MVNOs is a linear programming problem formulated as shown in Equations 3.23 & 3.24.

$$\mathcal{S}_{Net} = \mathcal{S}_{MNO} + \mathcal{S}_{MVNO1} + \mathcal{S}_{MVNO2} \tag{3.23}$$

$$\mathcal{R}_{Net} = \mathcal{R}_{MNO} + \mathcal{R}_{MVNO1} + \mathcal{R}_{MVNO2} \tag{3.24}$$

Accordingly, the optimization problem in each VRRMs decision window under consideration can be solved based on the following algorithm, which is segmented into four separate procedures to achieve better readability.

Algorithm 1 Resource allocation in VRRM

procedure STEP0
  $t \leftarrow$ current time
  $\mathcal{R}_{avail} \leftarrow$ available network resource at t
  $\mathcal{S}_t \leftarrow$ subscribers that require service at t
  $\mathcal{R}_{req} \leftarrow$ network resources required to serve $\mathcal{S}_t$ at t
top:
  if (Algorithm 2 $\wedge$ Algorithm 3 $\wedge$ Algorithm 4) **then**
    $\mathcal{S}_t \leftarrow \mathcal{R}_{req}$
    $\mathcal{R}_{t+1} \leftarrow \mathcal{R}_t - \mathcal{R}_{req}$
  else
    goto *top*.

In Algorithm 1, procedure STEP0 initiates the network for a given time instance. The procedure calculates the available network resources, identifies the number of subscribers requesting access to network and the resources needed to serve these subscribers.

These values then act as inputs to STEP1 through 3 of the algorithm to evaluate the various components. The RRM at various stages of this algorithm prunes the subscribers requesting access based on available resources computed in STEP1 (algorithm 2), profitability of the operator in STEP2 (algorithm 3), and the ability to manage subscriber load is calculated in STEP3 (algorithm 4). The RRM continues its evaluation till all the subscribers are serviced or the processing time reaches a timeout.

Algorithm 2 Data rate condition

procedure STEP1
  $i \leftarrow 0$
top:
  if $(t = t_{timeout})$ **then return** $false$.
  else
    if $(\mathcal{R}_{req}(t) \leq \mathcal{R}_{avail}(t))$ **then return** $true$.
    else
      if $(i == 0)$ **then**
        Block subscribers with SLA $= ES$
        $\mathcal{S}_t^{new} \leftarrow \mathcal{S}_t - \mathcal{S}_t^{ES}$
        $i \leftarrow i + 1$
        goto Algorithm 1 : *top*.
      else
        Drop subscribers with SLA $= ES$ from $t - i$ time
        $\mathcal{S}_{t-1}^{new} \leftarrow \mathcal{S}_{t-1} - \mathcal{S}_{t-1}^{ES}$
        $\mathcal{R}_{avail}(t) = \mathcal{R}_{avail}(t) + \mathcal{R}_{SLA=ES}(t)$
        $i \leftarrow i + 1$
        goto *top*.

The procedure STEP1 studies the state of network resources and calculates the difference between available free network resources and the amount needed to satisfy the needs of new subscribers. This is the first step of the pruning process to identity subscribers who cannot be serviced. If the available free resources is less than the resource requirement, the RRM of the virtual eNodeB attempts to scale up network capacity by marshalling additional compute and network resources. If the RRM detects that the network is running at peak availability, as a first step it blocks new subscribers with lower-tier SLA (ES) and reevaluate the network resource status. If the test still fails, then RRM will continue to drop ES users in the network to facilitate network access for PS users in an iterative fashion in a process known as load shedding.

Algorithm 3 Cost Condition

procedure STEP2
    $\mathcal{C}_{sub} \leftarrow$ cost to service one subscriber at t
    $\mathcal{C}_t \leftarrow$ cost to service all subscribers at t
    $\mathcal{P}_t^{SLA} \leftarrow$ price of service for user of a given SLA at t
top:
    $Revenue_t \leftarrow (\mathcal{S}_t^{ES} \times \mathcal{P}_t^{ES})$
        $+(\mathcal{S}_t^{PS} \times \mathcal{P}_t^{PS})$
            $-(\mathcal{S}_t \times \mathcal{C}_{sub})$
    if $(Revenue_t \leq 0)$ **then**
        $\mathcal{S}_t = \mathcal{S}_t - \mathcal{S}_t^{ES}$
        if (From [28], Algorithm 4) **then**
            Handoff $\mathcal{S}_t$ to neighbouring eNodeB **return** *false*.
        else
            return *true*.
    else
        return *true*.

The procedure `STEP2` studies dynamic pricing based on network demand and allocates resources to subscribers accordingly. In a cellular network, base stations can balance or transfer traffic among each other using a technique known as "Cell Zooming". The RRM tries to minimize its loss by trying to switch an eNodeB to idle state using techniques such as cell zooming and throttling down the network hardware responsible to serve a base station. The handoff decision is based on the Network-Initiated handoff presented in [28], and here PS subscribers can be referred to as primary users, and ES subscribers as secondary users. In the case of a network where a subscriber is in a long-term contract with an operator such profit-maximization techniques may lead to resource starvation for ES class of subscribers. As discussed in [28], opportunistic network access-based cellular network models provide avenues for profit-maximization and market-based network access pricing.

Algorithm 4 QoS condition

procedure STEP3
  $i \leftarrow 0$
  $\mathcal{CD}_t^{SLA} \leftarrow$ Call dropping rate at t
  $\mathcal{CB}_t^{SLA} \leftarrow$ Call blocking rate at t
  $\mathcal{W}_{CD}^{SLA} \leftarrow$ Weight of a Call drop
  $\mathcal{W}_{CB}^{SLA} \leftarrow$ Weight of a Call Block
  Q_{Obs} : function to measure QoS metrics (ideal value = 0)
  Q_{Th} : Threshold for QoS metric
top:
  $\mathrm{Q}_{Obs} = (\mathcal{CD}_{t-1}^{ES} \times \mathcal{W}_{CD}^{ES}) + (\mathcal{CB}_{t-1}^{ES} \times \mathcal{W}_{CB}^{ES})$
  $\quad + (\mathcal{CD}_{t-1}^{PS} \times \mathcal{W}_{CD}^{PS}) + (\mathcal{CB}_{t-1}^{PS} \times \mathcal{W}_{CB}^{PS})$
  if $\mathrm{Q}_{Obs} > \mathrm{Q}_{Th}$ **then**
    $i \leftarrow i + 1$
    if $(i == 0)$ **then**
      $\mathcal{S}_{t-1}^{ES}(\mathcal{CB}) \leftarrow$ Dropped.
      $\mathcal{S}_{t-1} \leftarrow \mathcal{S}_{t-1} - \mathcal{S}_{t-1}^{ES}(\mathcal{CB})$
      goto *top*.
    if $(i == 1)$ **then**
      $\mathcal{S}_{t-1}^{PS}(\mathcal{CB}) \leftarrow$ Dropped.
      $\mathcal{S}_{t-1} \leftarrow \mathcal{S}_{t-1} - \mathcal{S}_{t-1}^{PS}(\mathcal{CB})$
      goto *top*.
    if $(i == 2)$ **then**
      $\mathcal{S}_{t-1}^{ES}(\mathcal{CD}) \leftarrow$ Dropped.
      $\mathcal{S}_{t-1} \leftarrow \mathcal{S}_{t-1} - \mathcal{S}_{t-1}^{ES}(\mathcal{CD})$
      goto *top*.
    if $(i == 3)$ **then return** *false*.
  else
    return *true*.

The final step of the algorithm studies the state of QoS parameters of the base station. The network state is reconfigured based on the value calculated using the weighted sum of QoS violations. This is done by blocking some of the subscriber arrivals and dropping some ES subscribers. Along with the importance given to satisfy SLAs, the RRM also attempts to reduce the overall QoS violations by setting higher penalty to call drops than to call blocks in the following penalty order (high to low) : $\mathcal{CD}^{PS} > \mathcal{CD}^{ES} > \mathcal{CB}^{PS} > \mathcal{CB}^{PS}$.

Chapter 4

Simulations and Results

In this chapter, we discuss the hardware setup, simulation process, and evaluate the simulation results. First, we discuss the data collection process for spectrum utilization used in our research, including the data cleansing process. Second, we specify the design decisions that are adopted to carry out the simulations. Third, we discuss the results of various experiments that we conduct to study the network behaviour. Finally, we discuss the outcome of these experiments.

4.1 Dataset

Spectrum analyzers allow to selectively scan wireless channels for signals of different sources such as wireless networks, microwave radios, weather radar, etc. The dataset for our research was compiled using a HackRF One, a low-cost, open source SDR spectrum analyzer whose beta testing phase was funded by DARPA. The device can be used as a USB peripheral or programmed for stand-alone operation, and it is capable of scanning frequencies ranging from 1 MHz to up to 6 GHz. We use HackRF, as show in Fig. 4.1 for spectrum sensing and GNU Radio which provides predefined signal processing blocks for data collection.

Figure 4.1: HackRF One— a low cost, open source software defined radio

Gqrx, as shown in Fig. 4.2, is a SDR receiver powered by the GNU Radio SDR framework and the Qt graphical toolkit. It is free software and supports many of the SDR hardware available, including HackRF and USRP devices.

GNU Radio, as shown in Fig. 4.3, is free software that provides signal processing blocks to implement software radios. It can be used with readily available low-cost external RF hardware to create SDRs, or without hardware in a simulation-like environment. It is widely used to support both wireless communications research and real-world radio systems.

The process of data capture was carried out over a period of four weeks in the unlicensed 2.4 GHz frequency range. The raw data captured by HackRF consists of 512 continuous spectrum signal strength values measured every millisecond. The fineness of this data is reduced to one reading every millisecond by calculating the average. The final processed data consists of the timestamp of the instance and the

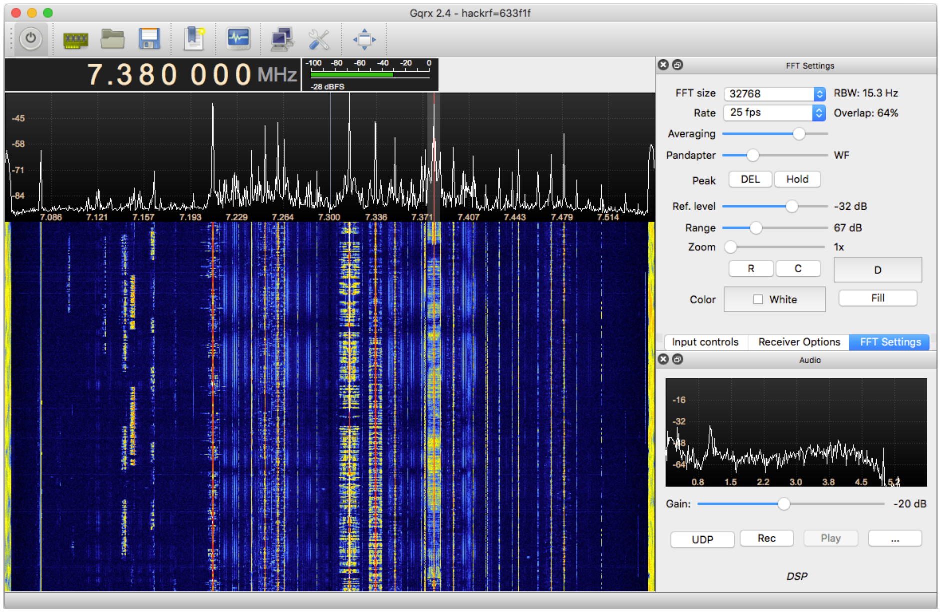


Figure 4.2: Gqrx Application—a free software application that interfaces with HackRF One

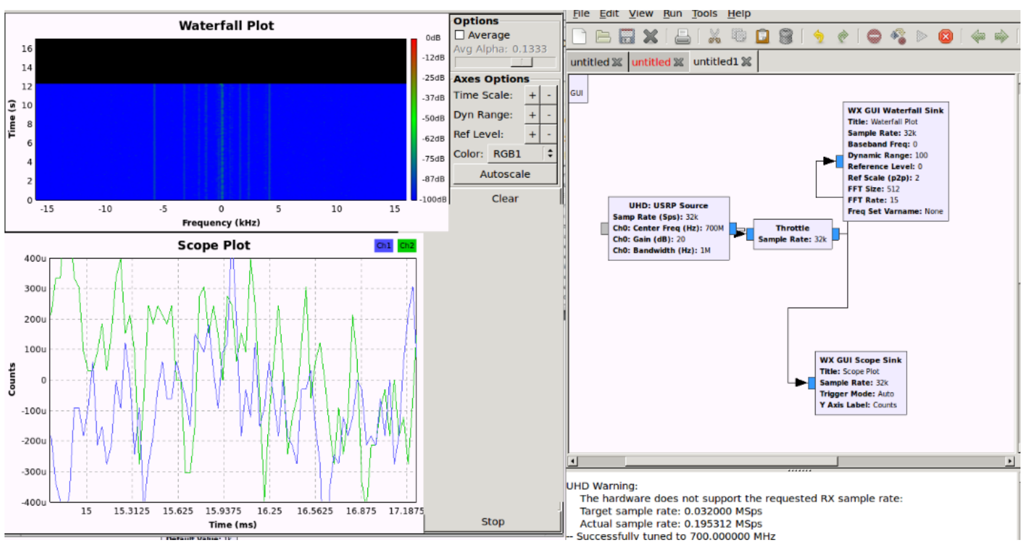


Figure 4.3: Screenshot of GNU Radio

spectrum strength value.

4.2 Test Scenarios

We evaluate our network model for two broad test scenarios: network utilization and revenue generation. In addition, we also test for how the proposed approach fares against SLA violations.

4.2.1 Network Utilization

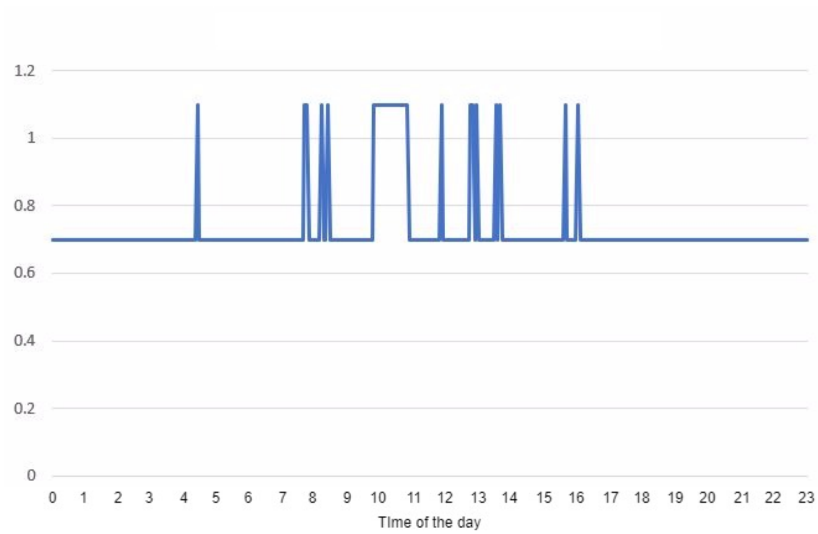


Figure 4.4: Maximum Available Network Resources

Unlike a traditional cellular network, an NFV-based one is elastic in nature as required hardware resources can be adjusted on-demand. The virtual network components can be shut down or powered up corresponding to subscribers needs. In this scenario, network access request pattern is input to the proposed model and the total available resources of the network is adjusted based on this input. The auto-scaling mechanism of virtualized eNodeB lets operators set target utilization levels for multiple resources. The elasticity of the modules is achieved using software rules to ramp up or down based on the average utilization of all scalable resources. This helps maintain optimal network performance and availability, even when user loads are periodic, unpredictable, or continuously changing. As displayed in Fig. 4.4, we use a simple two-step scaling mechanism which can be expanded to more complex scenarios.

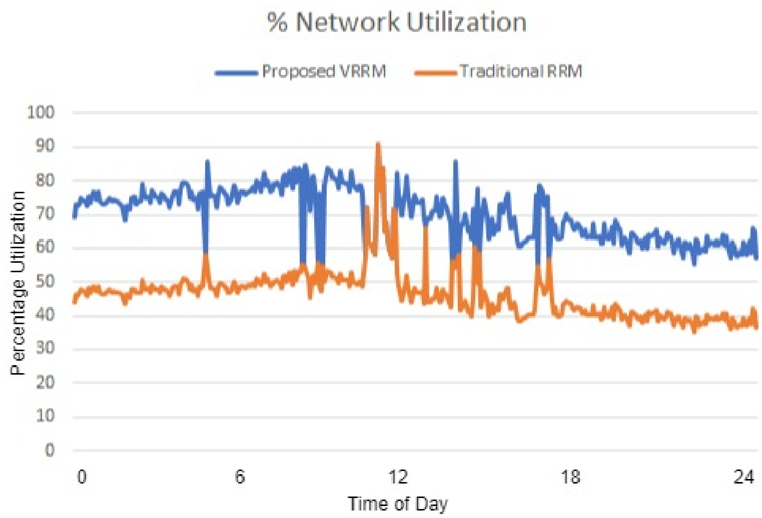


Figure 4.5: Network Utilization—Traditional vs. Proposed RRM

Fig. 4.5 depicts the percentage of network utilized in traditional and proposed NFV based RRM. It can be seen that the network resource utilization is higher in the proposed system due to the elastic nature of the RRM. There are several points of interest in the plot, where utilization measure of both the approaches are similar. This can be attributed to the passive nature of the elasticity mechanism as the model responds to higher user loads by scaling up the hardware resources. While the proposed model leads to an overall improvement in network utilization, inherent delays in resource activation may lead to call drops which we discussed in the following coming subsections.

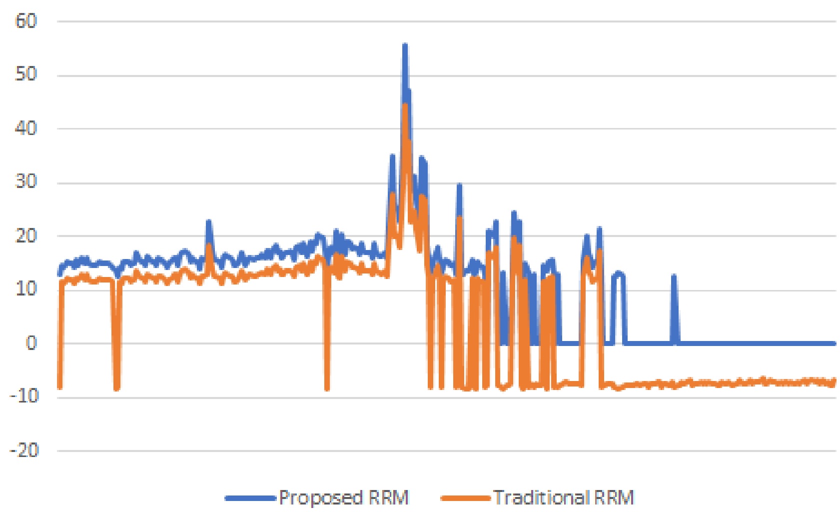


Figure 4.6: Operator Revenue based on Dynamic Pricing

4.2.2 Carrier Revenue

A carrier's revenue is based on three major components: First, the price at which a service is offered to the subscribers; second, number of subscribers active in the network; and, the average expenses incurred by the carrier to serve a single subscriber. In our proposed model, we use a dynamic pricing mechanism, as shown in Fig. 3.6, based on which the price of service varies depending on the demand for network resources. In the case of traditional networks, an operator has to maintain elements at peak availability irrespective of the network demand. On the other hand, our proposed NFV-based network model due to its elastic nature can scale down its infrastructure which leads to lower energy consumption, operational expenses, and in turn may lead to higher profits for the operators.

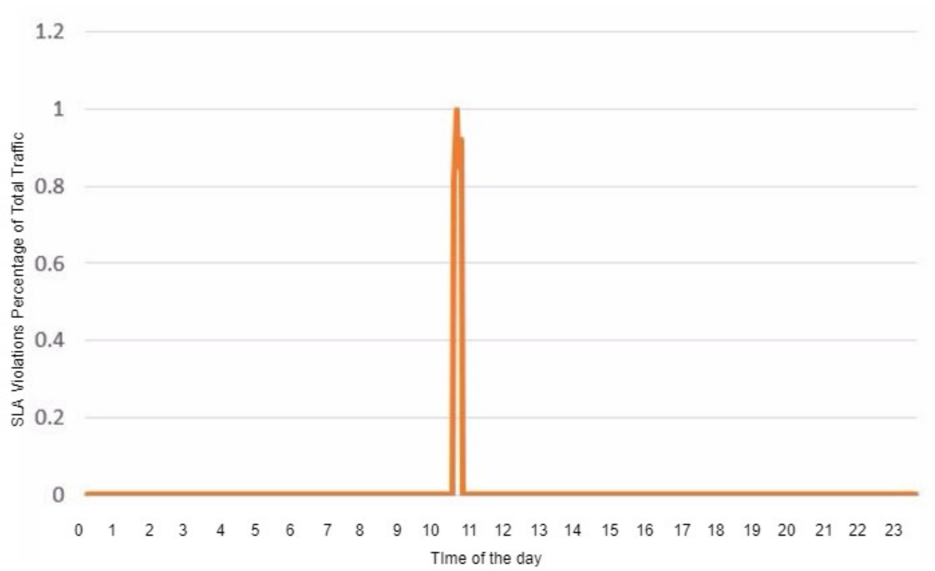
Dynamic pricing gives users the freedom to use the network at a price they are willing to pay. As a result, users with low-priority needs, such as, a class of M2M communications may opt not to use the network during daytime resulting in lesser congestion and increased prospects of better service quality. On the other hand, when prices are low, it may trigger higher user arrival rate. As a cellular network is a shared infrastructure, the application of dynamic pricing leads to a state of network usage that eventually attains equilibrium.

4.2.3 Violation of SLAs

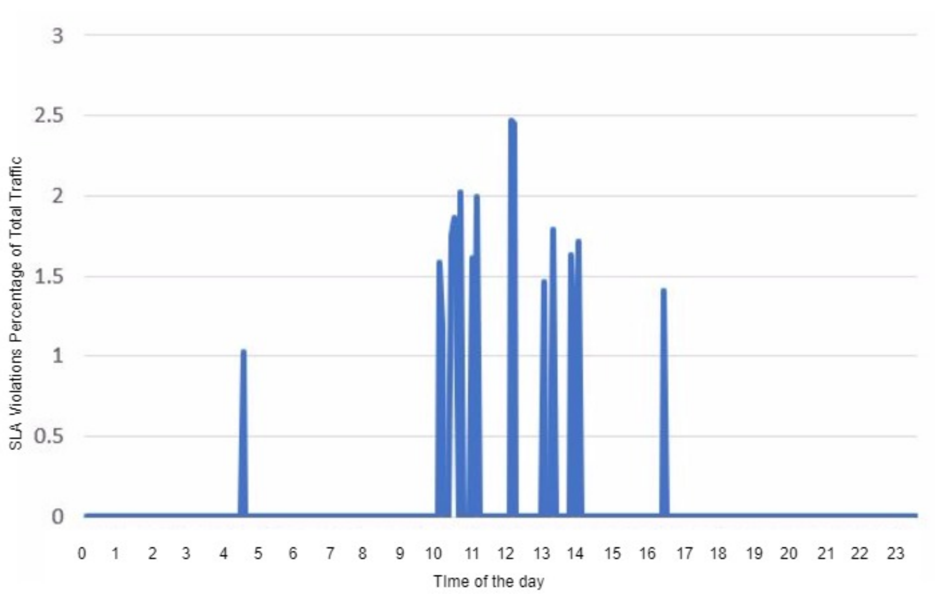
In real world networks, there are a greater number of service classifications with much complex SLAs. Our research studies the effects of changes in total network capacity when resources are scaled up or down for two types of SLAs as explained in Section 3.2.6. The RRM by design tries to provide access and service to PS subscribers. However, when there is an unanticipated rise in demand due to a large number of subsribers requesting network access, the operator is compelled to drop some users violating SLAs. When the various SLAs and their corresponding QoS parameters are assigned weights, the goal of the RRM is to minimize the overall violation weight. The call dropping of PS has the most weight followed by call dropping of ES, call blocking of PS, and finally call blocking of ES. The violation of services of proposed network compared to traditional networks is presented in Fig. 4.7.

We also study the spread of SLA violations for each of the QoS parameters considered, which follow the weights assigned to each QoS parameter violation. In Fig. 4.8, we observe that the maximum number of violations are concentrated in the call-blocking of SLA-ES, which can be attributed to the algorithm's preference to serve PS users over ES, and of call-dropping over call-blocking.

The proposed model while improves network utilization and reduces energy consumption is susceptible to violate SLAs when subjected to sudden spikes in network access requests. During times of sudden rise in user load, the model takes a reactive approach to solving the congestion problem. This can be avoided by using an



(a) SLA violation in traditional RRM based networks



(b) SLA violations in proposed RRM based networks

Figure 4.7: Percentage of SLA violations

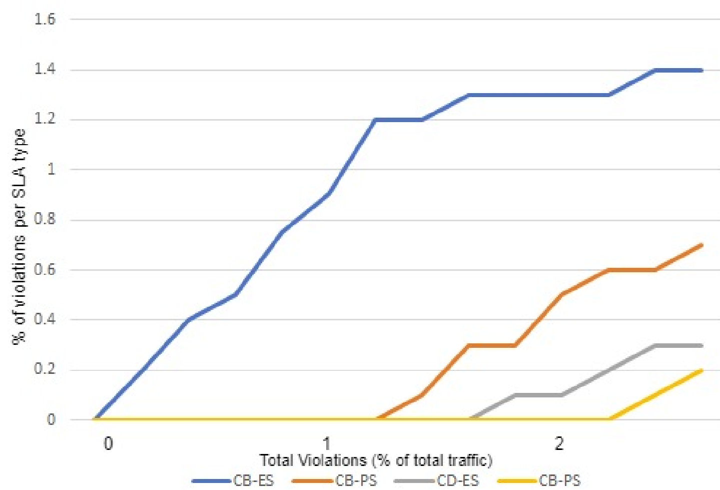


Figure 4.8: SLA violations per QoS parameter.

intelligent RRM, one with machine learning capabilities to effectively predict traffic patterns and preemptively allocate additional network resources for anticipated raise in user access demand.

Chapter 5

Summary and Future Work

5.1 Summary

In this thesis, we model and implement a virtual eNodeB component—Radio Resource Management (RRM)—and analyze the efficiency that can be achieved in the Call Admission Control function of the RRM. We define various service level agreement choices for the implementation and develop a model to allocate network resources based on three key factors: available data rate, carrier cost component, and service quality.

We evaluate the network efficiency that can be achieved using our model and find that our setup can potentially yield up to 33% reduction in energy consumption due to the elastic nature of the proposed NFV-based eNodeB. Our model leads to an increase in network utilization of as much as 60% compared to 40% when using a traditional RRM. We study the behaviour of the proposed setup under a dynamic pricing model which may generate as much as 27% increase in carrier revenue when compared to fixed access pricing . Finally, we put forth a weighted violation penalty metric to evaluate the model's QoS performance.

While our model may help increase carrier profits improves it is prone to violate user agreements during periods of sudden spike in network traffic. This can be attributed to the reactive nature of the elasticity of RRM.

5.2 Future Work

The work presented in this thesis can be extended and several variations of network resource allocation problem can be investigated in the future. Some of the topics are proposed below.

One possible area for research and extension is using machine learning techniques to perform predictive analysis on the network traffic patterns. This enables RRM to actively monitor traffic patterns and proactively allocate network resources anticipating demand in order to mitigate the occurence of SLA violations.

Finally, virtual RRM can be combined with cognitive cellular networks providing opportunistic network access to the users where they are not limited to be serviced by a single carrier. This could be a great value add in the design of next-generation mobile networks.

Bibliography

[1] S. Iyer, "Virtualisation of network functions and the SDN," *Australian Journal of Telecommunications and the Digital Economy*, vol. 2, no. 2, 2014.

[2] Ericsson, "Ericsson Mobility Report ," tech. rep., Ericsson, November 2016.

[3] Center for Energy-efficient Telecommunications, "The Power of Wireless Cloud," tech. rep., Bell Labs and University of Melbourne, April 2012.

[4] D. G. Ainscough, "The evolution and future of modern mobile communications," *blueadmiral.co.uk*, 2001.

[5] W. I. Forum, "What is software defined radio," *wirelessinnovation.org*, 2001.

[6] P. Dely, A. Kassler, and N. Bayer, "Openflow for wireless mesh networks," in *2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN)*, pp. 1–6, July 2011.

[7] L. E. Li, Z. M. Mao, and J. Rexford, "Toward software-defined cellular networks," in *Proceedings of the 2012 European Workshop on Software Defined Networking*, EWSDN '12, (Washington, DC, USA), pp. 7–12, IEEE Computer Society, 2012.

[8] J. Schulz-Zander, L. Suresh, N. Sarrar, A. Feldmann, T. Hühn, and R. Merz, "Programmatic orchestration of wifi networks," in *Proceedings of the 2014*

USENIX Conference on USENIX Annual Technical Conference, USENIX ATC'14, (Berkeley, CA, USA), pp. 347–358, USENIX Association, 2014.

[9] A. Gudipati, D. Perry, L. E. Li, and S. Katti, "SoftRAN: Software defined radio access network," in *Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking*, HotSDN '13, (New York, NY, USA), pp. 25–30, ACM, 2013.

[10] J. Vestin, P. Dely, A. Kassler, N. Bayer, H. Einsiedler, and C. Peylo, "Cloudmac: Towards software defined wlans," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 16, pp. 42–45, Feb. 2013.

[11] K.-K. Yap, R. Sherwood, M. Kobayashi, T.-Y. Huang, M. Chan, N. Handigol, N. McKeown, and G. Parulkar, "Blueprint for introducing innovation into wireless mobile networks," in *Proceedings of the Second ACM SIGCOMM Workshop on Virtualized Infrastructure Systems and Architectures*, VISA '10, (New York, NY, USA), pp. 25–32, ACM, 2010.

[12] X. Jin, L. E. Li, L. Vanbever, and J. Rexford, "SoftCell: Scalable and flexible cellular core network architecture," in *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies*, CoNEXT '13, (New York, NY, USA), pp. 163–174, ACM, 2013.

[13] J. Sachs and S. Baucke, "Virtual radio: A framework for configurable radio networks," in *Proceedings of the 4th Annual International Conference on Wireless Internet*, WICON '08, (ICST, Brussels, Belgium, Belgium), pp. 61:1–61:7, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008.

[14] Y. Zaki, L. Zhao, C. Goerg, and A. Timm-Giel, "LTE wireless virtualization and spectrum management," in *WMNC2010*, pp. 1–6, Oct 2010.

[15] X. Project, "The xen project, the powerful open source industry standard for virtualization," *http://www.xenproject.org/*, 2013.

[16] Y. Zaki, *Future mobile communications: LTE optimization and mobile network virtualization*. Springer Science & Business Media, 2012.

[17] L. Zhao, M. Li, Y. Zaki, A. Timm-Giel, and C. Görg, "LTE virtualization: From theoretical gain to practical solution," in *Teletraffic Congress (ITC), 2011 23rd International*, pp. 71–78, IEEE, 2011.

[18] M. Li, L. Zhao, X. Li, X. Li, Y. Zaki, A. Timm-Giel, and C. Gorg, "Investigation of network virtualization and load balancing techniques in lte networks," in *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, May 2012.

[19] B. Liu and H. Tian, "A bankruptcy game-based resource allocation approach among virtual mobile operators," *IEEE Communications Letters*, vol. 17, pp. 1420–1423, July 2013.

[20] V. D. Philip, Y. Gourhant, and D. Zeghlache, "Openflow as an architecture for e-node b virtualization," in *e-Infrastructure and e-Services for Developing Countries* (R. Popescu-Zeletin, K. Jonas, I. A. Rai, R. Glitho, and A. Villafiorita, eds.), (Berlin, Heidelberg), pp. 49–63, Springer Berlin Heidelberg, 2012.

[21] R. Sherwood, G. Gibb, K.-K. Yap, G. Appenzeller, M. Casado, N. McKeown, and G. Parulkar, "Flowvisor: A network virtualization layer," *OpenFlow Switch Consortium, Tech. Rep*, vol. 1, p. 132, 2009.

[22] M. Yang, Y. Li, D. Jin, L. Su, S. Ma, and L. Zeng, “OpenRAN: a software-defined ran architecture via virtualization,” in *ACM SIGCOMM computer communication review*, vol. 43, pp. 549–550, ACM, 2013.

[23] J.-M. Kelif, M. Coupechoux, and F. Marache, “Limiting power transmission of green cellular networks: Impact on coverage and capacity,” in *Communications (ICC), 2010 IEEE International Conference on*, pp. 1–6, IEEE, 2010.

[24] K. Son and B. Krishnamachari, “Speedbalance: Speed-scaling-aware optimal load balancing for green cellular networks,” in *INFOCOM, 2012 Proceedings IEEE*, pp. 2816–2820, IEEE, 2012.

[25] M. Barreiros and P. Lundqvist, *QoS-Enabled networks: Tools and foundations*. John Wiley & Sons, 2015.

[26] G. Schaffrath, C. Werle, P. Papadimitriou, A. Feldmann, R. Bless, A. Greenhalgh, A. Wundsam, M. Kind, O. Maennel, and L. Mathy, “Network virtualization architecture: Proposal and initial prototype,” in *Proceedings of the 1st ACM workshop on Virtualized infrastructure systems and architectures*, pp. 63–72, ACM, 2009.

[27] Radiocommunication Sector of ITU, “IMT traffic estimates for the years 2020 to 2030,” *The International Telecommunications Union*, July 2015.

[28] S. Raju, S. Boddepalli, N. Choudhury, Q. Yan, and J. S. Deogun, “Design and analysis of elastic handoff in cognitive cellular networks,” in *2017 IEEE International Conference on Communications (ICC)*, May 2017.

[29] S. Khatibi and L. M. Correia, "A model for virtual radio resource management in virtual rans," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, p. 68, 2015.

[30] S. Raju, S. Boddepalli, Q. Yan, and J. S. Deogun, "Achieving 5as in cloud centric cognitive cellular networks," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Dec 2016.

[31] J. Hou, J. Yang, and S. Papavassiliou, "Integration of pricing with call admission control to meet QoS requirements in cellular networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 13, pp. 898–910, Sep 2002.

[32] G. Boggia, P. Camarda, and A. D'alconzo, "Modeling of call dropping in well-established cellular networks," *EURASIP J. Wirel. Commun. Netw.*, vol. 2007, pp. 1:1–1:11, Oct. 2007.