ON-DEMAND OFFLOADING COLLABORATION FRAMEWORK BASED ON LTE NETWORK VIRTUALISATION

Subah A. D. Albinali

School of Computing, Science and Engineering
College of Science and Technology the University of Salford, Salford, UK

Submitted in Partial Fulfilment of the Requirements of the Degree of Doctor of Philosophy, June 2019

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Dr Omar Alani, for his continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me during my research work and the writing of this thesis. I could not have imagined having a better supervisor and mentor for my PhD study.

Last but not the least, I would like to thank my family: my wife, my children , my parents and my brothers and sisters, for supporting me spiritually throughout writing this thesis and in my life in general.

Abstract

Recently, there has been a significant increase in data traffic on mobile networks, due to the growth in the numbers of users and the average data volume per user. In a context of traffic surge and reduced revenues, operators face the challenge of finding costless solutions to increase capacity and coverage. Such a solution should necessarily rule out any physical expansion, and mainly conceive real-time strategies to utilise the spectrum more efficiently, such as network offload and Long-term Evolution (LTE) network virtualisation. Virtualisation is playing a significant role in shaping the way of networking now and in future, since it is being devised as one of the available technologies heading towards the upcoming 5G mobile broadband. Now, the successful utilisation of such innovative techniques relies critically on an efficient call admission control (CAC) algorithm. In this work, framework is proposed to manage the operation of a system in which CAC, virtualisation and Local break out (LBO) strategies are collaboratively implemented to avoid congestion in a mobile network, while simultaneously guaranteeing that measures of quality of service (QoS) are kept above desired thresholds. In order to evaluate the proposed framework, two simulation stages were carried out. In the first stage, MATLAB was used to run a numerical example, with the purpose of verifying the mathematical model of the proposed framework in air interface level. The second stage involved of using open source applications such as, Emulated Virtual Environment (EVE) and Wireshark, for emulating the traffic in the network for different scenarios inside the core network. The results confirm the effectiveness of the proposed framework.

List of Abbreviations

1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
AIPN	All IP-Network
AS	Access Stratum
CAC	Call Admission Control
CAPEX	Capital Expenditure
CDMA	Code Division Multiple Access
CN	Core Network
D2D	Device to Device
eNB	eNodeB
EPC	Evolved Packet Core
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
FDMA	Frequency Division Multiple Access
GERAN	GSM EDGE Radio Access Network
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communication
GSM-EFR	GSM – Enhanced Full Rate
HSPA	High-Speed Packet Access
HSS	Home Subscriber Server

IaaS	Infrastructure as a Service
IMS	IP Multimedia Subsystem
IMT	International Mobile Telecommunication
InP	Infrastructure Provider
IoT	Internet of Things
ITU	International Telecommunication Union
LBO	Local Break Out
LIPA	Local IP Access
LTE	Long Term Evolution
M2M	Machine to Machine
MAC	Medium Access Control (layer)
MME	Mobile Management Entity
MVNE	Mobile Virtual Network Enabler
MVNO	Mobile Virtual Network Operator
MVNP	Mobile Virtual Network Provider
NaaS	Network as a Service
NAS	Non-Access Stratum
NRT	Non-Real Time
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OPEX	Operating Expenses
OTT	Over the top operator
PCRF	Policy and Charging Rules Function
PDN	Packet Data Network
PRB	Physical Resource Block
PS	Packet Switching

QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RT	Real Time
RWP	Random Way Point
SAE	System Architecture Evolution
SDN	Software Defined Network
SIPTO	Selective IP Traffic Offload
SNR	Signal to Noise Ratio
SP	Service Provider
UE	User Equipment
UMTS	Universal Mobile Telecommunications Service
UTRAN	UMTS Terrestrial Radio Access Network
VLAN	Virtual Local Area Network
VN	Virtual Network
VOIP	Voice over Internet Protocol
Wi-Fi	Wireless Fidelity
WLAN	Wireless Local Area Network
Xaas	X (Anything) as a Service

ON-DEMAND OFFLOADING COLLABORATION FRAMEWORK BASED ON LTE NETWORK VIRTUALISATION

Contents

List of figures

List of Tables

Chapter 1

1 Introduction

1.1 Background

Recent technical reports show a tremendous increase in data traffic on mobile broadband networks [1], due to the growing numbers of users and the average data volume per user. Figure 1 illustrates such increase between the fourth quarters of 2011 and 2016; data traffic grew 55% and that the data traffic significantly overpasses voice traffic. In addition, worldwide mobile broadband users raised from 268 million in 2007 to 2.1 billion in 2013 [2]. This corresponds to an average yearly growth rate of 40%, making mobile broadband the most active market in Information and Communications Technology (ICT). Fulfilling this growing traffic poses an operational and cost challenge for mobile service providers considering that, (i) the spectrum dedicated to mobile communications is limited; in fact, nowadays is very difficult to obtain a new spectrum band since most have already been assigned (e.g. TV, satellite, Privet Mobile Radio) [3], and (ii) deploying a new network element such as a base station is difficult in some cases due to site issues, which create additional expense or even QoS degradation as a result of possible interference.

The following two examples may give a notion of the cost of obtaining a new spectrum license. In December 2011, the French spectrum regulator known as ARCEP published the provenance of 4G in the 800 MHz band, where a 30 MHz duplex was valued at 2.639 billion euros, while a 70 MHz duplex in the 4G 2.6 GHz band was appraised at 0.94 billion euros [3]. In addition, the regulator requires service provider operators to guarantee coverage of up to 99.6% of the country's population. In a similar way, the United Kingdom, through the spectrum regulator known as Ofcom (Office of Communication), held an auction between the end of 2015 and the beginning of 2016 to make the new spectrum available for mobile broadband with a reserved price of between £2.5 million and £5 million per lot (10 MHz) for the 2.3 GHz spectrum, and £1 million for the 3.4 GHz lot (5 MHz) spectrum, leading to a total value of £50 million and £70 million. Yet more spectrum is required, together with the new technology needed to use the spectrum in a more efficient and effective way to cope with the tremendous increase in data [4].

From an operational standpoint, high data rate, enhanced performance, QoS and end-user quality of experience (QoE) will generally be the key performance indicators (KPIs) for the evolution of future mobile networks and applications [5, 6]. In particular, guaranteeing QoS is a challenge due to the different types of traffic, such as real-time applications and non-real time applications, different customer classes (subscriptions classes, emergency calls, priorities calls), and the need of simultaneously adjusting between two kinds of probability, namely a dropping probability and a blocking probability [7]. Besides, the average revenue per user (ARPU) is decreasing as a result of the flat rate business models. In addition, significant investments in the physical expansion of 4G networks may be discouraged by the fact that the market is headed to the era of the fifth-generation (5G) mobile wireless networks, which promise higher data rates and spectrum efficiency, enhanced QoE and QoS, reduced jitter latency and energy consumption, among other things. Therefore, operators are in crucial need of costless alternative solutions to cope with the increasing demand in data traffic, which will require more efficient use of the spectrum, more aggressive frequency reuse and the collaboration of several enabling technologies [8].

To this end, the wireless virtualisation technique has become a good alternative, attracting the attention of industry and academia as one of the critical enabling technologies towards the fifth generation (5G) [9]. In addition, the selective offloading strategy will deal with the demand for intensive bandwidth applications in a cost-effective manner, by bypassing the operator core network and lowering congestion [10].

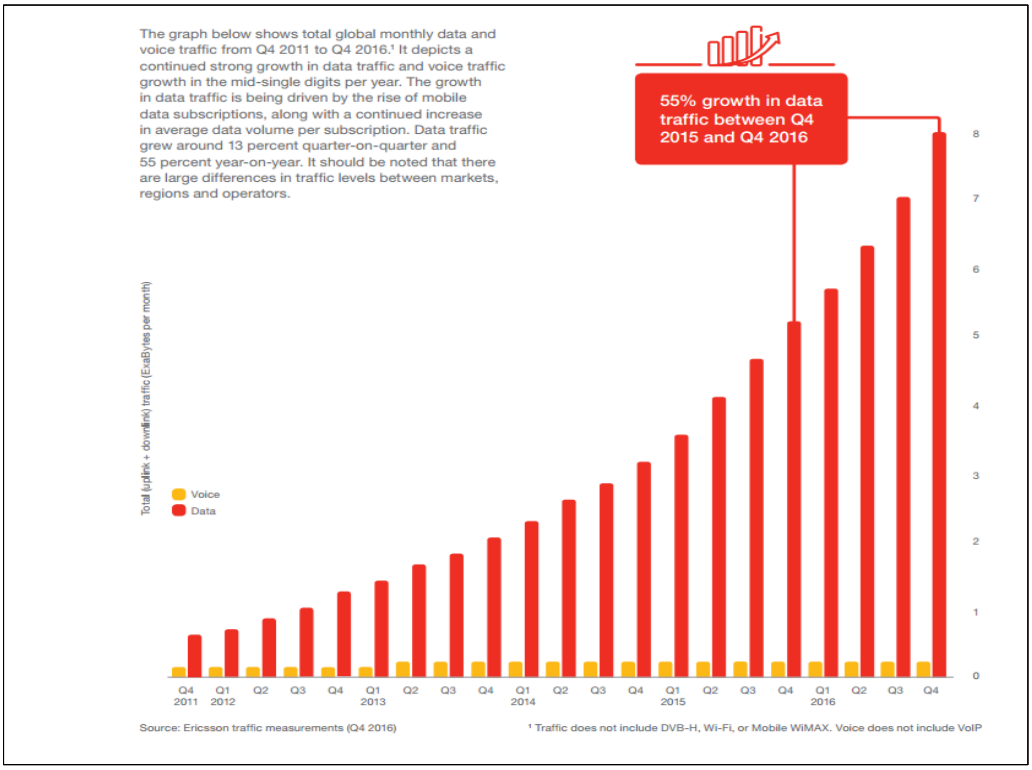


Figure 1 Voice And Data Traffic In Mobile Networks Between Q4 2011 And Q4 2016 [11]

1.2 Problem Statement

From the perspective of a mobile network operator, there are two key QoS measures, namely the rate of blocked new call requests, and the rate of calls dropped in a handoff attempt due to lack of radio resources. Nowadays, in a context of traffic surge and reduced revenues, operators face the challenge of finding costless solutions to reduce both measures. Such a solution should necessarily rule out any physical expansion as a way to increase the capacity of the network and mainly devise real-time strategies to utilise the spectrum more efficiently. Some of these strategies might include (i) mobile network offload, either by employing underutilised resources from alternative operators or using other types of networks (e.g. Wi-Fi, WiMAX), or LTE network virtualisation, a technique that allows various virtual operators to use the same physical infrastructure simultaneously.

1.3 Motivation

Most of the previously reported research approaches and techniques in call admission control (CAC) have addressed the problem either degrading the quality of service to admit more calls or implementing costly solutions, such as installing additional base stations to increase capacity. In addition, typically a higher priority is given to handoff calls over new calls, which became a fundamental issue from a user perspective [12].

The spectrum dedicated to mobile communications has been licensed to different service providers or operators. Such spectrum remains to be somewhat limited, and it seems that the physical capacity required to fulfil the projected surge in mobile data traffic is higher than what the air interface can provide under the actual operating strategies.

Some large-scale spectrum occupancy measurement studies reveal an unbalanced usage of the bands licensed to those operators in several locations. Such an imbalance is mainly due to the mismatch between static spectrum allocation and dynamic spectrum demand [13-15]. This indicates that there is room for enhancing the

QoS, which will require improving not only the spectral efficiency at the radio link level but the overall network efficiency. In a few words, it is necessary to find new methods and solutions to efficiently utilise the spectrum and manage such an increase in data traffic.

1.4 Objectives

General Objective

Propose a CAC algorithm to manage a collaborative strategy that combines network virtualisation and offloading to avoid network congestion and guarantee the quality of service.

Specific Objectives

- Build a mathematical model to determine the average number of resource blocks required for voice and data calls and the blocking probability of both types of calls.
- Evaluate the performance of the mathematical models by means of a simulation using MATLAB.
- Propose the CAC algorithm.
- Evaluate the performance of the proposed CAC algorithm by means of an emulation using open source applications, such as EVE, and Wireshark.

1.5 Contributions

Regulatory bodies are increasingly considering spectrum sharing through upgraded statistical multiplexing techniques. For instance, the 3GPP has looked upon data offloading, in particular using different networks (e.g. Wi-Fi) and collaboration techniques such as the coexistence (sharing) of different mobile service providers. Other alternatives for data offloading are LTE network virtualisation and Local Break Out (LBO) procedures implemented by the local operator. It is important to remark that any strategy to tackle this problem should be suitable for real-time functioning; this additional issue further complicates the solution of the problem.

In order to address the above-mentioned problem of traffic offloading, this work proposes an adaptive collaborative strategy among mobile network operators, by implementing the concept of Virtual Mobile Network Operator(VMNO) using a Virtual Network Enabler (VNE) element in conjunction with an LBO technique which is mainly the over the top service provider. To the best of our knowledge, most of the work carried out in this area has implemented a single approach.

The proposed strategy takes into consideration the possibility that one operator takes advantage of underutilised resources from other operators, as an opportunity to compensate a shortage in its resources and thus serve unbenefited traffic which may congest the core network (e.g. video stream and gaming), by diverting it to the LBO.

This should result in an essential reduction of the blocking and dropping rates, lead to better utilisation of the spectrum.

The applicability of virtualisation and offloading techniques in mobile communication technology is a relatively new area of research in LTE, on the context of collaboration and sharing. More studies are essential to assess the feasibility and the impact of such usage [16] .

1.6 Research Methodology

The research methodology comprises the following phases:

- Phase 1: Review the relevant literature

The literature review focused on topics related to this research, which include mobile networks standards (e.g. LTE and LTE-Advanced), elements of wireless virtualisation (Mobile Virtual Network Operator and Virtual Network Enabler (MVNO, VNE), Infrastructure Provider (InP), Service Provider (SP)), infrastructure and air-interface resources' sharing, LBO, data offloading methods (LIPA, SIPTO, Wi-Fi offload) and its benefits, and CAC algorithms.

- Phase 2: Build a mathematical model

A mathematical model is built to describe the on-demand resource sharing (via virtualisation/offloading) for voice and data calls to operators. This model should include the calculation of average resource blocks required for voice and data calls, and the blocking probabilities of both types of calls using a Poisson distribution model. Also, a mobility model of users should be considered.

- Phase 3: Set up the simulation (first stage) and the emulation (second stage)

Regarding the first stage, MATLAB will be used to verify the proposed framework. This comprises an LTE simulator, to determine the average number of LTE Physical Resource Blocks required for voice and data calls, and a Poisson Distribution simulation, to determine the overall voice and data call blocking probabilities of the CAC algorithm which is a part of the framework for all operators in the air interface level.

- Phase 4: Evaluate the proposed CAC algorithm

The simulation will be used to verify the mathematical model through a numerical example, while the emulation will be used to validate the performance of the proposed framework.

- Phase 5: Document the work

The complete work is documented in a thesis structured in four chapters, as described below.

Figure 2 illustrates how the phases are sequentially carried out.

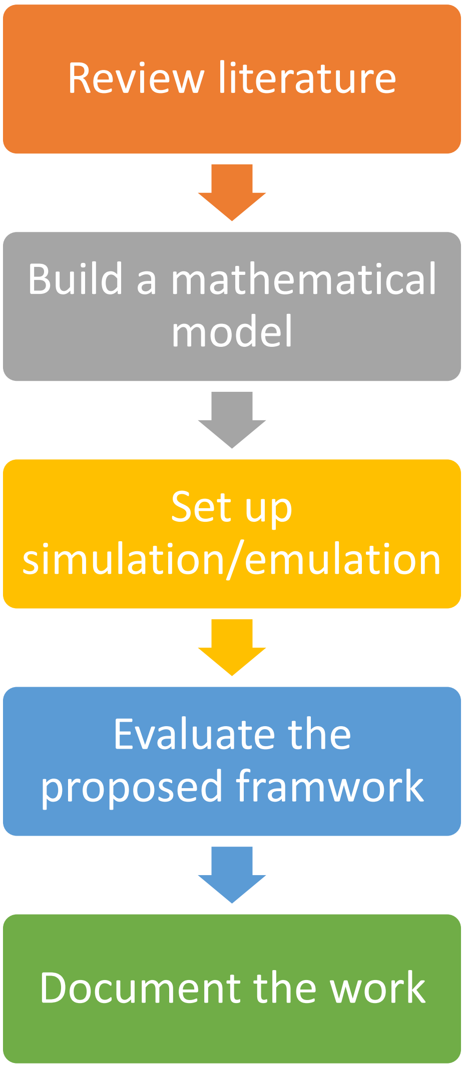


Figure 2 Research Methodology Phases

1.7 Thesis Structure

The rest of the thesis is organised as follows:

Chapter 2 gives the underpinning concepts related to the topic of the thesis. It starts describing the evolution of the generations in mobile networks, emphasising in LTE and LTE-Advanced. Then, it describes offloading techniques such as virtualisation and LBO, highlighting business models of Mobile Virtual Network Operators (MVNO), Virtual Network Enablers (VNE), Infrastructure Providers (InP) and Service Providers (SP). At last, it gives concepts related to the CAC, in particular, pros and cons of most of the CAC strategies and the challenges in CAC designs, including QoS parameters together with QoS of bearer in the evolved packet switch (EPS) network of the LTE and LTE-advanced system the End-to-End service delivery in LTE/LTE-Advanced architecture.

Chapter 3 describes the mathematical model within the proposed framework for on-demand sharing of air interface resources among mobile operators. This model constitutes the support of the applications that will be used to emulate real traffic situations in mobile networks in different scenarios. The average numbers of physical resource blocks (PRB) required for voice and data calls are computed using an LTE simulator. Then, the required PRBs are allocated both in the standard and virtualisation scenarios. In addition, a Poisson Distribution Simulator (PDS) is used to compute the voice and data calls blocking probabilities in the virtualisation scenario. At last, the Random Waypoint mobility model, which is used in the simulation, is explained.

Chapter 4 presents the results obtained from simulations, which were carried out in two stages. In the first stage, MATLAB was used to run a numerical example for the purpose of verifying the mathematical model and the proposed framework. The second stage consisted of using commercial applications for emulating real traffic in the network for different scenarios, to validate the performance of the proposed framework utilising Offloading and Virtualisation.

Chapter 5 gives the overall conclusion of the thesis, highlighting all the main points and achievements. Finally, an outlook concerning future work is given.

1.8 Literature Review

Network virtualisation enables multiple network operators to share a common physical infrastructure (including core network, transport network and access network), in order to simultaneously reduce implementation costs and improve the overall performance. In recent years, substantial research efforts have focused on building virtual networks above the same physical infrastructure towards the future of the Internet [10]. Mobile network virtualisation is an essential technique which has recently attracted more research attention. For instance, Zaki [16] proposed an LTE virtualisation scheme, mainly focusing on addressing the benefits (in terms of capacity) that can be achieved by sharing the spectrum resources between different mobile network operators; more practical scenarios were investigated in [17].

From another perspective, this work searches for further improvements to take advantage of the benefits of virtualisation regarding radio resource management (RRM) elements, such as the load-balancing technique [18], and modulation and coding schemes (MCS) [19]. A framework for wireless resource virtualisation in LTE is proposed, which allows the MNOs to modify the scheduling policies to cope with the services required among the users and the business model, by formulating the virtualisation problem as a binary integer programming (BIP) problem. An algorithm is presented to solve the BIP with less computational overhead. All in all, the outcome shows the efficient use of the resources among the operators and preserves the sharing condition agreement. In [20] a collaborative spectrum-sharing framework was proposed with minimal modifications in the radio resource manager (RRM). Motivated by the virtualisation technology, a framework based on the 3GPP LTE is developed, in which the spectrum can be used by more than one operator in air interface level, taking into consideration the enhanced intercell interference cancellation (eICIC) feature, which is used as an isolation technique in the proposed virtual RMM. The network simulation parameters are based on the power and the radio propagation characteristics. The result verified the benefit of sharing compared to a standalone operator.

Similar work was applied in [21] to limit the mobile virtual networks (MVNs) embedded in the physical network, and to make use of the long- and short-term physical resource allocation leasing, with the formulation of the MVN admission control problem as a robust optimisation problem. Then the two-stage admission control has been converted to convex problems which have been solved effectively according to the simulation results and confirmed the usefulness of the proposed scheme.

As a result, the importance of considering the RRM elements is a productive part of the mobile network virtualisation process. In this context, CAC strategy is one of the essential elements of radio resource management, similarly to the load balancing (LB) techniques and the modulation and coding schemes (MCS). Therefore, few works in the literature have dealt with the resource allocation problem within the wireless virtualisation environment. For this reason, the problem of resource allocation is a challenging issue, when it comes to the slicing and assignment process of physical resources to the mobile virtual network operators (MVNOs) to fulfil the dynamic needs of users, while satisfying the requirements of efficient resource allocation.

In general, there are two types of implementation scheme for resource allocation in wireless virtualisation [10]. In the first type, the infrastructure provider (InP) or mobile network operator (MNO), performs the central role in allocating the physical resources to users of different MVNOs according to specified requirements (e.g. pre-determined resource-sharing ratios). In the second type, the MVNOs are also involved in the resource allocation process to their end-users, making such allocation a hierarchical problem. In this case, the InP is only responsible for allocating the resources to each MVNO through the virtual network enabler (VNE), while each MVNO manages the resource allocation for its own users. Most of the existing work on resource allocation for wireless virtualisation can be categorised into the first type. Specifically, optimisation-based dynamic resource allocation schemes were proposed in [22-24], and a stochastic game-based scheme was proposed in [25]. In [22] an algorithm is developed to support the optimisation process for the shortest virtual path embedded in the network, in order to allow the efficient sharing in MVNOs by utilising a common infrastructure. The proposed algorithm considers unpredictable parameters such as user mobility and data traffic utilisation, to allow effective mapping for mobile networks. The performance has been evaluated comparing it with other existing works based on fixed parameters. Numerical results show that the proposed algorithm can yield its purpose, by effectively managing and controlling the adjustment between utilisation and no utilisation of resources. Gao, L in [23] state that most of the work done in management techniques for wireless virtualisation focus in isolation, partitioning and resource, and just few researchers tackle the area of the business model between the service providers (SPs) and the infrastructure providers (InPs). As a result, a framework has been developed which consists of pre-set parameters and an algorithm based on game theory, namely the Vickrey-Clarke-Groves (VCG) mechanism, to benefit from resources utilisation by modelling an auction game between the InPs and SPs, which are the MVNOs in this case. Zhang ,Zhao, ,Lopez, & Chen in [24] cited the importance of resource allocation in orthogonal frequency division multiple access (OFDMA) systems, which led to the efficient use of the energy because the focus was mainly placed in the power consumption and in maximising the data rate using a virtual resource allocation algorithm between the MNOs and MVNOs. The simulation shows that the performance in energy usage has been increased by 50 percent for the virtualisation approach, compared to the normal approach.

In [25] a new framework of wireless virtualisation is used to decouple the responsibilities of network entities. For instance, the network operator is responsible for resource allocation and the SPs for QoS. In addition, a VCG mechanism is used to verify that spectral efficiency can be reached, while the RRM and QoS provisioning can be decoupled from each other.

The proposed schemes and frameworks in these works can achieve high resource utilisation. On the other hand, a virtual embedding network algorithm was proposed in [26], based on opportunistic spectrum sharing allocation. However, since the VMNOs are not involved in the resource allocation, the capability of intra-slice customisation for each MVNO cannot be easily achieved. Besides, the computation complexity for InP is high, considering that the optimal resource allocation should be directly obtained for all users.

A few works considered the problem of resource allocation to VMNOs. For instance, an opportunistic sharing-based resource allocation scheme was proposed in [27], also driven by the virtualisation concept in wireless networks to overcome the shortage in spectrum by utilising it in more efficient and optimal manners. First, the problem of resource allocation has been formulated as an “NP-Hard integer program”. Then, two algorithms have been proposed: (i) a dynamic algorithm which mainly deals with resource sharing, by determining the most suitable process to be implemented to increase the revenue and resource utilisation rate, while reducing the cost of virtual networks and (ii) a heuristic algorithm to afford a simple but effective process which is easy to execute. At last, simulation results show the benefits of the proposed scheme. On the other hand, a bankruptcy game was proposed in [28] for dynamic wireless resource allocation among multiple operators. However, in this work, the users were not involved.

Moving towards fifth-generation mobile networks, [29] proposed a wireless virtualisation scheme, which is based on a hierarchical combinational auction mechanism, and dealt with the resource allocation problem and the roles of the infrastructure providers (InPs) and mobile virtual network operators (MVNOs); they also consider the massive MIMO technique as a key enabler for 5G networks. In summary, most of the existing work on wireless virtualisation does not consider offloading techniques as an active element of an alternative to face the tremendous demand in traffic, and the associated degradation of the service quality of mobile operators. The primary goal of offloading is to avoid congestion in the core network, by redirecting the hungry bandwidth applications (e.g. video streaming/gaming) and low-priority traffic to an alternative low-cost path [3]. Furthermore, solutions that focus on network capacity alone are not really tackling the root of the problem [30, 93].

On the other hand, the offloading strategy has also been used to decongest a network, by partially transferring the traffic to other types of networks, such as WiFi and WiMAX, among others. In [31] the authors proposed an algorithm for offloading LTE networks to WiFi networks, while in [32-34] show advances in the use of offloading to improve QoS in congested LTE networks. Now, as was previously mentioned, most of the reported works employ either virtualisation or offloading, i.e. they have been used separately.

Table 1 classifies some reported work among virtualisation or offloading. These works propose a collaborative strategy which combines virtualisation and offloading to improve QoS.

There are some recent work is done in this capacity ,by proposing a virtualisation, offloading strategies too, for improve QoS and manage resources in a collaborative way , For example in [76] by implementing the cloud technology .Author's in[77,78] represent the virtualised SDN network .From comparison point view ,the disadvantages of these work is the complexity . For instant, virtualised SDN network can deal with many applications, services, different plate forms, technology and awareness capability of the end-user's behaviours and yet, its sill promising on going technology with the upcoming 5G. The same thing goes with the cloud technology. More details are given in the future work section, to support the next generation 5G.

Table 1. Classification of Previously Reported Works Among Virtualisations and Offloading Techniques

Paper authors	**Virtualisation**	**Offloading**
C. Liang and F. R. Yu	X	
M. Kalil, A. Shami, and Y. Ye	X	
X. Wang, P. Krishnamurthy, and D. Tipper	X	
C. Liang and F. R. Yu	X	
G. Chochlidakis and V. Friderikos	X	
L. Gao, P. Li, Z. Pan, N. Liu, and X. You	X	
Y. Zhang, L. Zhao, D. Lopez-Perez, and K. Chen	X	
F. Fu and U. C. Kozat	X	
M. Yang, Y. Li, D. Jin, J. Yuan, L. Su, and L. Zeng	X	
K. Zhu and E. Hossain	X	
Y. Zaki, L. Zhao, C. Goerg, and A. Timm-Giel	X	
K. Samdanis, T. Taleb, and S. Schmid,		X
Aiping Huang et al.		X
Mojdeh Amani, et al.		X
Pavel Masek et al.		X

Chapter 2

2 Underpinning Concepts

2.1 Introduction

The bandwidth requirements in mobile networks have significantly increased in recent years. In the last few years, the usage time of smartphones has grown by 75 % and the usage time of tablets exhibits a similar trend [31].

In this context, mobile operators have been urged to offer higher transmission velocities, while simultaneously keeping the QoS to an increasing number of users. This necessarily involves increasing the capacity of the system to accept and maintain voice and data calls.

In addition, this situation has also motivated the research in the area, and the 3GPP has successively increased the capacity of the system developing the second (GSM+GPRS), third (WCDMA+HSDPA) and fourth (LTE and LTE Advanced) generation of mobile networks, and currently the fifth generation.

Another expected innovation is the Internet of Things (IoT), which will incorporate several devices (such as vehicles, appliances, among others) to the network. This will also require new alternatives (technologies, algorithms and configurations) that increase the capacity of the system. One of these alternatives is the Virtualisation, which implements different mobile networks using the same physical infrastructure. Other strategies to avoid network congestion include offloading and LBO.

Now, the implementation of these strategies demands more complex algorithms to control the traffic flow in the network. This work proposes a CAC algorithm to manage the whole process on an LTE network.

2.2 LTE and LTE-Advanced

Beginning in 1921 in the United States, the Constabulary Department Experimental Mobile Radio Communications started operations above the AM radio broadcast band. In June 1946, in Saint Louis, AT&T and Southwestern Bell introduced the first American commercial mobile phone service (typically in vehicles), the basic concept of mobile phones was developed as early as 1947 [38]. In the mid-1960s, the Bell system introduced the improved mobile telephone services (IMTS), which markedly improved the mobile telephone systems [38]. An antenna was installed on top of the Southwestern Bell's central office, for paging mobiles and providing wireless cell phone traffic. With the installation of that small cell with the technique of frequency reuse, the traffic capacity was substantially increased; however, the technology did not exist at that time. The concept was later perceived in the sixties and seventies and was then used to improve the capacity and efficiency of mobile systems. The mobile network is equipped with a base station and several radio channels which assigned according to the transmission power constraints and the bandwidth available of each cell. A channel can be a time slot, frequency slot or a code sequence. Any user equipment (UE) within the cell area can interconnect through a channel, by establishing a radio link with the base station that communicates with the Mobile Switching Centre (MSC), which is coupled to the Public Switched Telephone Networks (PSTN). From the late 1980s until now, an enormous interest emerged in mobile systems which promised higher capacity and higher quality of services at reduced costs, mainly driven by the improvement of digital and micro-processing computing technologies. Historically, mobile technology has undertaken four evolution stages or generations.

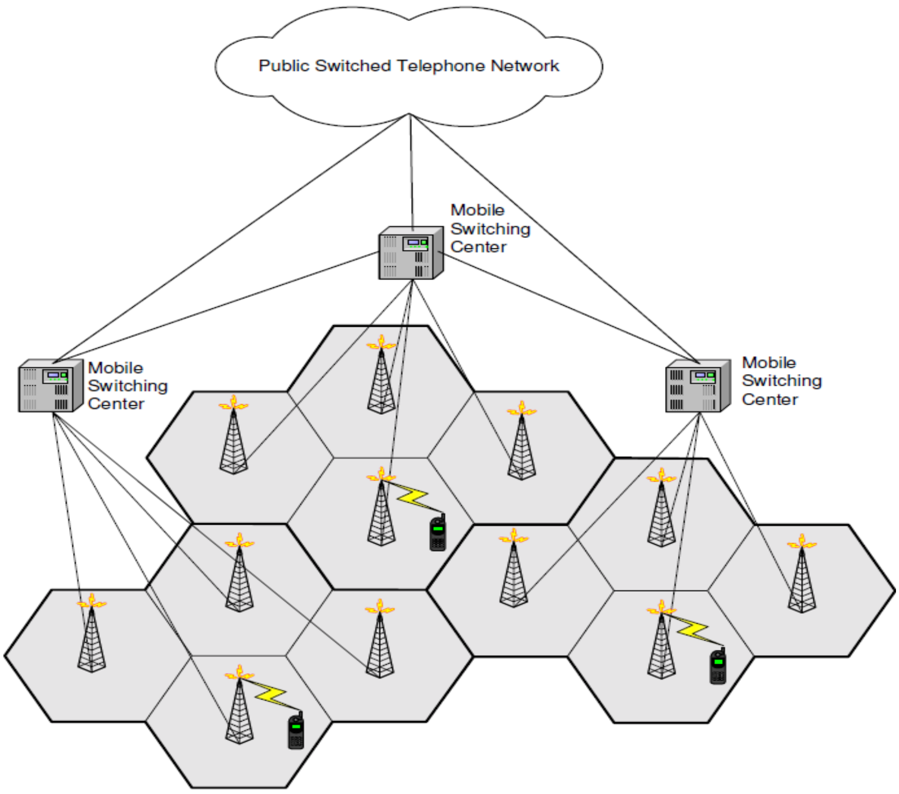


Figure 3 A High-Level Mobile System Architecture In 2nd And 3rd Mobile Generation [38]

The first generation mainly involved analogue mobile systems and the primary service delivered was voice [39]. The second-generation mobile systems apply digital technology to furnish a better quality of service for both voice and data, but data was still limited. The third generation appeared to fill this gap, offering higher system capacity, multimedia transmission, global roaming across a homogeneous wireless network, and bit rates ranging from 384 kbps to several Mbps. Figure 3 shows the architecture of second and third generation networks. At last, the fourth-generation wireless networks, which have attracted a growing interest of the market and the research community [12,39], provide global roaming across heterogeneous wireless and mobile networks. QoS provisioning in wireless networks is an arguing problem due to the scarceness of wireless resources (i.e. Bandwidth), and the mobility of users. Figure 4 illustrates the evolution of the different generations of mobile networks. In addition, Figure 5 illustrates the evolution of the system architecture from GSM and UMTS to LTE.

In order to avoid network congestion and QoS degradation for the served users, a CAC mechanism became a necessity, since it can limit the access to the network resources based on the availability and simultaneously can provision the QoS of the service provided or will be provided to the end users [38]. According to the network layer architecture, different QoS parameters are involved in different layers.

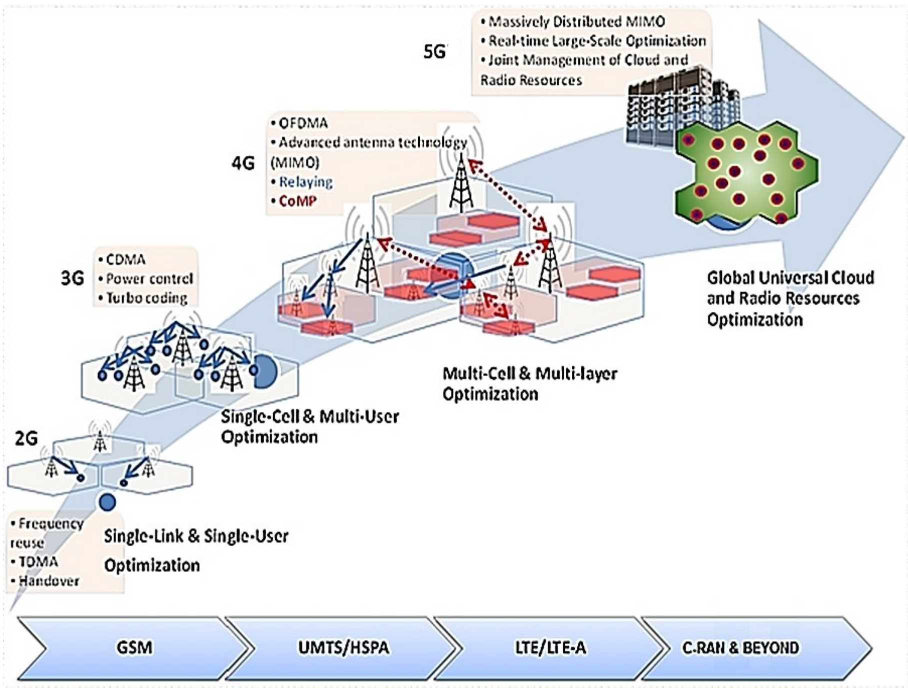


Figure 4. Evolution of Generations in Mobile Networks

Motivated by the ITU's requirements for IMT-Advanced, 3GPP started to enhance the capabilities of LTE by addressing the specification of a new system known as LTE-Advanced [7], defined mainly as a product of the operators which makes its deployment more favourable. LTE-Advanced was designed to deliver a peak data rate of

1 Gbps in the downlink and 500 Mbps in the uplink [7,40], and in the long run deliver peak data rates up to 3000 and 1500 Mbps respectively, using a total bandwidth of 100MHz that is made of five bandwidth components of 20MHz [40]. The specification also includes aims for the spectral efficiency in specific scenarios. LTE-Advanced is designed to be backwards compatible with LTE, i.e. an LTE mobile terminal can operate with a base station that is operating LTE-Advanced and vice-versa [41]. The introduction of Relay Stations (RS) and other techniques are the major differences between LTE and LTE-Advanced [42,43]. The use of OFDMA, MIMO and HARQ technologies allow LTE-Advanced to configure its bandwidth according to available frequencies dynamically and can support high mobility environments such as 350km/h in the case of high-speed rails [42].LTE-Advanced networks enable enormous flexibility in higher data rate provision and better QoS guarantee, but at the same time it is more complicated to gain precise quantitative insight into the capability of the system under different service provision scenarios, especially for practical LTE-Advanced network deployment, which requires professional capacity planning for improved user experience and reduced cost [43].

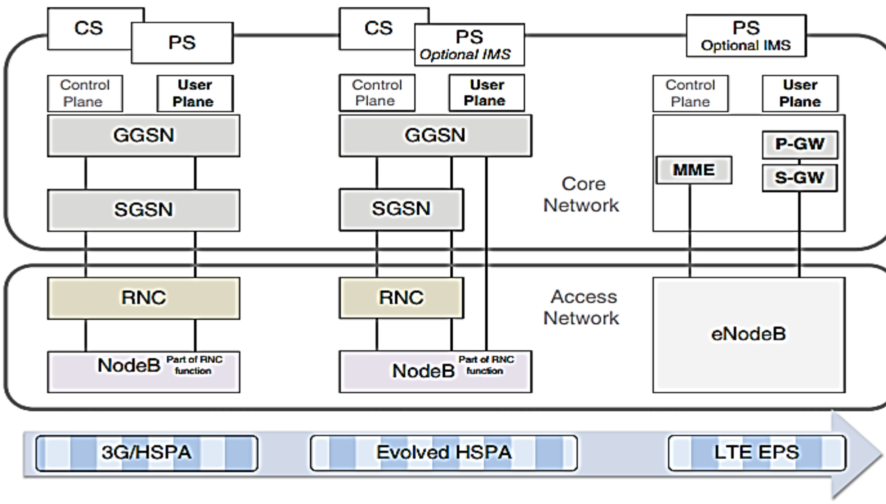


Figure 5 Evolution Of The System Architecture From GSM And UMTS To LTE [44]

Along with improvements in the access technology for LTE, the overall system architecture of both the Radio-Access Network (RAN) and the Core Network (CN) was revisited. This also means that the split of functionality between the two network components was reconsidered. This work was known as the *System Architecture Evolution* (SAE), which resulted in a flat RAN architecture, while the *Evolved Packet Core* (EPC) emerged as the new core network architecture [45].

The RAN handles radio related functions including radio-resource handling, retransmission protocols, scheduling and control of various multi-antenna schemes. The EPC provides a complete mobile-broadband network by means of functions such as authentication, charging functionality and setup of end-to-end connections [45]. Handling these functions separately, instead of integrating them into the RAN, is beneficial as it allows for several radio-access technologies to be served by the same core network. For example, the Radio Access Networks for 3G-HSPA and 4G-LTE could be served by the same core network [45].

Note that the Evolved Packet Core supports access to the packet-switched domain only, with no access to the circuit-switched domain. All voice and data accesses are done in the packet-switched domain only. The EPC consists of several types of nodes, which are described below.

The Mobility Management Entity (MME) is the control-plane node of the EPC, which handles the connection/release of bearers to a terminal. The functionality operating between the EPC and the terminal is sometimes referred to as the Non-Access Stratum (NAS), as opposed to the Access Stratum (AS) which handles functionality operating between the terminal and the radio-access network [45].

The Serving Gateway (S-GW) is the user plane node connecting the EPC to the LTE RAN, which acts as the mobility anchor when the UEs move between eNodeBs in the LTE network, and even between other 3GPP technologies and the LTE network. This means that the S-GW will be the same even as the eNodeBs change due to the mobility of a terminal, i.e. the S-GW acts as an anchor for the mobility of the terminal.

The Packet Data Network Gateway (PDN Gateway, P-GW) connects the EPC to the internet. The P-GW handles allocation of the IP address for a specific terminal. The Policy and Charging Rules Function (PCRF) is responsible for QoS handling and charging, and the Home Subscriber Service node is the database containing subscriber information [45].

The network architecture of the 4G, 3.5G and 3G systems along with femtocell deployments is shown in Figure 6.

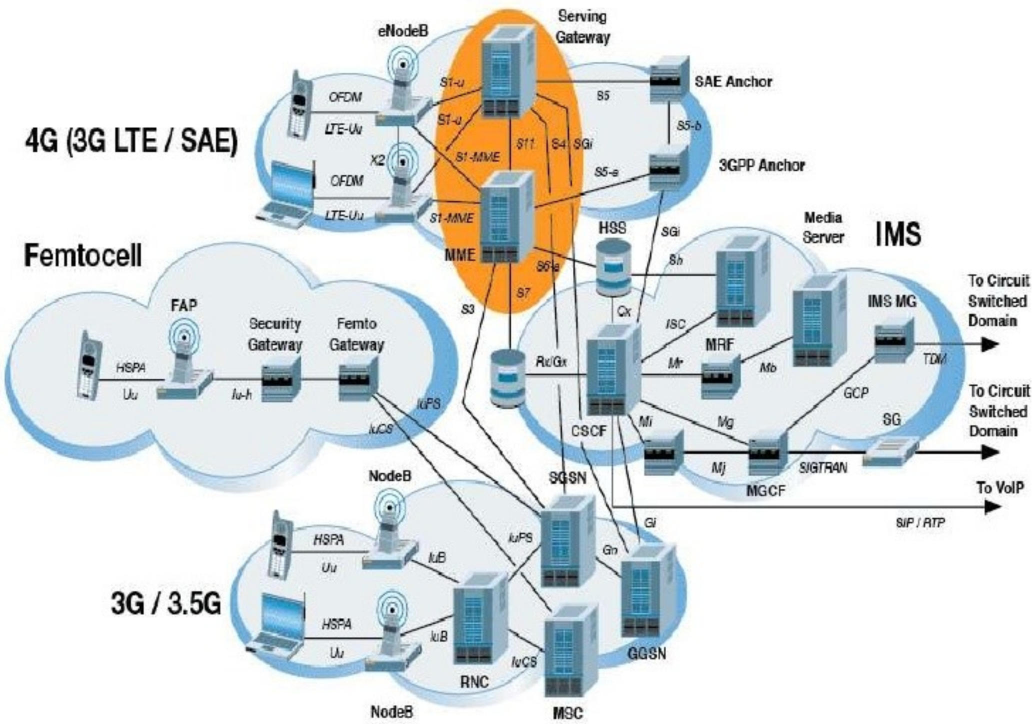


Figure 6 System Architecture Evolution [46]

The salient features of this evolution are [46]:

- System Architecture Evolution (SAE) is the core network architecture of the 3GPP's LTE wireless communication standard
- SAE is the evolution of the GPRS Core Network for LTE, with some differences:
    - Simplified architecture
    - An all IP Network (AIPN)
    - Support for higher throughput and lower latency radio access networks (RANs)
    - Support for multiple heterogeneous access networks, including E-UTRA (LTE and LTE-Advanced air interface), 3GPP legacy systems (GERAN or UTRAN), also non-3GPP systems like WiMAX. The mobility of the terminal between above systems is supported.

2.2.1 LTE Frame Structure

The content of this section is mainly based on [47]. There are two types of a frame structure in the LTE standard. Type 1 uses Frequency Division Duplexing (FDD) (uplink and downlink separated by frequency), and Type 2 uses Time Division Duplexing (TDD) (uplink and downlink separated in time). This section covers both LTE FDD Type 1 signals and LTE TDD Type 2 signals described in the LTE standards. First, an introduction to some of the terms used in describing an LTE Frame is given. There are six-time units: frame, half-frame, sub-frame, slot, symbol, and the basic time unit (T_s), as shown in Table 2.

The smallest unit of resources that are allocated to a user is known as a resource block (RB). In frequency the RB is 180 kHz wide, being either 12 x 15 kHz or 24 x 7.5 kHz subcarriers wide. In the time it is 1 slot long. Frequency units are expressed as several subcarriers or RBs. For example, as shown in Table 3, 1.4 MHz bandwidth in Downlink could be described as 6 RBs or 73 subcarriers wide [47].

The underlying data carrier for an LTE frame is the resource element (RE). which is the smallest part of the frame with 1 subcarrier x 1 symbol, and contains a single complex value representing data from a physical channel or signal (See Figure 7).

Table 2 LTE FDD Frame Boundaries [48]

Time Unit	Value
Frame	10 ms
Half-frame	5 ms
Subframes	1 ms
Slot	0.5 ms
Symbol	(0.5 ms) / 7 OFDM symbols for normal Cyclic Prefix (0.5 ms) / 6 OFDM symbols for extended Cyclic Prefix
T_s	1/ (15000 × 2048) sec » 32.6 ns

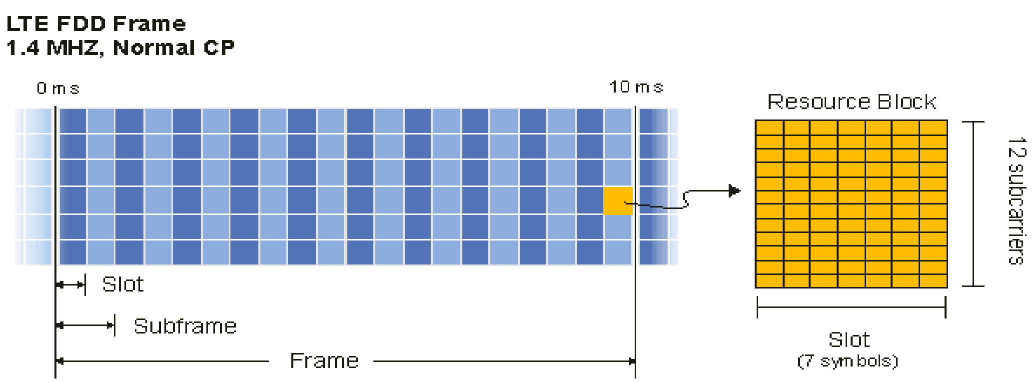


Figure 7 LTE FDD Frame [49]

2.2.2 Bandwidths

The bandwidths of an LTE channel defined by the standard are 1.4, 3, 5, 10, 15, and 20 MHz. Table 3 shows how many subcarriers and resource blocks are there in each bandwidth for uplink and downlink.

Table 3 Bandwidth vs No. of subcarriers [49]

Frequency measures			
Bandwidth	**Resource Blocks**	**Subcarriers (downlink)**	**Subcarriers (uplink)**
1.4 MHz	6	73	72
3 MHz	15	181	180
5 MHz	25	301	300
10 MHz	50	601	600
15 MHz	75	901	900
20 MHz	100	1201	1200

2.2.3 Road to 5G

3GPP continues to expand the LTE platform to new services while improving its efficiency to meet the increasing mobile broadband demand. To address the expanded connectivity needs of the future, 3GPP has started to work on the standardisation of the next generation mobile technology, also known as the fifth-generation technology or 5G [50].

The 5G standard is looking to become a rather large, all-encompassing wireless communication system that not only caters for faster data speeds but also supports many more interconnected devices online at the same time with greatly reduced latency [51].

Although there is not a definitive standard yet for a 5G technology, the groups working on the early trials have defined several key requirements going forward. Here are some of the most important ones [46]:

- 1Gbps to 10Gbps connections for peak data rates
- 100Mbps cell edge data rate (mobile data speeds)
- 1-millisecond end-to-end latency
- 1000x bandwidth per unit area
- 10-100x number of connected devices
- 90% reduction in network energy usage
- Full coverage.

Despite the advancements in technologies that enable high data rates over the wireless medium, the demand for data keeps increasing. Although more and more mobile spectrum is being allocated to operators to meet the customer’s insatiable expectations of bandwidth, it keeps falling short. Offloading of data (for example, via Wi-Fi) provides a way to meet this demand. Mobile devices have built-in Wireless LAN transceivers, that can connect directly to the Wi-Fi access points when mobile access is unavailable or is congested, resulting in slow download speeds. Data offloading via Wi-Fi or via femtocells helps decongest data traffic. A drawback of offloading is that a mobile operator cannot monitor data traffic that does not go through its core network [46].

2.3 Virtualisation

This section summarises the virtualisation technique applied to the LTE system. The virtualisation concept was first mentioned in 1959 when Strachey published his paper [16], which was focused on the concept of a multi-programming concept.

Formerly, in the mid-1960s, the project IBM M44/44X was introduced, where the expression Virtual Machine (VM) was announced for the first time [51]. Virtualisation is the process of creating virtual layers of physical resources that duplicate the same physical characteristics. It is frequently deployed in the information technology field to convert the physical resources into some virtual parts, for instance, virtual memory, partitioning the hard disk, virtual machine.

Today, Server Virtualisation refers to the creation and maintenance of virtual machines. Since computers were very expensive in the past, at the beginning of this project the aim was creating several virtual machines out of one mainframe computer, to enable multi-task processes such as running applications and processes at the same time on one computer. Network virtualisation is the procedure of joining different virtual network resources into a Virtual Network. Individual virtual networks can contain operator specific protocols and architectures, which could be totally different from other coexisting virtual networks [52].

Moreover, network virtualisation also offers full flexible end-to-end control for the operators over their virtual networks [16]. Many research activities are focusing on the Future Internet architecture, have been launched around the world, such as (FP7 4WARD PROJECT) in Europe [53], VINI (Virtual Network Infrastructure) [54] in the U.S. and AKARI (Architecture Design Project for New Generation Network) [55].

The virtualisation concept can be applied to different broad areas. However, from an information technology perspective, there are three main areas in which virtualisation techniques can be used [52]: network, storage and server.

In this work, a novel virtualisation collaborative structure will be adapted to allow mobile network operators to share the mobile spectrum and the infrastructure, such as the eNB hardware entity. The proposed structure targets the concepts of wireless virtualisation applied within the 3GPP LTE system, which represents one of the latest mobile communication systems technologies that are yet entering and still developing in the telecommunication market. Thus, such a technique will be used in an LTE network simulator with some changes, to cope with the proposed framework.

2.3.1 Requirements for Wireless Virtualisation

The requirements for carrying out wireless network virtualisation are the following [10]:

1) Isolation: Network virtualisation means the creation of individual virtual networks on top of shared physical network resources. These networks should behave independently of each other, i.e. they should be isolated. Security issues in one virtual network should not affect the other virtually created networks. A malfunctioning virtual network could end up consuming most of the resources of the underlying physical infrastructure of network elements. This should be prevented by having a limit of the resource consumption by individual virtual networks. In the wireless scenario, isolation would involve monitoring interference between the virtual networks, which makes isolation more complex in wireless networks compared to the wired counterparts.
2) Programmability: Network virtualisation should support both control and data plane programmability, to provide flexibility and the possibility to evolve networks using new control schemes and new data processing capabilities. The created virtual

networks should be programmable by their users; in case the users are service providers, they should be able to manage configuration and allocation of virtual networks, e.g., a routing table and virtual resource scheduling among others. Each logical isolated network partition (virtual network) should support the free deployment of control schemes or network architecture independent of other virtual or physical networks.

3) Coexistence: All virtual networks should be able to coexist on the same physical substrate network. Multiple virtual networks will have different QoS requirements, topology, security level, the behaviour of users etc.
4) Other requirements: Wireless virtualisation has some unique characteristics such as limited resource usage, signal interference, etc., that do not appear in wired networks. One of the biggest challenges is virtualisation of the wireless links because establishing a wireless link requires configuring parameters of the air channel between transmitter and receiver, such as the channel of operation, appropriate setting of transmitting power and receiver sensitivity, among others. In order to create two separate virtual networks that coexist on the same hardware, communication activities from one virtual network should not affect reception behaviour on the other virtual network in any form.

2.3.2 Virtualisation in Mobile Communication

Mobile Virtualisation has not yet received the appropriate attention, and there have been few works in this area [16]. Applying virtualisation techniques in mobile networks in order to share the limited resources, should lead to more effective utilisation of such resources. Furthermore, network virtualisation can reduce the cost by reducing the number of base station deployments, thus decreasing energy consumption and the overall investment and operational costs [16]. Network virtualisation also gives the opportunity that small mobile operators can join the market and provide new services to their customers, by using existing physical infrastructure.

Furthermore, the idea of being able to share the frequency resources among multiple operators is a very interesting issue, which gives them full control to scale up/down the infrastructure and spectrum resources they use [16,52]. Thus, there are three main requirements to enable the virtualisation of mobile networks [19]:

1. Isolation between Mobile Network Operators (MNOs): Isolation is the capability of preventing the impact of one MNO on another, despite them sharing the same physical substrate. For instance, any modification in the traffic load or channel quality for any specific operator should not affect the others.
2. Customisation: MNOs have the ability to implement different custom scheduling policies, aiming to maximise profits and fulfil user requirements. Thus, different MNOs may have different services, QoS requirements, and business models. Radio resources are allocated to users based on scheduling policies, and every MNO have to be offered the flexibility to implement its own scheduling policy to achieve its objectives.
3. Efficient radio resource utilisation: Efficient use of the radio resources have to be maintained to the possible extent.

2.3.3 Business Model of Wireless Virtualisation

In wireless network virtualisation, physical resources are owned by some parties, and virtual resources are utilised by some other parties. Business models can describe the constitution of the roles in the wireless network market and the main functions of these roles [10].

Generally, there are two logical roles following wireless network virtualisation, namely Mobile Network Operator (MNO) and Service Provider (SP) ([56-58]). All the infrastructure and radio resources of the physical substrate of the wireless network, including the licensed spectrum, radio access networks (RANs), backhaul, transmission networks (TNs), and core networks (CNs), are owned and operated by MNOs.

MNOs execute the virtualisation of the physical substrate networks into some virtual wireless network resources, which, for brevity are called virtual resources. SPs lease operate and program these virtual resources, in order to offer end-to-end services to end users. In some papers (e.g., [10]) the MNO becomes Infrastructure Providers (InP), which is only responsible for owning and leasing wireless network resources to SPs, who create and deploy virtual resources by themselves, based on the leased and allocated resources, to satisfy the requirements of end-to-end services.

The roles in the below business models are further decoupled into more specialised roles, including InP, a mobile virtual network provider (MVNP) and in some references such as in [10] refer to it as virtual network enabler (VNE), the mobile virtual network operator (MVNO) and SP. The functions of each of them are described in the following [10]:

- InP: Own the infrastructure and wireless network resources, i.e. they own the spectrum, base stations and other network elements. Spectrum resources may or may not be owned by the InP.

- MVNP: Responsible for the creating process of the virtual resources and the leasing process of the network resources, because in some regulation cases the MVNP have the rights to own a licensed spectrum, so no need to request and utilise spectrum resources from the InP.
- MVNO: Operates and assigns the virtual resources to SPs. In some approaches, MVNOs performs the roles of both MVNOs and VNE. This model fits the emerging concept of the so-called providers of Anything-as-a-service (XaaS) [42] in cloud computing. InPs provide the Infrastructure-as-a-service (IaaS), while MVNOs provide networks-as-a-service (NaaS).
- SP: Concentrates on providing services to its subscribers based on the virtual resources provided by MVNOs.

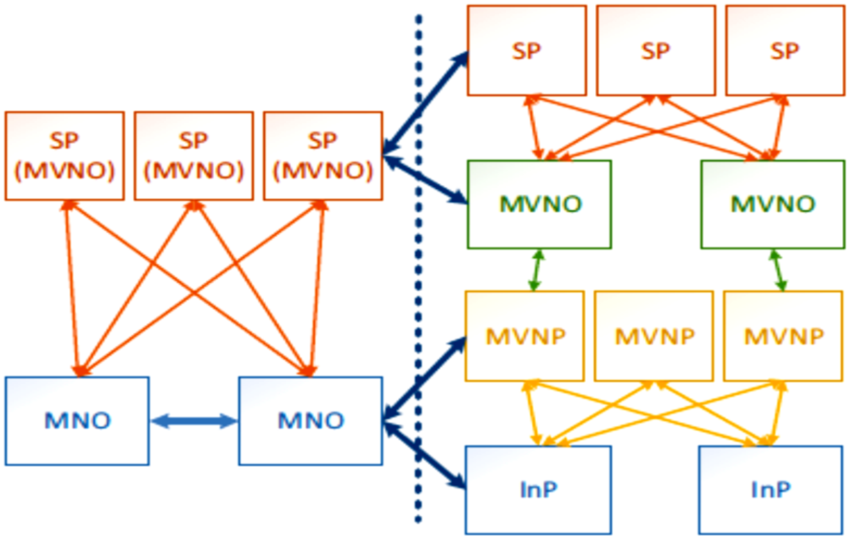


Figure 8 Business Models of Wireless Network Virtualisation [10]

Figure 8 depicts the described two different business models, showing the placement of the different elements and the relationship between them. In a few words, virtual resources are requested by SPs, managed by MVNOs, created by MVNP/VNE, and run on the physical resources owned by the InPs. Obviously, this four-level model can create more opportunities in the market and intuitively simplify the functions of each role. However, it requires more coordination mechanisms and interfaces, which may significantly increase the complexity and latency.

2.4 Offloading Techniques

Internet gaming, video and social media has become very popular on new devices such as tablets and smartphones, creating a surge of network data traffic. In addition, device to device connectivity commonly known to as machine to machine (M2M), is expected to give rise to a new set of applications that will require even more network capacity. Consequently, data traffic will grow significantly, thus urging operators to expend capacity, while data revenues are expected to only grow slowly creating a considerable gap accordingly. In this context, telecom operators need to continuously review their data traffic patterns and implement offloading mechanisms, which can help them manage their load and capacity more efficiently. Over-the-top (OTT) players are capturing a growing share of the value of the data that flows over the network. Combined with decreasing gain margins of telecommunication companies, requires that all this should be done in a more cost-efficient way. Some of the means that operators have to consider it in their network [59]:

1) **Efficient Radio Spectrum Utilisation**. The spectrum for operators is both limited as well as very expensive. To cope with the tremendous increase in data traffic, operators need to plan the effective utilisation of their radio

resources by offloading data between licensed and unlicensed spectrum, as the number of connected devices continues to increase.

2) **Controlling CAPEX**. Operators are already investing heavily in upgrading their networks, to be able to meet the growing data needs of their consumers. However, their revenues do not increase as much as the increase in the data traffic. Hence, operators need to focus on core and access network investments only in those areas that offer the most active potential returns.
3) **Backhaul Network Optimisation**. Besides causing strain to the operator radio access networks, the growth in data traffic is also generating backhaul bottlenecks. Therefore, operators also need to design and implement an efficient backhaul system to transport the data from the access to the core network.
4) **Transactional Load Management**. It is crucial for operators to keep the signalling and transactional load to a minimum so that bandwidth optimised. With the increasing number of devices, this becomes a firm requirement.

Moreover, data traffic patterns depend upon the type of device, its form factor, time of the day, type of application and even the density of users in a location. For example, in 3G devices with the largest data consumption is the routers, it can reach up to 16 GB per month. Then PCs with 7 GB per month on average, followed by tablets at 1600 MB and mobile phones at 230 MB approximately; the M2M traffic is around 10 MB per subscription. The aforementioned factors and the increasing device diversity make network capacity planning and load management even more complex [59].

Data traffic offload is an alternative that operators have at hand to reduce the traffic on their radio spectrum and lower the operating load on base stations. It also represents an opportunity for service providers to charge users to offload solutions such as small cells and help customers to reduce their usage costs by offloading data to alternate networks. In few words, a robust data offload solution can provide operators with flexibility to control data flow across the network based on traffic patterns, class of service and type of customers, thus achieving a better QoS.

There have been numerous research initiatives to explore potential solutions to efficiently utilise network resources, while simultaneously maintaining a high QoE for subscribers. Offloading is one specific family of solutions among the several possible solutions. The primary goal of offloading is to avoid transporting low priority traffic in costly networks, which may be done to avoid degrading the perceived QoE. The traffic that causes network congestion without creating any additional revenue should be targeted by mobile operators. Traffic redirection from parts of the network where congestion could occur to other low-cost parts of the network where capacity is available and less expensive is the basis of offloading. There are six different alternatives to offload data from the mobile network at either the access or the core network level. Each of them can co-exist, and the operator will have to determine which is the best based on multiple factors such as current infrastructure, customer usage patterns, associated costs, deployment and maintenance complexities and user density in a location. These alternatives are [59]:

1. Wi-Fi Hotspot

2. LTE Small Cells / Relay Nodes

3. Integrated Femto / Wi-Fi

4. Direct Tunnel

5. Internet offload Gateway (IOGW)

6. M2M Gateway

In LTE systems, the idea behind traffic offloading is to free up the costly loaded paths, such as 3GPP Radio Access Network (RAN) and the Mobile Packet Core Network (MPCN), by redirecting part of the traffic to alternate, cost-effective paths. This could also be done by enabling the direct communication between nearby UEs via a D2D solution [51]. Several alternate offloading paths are standardised within 3GPP, such as the Interworking Wireless Local Area Network (I-WLAN) that integrates non-3GPP access (Wi-Fi) with the MPCN [52].

Because of the use of Wi-Fi instead of 3GPP access, data does not have to go through the traditional expensive 3GPP RAN path, thus reducing the cost of the service delivery since Wi-Fi is license free. However, the use of Wi-Fi is accompanied by several challenges pertaining to the continuity of the session and the mobility of the service ([53], [54]). Another path at the RAN is possible to compensate the macro coverage holes when implementing small cells. Small cells are connected directly to the MPCN via the fixed network, resulting in a reduction of the load at the RAN. The base stations of the small cells are 3GPP compatible equipment, which eliminates the interoperability challenges, and thus the desired QoS is also supported. The disadvantage of using small cells is that they create interference problems with the macrocell users that should be properly handled.

2.4.1 Data Offloading Via Mobile Pathways

Direct tunnelling is a method in 3G networks, in which data flows from the base station to the radio network controller GGSN and then directly to the Internet, thus avoiding the SGSN network element. This implies that service providers utilising this offloading solution would require much lesser SGSN nodes, which would result in reduced capital and operating expenses [59].

The Internet Offload gateway solution gives the possibility to selectively offload data traffic in a 3G network, between the Radio Network Controller and the SGSN. The non-profitable applications such as streaming media and surfing the internet are offloaded from the RNC to the internet bypass the core network, while the profitable ones are pass through the core network SGSN and GGSN [59].

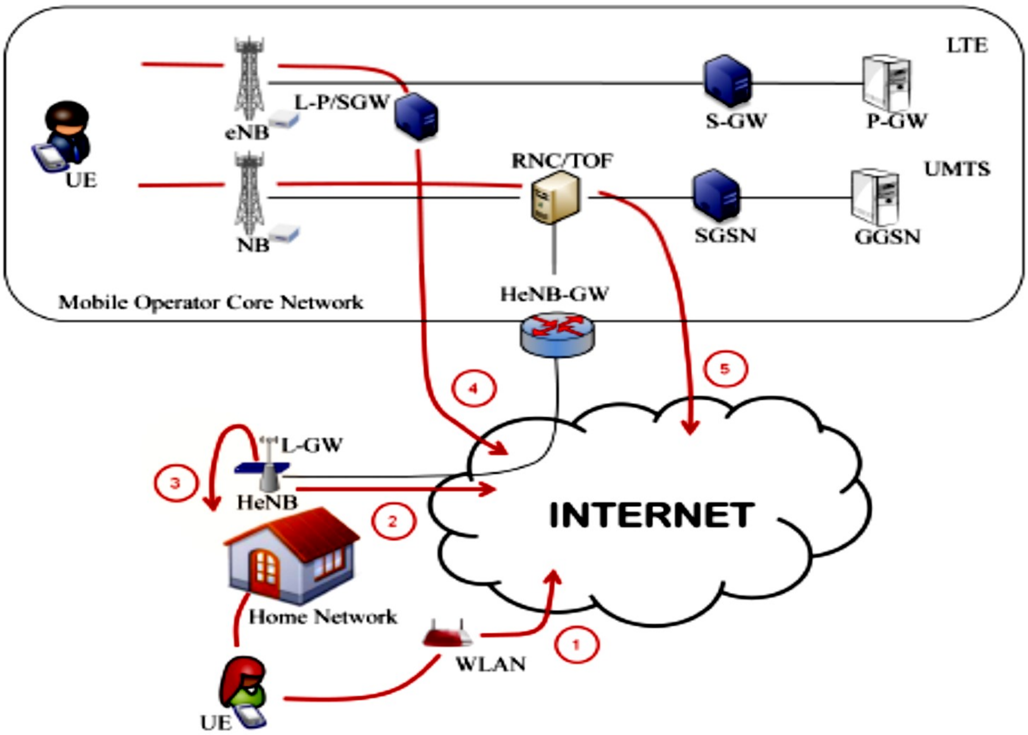


Figure 9 Data Offloading Via Mobile Pathways [59]

Implementing this solution in the network, it would not be necessary for the operator to upgrade SGSN/GGSN in order to serve the important data growth because both are offloaded. However, the challenge of a restricted radio spectrum continues, since this solution is implemented in the core network [59].

Data may be offloaded via several pathways, within the mobile network itself (through same or different operators) or outside of the mobile network (for example, through Wi-Fi). Various data offloading pathways are illustrated in Figure 9.

2.4.2 Homogeneous Deployments

Multiple mobile operators co-exist in the same geographical area to provide services to their respective users, and they may have an unutilised spectrum which may be shared. Wireless virtualisation enables the dynamic sharing of physical infrastructure and air interface resources amongst the operators, based on user demand. In this way, the operating and capital expenses of the operators are reduced significantly. The users are unaware that they may be using the resources of other operators via virtualisation, so they are seamlessly connected to the LTE network [60].

Air interface virtualisation means the sharing of air interface resources among the mobile operators, for the benefit of their users. The LTE transmission frequencies are shared so that the user calls are not blocked or dropped as they move in and out of cells. Once the air interface resources are shared (obtainable via a Mobile Network Enabler), the call may use the Home Operator’s core network for reaching the Internet or may use the same Partner Operator (that was used to virtualise the air interface resource) for this purpose [60].

Resources may also be shared through an LBO, in which both air interface and core network resources are obtained from a partner operator. In fact, the voice/data call is transferred completely to the partner operator. Therefore, the LBO is different from virtualisation since even the access to the internet is through another partner operator.

A network function virtualisation has been proposed as a promising way to allow multiple heterogeneous virtual networks (VNs) to coexist on a shared infrastructure. A major challenge in this respect is the VN embedding problem, which deals with the efficient mapping of virtual nodes and virtual links onto the substrate network resources. Since this problem is known to be *NP*-hard, research has been focused on designing heuristic-based algorithms which had a clear separation between the node mapping and the link mapping phases [61].

2.4.3 Heterogeneous Deployments

A mobile operator may operate via high power base stations (macro-cells) or low power base stations (femto or pico cells, also known as small cells), even with overlapping coverage areas. Calls of mobile users might be routed through the small cells rather than through the macro-cells if routing through the core network is not required. The former may be connected directly to the internet to offload the data [60].

Wi-Fi and small cells technology in LTE are almost the same, except in coverage where the Wi-Fi can cover larger area due to its connection capabilities , where it can be linked with the MNO ,additionally it can operate in all kind of the spectrum (licensed and the unlicensed) , avoiding the core network congestion by offloading the data traffic and saving the MNO resources .The advantage of this technology its suitable for the high traffic and user density area , and it could very useful long-term strategy if it can manged and operated by the MNOs , but with some challenges in deployments that may affect the CAPEX and OPEX , on the other hand small cells in LTE technology can be deployed easier, which consider as one of the advantage in this technology [60].

2.5 Offloading (LBO)

As mentioned earlier, the rapid growth of mobile data traffic has posed challenges for mobile operators, who are looking for more cost-effective solutions to deliver mobile data without significantly compromising the QoS. To achieve CAPEX savings, 3GPP has offered some options to selectively offload certain mobile data to another path, without passing through Evolved Packet Core (EPC). Despite the fact that mobile operators have already deployed the LTE/LTE-Advanced networks that support high downlink/uplink peak rate as compared to legacy systems, the increasing network capacity still cannot keep up with the data traffic growth [67].

A large percentage of macro-mobile traffic initiates from local environments, such as home and enterprise networks, along with users that use bandwidth-intensive applications. Offloading selective mobile data traffic by low-cost fixed access networks and the Internet, thus avoiding the operator core network, is a promising solution [67,62].

Mobile network vendors and operators are interested in standardising the solution via the 3rd Generation Partnership Project (3GPP). Such a solution is expected to cost-efficiently support the data traffic growth by optimising network usage and lowering network congestion while reducing CAPEX investments. Two data offload solutions are currently offered by 3GPP's System Architecture 2 (SA2) working group, namely LIPA (Local IP Access) and SIPTO (Selective IP to Offload) [67,62].

2.5.1 Difference Between LIPA and SIPTO

LIPA is mainly for end users to access their intranet locally through a local 3GPP access point such as indoor femtocells or picocells. In other words, whenever a user has his own internet access in his/her residential area by either a femtocell or a picocell, mobile devices can use LIPA to access other devices connected to the local network, such as a local computer. LIPA is subscription-based. A mobile device can use LIPA in its own network or in a visited network, subject to roaming agreements between mobile operators. It is up to the mobile operator to enable or disable Local IP Access for user subscriptions, per Closed Subscriber Group (CSG) for each LIPA Access Point Name (APN). The key benefits that mobile operators get from LIPA include [68]:

- Reduced network congestion for Local IP access
- Better quality of experience for services delivered through LIPA
- Improved revenues due to increased usage of operator networks for local access, which otherwise would have used WiFi, WLAN, etc.

SIPTO allows managing Internet traffic optimizing the cost and is valid for both femtocell and macrocell. It enables routing of selected IP traffic through either the most optimal path in an operator's core network or bypassing it completely. The type of traffic to be offloaded is determined by the operator. Thus, the operator may decide to offload only selected IP traffic from a mobile device, while not offloading other IP traffic on the same mobile device associated with other defined IP networks. The key benefits that mobile operators get from SIPTO are [68]:

- Reduced network congestion by offloading certain services
- Reduced CAPEX and OPEX with fewer backhaul requirements, which are typically leased by most operators.
- Improved Quality of Experience for subscribers, thus improved consumer satisfaction, leading to better revenues

2.5.2 Network Coupling

Networking is about several layers of networks working together, which is considered as a heterogeneous environment. The relationship between networks of different layers can be roughly described as a network coupling subject. Although there are no precise definitions yet, loose network coupling usually refers to two networks which keep independence, while tight coupling usually means a high level of integration between two networks, e.g. 3GPP LTE and WiFi collaboration. In this scenario, loosely coupled networks are, for instance, a WiFi access network and a 3GPP core network that only share the AAA server for authentication; each one operates independently, and none of them has to modify its architecture or protocol stack, according to the 3GPP specification [19]. However, since there is no support for service continuity during handovers, a user may face longer handover latency and a certain amount of packet loss. In a tight-coupling scenario, a WiFi data flow through the 3GPP core network from ePDG to the 3GPP packet data network gateway (PDN-GW), occurs in the same way as mobile data does. Now, WiFi users can also utilise the 3GPP services to obtain QoS even if they move out if the coverage. Other 3GPP services may also be enabled for WiFi users through the tight coupling [60].

In a broader sense, network coupling is about the level of integration between two networks. The femto base station is another good example of tight coupling. In addition to the tight coupling at the core network level described above, a heterogeneous network may also create a tight coupling at different levels, for instance between two radio access networks, and allow one of them to take over the control of the other in the radio resource management control scheme.

An application example is the CDMA-LTE handover model, which is considered a tight coupling case at the radio link layer, offering the same level of fast handover performance [20].

In LTE-Advanced and beyond, this type of tight coupling can be extended to the air interface of the physical layer PHY and MAC layer, to allow the processes of joint radio resource management and traffic scheduling. In LTE networks, more tight-coupling issues in the context of the heterogeneous environment have been already implemented, such as the X2-link interface between a macro base (eNB) station and a femto base station (HeNodeB) [60].

Network coupling deployment is a difficult task and a hard business decision too, driven by the trade-offs of complexity, cost, security and performance. A mobile network operator may choose loose-coupling over tight-coupling, despite its potential higher performance; therefore, it all depends on the situation [60].

Table 4 presents a summary of the coupling types, the general design characteristics and the associated benefits and complexity. From a perspective of standards development, the communication industry leans towards tightly integrated solutions with secure data offloading and service continuity. These solutions will continue to advance in different markets and in different environments too, such as homes and enterprises [60].

Table 4 Types of Network Coupling [60]

Type of Network Coupling	**Exam. Design Characteristics**	**Benefits and Complexity**
No Coupling	• Two networks operate independently • Mobile device connection manager coordinates wireless connectivity.	• No change to existing networks, suitable for roaming. • The operator may perform offline billings consolidation for a single bill.
Loose Coupling	• Two networks share user credential and AAA. • Data traffic goes through separate core networks.	• Common users/device credential for the two networks; potentially eliminate user intervention for WiFi access. • Core network traffic offload to the internet.
Tight Coupling	• User traffic goes through the mobile operator core network, (e.g., home femtocells connected through a Femto gateway)	• Opportunity for operators to offer value-added services. • Opportunity to offer service continuity or even seamless handover.
Very Tight Coupling	• Two access networks are directly interconnected. • Real-time radio resource management across networks, (e.g., picocell connected through X2 interface)	• Opportunity to implement carrier aggregation and interference coordination. • Requires significant design and implementation efforts.

2.6 CAC Concepts

Mobile Network Operators (MNO) provides voice and data connectivity services in the certain area named coverage. The target of MNO is to simultaneously offer high-quality services to the maximum possible number of users while guaranteeing

(i) Accessibility: if a user makes a request for a call or data session, he/she gets connected to the network.

(ii) Mobility: if a user is moving through the coverage while in the middle of a call or data session, he/she can keep the session even if a handover is necessary.

(iii) Retention: all sessions must only be released by the decision of the user.

In the third generation (3G) of mobile network technology, calls are known as RAB, with voice sessions being called CS RAB and data sessions PS RAB. Basically, there is two Key Performance Indicator (KPI) to measure the operation of the network: CALL DROPPED (CD) and CALL SETUP FAILURE (CF); the drop of both voice and data sessions are considered RAB DROP.

Fundamentally, the CAC constrains the access to the mobile network based on different criteria that primarily depend on the availability of the channel, in order to avoid, network congestion and service degradation, to fulfil the desired QoS parameters [12,64,65]. Figure 10 illustrates the concepts.

A single user can link to a certain cell by means of a new call or a handover. In either case, a CAC procedure is triggered based on admission criteria, to determine whether or not to accept the call. As a result, if the admission criteria are activated the call will go through; otherwise, the new call will be blocked, or the handover call will be dropped. The performance of CAC techniques has a direct effect on the performance of both a single user and of the total network [66]. Beyond third-generation networks, CAC schemes are become more complex due to the evolution of wireless systems, which provide multimedia services with various QoS levels [56].

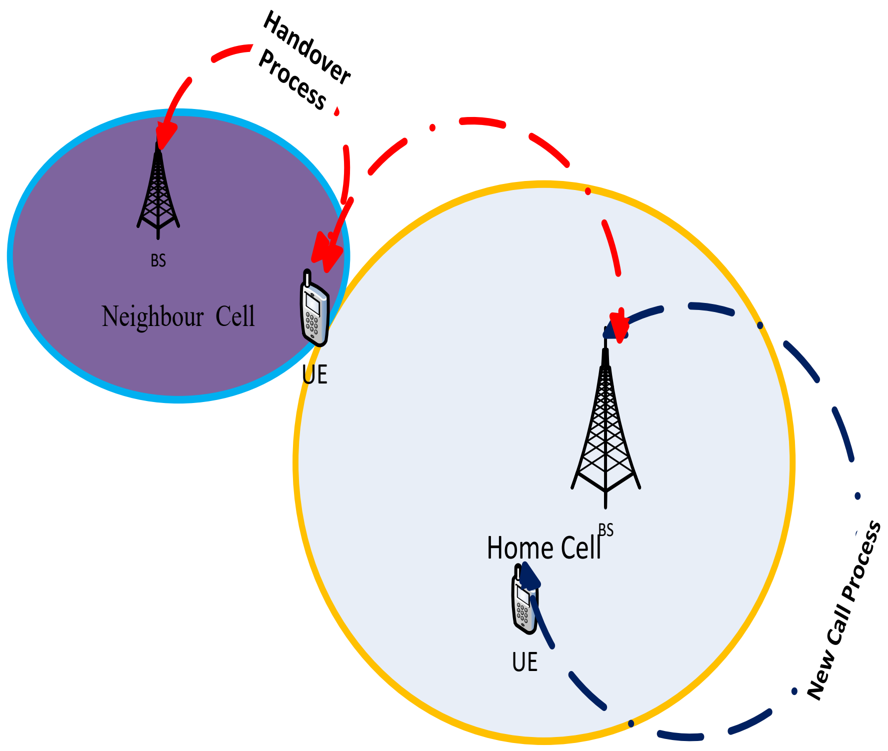


Figure 10 New Call and Handover Call Process

In other words, more sophisticated CAC schemes are being developed to cope with these changes. For example, CAC schemes are classified based on the network entity decision maker as centralised and decentralised [12]. In centralised networks, schemes such as the Mobile Switching Centre (MSC) take over the admission process in the whole network as in 3G [69], while in decentralised (distributed) CAC is performed in each cell by the base station as in LTE systems [70]. A Central CAC scheme is more effective because of the availability of global information, but it is more complex than distributed schemes. However it is not used in the LTE or LTE-A cases due to the network hierarchy infrastructure which is considered as distributed, and the global information is still implemented in a sort of centralised manner through X2 interface link among base stations [71].

In heterogeneous networks, as a result of the collaboration between different radio access technologies (RATs) such as Wi-Fi and mobile networks, the seamless connection and global information mobility requirements need a type of collaboration CAC scheme, such that a call in one network has to be capable of moving and being handed over a different network technology pervasively; this is the so-called vertical handoff, and several related issues have to be considered [72].

A CAC approach based on signal quality parameters to trigger the handoff process will not perform satisfactorily if such parameters may be not measured reliably; furthermore, other system parameters have to be considered, such as the congestion level in the system in terms of channel availability. For instance, non-real time applications can be offered by WLANs in order to mitigate the congestion in mobile networks. All these factors will impact the distribution of the call holding time in both networks, which leads to increased dropping and blocking probabilities, especially in peak hours [72].

Generally, the data traffic varies along the day. For example, traffic reaches high levels during the daytime in the downtown area of a city, and alternatively high-level consumption at night in residential areas [57]. Moreover, such variation mainly depends on the geographic location and data dynamics, which is subject to the user movement and data consumption behaviour [57].

Therefore, this traffic difference among times and locations will lead to an imbalance usage, due to the static bandwidth and the dynamic demand by users.

Current mobile systems deploy different types of base stations and relay nodes to accommodate peak hour traffic [58]. As a result, many base stations are required, which may affect the network performance in terms of highest allowed interference level and may lead to increases in CAPEX and OPEX. As mentioned earlier, single class CAC schemes are no longer appropriate to face growing requirements of data and multimedia services; multiple-service/class CAC schemes are more applicable, especially in the third-generation networks and beyond.

The design of multiple service/class CAC schemes is more challenging since some critical issues, such as service prioritisation, fairness, and resource sharing policy, have to be considered. Optimal CAC schemes as in [73] are always preferred, but not certainly achievable in a particularly realistic problem scenario.

Additionally, Cognitive Radio (CR) technology assists the progress of an intelligent and adaptive wireless communications system, which is basically aware of the radio frequency environment. So, the communication parameters (such as carrier frequency, bandwidth and transmission power) can be dynamically chosen, which lead to an efficient spectrum utilisation [74].

However, this technology faces many technical challenges, including protocol design, interference characterisation, environmental awareness, as well as the development of distributed algorithms, distributed measurement techniques, and quality of service (QoS) guarantees [75].

In [41] the author investigates the performance of spectrum renting solutions; to improve utilisation among different wireless networks for spectrum management. In addition, a heterogeneous architecture consisting of multiple mobile networks that rent spectrum from each other according to their availability is also considered.

Similarly, [43] presents a study of mobile wireless networks which lease spectrum bands from another system. Both [41] and [43] show the benefits of spectrum renting in heterogeneous architectures and address the trade-offs between different CAC policies. These works consider a reserved band approach to mitigate the blocking and dropping probabilities

2.6.1 QoS Parameters

This section describes the QoS parameters that can be considered as admission criteria in CAC, to guarantee performance with acceptable KPI in a scenario with many users. This section and (2.6.2) cited from [58,66]

1. Signal Quality: This parameter is used to guarantee acceptable interference levels in wireless networks. In few words, more congestion will result in a network with the more degraded signal quality for the users in terms of the interference level. Therefore, the possible admission criteria could be the following:

    (i) Number of users in a cell

    (ii) Interference level

    (iii) Transmitted power from the base station

    (iv) Received power from the base station or the user terminal

The call is granted admission if and only if it can keep a minimum signal quality level without interfering with existing calls; this applies to new and handoff calls.

2. Call Dropping Probability: CAC is used in limited bandwidth wireless networks to minimise the handover failure probability, since dropping an on-going call is commonly more bothersome than blocking a new call. This can be avoided by

reserving a portion of the bandwidth for handover calls use only. The corresponding admission criteria may be:

(i) Nnumber of users per class in a multiple-class system.

(ii) Approximate handover failure probability.

(iii) Availability of resources.

3. Packet-Level Parameters: When a wireless network provides packet-oriented services, overloading can cause an unacceptable level of packet delay or end-to-end system delay (delay jitter) by the system service bearer. Therefore, CAC should be deployed to manipulate the packet level threshold to warrant the required QoS parameters (packet delay, delay jitter, and throughput). Under this circumstance, the admission criteria may be:

(i) Number of users.

(ii) Availability of resources.

(iii) Estimation of packet-level QoS parameters.

4. Transmitting Rate: CAC schemes are utilised in wireless networks through data services that intend to guarantee a minimum transmission rate. Nevertheless, it is more complex in wireless networks due to user mobility, limited bandwidth, channel characteristic and channel noise.

2.6.2 Other QoS Strategies

1- Revenue-Based CAC: From the perspective of the service provider, utilising full network resources may be considered a good way to increase revenues. However, accepting calls based only on the capacity of the network, especially admitting additional new calls, may have some undesirable consequences. In particular, in a highly congested network, this could result in a drop of the QoS offered to calls already active and may lead to a higher dropping rate in calls. Hence, CAC can be used to increase the network revenue function based on the potential profit and the consequences of admitting new calls. In this situation, the admission criteria can be:

(i) Number of users.

(ii) Estimate of the probability of QoS deterioration.

2- Prioritise some services or classes: One of the purposes of a CAC scheme may be assigned a higher priority to some services or classes.

3- Fair Resource Sharing: Establish equality between different users in the same class and between users of different classes, based on criteria such as channel condition and mobility characteristics. A CAC can be employed to accept or reject users based on the allocated resources, such that no user class controls the network resources.

2.6.3 Challenges in CAC Design

The different QoS requirements for multimedia applications and the occurrence of different wireless access technologies (RAT) add greater challenges in designing efficient CAC algorithms for 4G wireless networks. The traditional CAC algorithms for homogeneous wireless networks determine whether a user may be allowed into the network; CAC algorithms developed for homogeneous wireless networks are reviewed in [6–8]. All the same, homogeneous CAC algorithms do not offer a single result to address the heterogeneous architectures which characterise the next generation wireless network [9]. This limitation has led to the development of new CAC algorithms for heterogeneous wireless networks.

The aim of CAC is to help service provider to ensure that their customers are satisfied with appropriate QoS parameters while utilising the limited network resources in a more efficient way.

The quality of service class identifiers (QCI) shown in Table 5 is the most important parameters used to measure the QoS of a bearer in an LTE network. With respect to Table5, the following quantities can be defined [73]:

(i) Resource type: Guaranteed bit rate (GBR) or non-guaranteed bit rate (Non-GBR).
(ii) Packet loss rate: Indication for the part of packets that are lost as a result of errors in the transmission and reception processes.
(iii) Packet delay budget: Indication, with 98% assurance, for the delay that a packet experiences between the mobile and the packet data network (PDN) gateway.
(iv) QCI priority level helps in the scheduling process.

As aforementioned, the GBR and non-GBR are the two types of bit rate bearers involved in Evolved Packet System (EPS). The GBR value is predefined permanently and linked with an EPS bearer by an admission control function executed at eNB level; else, an EPS bearer is considered as a Non-GBR [73].

The GBR value is controlled by the scheduling scheme at eNB level for the channel allocation process and determines the necessary number of channels (Physical Resource Blocks) to achieve the required bit rate.

Table 5 Standardised QCI characteristics [78]

QCI	Resource Type	Packet error/loss rate	Packet delay budget (ms)	QCI Priority	EXAMPLE services
1	GBR	10^{-2}	100	2	Conversational voice
2		10^{-3}	150	4	Real-time video
3		10^{-3}	50	3	Real-time games
4		10^{-6}	300	5	Buffered video
5	Non- GBR	10^{-6}	100	1	IMS signalling
6		10^{-6}	300	6	Buffered video, TCP file transfers
7		10^{-3}	100	7	Voice, video, real-time games
8		10^{-6}	300	8	Buffered video, TCP file transfers
9		10^{-6}	300	9	Buffered video, TCP file transfers

Each EPC bear is associated with one QoS class identifier (QCI) together with the allocation and retention priority (ARP) bearer level. The ARP is mainly used in the bearer decision establishment and modification process based on the availability of resources [68]. The priority level in the ARP is used to differentiate bearers to ensure the prioritisation among them. In addition, in case of a congested network eNB can use the information contained in the ARP to eliminate bearer due to resource limitations and assign it to another highly important activities such as a handover process [73,74].

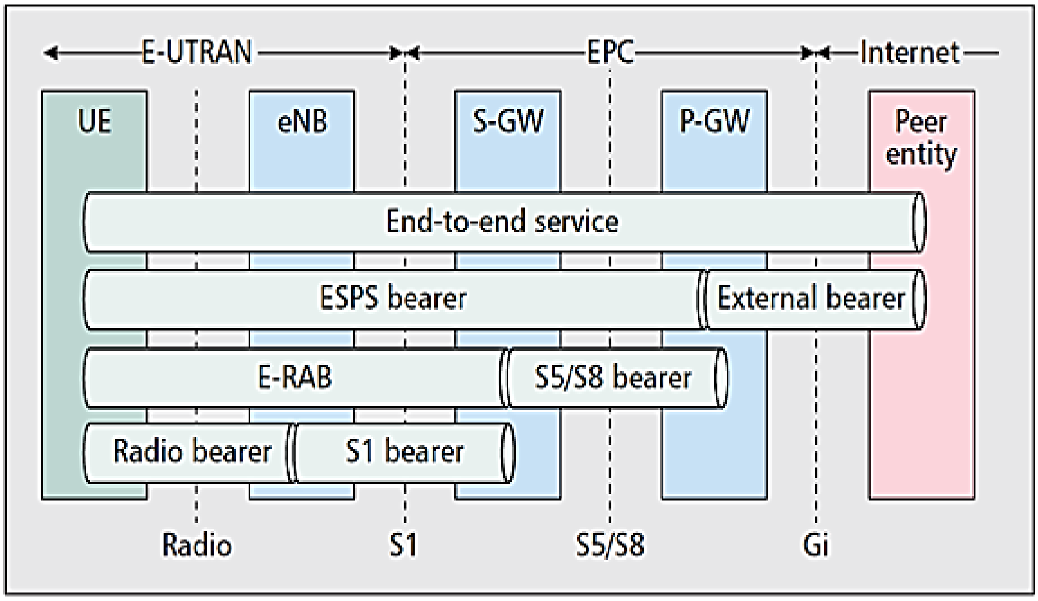


Figure 11. End-to-End service delivery in LTE/LTE-Advanced architecture [44]

Figure 11 shows the elaboration on the different bearers employed in end-to-end service delivery in an LTE system: the radio access network, which is the E-UTRAN, the core network, which is the evolved packet core level (EPC) and the multimedia service level, which is the peer entity [73,74]. The new architecture has been designed as part of two 3GPP work objects: system architecture evolution (SAE), which covers the core network section, and LTE, which covers the radio access network section, air interface and the user terminal (UE). LTE has become the acronym for the entire system and is frequently used as such by 3GPP. Officially, the whole system is known as the 'Evolved Packet System' (EPS), while the acronym LTE refers only to the evolution of the air interface [73].

2.6.4 EPS Bearer QoS

Voice and multimedia are a mixture of real time and non-real time applications; if delay problems occur, the traffic flow must be coordinated. The process of classifying, scheduling, and forwarding traffic based on the destination address are under the responsibility of the packet switches, in addition to the type of media which have been transported. This process becomes feasible through QoS-aware systems [75]. The QoS for data radio bearers is delivered to the eNB by the MME using the standardised QoS attributes (refer to Table 5), and according to these attributes in the EPS domain, the protocols between the UE and eNB can control the current scheduling of uplink and downlink traffic flow. Many parameters are adapted to control and identify the QoS; such parameters are shown in Figure 12.

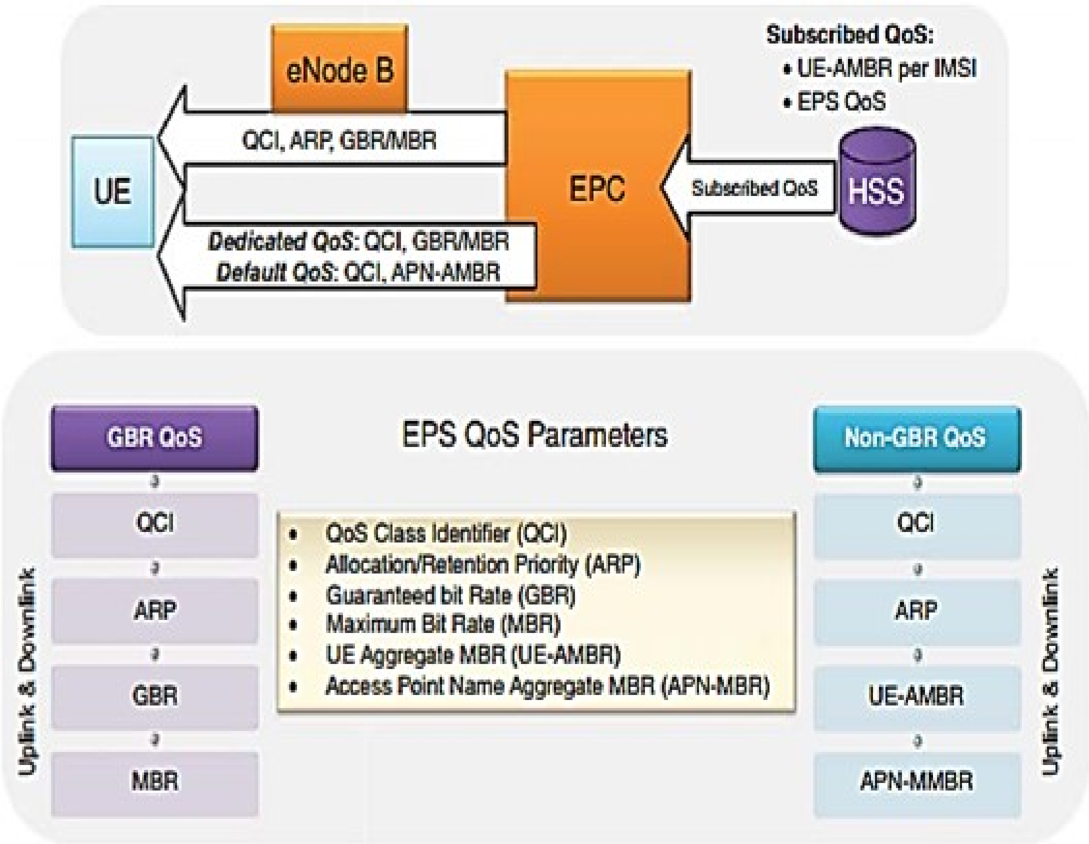


Figure 12 EPS QoS definitions and parameters [44]

EPS bearer QoS depends on the resource type: GBR or non-GBR. By default, the dedicated radio bearer (DRB) is set up as a non-GBR, but it can be either GBR or non-GBR [70]. As shown in Figure 12, the GBR-based EPS bearer contains two different parameters, namely the GBR and Maximum Bit Rate (MBR). The GBR is the bit rate that could be likely to be delivered by a GBR-based bearer, whereas the MBR limits the bit rate that could be expected to be supported by this EPS bearer. The GBR-based QoS parameters provide the eNB with information on the uplink and downlink rates for the E-RAB (E-UTRAN Radio Access Bearer), which is responsible for delivering the packets between an EPS bearer and the UE and guarantee a high DL/UL bit rate. Non-GBR EPS bearers are subject to control through an Aggregate Maximum Bit Rate (AMBR), one for the User and another for the Access Point Name (APN) linked with the user. They are defined as [41]:

- UE-AMBR: the maximum value that could be added to the total bit rate allocated to the UE for all non-GBR services. The UE-AMBR limits the aggregate bit rate across all non-GBR bearers of a UE.
- APN-AMBR: the maximum value that could be added to the total bit rate allocated to a group of users associated with a specific APN. The APN-AMBR limits the aggregate bit rate across all non-GBR bearers, and across all PDN connections of the same APN (extra traffic may get dropped).

The priority control in bearer establishment, modification or release is done by the allocation and retention priority (ARP). It also may be used to show which bearers are dropped when congestion in the network occurs. The parameter ARP can be used for GBR or non-GBR QoS [41], and its priority level ranges from 0 to 15, with 15 corresponding to no priority, 1 to the highest level of priority, and 0 being held in reserve [41]. The QCI is another QoS parameter in GBR and non-GBR EPS bearers. It provides a mapping from an integer value to precise QoS parameters that control how bearer level packets are sent. The QCI is responsible for packet forwarding, i.e. scheduling, admission thresholds, queueing thresholds, and link layer protocol configuration. Typically, the operator controls QCI values for an E-RAB in advanced configuration [41]. As mentioned before, the QCI is characterised by the nine different indicators shown in Table 5.

Next chapter develops a mathematical model of the on-demand sharing of air interface resources among mobile operators. It starts with the calculation of the number of physical resource blocks required for ongoing (or demanded) voice/video calls, allocating call resources if available.

Chapter 3

3 Framework Fulfilments

3.1 Introduction

This chapter describes the mathematical model required to develop the proposed framework for on-demand sharing of air interface resources among mobile operators. This model constitutes the support of the applications that will be used to emulate real traffic situations in LTE core networks in different scenarios.

The average numbers of physical resource blocks (PRB) required for voice and data calls are computed using an LTE simulator. Then, if available, the required PRBs are allocated based on these average numbers, both in the normal and virtualisation scenarios. In addition, a Poisson Distribution Simulator (PDS) is used to compute the voice and data calls blocking probabilities in the virtualisation scenario.

3.2 Derivation of The Average Number of Resource Blocks

This section presents a mathematical description of how operators share PRBs among users and determine δ_{ave_voice} and δ_{ave_data}. If virtualisation is not being used, the PRBs of an operator will be used only for its own users. Otherwise, such resources may be scheduled for users of a different operator.

A model of the LTE virtualisation system, and the corresponding simulator coded in MATLAB for a real scenario (i.e., in a mobile environment), were used to determine δ_{ave_voice} and δ_{ave_data}, assuming that there are several mobile users (User Equipment's (UEs)) subscribed to N operators. A Random Way Point (RWP) mobility model is used for describing the movement of users in the cell. According to such model, a user may be located at any randomly chosen location in the cell at the start of the simulation (time t = 0), and then moves along a straight line with speed u for a time Δt; Δt is uniformly distributed between two chosen values, and u is a randomly chosen value less than a maximum. The angle η between the straight line and the horizontal is a random variable uniformly distributed between 0 and 2π. After the user has moved for time Δt, it stops for time Δt_2 at the new location. Then it starts moving again, and the process is repeated but with new random values of Δt and η. If during movement a user is headed to cross the cell boundary, its path would be reflected into the cell at such boundary.

It is assumed that the eNodeBs of all N operators are located at the centre of the cell. A Hypervisor module between the PHY and MAC layers at the eNB allocates the LTE Physical Resource Blocks (PRBs) to the users from either the Home Operator or from the available pool of virtual operator resources. Such allocation requires knowing the channel conditions / SNRs of the users as seen by the eNodeBs. The SNR at the UE's is calculated using the Path Loss, Shadowing and Jakes' Fast Fading channel model.

The number of PRBs is calculated such that there is a reliable transmission of data between each UE and the operator it is allocated to (which may be the Home Operator or the Virtual Operator). At last, all the previous information is used to compute δ_{ave_voice} and δ_{ave_data} over time, across UEs and operators.

3.3 Channel Model of The LTE Simulator

The model of the channel implemented in MATLAB for the LTE simulator comprises three parts:

1- Path Loss Model: Path loss is described according to the distance-dependent path loss model given by the 3GPP

$$PL = 128.1 + 37.6log(R) \quad (1)$$

where PL is the path loss factor in dB, and R is the distance between the user and the eNB in km. Clearly, the loss in SNR at the UE increases with the distance from eNB.

2- Slow Fading Model: Slow fading is typically modelled using a log-normal distribution, with zero mean and constant variance (taken as 9). In addition, the time correlation between the slow fading values needs to be also considered. In this model, a moving user starts at an initial point P, where the slow fading value is randomly generated using the log-normal distribution. The shadowing at points at distances δ, 2δ, 3δ from P is determined as

$$S(n\delta) = S\big((n-1)\delta\big)e^{-\delta/X_c} + V_i \quad (2)$$

where $S(n\delta)$ represents the slow fading value at a distance $n\delta$ from P, X_c is the de-correlation distance, and V_i are independent and identically distributed normal random variables with zero mean and variance

$$\delta 2^2 = \delta 2\big(1 - e^{-2\delta/X_c}\big) \quad (3)$$

The $S(n\delta)$ are obtained recursively from $S\big((n-1)\delta\big)$ using equation (2), starting from values S (0) which are generated according to the log-normal distribution specified above, and are different for all mobiles. The de-correlation distance X_c is the distance up to which the shadow fading values are strongly correlated and is set as 4.5 km for the urban, OLOS scenario.

3- Fast Fading Model: Fast fading is implemented using a Jakes' like method. A frequency-selective (multiple paths) fading channel object that models each discrete path as an independent Rayleigh fading process, is created using the MATLAB command

$$\text{Chan} = \text{Rayleighchan (TS, FD, TAU, PDB);} \tag{4a}$$

where TAU is a row vector of path delays in a sec, PDB is a row vector of average path gains in dB; TS is the sample time of the input signal in a sec and FD is the maximum Doppler shift in Hz. The effect of the channel CHAN on a signal X is modelled using the command

$$\text{Y = filter (CHAN, X);} \tag{4b}$$

In the proposed model, channel tap values are required for all operators, for all mobiles and at Hypervisor resolution times of 1 sec. Since for the selected Doppler frequency of 300 Hz [8] the channel varies significantly in 1 second, independent values are taken rather than going through the Doppler Effect, which is computationally demanding. Therefore, every 1 sec CHAN is called again, and new independent values are generated.

3.4 Summary of The LTE Simulator

Then, the LTE simulator is used to obtain the average number of PRBs occupied by a voice/data call. The simulation parameters are shown in Table 6.

Due to the Random Way Point mobility model, the different mobiles (User Equipment's (UEs)) will be at different locations in the cell, moving at their own speeds or static. Therefore, the SNRs of their received signal from different operators would vary, due to different path losses, shadow fading, channel condition and location of the UE. Correspondingly, the PRBs may or may not transmit in different constellation sizes and coding rates. Hence, to be able to transmit a fixed number of bits of the voice/data call every fixed interval of time, varying number of PRBs may be required by the users.

Table 6 Features of the LTE Simulator [16,76]

The radius of the circular cell	R = 375 m (Equivalent square cell has its side, a = $\sqrt{\pi}R$)
No. of PRBs of LTE	25 (Total 5 MHz bandwidth)
Transmit Power	29 dBm / PRB
Mobility Model	Random Way Point (max. speed of 120 km/hour)
Channel Model	Path Loss + Shadow Fading + Jakes' Model Fast Fading
The required data rate of voice call at the Physical layer	300 bits every 20 ms (GSM-EFR codec)
The required data rate of video call at the Physical layer	356 bits every ms

Thus, the procedure adopted to calculate the average number of PRBs required for a voice call and similarly for a data call is the following: Run the simulator with 50 voice call UEs, to obtain the mean bandwidth occupied over time. Divide this value by $200\times10^3\times 50$ (200 kHz is the bandwidth of the PRB and there are 50 UEs), to get the average number of PRBs required by a voice call. For a voice call, an average of 0.12 PRBs was obtained. Following a similar procedure, an average of 2.5 PRBs for data calls is obtained when 4 data callers are simulated.

3.5 Proposed Framework

Wireless Network Virtualisation can be considered as an extension of Wired Network Virtualisation, as it is motivated and based on the substantial benefits obtained for Wired Networks. It can be said that Virtualisation of Networks and Computer Systems has become a technological trend. However, this trend is not as highly utilized on Infrastructure based wireless networks [77]. Nonetheless, it has recently attracted a significant attention in both Academia and Industry, since it is considered one of the advanced research areas in Computer Science [77]. In virtualisation terminology, the expression Wireless Virtualisation may refer to Wireless Access, Wireless Network, Wireless Infrastructure and Mobile Network Virtualisation [77].

To clarify, as compared to Wired Network virtualisation, Wireless Network Virtualisation may generate logical channels, RAN, switches and routers. Thus, Wireless virtualisation can potentially facilitate the separation of traffic to increase flexibility in terms of QoS, enhanced security and manageability of networks. A solid network management mechanism is crucial in emergent heterogeneous networks [77].

As mentioned in chapter 1, a virtualisation technique is a great approach to increase the utilisation of networks resources. This is important from the infrastructure and spectrum virtualisation perspectives. It also expands the possibilities for the notion of IaaS, such that one operator is able to use its own infrastructure or other operator's underutilised equipment such as BSs on the congested areas.

Depending on the type of virtualised resources and the impartial of virtualisation, three different general frameworks can be identified for wireless virtualisation [77]:

- Flow-based virtualisation
- Protocol-based virtualisation
- RF frontend and spectrum-based virtualisation

The focus of this work will be the flow-based virtualisation framework, since it is used for the isolation, scheduling, management, and service differentiation between traffic flows and streams of data sharing a common nature. It is inspired by the flow based SDN and network virtualisation, which will be used in wireless virtualisation [77]. Therefore, it requires wireless special fulfilment, such as the CAC and the VNE, to support QoS and SLA over the traffic flows in which the LBO is used.

From the perspective of implementation, the flow-based virtualisation can be considered a feasible approach compared to others approaches with instant benefits, as it joins virtual resources and provides more flexible and efficient traffic management. Assumptions and considerations for the proposed framework are as follows:

- Network infrastructure: An LTE network infrastructure is utilised because it is an IP infrastructure designed to avoid interoperability and protocol mapping challenges in legacy systems.
- Call type: New calls are considered; since priority is always given to the handover calls, new call requests will most likely be blocked during peak hours and in high-demand situations. This also allows operators to accept new calls if the QoS can be maintained above the desired level [12].
- Call traffic class: Non-real-time traffic and real-time traffic; the virtualisation part will be activated when required.

- Offloading strategy: Selective Internet protocol (SIPTO) offloading will be considered, as it is more applicable to MNO scenario [51].

The proposed business model is described in terms of different roles and entity functions within the virtualisation terminology (See Section 2.3.3). Such a model can be further decoupled, as shown in Figure 8, into more specialised roles [10].

Nevertheless, more coordination mechanisms and interfaces should be used, which may significantly increase the complexity and latency. For this reason, coordination mechanisms are used in the minimal matter, meaning that the virtualisation will be triggered in a specific condition such as a shortage in the available physical resource blocks and the VNE is responsible in this part, without a trade-off between degradation in QoS in terms of accepting more calls.

A related scenario would be the unbalanced usage among different mobile operators who are in the same geographical area, due to unpredictable user data consumption behaviour and differences in peak hours. This will create a window of opportunity to benefit from the virtualisation concept. On the other hand, this involves the offloading technique, which handles the problem of congestion in the core network. To describe the proposed framework as an on-demand need and in a collaborative manner, the framework will be split into the following three phases, for evaluation later by using MATLAB and Emulation in different scenarios:

- Phase One: A standalone operator not involved in the virtualisation and offloading strategy in terms of over-the-top service provider agreement.

- Phase Two: An MNO with multiple VMNOs (one-dimensional model), where the mobile operator deals with the shortage in capacity and makes explicit the benefits of sharing.
- Phase Three: The proposed framework where multiple MNOs/VMNOs are involved in the scenario with different capabilities. For instance, partnership with an OTT service provider who can benefit from this agreement and expand it to other parties in the network; or as a foreign MNO/VMNO where they can achieve a win-win situation. Figure 13 provides a network overview.

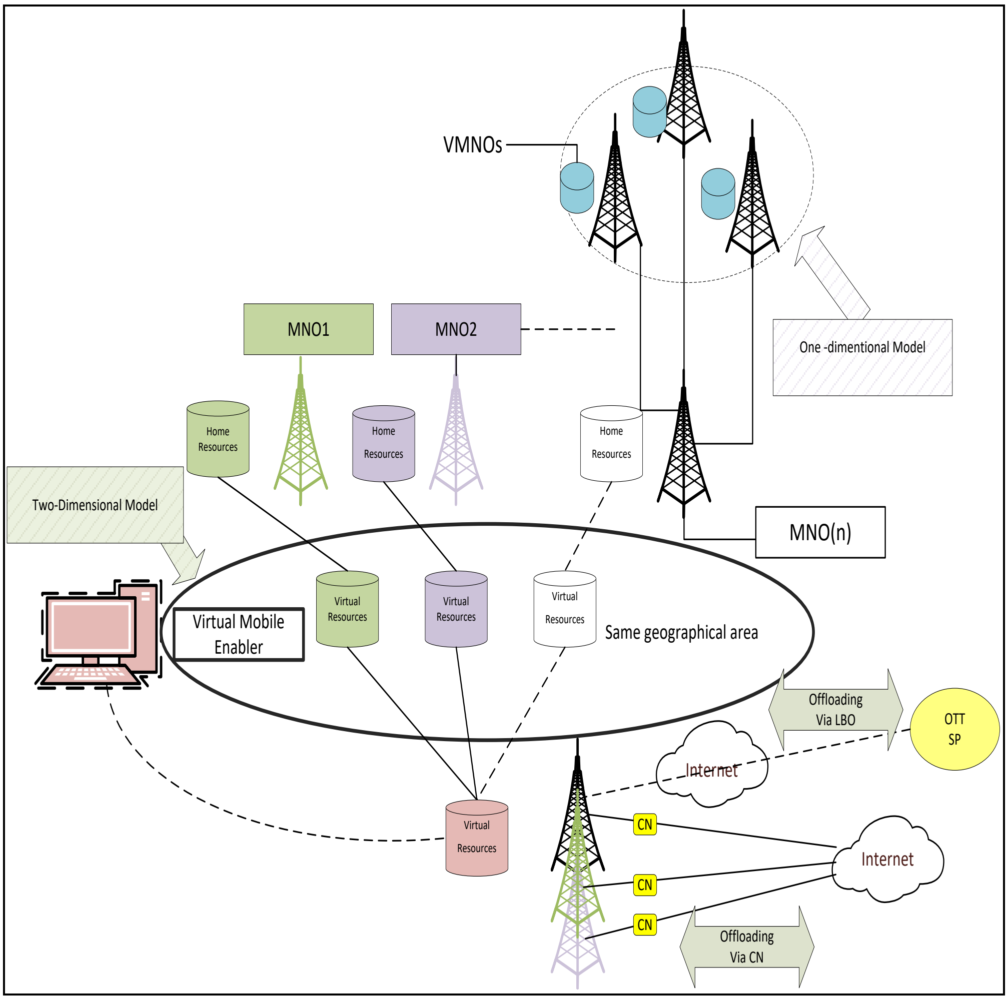


Figure 13 High Level- Network Layout

3.6 Symbology

The symbols included in Table 7 are used in the analytical model.

Table 7. Symbology

Symbol	Definition
N_{SPs}	Number of service providers or mobile operators
q	Home operator index, $q \in [1, N_{SPs}]$
l	Another operator index, $l \in [1, N_{SPs}]$
m	Number of voice calls
n	Number of data calls
N_{RBs}^{q}	Number of Resource Blocks (PRBs) available with Operator q
$\delta_{ave_{voice}}$	Average number of resource blocks for a voice call
$\delta_{ave_{data}}$	Average number of resource blocks for a data call
α^{q}	The fraction of total resources available for new call requests. Here, $(1-\alpha^{q}) \times N_{RB}^{q}$ are reserved for handover calls.
RSC^{q}	Resource blocks service capacity of operator q: Number of resource blocks of operator q.
$RSCV^{q}$	Resource blocks of operator q available for sharing with other operators.
$\eta_{Rcv}^{q}(t)$	Number of resource blocks of operator q occupied by received voice/data calls at time t.
$n_{voice}^{q}(t)$	Number of active voice calls of operator q at time t

$n_{data}^{q}(t)$	Number of active data calls for operator *q* at time t
$\eta_{Bl}^{q}(t)$	Number of calls blocked by operator *q* within time duration t.
β^{q}	Resource utilisation/loading factor for operator *q*. It is proportional to the number of calls being served by operator *q*. An operator may be partially or fully loaded, but it can never load more calls than the permitted by its available resources. Hence, $\beta^{q} \leq 1$. $$\beta^{q} = \begin{cases} \dfrac{\eta_{Rcv}^{q}(t)}{RSC^{q}}, if\ \eta_{Rcv}^{q}(t) < RSC^{q} \\ 1, otherwise \end{cases}.$$
γ_{voice}^{q}	The fraction of available resources with operator *q* that meets the QoS requirement for voice calls, when shared with another operator (in proposed CAC).
γ_{data}^{q}	The fraction of available resources with operator *q* that meets the QoS requirement for data calls when shared with another operator (in proposed CAC). This is equal to 1 since data is an NRT application.

3.7 Mathematical Model

Essentially, two types of the call can occupy the physical resources block: handover calls and new calls. In both cases, a CAC procedure is activated to determine whether or not to accept the call. Specifically, the CAC constrains access to the mobile network based on different criteria – primarily channel availability, to avoid network congestion and service degradation, in order to fulfil the desired QoS parameters as far as possible. For a given time, duration *t*, the blocking probability of voice calls is given by

$$P_{bl_{voice}} = \frac{N_{VoiceBlocked}}{N_{VoiceReceived}} \tag{5}$$

where $N_{VoiceBlocked}$ is the number of blocked voice calls and $N_{VoiceReceived}$ is the number of received voice calls. Similarly, the blocking probability of data calls is given as

$$P_{bl_{data}} = \frac{N_{DataBlocked}}{N_{DataReceived}} \tag{6}$$

where $N_{DataBlocked}$ is the number of blocked data calls and $N_{DataReceived}$ is the number of received data calls.

The Resource Block Service Capacity RSC^q of an operator q, which is the total number of its own resource blocks that are available to serve calls, is given by

$$RSC^q = \alpha^q \times N_{RBs}^q \tag{7}$$

where α^q is the fraction of resources available for new call requests and N_{RBs}^q is the total available resources. The number of resource blocks of operator q occupied by received voice/data calls at time t is given by

$$\eta_{Rcv}^q(t) = \delta_{ave_{voice}} \times n^q{}_{voice}(t) + \delta_{ave_{video}} \times n^q{}_{video}(t) \tag{8}$$

where $n_{voice}^q(t)$ and $n_{data}^q(t)$ are, respectively, the number of active voice calls and the number of active data calls, of operator q at time t.

An arriving voice/video call is blocked if the available resources are less than that required by the call, i.e. if

$$\left(RSC^q - \eta_{Rcv}^q(t)\right) < \delta_{ave_{voice}} \tag{9}$$

for a voice call, or if

$$\left(RSC^q - \eta_{Rcv}^q(t)\right) < \delta_{ave_{data}} \quad \text{for a data call.} \tag{10}$$

To this extent, the blocking probability has been calculated only for the regular case, i.e. where each operator is limited to use its own available resources. Moving forward to the proposed algorithm, operators share resources via virtualisation. The spare resources of an operator are the resources available after accounting for the active calls to be usable. Resources have to meet the quality of service requirements. Therefore, for any operator l, the available spare resources are given by

$$Spare\ resources = (1 - \beta^l) \times N_{RBs}^l \tag{11}$$

where β^l and N_{RBs}^l are, respectively, the resource utilisation factor and the total available resources for operator l.

Then, the resources of operator l available for sharing are given by

$$RSCV^l = \alpha^l \times (1 - \beta^l) \times N_{RBs}^l \times \gamma^l \tag{12}$$

where the available spare resources are multiplied by α^l and γ^l, which are, respectively, the fraction of total resources available for new call requests and the fraction of available resources with operator l that meets the QoS requirement for data calls, when shared with another operator.

At any particular time, the total resources available to be assigned to other operators on demand, according to the proposed algorithm, is known as a virtual resource and is given by

$$Virtual\ Resource = \sum_{\substack{l=1 \\ l \neq q}}^{N_{SPs}} RSCV^l = \sum_{\substack{l=1 \\ l \neq q}}^{N_{SPs}} \alpha^l \times (1 - \beta^l) \times N_{RBs}^l \times \gamma^l \tag{13}$$

Operator q will utilise a fraction of this virtual resource through virtualisation and resource offloading, depending on its call service requirement. Let us define this fraction as $\lambda^q \in [0,1]$, which in the MATLAB simulator is set as $1/(N_{SPs} - 1)$ for all q, i.e. virtual resources are being equally shared among all operators (including operator q). Hence, the virtual resource available for utilisation by operator q is

$$VRAU^q = \left(\frac{1}{N_{SPs}-1}\right) \times \sum_{\substack{l=1 \\ l \neq q}}^{N_{SPs}} \alpha^l \times (1 - \beta^l) \times N_{RBs}^l \times \gamma^l \tag{14}$$

Therefore, in the sharing scenario (SS), the total resources that operator q has available to distribute between new voice and data calls, are

$$RSC^q + VRAU^q = \alpha^q N_{RBs}^q + \left(\frac{1}{N_{SPs}-1}\right) \left[\times \sum_{l \neq q}^{N_{SPs}} \alpha^l (1 - \beta^l) N_{RBs}^l \gamma^l\right] \tag{15}$$

Since occupied resource blocks $\eta_{Rcv}^q(t)$ at time *t* are still given by (7), an arriving voice call is blocked if the available resources are less than that required by the call, i.e.

$$\left(RSC^q + VRAU^q - \eta_{Rcv}^q(t)\right) < \delta_{ave_{voice}}. \tag{16}$$

Similarly, an arriving data call is blocked if the available resources are less than that required by the call, i.e.

$$\left(RSC^q + VRAU^q - \eta_{Rcv}^q(t)\right) < \delta_{ave_{video}}. \tag{17}$$

Now the exact voice and data call blocking probabilities are determined for the normal and virtualisation scenarios using a PDS. In the latter, the traffic of calls would now be shared between operators.

3.8 Mobility Model

It is important to evaluate its performance through simulation. One key parameter in the simulation for a mobile communications network is the mobility model, which describe the movement of users, i.e. how their location, velocity and acceleration change over time. Such models are used for predicting future user positions. In this work, the Random Waypoint (RWP) Mobility Model was used.

In the RWP model, a node moves in a convex domain $A \subset R^2$ along with straight line segment from one waypoint to the other. The waypoints, represented by Pi, are uniformly distributed in A. Transition from P_{i-1} to P_i is referred to as the ith leg, and the velocity of the node on ith leg is given by random variable vi. Particularly, in the RWP model, it is assumed that $Pi,'s$ and $vi's$ are all independent. With the concept of the process of a single node and defined by the infinite sequence of triples [78]

$$\{(P0, P1, v1), (P1, P2, v2), \ldots\} \tag{18}$$

The average number of clients moving from cell i to cell j per time unit is equal to the number of customers moving from cell j to cell i per time unit [73]. Besides, RWP process is time reversible in the sense that any path along the waypoints $P0, P1, \ldots, Pn$ is equally likely to happen as the time reversed path $Pn, Pn - 1, \ldots, P0$. A direct consequence of this is the fact that arrival rates across any line segment or border are equal in both directions[78].

Now, let $a1 = a1(r, \varphi)$ represent the distance from point $r \in A$ to its border in direction φ. Equally, let $a2$ represent the distance to the border in opposite direction, i.e., $a2(r, \varphi) = a1(r, \varphi + \pi)$. Define [78]

$$h(r, \varphi) = \frac{1}{2} \cdot a1 . a2(a1 + a2). \tag{19}$$

The fixed distribution of a node moving according to RWP model is given by [76],

$$f(r) = \frac{1}{C} \int_0^{2\pi} h(r, \varphi)\, d\varphi , \tag{20}$$

where parameter C is the normalisation constant [78],

$$C = \bar{l}A^2 \tag{21}$$

where $\bar{l}$ is the mean length of leg and A is representing the area of its domain. Hence, the mean leg length can be obtained by normalisation [78],

$$\bar{l} = \frac{1}{A^2}\int_A \int_0^{2\pi} h(r,\varphi)\, d\varphi\, dA \tag{22}$$

A typical choice for the distribution of the velocity is the uniform distribution from $vmin$ to $vmax$, which yields [78].

$$E[T] = \frac{\bar{l}.ln(\frac{vmax}{vmin})}{vmax - vmin} \tag{23}$$

3.9 Poisson Process

Another key parameter in the simulation of a network protocol is the traffic model, which is used to estimate the network performance in several situations. The Poisson process is an important concept in traffic modelling. This process gives a surprisingly good description of many real-life processes, because of its notable randomness. In general, more complex systems are better described by a Poisson process [79].

3.9.1 Characteristics of the Poisson Process

The fundamental properties of the Poisson process are:

(i) Stationarity: According to this property, for an arbitrary $t2 > 0$ and every $k \geq 0$, the probability that there are k arrivals in the interval $[t1, t1 + t2]$ is independent of $t1$, i.e. for all t, k [79]

$$P\left(N_{t_1+t_2} - N_{t_1} = k\right) = P\left(N_{t_1+t_2+t} - N_{t_1+t} = k\right) \quad (24)$$

(ii) Independence: is the requirement that the future evolution of the process only depends upon the present state [79]. This property states that the probability that k (integer and nonnegative) events take place in the interval $[t1, t1 + t2[$ is independent of events before time $t1$, i.e.

$$PNt2–Nt1=kNt2–Nt0=n=PNt2–Nt1=k \quad (25)$$

(iii) Regularity: A point process is called regular if the probability that there is more than one event at a given point is zero [79], i.e.

$$p\ \{Nt + \Delta t\ -\ Nt\ \geq\ 2\}\ =\ o(\Delta t). \quad (26)$$

From the previous properties, additional ones may be derived that are sufficient for defining the Poisson process. The two most important ones are [79]:

• Number representation: The number of events within a time interval of fixed length is Poisson distributed. Therefore, this process is named Poisson process.

• Interval representation: The time distance Xi between consecutive events is exponentially distributed.

3.9.2 Poisson Distribution

The Poisson distribution is a probability pattern of events in a Poisson process. The principal considerations are:

Let $p(\nu, t)$ denote the probability that ν events occur within a time interval of duration t. The above model may be mathematically formulated as [79]:

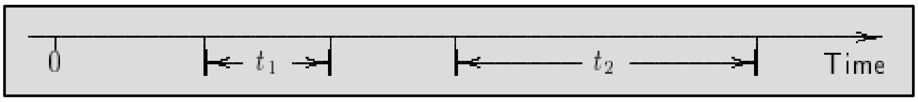


Figure 14. Two Non–Overlapping Time Intervals of Duration $t1$ And $t2$, Respectively [79].

1. Independence: If $t1$ and $t2$ are two non–overlapping intervals (Figure 14), based on the independence assumption:

$$p\ (0, t1)\ \cdot\ p\ (0, t2)\ =\ p\ (0, t1\ +\ t2) \tag{27}$$

2. The mean value of the time interval between two successive arrivals is $1/\lambda$:

$$\int_0^\infty\ p(0, t)\ \cdot\ dt\ =\ 1\,/\lambda\,,\ 0\ <\ 1/\,\lambda\ <\ \infty \tag{28}$$

where $p(0, t)$ is the probability that there are no arrivals within the time interval $(0, t)$, which is also identical to the probability that the time passed until the first event is larger than t (the complementary distribution). The mean value $1/\lambda$ can also be interpreted as the area under the curve $p(0, t)$, which is a never–increasing function decreasing from 1 to 0.

3. Item 1 implies that no arrivals within the interval of length 0 is an event, sure to take place [79]:

$$p(0,0)\ =\ 1 \tag{29}$$

4. Item 2 implies that no arrivals within a time interval of length ∞ is an event which never will take place [79], i.e.

$$p(0, \infty)\ =\ 0 \tag{30}$$

Poisson distribution is a good representation of most of the telecommunication events. The arrival of calls to a call center is usually associated with a Poisson process.

However, the 4G telecommunication technology is establishing new levels of user experience and multi-service capacity, integrating all previous mobile communication technologies. In this scenario, the arrival of a user request for service is not just for a traditional voice call, but also can be for a data browsing session or a video call, among others [80]. The session arrival process may be approximated by Poisson distribution, under one or more of the following assumptions: (i) There are a lot of users getting access to a certain site, (ii) users are assumed to be independent, (iii) the situation is similar to a telephone network, where Poisson assumption holds [81], (iv) at time zero, the event is always zero, (v) it is stationary and (vi) it is memoryless.

In particular, the user model in LTE is called finite buffer, which is characterized by a finite amount of data that the user transmits or receives. Once the data transmission has finalized, the session is terminated, and the user is eliminated from the system. For simplicity, the session is typically considered as constituted by a unique useful load of data of a specific size. Therefore, the effects of upper layer protocols such as TCP and HTTP [82], in which the traffic is typically modeled as nonstationary and thus Poisson does not apply, are neglected. User arrival pattern in live LTE networks is affirmed to follow a Poisson process similar to that of legacy networks [80].

On the other hand, the Poisson distribution is widely used to model systems with a high probability of call drops like DTN (Disruption Tolerant Networking). Recent works using well-known mobility models, such as random waypoint, random direction and real-life mobility traces, have confirmed that the individual contact time follows an exponential distribution. Therefore, it is reasonable to model traffic by the Poisson process with different rate $\gamma i, j$ [83].

3.10 Poisson Distribution Simulator

In the Poisson distribution MATLAB simulator, voice and data calls are generated assuming a Poisson process for call arrivals, with arrival rates λc and λv, respectively. The inter-arrival times between successive arrivals of voice calls and data calls, given as iatc and iatv, respectively, are random variables which are distributed exponentially with parameters 1/λc and 1/λv. The call holding times are random variables which are also distributed exponentially with parameters 1/μc and 1/μv, where μc and μv are the mean service times of voice and data calls, respectively. The calls start and end times are recorded suitably in arrays in the MATLAB simulator.

Figure 15 indicates the call start and end times, the inter-arrival time iat (2,3) between calls 2 and 3, and the holding time cht (1) of call 1. Essentially, the start and end time of all calls and the inter-arrival times between the start of consecutive calls are generated according to the underlying random processes and recorded. Whenever a call starts, it requests the calling resources from the operator. The call starts if either the home or virtual operator resources are available. Otherwise, it is blocked. The simulator runs in this manner for a long enough time, and the ratio of blocked calls to the total number of calls is calculated for both voice and data calls to compute the corresponding call blocking probabilities.

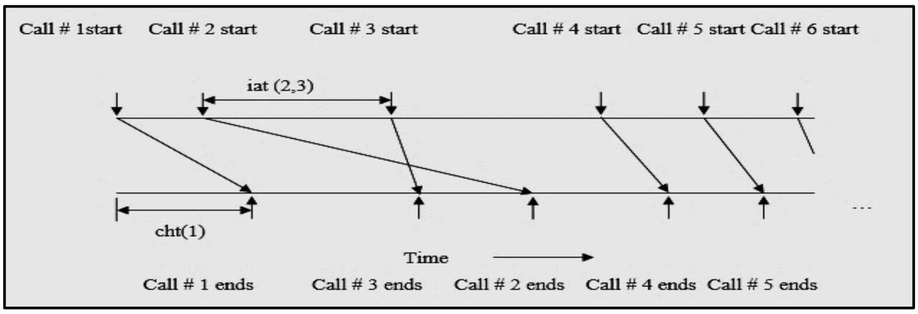


Figure 15 PDS Start and End of Calls

As it can be seen in Figure 15, a call that started later can end before a call that had started earlier. In the Poisson distribution MATLAB simulator, the numbers $n_{voice}(t)$ of voice calls and $n_{data}(t)$ of data calls at any time t for a particular operator, can be calculated by adding the total voice/data calls arrived up to time t, and subtracting the calls that have been serviced at time t. If the home operator does not have the resources to support the call, other operators combined resources are provided to the call. The resources required for a new call are taken from, if required, multiple operators. The simulator is run for a long enough time to calculate the total number of arriving calls and rejected calls corresponding to a particular operator.

3.11 Simulation and Emulation

Simulations were carried out in two stages. In the first stage, MATLAB was used to run a numerical example for verifying the mathematical model and the proposed framework.

The second stage consisted of using Emulated Virtual Environment (EVE) and Wireshark, for emulating the traffic in the network for different scenarios.

Chapter 4

4 Results and Discussions

4.1 First Stage: MATLAB

This section includes the results obtained for various scenarios regarding voice and video call blocking probability, and bandwidth utilisation as well. The proposed CAC algorithm is executed in the following two ways, to illustrate the notion of the proposed on-demand framework:

(i) **Simple CAC**: This basically means a single domain; it may also be referred to here as a single MNO within its own VMNOs. In this case, when the home operator resources are not available to be used by a mobile user, the virtual resources of any another VMNO are provided, if available. As a result, the call admission is served by the MNO itself.

(ii) **Distributed CAC**: When the home operator resources are not available to be used by a mobile user, the combined virtual resources of other MNOs are provided, if available. Therefore, in this case the VNE is required on demand, and the call admission process belongs to such VNE.

4.1.1 Proposed CAC Algorithm

By default, a user will always make a call to its own operator (the Home Operator), who will decide how to assign the resources required for such call. These resources include air interface and core network resources. If the home operator is not able to provide the resources itself, the call would go to the VNE. The VNE is responsible for obtaining the air interface resources required for the call from all operators in a combined fashion. Such resources would be provided through the VNE, by the operator with which the home operator has service level agreements.

The proposed CAC algorithm for handling voice and data calls of operators is illustrated in the flowchart shown in Figure 18, which depicts the sequence of actions the operator would undertake when a user makes a voice or data call. It can be observed that the sequence of decisions is different for non-real time and real-time calls. If resources are not available the call is rejected, and other real-time applications will use the same path decision to benefit from the LBO functionality and agreements. These matters will be described in flowcharts of different scenarios (see section 4.3).

4.1.2 Numerical Results

For illustration, assume that the resources available with N = 3 operators are [1.1 0.5 2.0] PRBs. An arriving video call to any of the operators is rejected because the required number of PRBs for a video call is 2.5. Even if virtualisation is enabled with the Simple CAC algorithm, the call is rejected as none of the operators has 2.5 PRBs available. However, with the Distributed CAC in operation the call is accepted, since the total available PRBs are 3.6, which is greater than 2.5.

The simulations were carried out for the Simple CAC and the Distributed CAC algorithms for N = 3 operators, with the Number of Physical Resource Blocks PRB = 25, α = 1, $\delta_{ave_{voice}}$= 0.12 and $\delta_{ave_{video}}$ = 2.5. The voice call arrival rate is λ_c = 0.5/sec and the voice and video call service rates (μ_c and μ_v) are both 0.02/sec. Values of zero for both γ_c and γ_v imply that the simulation is being performed for the normal (non-virtualised) case, while values of γ_c = 0.9 and γ_v = 1 imply the virtualised case. The reported blocking probabilities are equal for all operators.

Table 8 shows the results of the PDS for N = 3, PRB = 25 and α = 1. It can be seen that, as expected, the voice call blocking probability is less than the data call blocking probability.

Table 8 PDS Numerical Results

Input Parameters			Results of the Poisson Distribution Simulator (with Simple CAC algorithm)		Results of the Poisson Distribution Simulator (with Distributed CAC algorithm)	
λ_v	γ_c	γ_v	$P_{Bl_{voice}}$	$P_{Bl_{video}}$	$P_{Bl_{voice}}$	$P_{Bl_{video}}$
Video call arrival rate (s^{-1})	Fraction of virtual resources satisfying QoS (voice)	Fraction of virtual resources satisfying QoS (video)	Voice call Blocking Prob.	Video call Blocking Prob.	Voice call Blocking Prob.	Video call Blocking Prob.
0.1250	0	0	0.0048	0.1219	0.0048	0.1219
0.1250	0.9	1	$1.6e^{-5}$	0.0334	0.0007	0.0237
0.1375	0	0	0.0066	0.1579	0.0066	0.1579
0.1375	0.9	1	$2.2e^{-5}$	0.0632	0.0017	0.0446
0.1500	0	0	0.0099	0.1892	0.0099	0.1892
0.1500	0.9	1	$3.2e^{-5}$	0.0955	0.0026	0.0759
0.1625	0	0	0.0115	0.2275	0.0115	0.2275
0.1625	0.9	1	$9.6e^{-5}$	0.1314	0.0039	0.1035

The above numerical results are also plotted in Figure 16, which shows the blocking probability for voice and data calls in the normal and virtualised cases for the following parameters: Voice call arrival rate $\lambda_c = 0.5$ s^{-1}, Voice call service rate μ_c and Video call service rate μ_v both equal to 0.02 s^{-1}, Fraction of virtual resources satisfying QoS (voice) $\gamma_c = 0.9$, Fraction of virtual resources satisfying QoS (video) $\gamma_v = 1$. The varying parameter is the Video call arrival rate λ_v, which is given in s^{-1}.

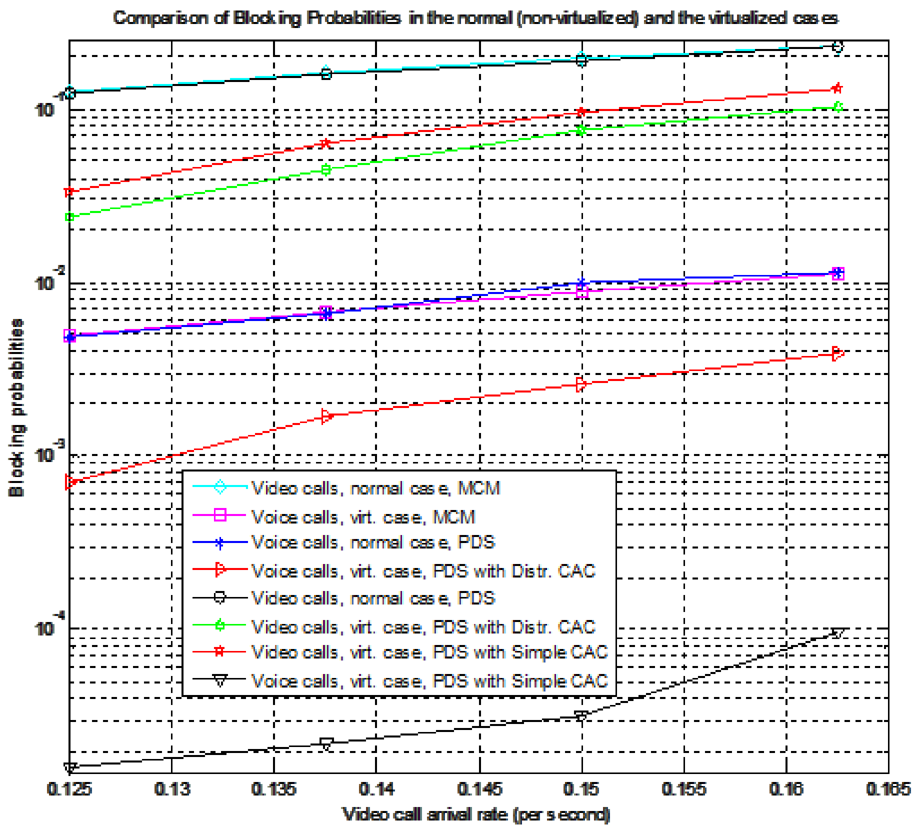


Figure 16 Blocking Probability of voice and data calls for the normal and virtualised cases

It can be observed that a difference of approximately one order of magnitude exists between the data call blocking probability in the normal (green curve) and virtualised (black curve) cases with distributed CAC, for $\lambda_v = 0.125$ s^{-1}. The difference between the voice call blocking probabilities in the normal (blue curve) and virtualised (red curve) cases with distributed CAC, is also of approximately one order of magnitude for $\lambda_v = 0.125\ s^{-1}$.

The results of the PDS simulator for the simple CAC are worth noting. With the simple CAC, the data call blocking probabilities increase compared to the distributed CAC. This is expected as the resources in the simple CAC are not available in a combined manner. When any single virtual operator has the resources to support the data call (which takes 2.5 PRBs), the call is accepted. However, the voice call blocking probabilities are reduced by over two orders of magnitude when compared to the typical case. This is because more resources are now available for voice calls, as data calls are rejected with higher probability, and the mean number of PRBs required for a voice call is only 0.12. In the following, unless otherwise stated, the CAC algorithm will refer to the Distributed CAC algorithm.

4.1.3 Bandwidth Utilisation

Figure 16 shows the utilised bandwidth of N = 3 operators for the standard (non-virtualised) scenario, when the PRBs for such operators are [25 25 25], the inter-arrival times of voice calls are [2 2 2] sec, and the inter-arrival times of data calls are [8 8 8] sec. It can be seen that the maximum utilised bandwidth is 5 MHz (corresponding to 25 PRBs), and the mean utilised bandwidth is 3.32 MHz. On the other hand, Figure 17 shows the utilised bandwidth for N = 3 operators for the virtualised scenario, with PRBs, inter-arrival times of voice calls and inter-arrival times of video calls equal to the previous case. The mean bandwidth utilised by all operators is 3.66 MHz, which is 0.34 MHz greater than the corresponding to the non-virtualised case.

It is important to mention that α has been used to indicate the fraction of resources allotted for new calls, and (1 – α) is the fraction kept for handover calls. Since in this scenario handover calls are not considered, α = 1 for all operators

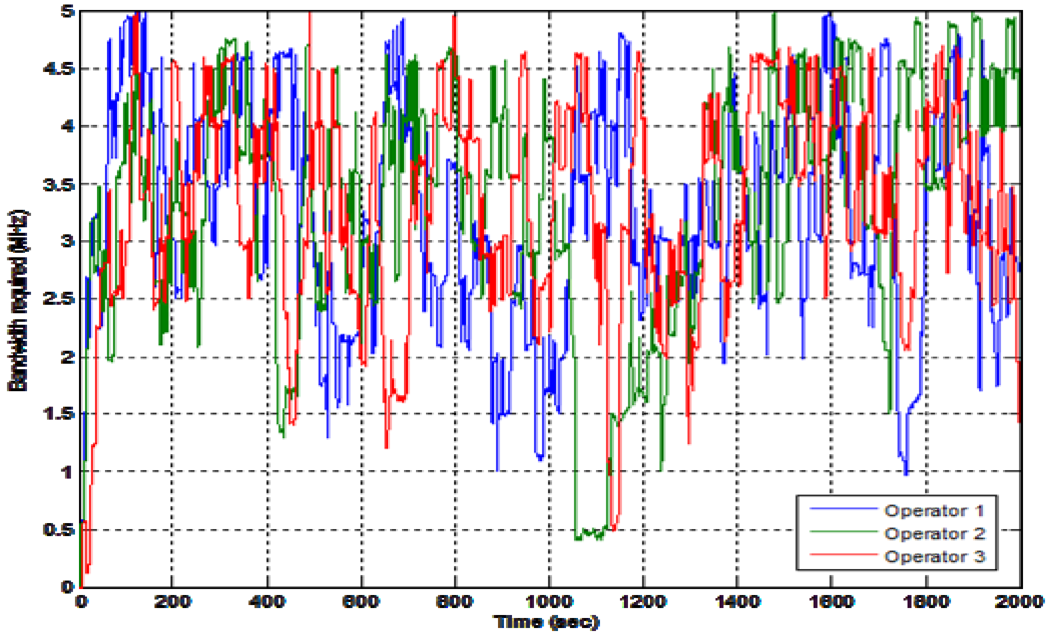


Figure 16 Bandwidth Utilisation (non-virtualisation)

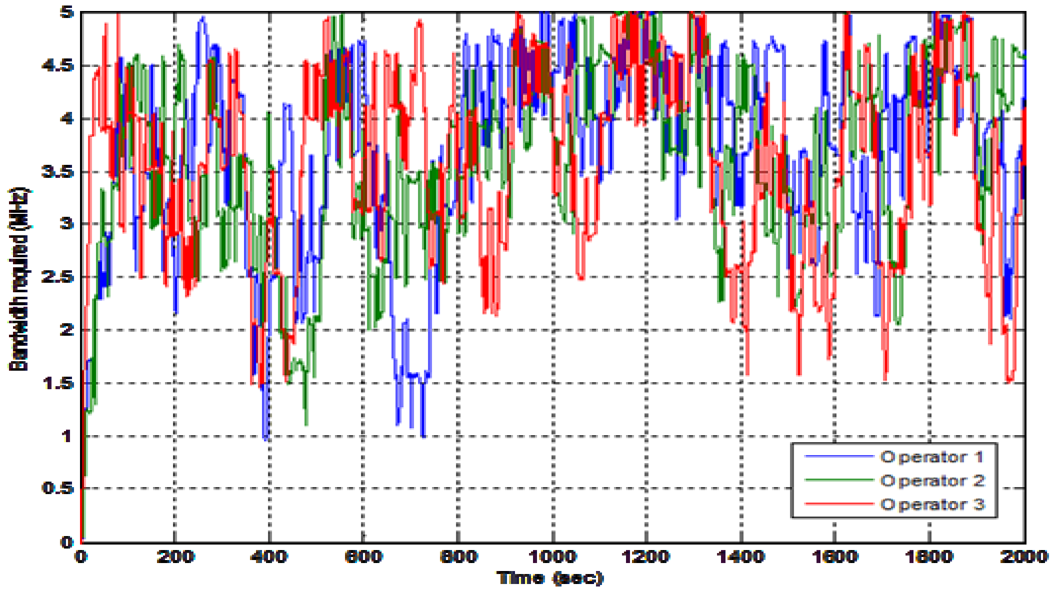


Figure 17 Bandwidth Utilisation.

4.2 Second Stage: Emulation EVE and Wireshark

The main purpose is to emulate real traffic scenarios in order to validate the performance of the proposed offloading and LBO strategies.

i) EVE, which are considered the most advanced realistic emulators, were used in the final emulation to test a real-life scenario of networking with different vendors (routers, switches, servers, etc.) [84].

ii) Wireshark was used to show real live traffic at the packet level in networks, and to see TCP frames with details. It was also used to conduct a live statistical analysis of the traffic flows with the RTT to it (all Live while establishing traffic). It is one of the finest protocol analysis tools that can be used to obtain information on both the packet and frame levels [85].

4.2.1 Emulation preview

The purpose of this emulation is to explain how the Virtual Network Enabler will operate in an automated way between two service providers MNO1 and MNO2, to switch excess traffic from one to the other if the former is congested. This switching is performed based on predefined thresholds for congestion and delay, as illustrated in the flowchart shown in Figure 18. It also explains the LBO behaviour for MNO1, such that this LBO carries a specifically identified traffic (an application traffic with marking AF22 was used in this emulation). The devices used in the emulation include (i) Routers, (ii) Layer 2 Switches, (iii) MLS (Multi-Layer switches) and (iv) Servers.

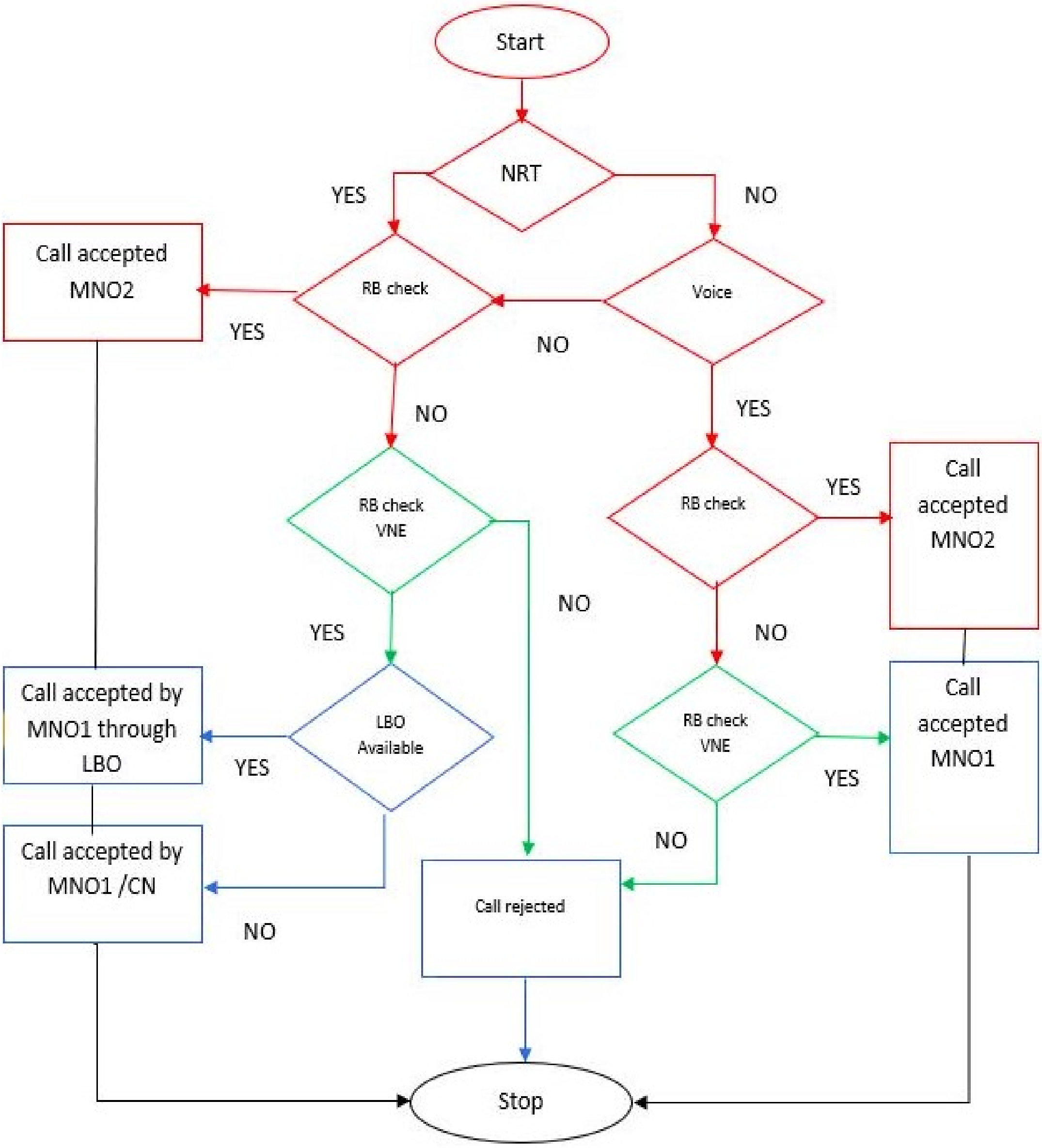


Figure 18 Proposed Framework Flow Chart

In the emulation, the Serving GPRS Support Node (SGSN), MME, Home Subscriber Server (HSS), Gateway GPRS Support Node (GGSN), Packet Data Network Gateway (P-GW), PCRF, Serving Gateway (SGWY) are represented as the servers. On the other hand, the Gateway with LBO, Gateway and the Virtual Network Operator are emulated with routers and switches. The protocols used in the emulation were: (i) BGP, (ii) EIGRP, ISIS, (iii) MPLS, VRF, (iv) Cisco Event Manager Script (EEM), (v) Static routing and (vi) QOS (Marking, Classifying & policing).

4.2.2 Devices Topology Explanation:

Figure 19 shows the Emulation Diagram Created on EVE, in which the router Family ISR4k (4000 Integrated Services Router) from Cisco was used. The routers in this family have a processor and memory functions that provide convergence abilities, security and high performance appropriate to be used by many service providers in complicated networks. For instance, features offered by this series include (i) a performance throughput than can reach up to 2 Gbps, (ii) a high-bandwidth module-to-module communication with a backplane architecture that allows it to deliver up to 10 Gbps, and (iii) a software bundle with advanced routing and network monitoring services. In addition, one of its finest features is a dual integrated power supply for backup purposes. In fact, all the ISR 4000 Family supports optional power supply that can deliver extra power over ethernet (PoE) to distensions [86].

The high-performance gateway of the ISRs allows VoIP to run over less expensive Session Initiation Protocol (SIP) trunks. Additionally, it contains an integrated IP PBX and a Session Border Controller. Correspondingly, all features and performance-on-demand licensing flexibility is supported by Cisco together with third-party vendors, by a universal software which is used as a manageable tool and can also support multi programming languages [86].

For the Switching part, the series 3850 Stack switches was used , where each base station has at least 2 links connected to each stack switch member in a redundant way (one link is active, and the other link is passive until the active link goes down); each stack switch can handle 48 base stations. The Catalyst 3850 switch gives you smart, simple, and highly secure converged access and support PoE [86].

The ISR series was used in the simulation instead of the ASR, due to the limitations of resources (RAMs, Processor, etc.). However, in real life, service providers use ASR 5K, 9K and 1k series, as recommended by Cisco [87-91]. Note that the connectivity in real life is different than in this lab; the main purpose of the work is to emulate the routing and decision making by LBO & VNE.

Regarding the LBO Diagram, this topology is chosen since there has to be a device, the Gateway-with-LBO in this case, that handles the excess traffic. In addition, the switch 5 (3850 stack switches) is placed before the Gateway-with-LBO, as this switch is known for its high density port to adapt many Enodes, each with fully redundant connectivity (one link is active and the other link is passive till the active link goes down); each stack switches can handle 48 Enodes. The Catalyst 3850 switch gives you smart, simple, and highly secure converged access, and supports PoE [91].

4.2.3 TCP Retransmissions

In a TCP retransmission, an associated sequence number is tagged to each byte of data sent, on the field of the TCP header. Such acknowledgement number is used to indicate that the segment of data has been received. When the receiving socket detects an incoming segment of data, the source will set a retransmission timer of variable length after a packet is sent. If the acknowledgment response has not been received before the timer expires, the source will retransmit the segment assuming it has been lost on the way to its destination; this will cause a delay to the process. The TCP retransmission mechanism is used to ensure that data is truly sent from the source to the destination without any error [92].

that the segment of data has been received. When the receiving socket detects an incoming segment of data, the source will set a retransmission timer of variable length after a packet is sent. If the acknowledgment response has not been received before the timer expires, the source will retransmit the segment assuming it has been lost on the way to its destination; this will cause a delay to the process. The TCP retransmission mechanism is used to ensure that data is truly sent from the source to the destination without any error [92].

that the segment of data has been received. When the receiving socket detects an incoming segment of data, the source will set a retransmission timer of variable length after a packet is sent. If the acknowledgment response has not been received before the timer expires, the source will retransmit the segment assuming it has been lost on the way to its destination; this will cause a delay to the process. The TCP retransmission mechanism is used to ensure that data is truly sent from the source to the destination without any error [92].

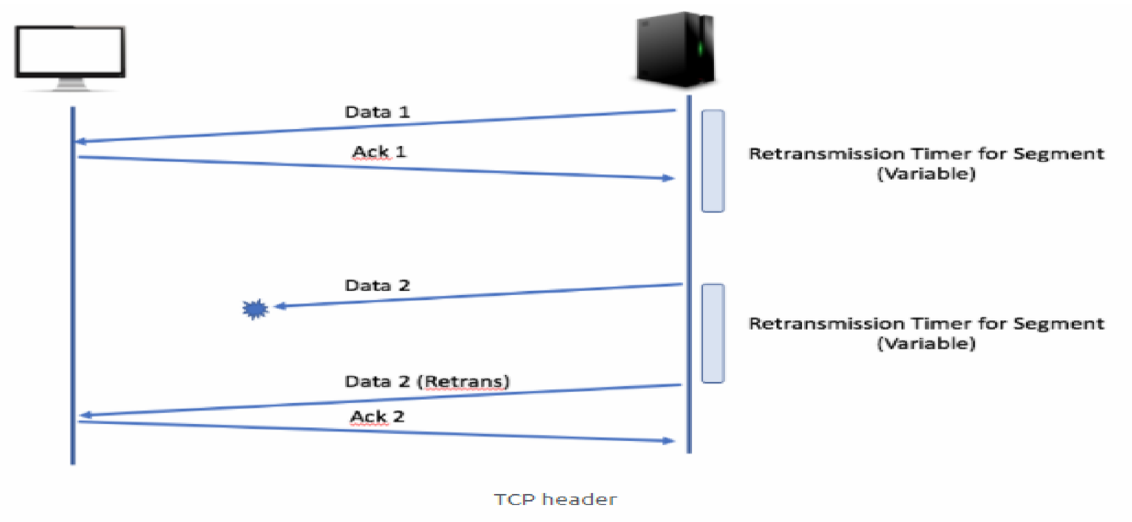


Figure 19. TCP Retransmission [92]

4.2.4 TCP Duplicate/Selective Acknowledgments

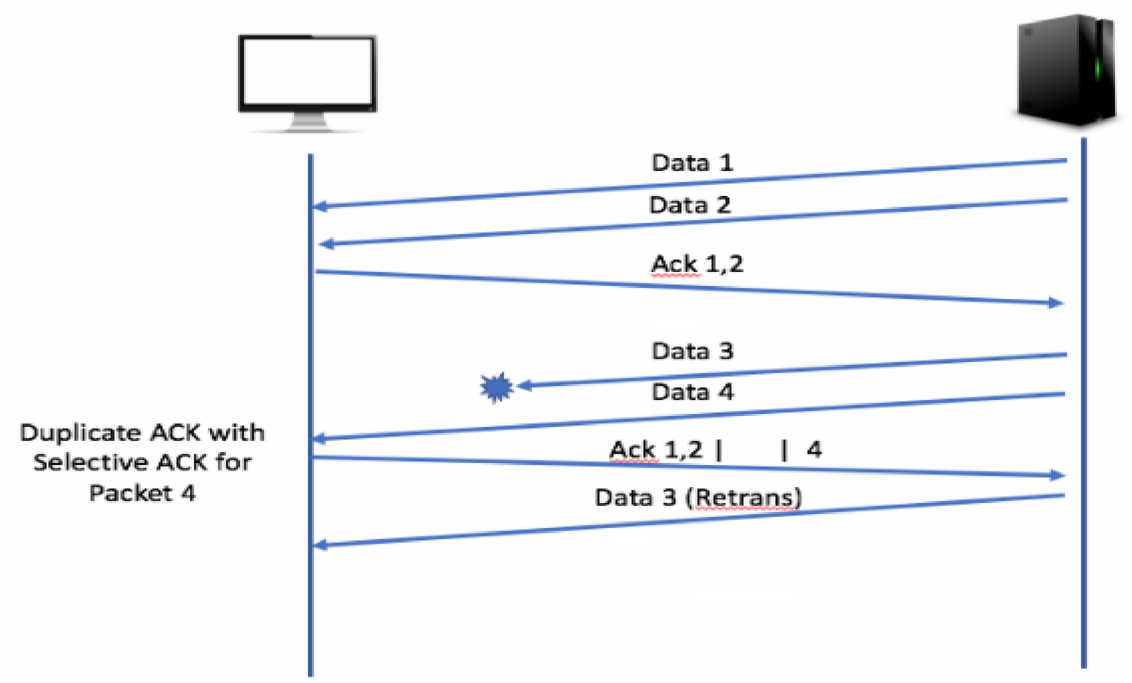


Figure 20. TCP Duplicate/Selective Acknowledgments [92]

The detection of two packets with the same ACK numbers is known as a Duplicate Acknowledgment condition, which indicates that one or more data packets have been lost in the transmission process and the source is trying to recover it [93]. Therefore, SACK and the Duplicate Acknowledgments are a common indicators of packet loss.

The selective acknowledgment (SACK) function is promoted by each station at the beginning of the TCP connection. A SACK occurs when several data packets from the source (TCP socket) are sent to the destination, rather than sending it packet by packet and waiting for an acknowledgement. In this situation, the destination (receiving socket) can identify the lost packets during this process by using SACK. As a result, the incoming data will flow to the destination, and at the same time it informs the source of the missing data in the segments [93].

As shown in Figure 20, the SACK will use the ACK number in the TCP header to specify which packet was lost. At the same time, the receiver can use the SACK option in the TCP header of these ACK packets, to show which have been successfully received after the point of loss.

Then the higher the latency between the source and the destination results in more duplicate acknowledgement connections. In this work, Wireshark is used for illustration purposes. As shown in Figure 23, when a segment is lost a retransmission is required; as a result, more time elapses, and the jitter will increase.

Depending on how many retransmissions occur in the process and how fast the endpoints can recover the missing packets, the application performance and the overall process can be significantly impacted.

4.3 The *Emulation* Scenarios

The scenarios described in this section are based on the Emulation diagram shown in Figure 21. All possible scenarios are driven from the proposed framework (Figure 18) to show the importance of the collaboration and the offloading techniques.

Figure 21 shows the emulation diagram created by EVE, which illustrates the layout of the network with two MNOs, an internet service provider, two VNE with LBO agreements for each MNO and the LBO. Network topology can be considered as overlaying network, meaning VPNs, peer-to-peer (P2P) networks and voice over IP services are examples of overlay networks which no cause any changes to the underlying network infrastructure [77].

Table 9. Emulation Scenarios

Scenarios	Brief	Results
Scenario 1	There is no VNE agreements between MNO1 and MNO2	Excess traffic beyond the identified threshold will be dropped
Scenario 2	There is a VNE agreement between MNO1 and MNO2	Once traffic exceeds the defined threshold, excess traffic is automatically shifted to Gateway-with-LBO towards MNO2 (will not be dropped)
Scenario 3	MNO1 has an LBO Agreement	Download (FTP) Traffic will go to the LBO cloud and the rest of the traffic will go through the MNO1
Scenario 4	MNO1 has no LBO Agreement	Download (FTP) Traffic will go to the MNO1 and delay will increase in MNO1
Scenario 5	MNO2 does not have LBO but it uses LBO by means of an agreement with MNO1	Download (FTP) Traffic will go to the LBO of MNO1

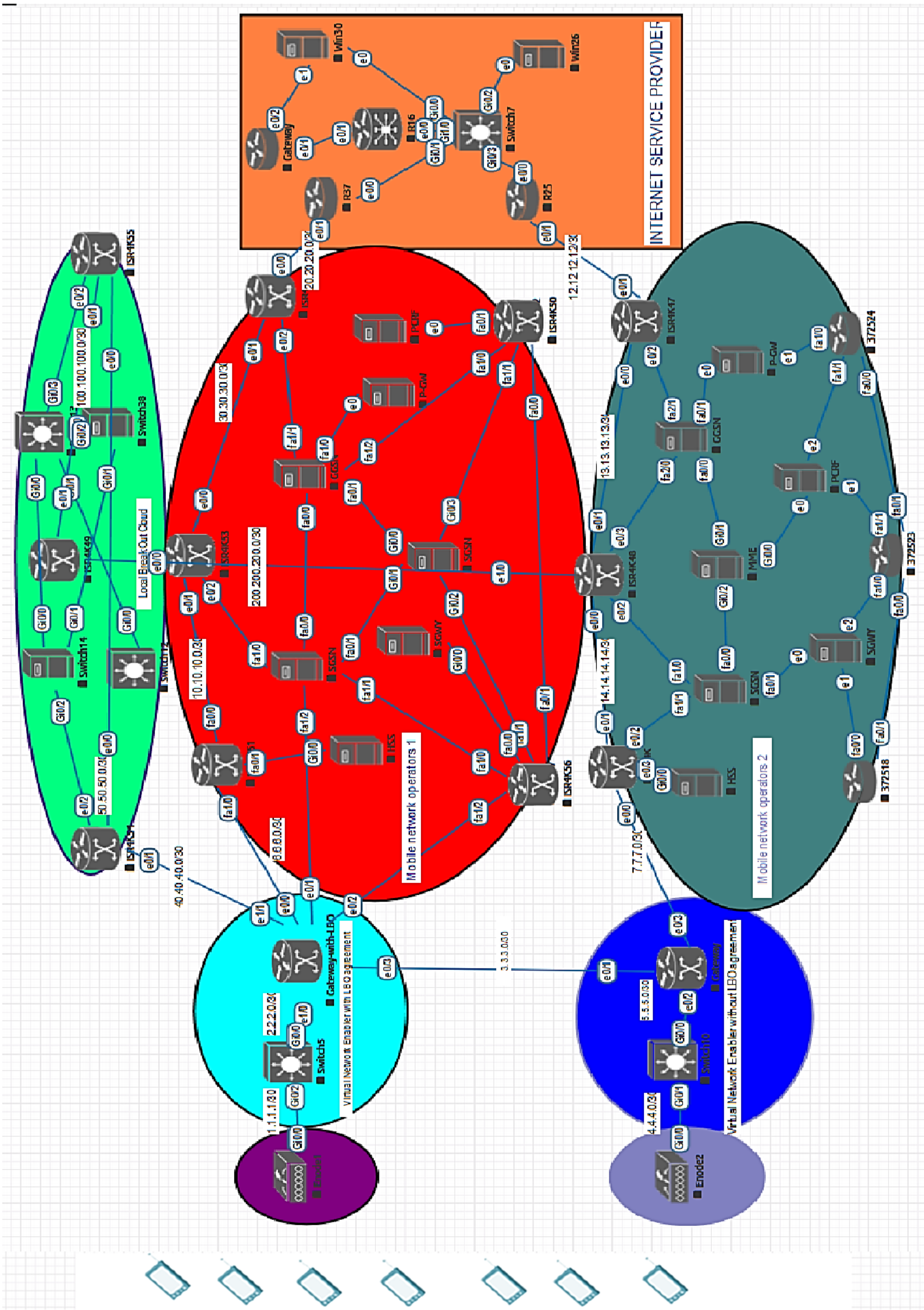


Figure 21: Emulation Diagram Created on EVE

4.3.1 First Emulation Scenario

This scenario considers that there is no Virtual Network Enabler agreement between MNO1 and MNO2. This was emulated by shutting down interface E0/3 of the Gateway with LBO agreement besides MNO1 (in Figure 21). Then, traffic towards the Internet was generated with the Emulator, using TCP to generate Traffic on port number 5001 (as a user application or gaming). Figure 22 shows the logic flowchart of the first scenario.

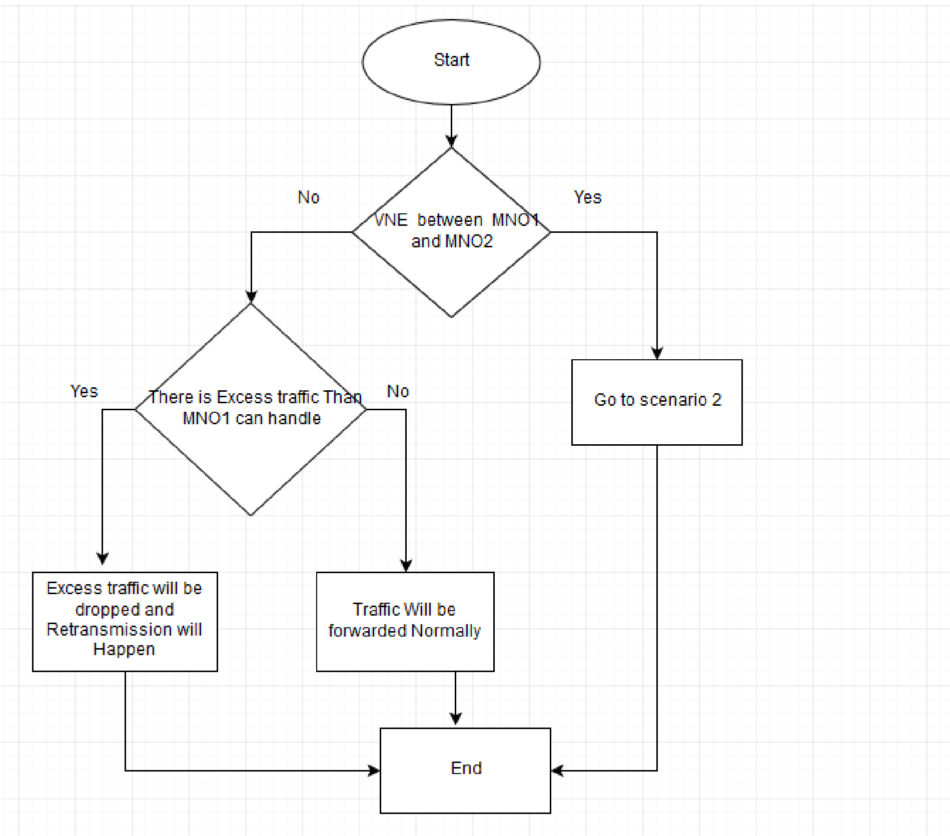


Figure 22. Logic flowchart of the first scenario

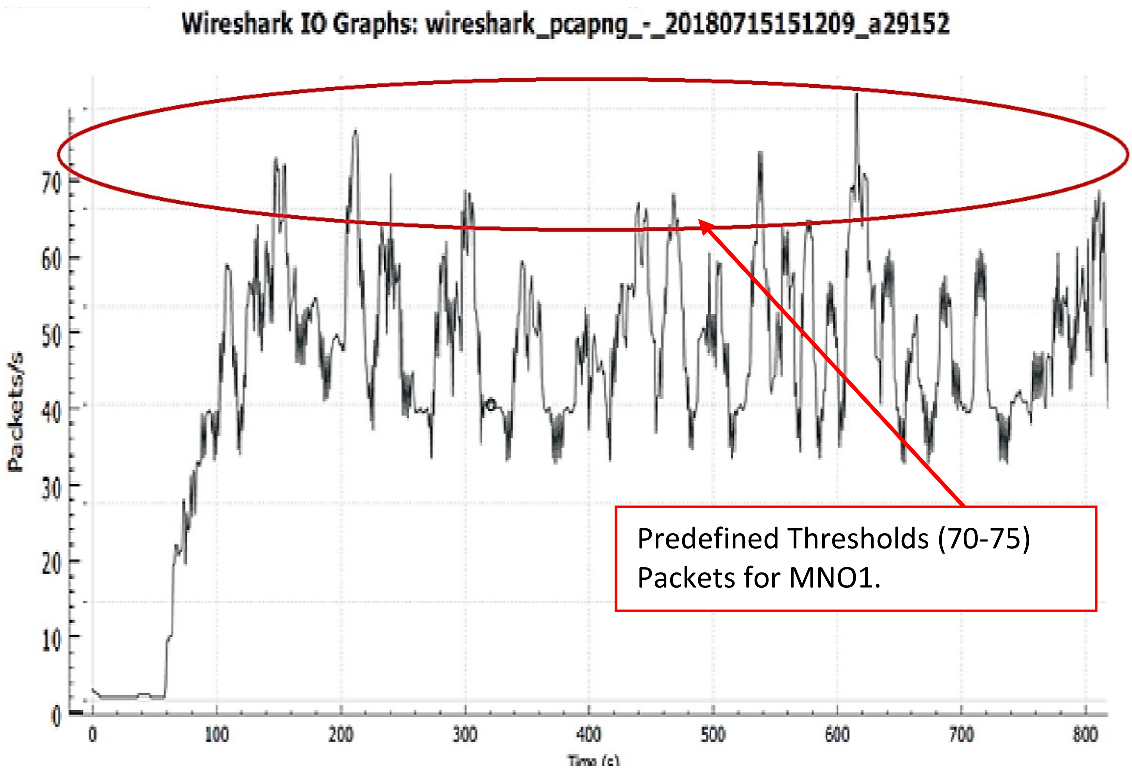


Figure 23. Emulation Results for the First Scenario

Figure 23 shows that MNO1 can fulfil limited maximum traffic, and the excess traffic beyond that threshold will be dropped as there is no VNE agreement between MNO1 and MNO2. Figure 23 also illustrates the number of packets in a time frame, where it can be seen that when the user tries to send large number of packets above the MNO1 predefined thresholds (more than 70-75 packets), the excess traffic will be dropped. As a consequence, the graph statistics is not smooth since MNO1 will drop a bulk of packets above 70-75, and then the users try to send more packets and again the MNO1 will successively drop the excess traffic (beyond 70-75 packets).

This is also clearly illustrated in Figure 24, which shows significant retransmission errors and packet drops (highlighted in Black) in the Wireshark results. These drops will prevent excess traffic from entering the MPLS backbone of MNO1 and traffic will be dropped, since there is no VNE enabler agreement.

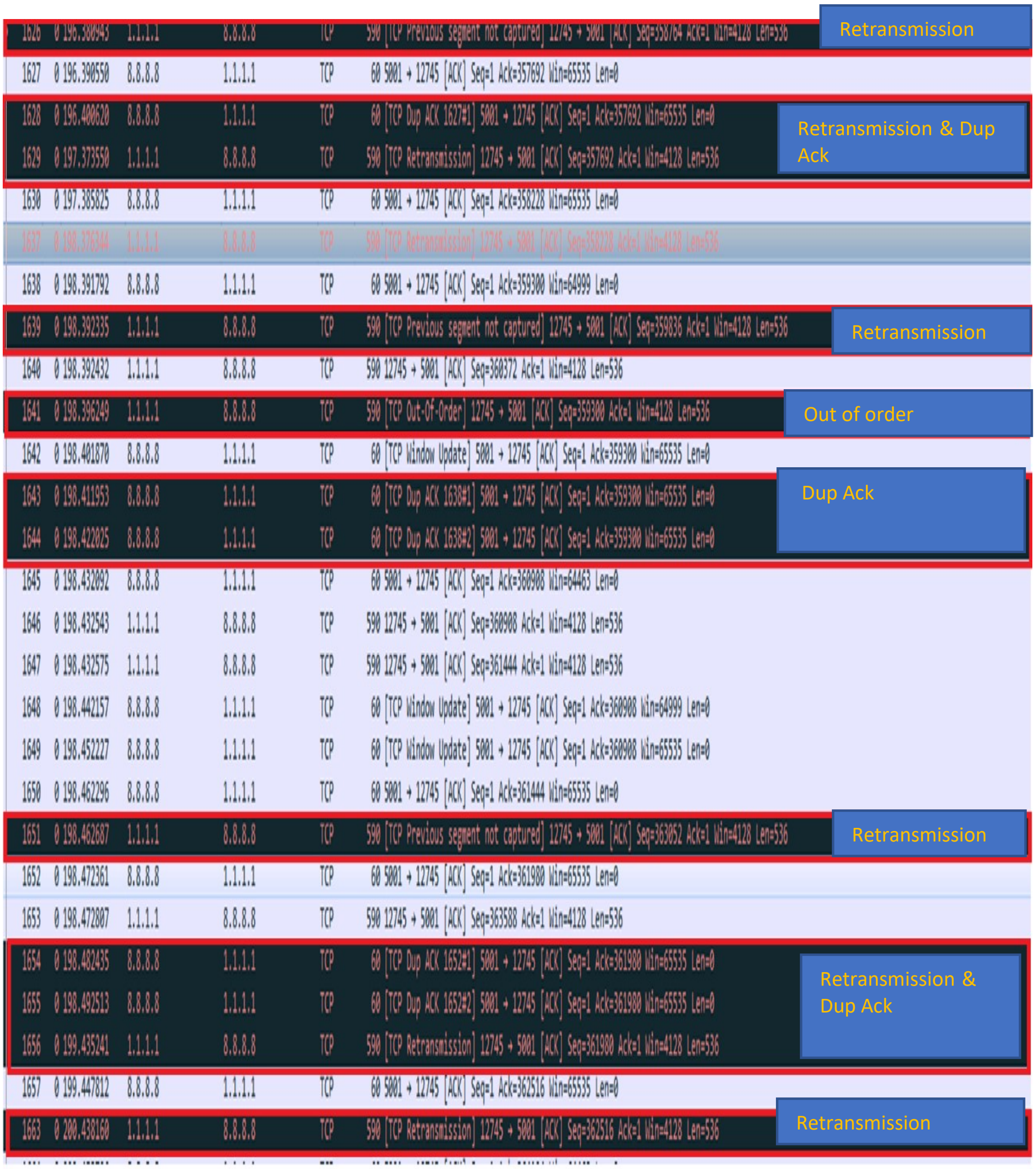


Figure 24. TCP Retransmission Due to Network Congestion for the First Scenario

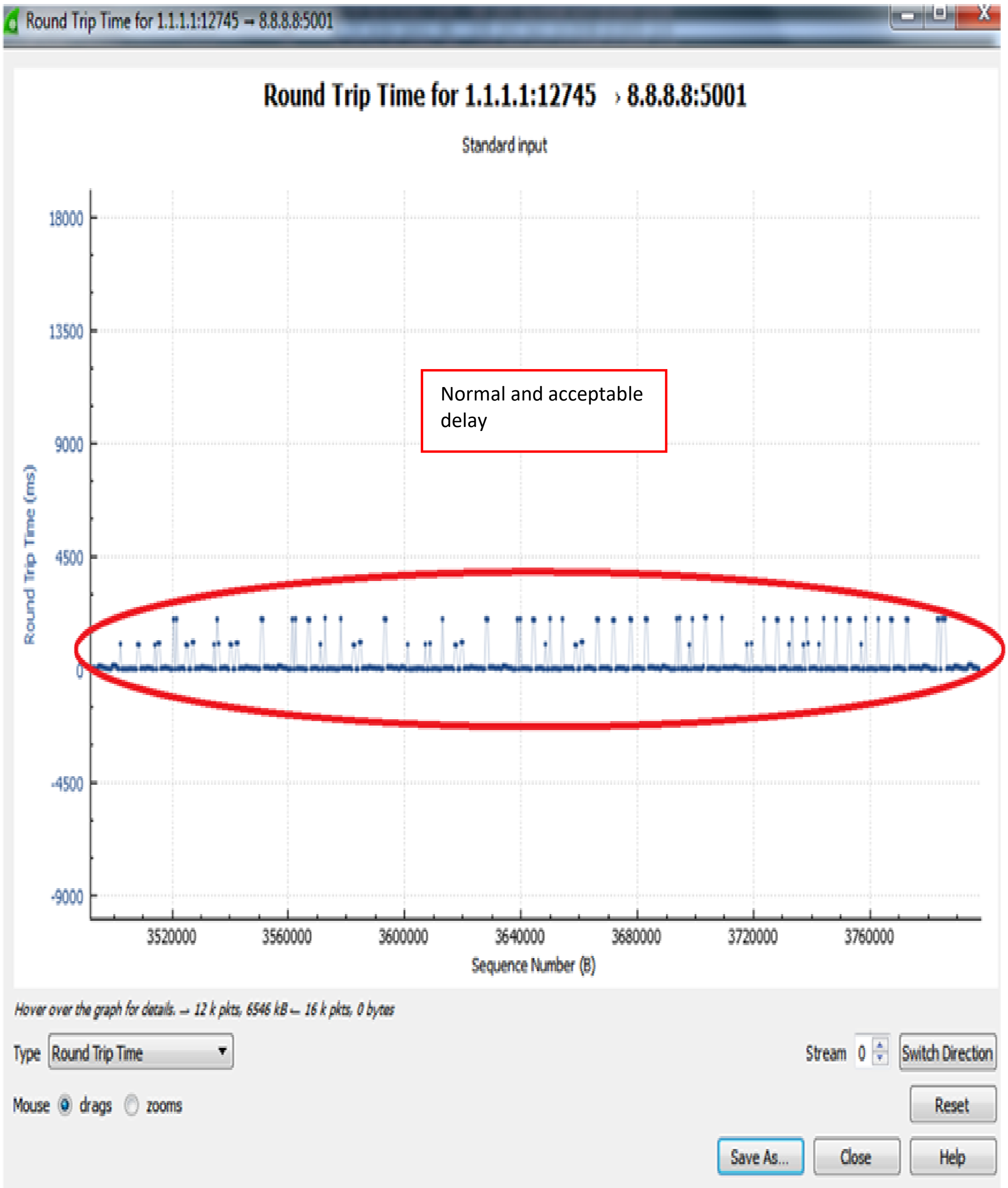


Figure 25 Normal RTD for Traffic that will pass to MNO1 in the First Scenario

Figure 25 shows statistical information for regular traffic only. It can be seen that the results are as expected, including a delay which is acceptable because there is no excess traffic received yet on the configured threshold of MNO1. If the users send a number of packets greater than the configured threshold MNO1, the delay will be significantly increased as it is shown in Figure 26.

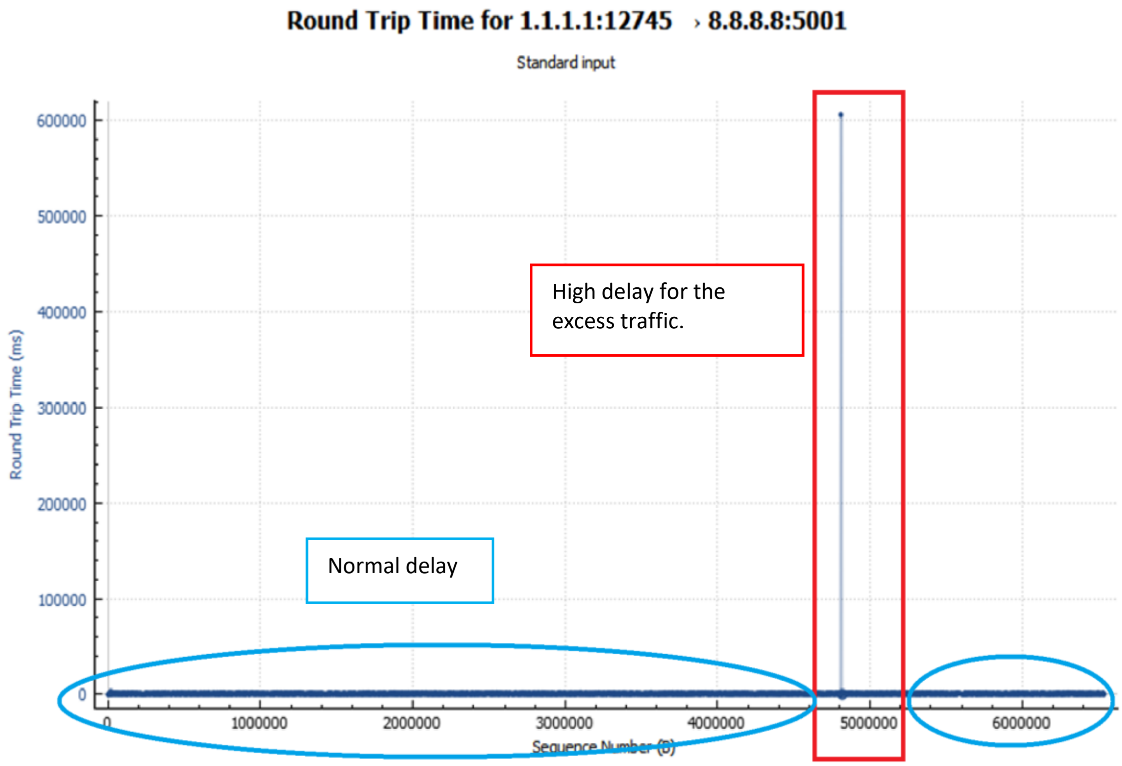


Figure 26. Significant Drops & Delay for the Excess Traffic for MNO1 in the First Scenario

Figure 26 shows statistical information corresponding to normal and excess traffic combined. Note that excess traffic experiments significant delay and drops, as can be seen in a sequence number close to 5000000, while the rest of the traffic accepted by MNO1 (regular traffic) does not exhibit significant delays. At this sequence number, the users sent large number of packets beyond the capacity of the MNO1, which caused an increase in the delay compared to the rest of the sequence numbers for other packets. Table 10 contains detailed information about the behaviour of the packets. Overall, 240 packets were dropped, 228 packets shown in the table and other 12 were dropped due to unknown protocol issues. Consequently, the QoS falls due to congestion, as there is no LBO agreement between MNOs. Then, it can be concluded that the excess traffic (beyond what MNO1 can serve) will be dropped, as there is no VNE between MNO1 and MNO2.

Table 10. Detailed Information about the Behaviour of the Packets in the First Scenario

Bearer Type	Priority	Example	Packet Delay	Marking Used in the Emulation	Packets in Emulation	Packet Drops with LBO
GBR	2	Voice Call	100 ms	DCSP EF	1800900 bps	2
GBR	4	Video Call	150 ms	DSCP AF41	130300 bps	10
GBR	3	Online Gaming (Real Time)	150 ms	DSCP AF31	203000 bps	34
Non-GBR	6	TCP based services e.g. email, chat, ftp	300ms	DSCP Af21	90000 bps	82
Non-GBR	7	Interactive gaming	100 ms	DSCP AF11	70000 bps	100

4.3.2 Second Emulation Scenario

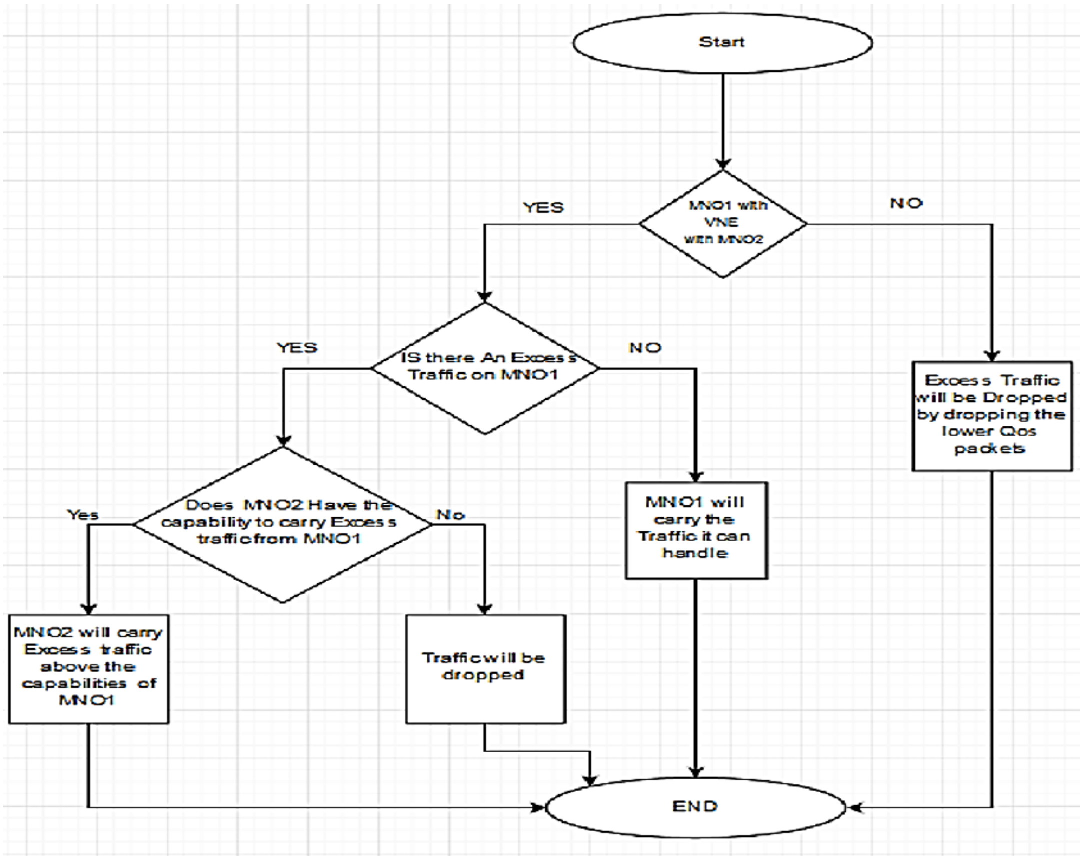


Figure 27. Logic Flowchart of the Second Scenario

This scenario considers that there is a VNE agreement between MNO1 and MNO2, which operates according to the logic flowchart in Figure 27. First, TCP was used to generate traffic towards the Internet on port number 5001 (as a user application or gaming). Traffic will then go to Gateway-with-LBO towards MNO1. Once traffic exceeds the defined threshold, excess traffic is automatically shifted to Gateway-with-LBO towards MNO2.

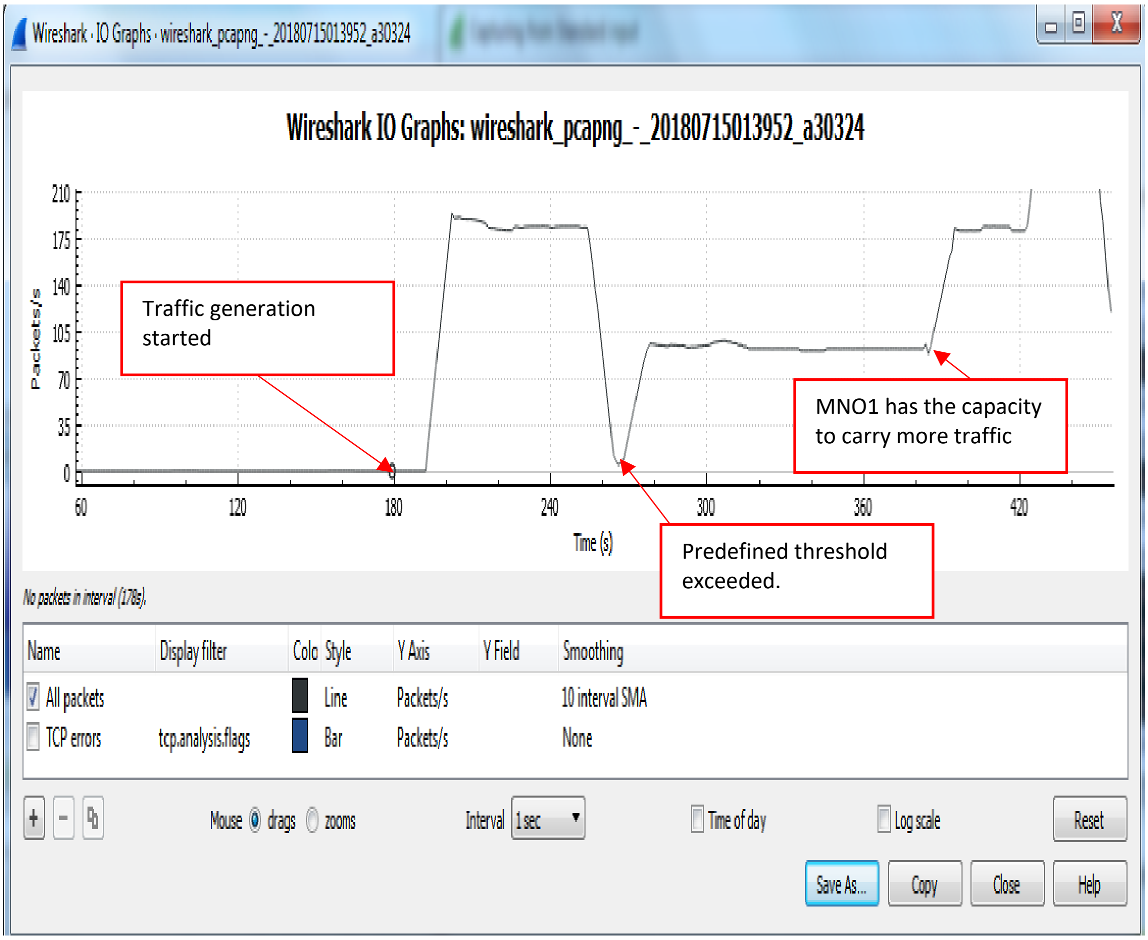


Figure 28. Traffic Analysis Towards MNO1 In the Second Scenario

Figures 28 and 29 analyse the traffic towards MNO1 and MNO2, respectively; note that there is some delay between these two graphs due to processor delay.

Since there is capacity available in its network resources, MNO2 start a TCP session to carry the excess without affecting the end users. At 380 sec, MNO1 has more capacity available and traffic is automatically shifted back from MNO2, to take MNO1 as the primary path again since it can adapt and take more traffic with less delay. Figure 28 plots the Round-Trip Time (RTT), which shows congestion in MNO1. Figures 31 and 32 show that all traffic is being served between MNO1 and MNO2.

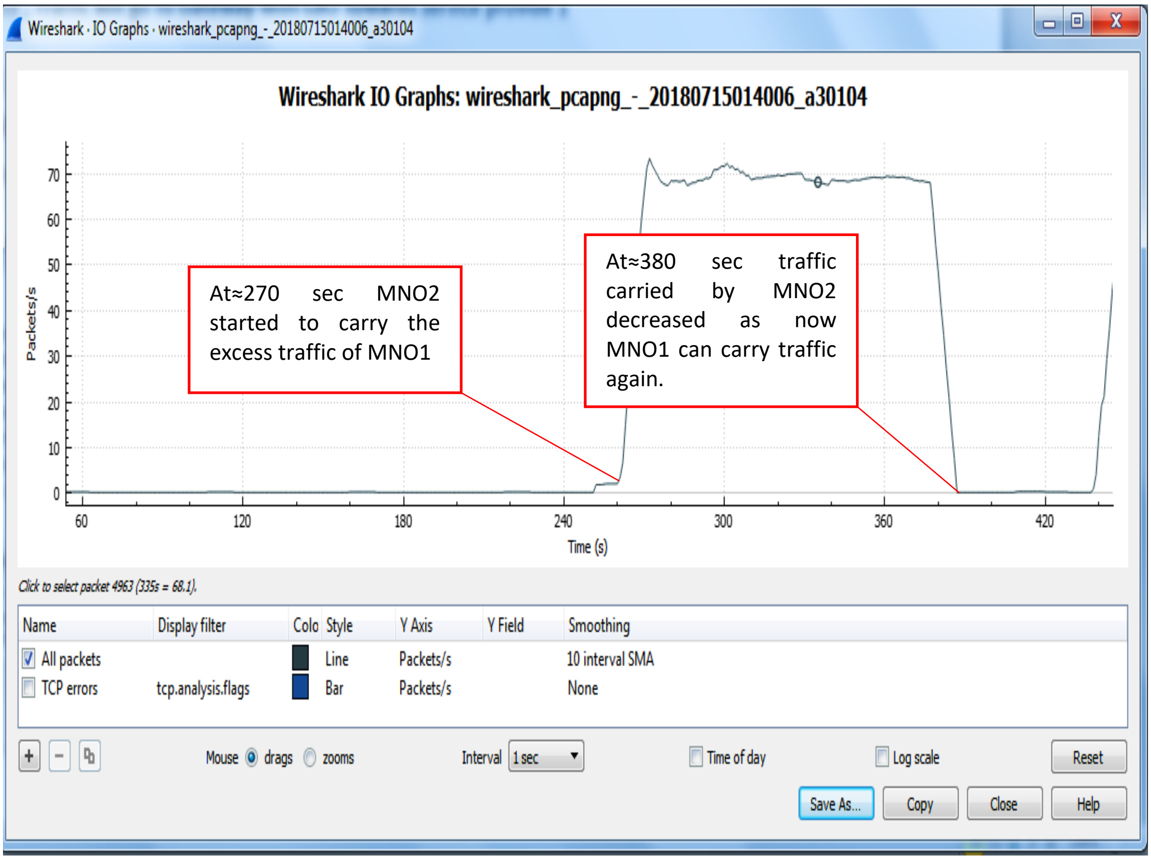


Figure 29. Traffic Analysis Towards MNO2 In the Second Scenario

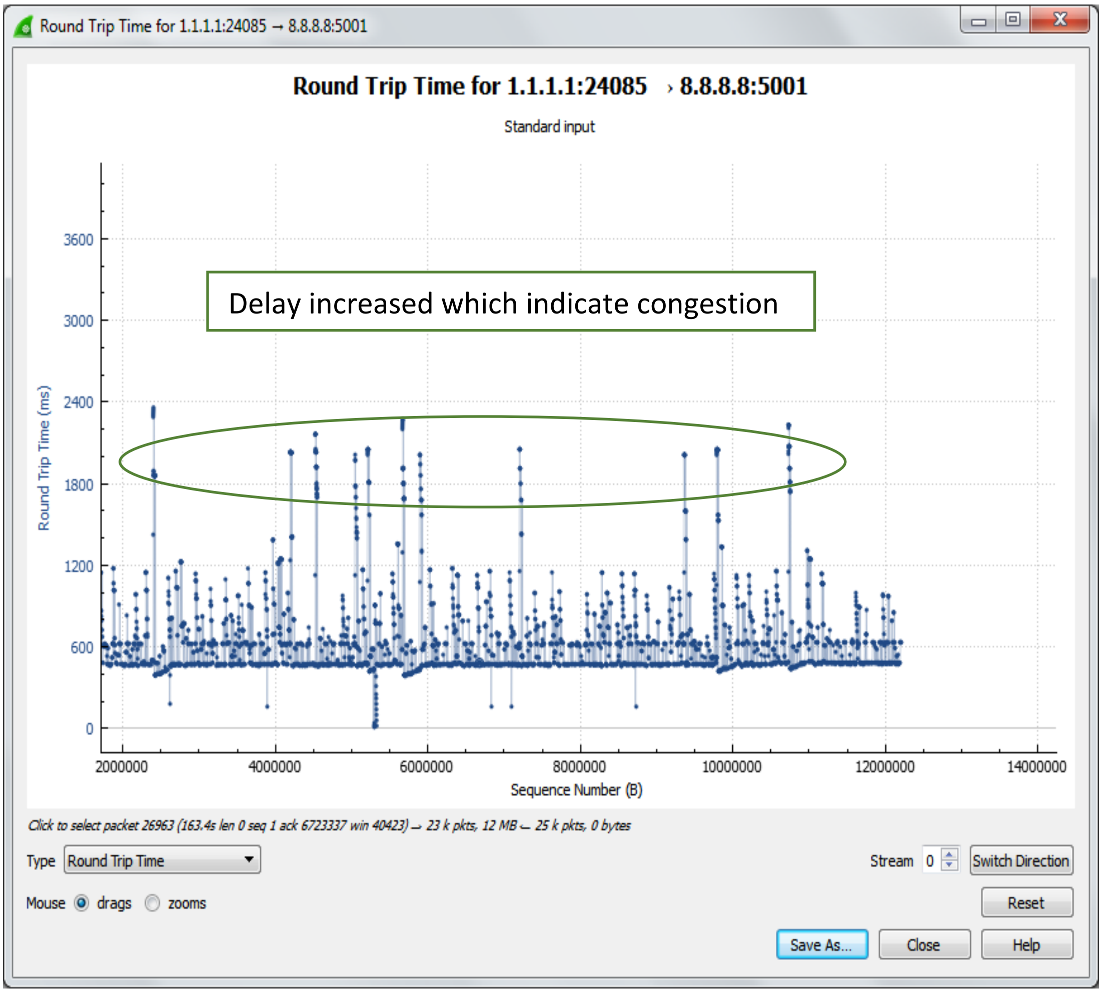


Figure 30. RTT shows congestion in MNO1 in the Second Scenario

In Figure 30 shows that the delay for MNO1 remains in an acceptable range, as there is a VNE agreement between MNO1 & MNO2. It increases when the identified threshold is reached at MNO1, and then decreases as MNO2 carries the excess traffic beyond the MNO1 configured threshold.

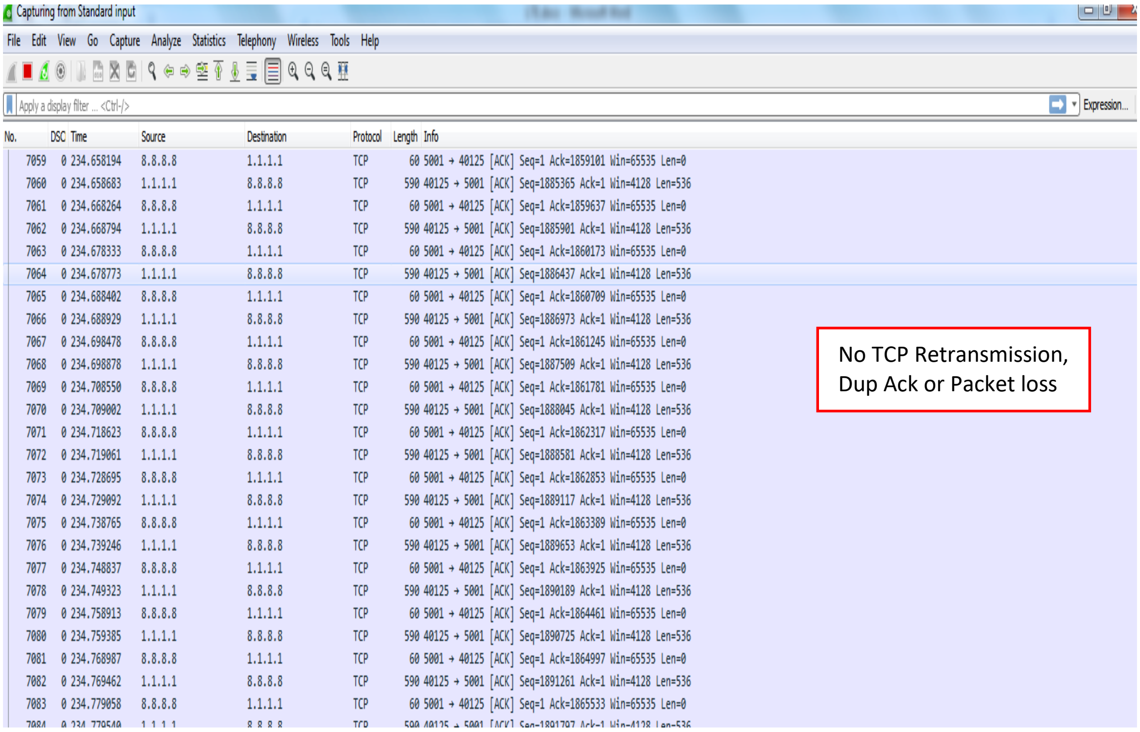


Figure 31. TCP Capture of Live Traffic, MNO1

Figure 31 shows that there are no packets lost at MNO1 (no TCP retransmissions or TCP duplicate ACK), due to the VNE agreement between the two operators and vies reserve in Figure 32.

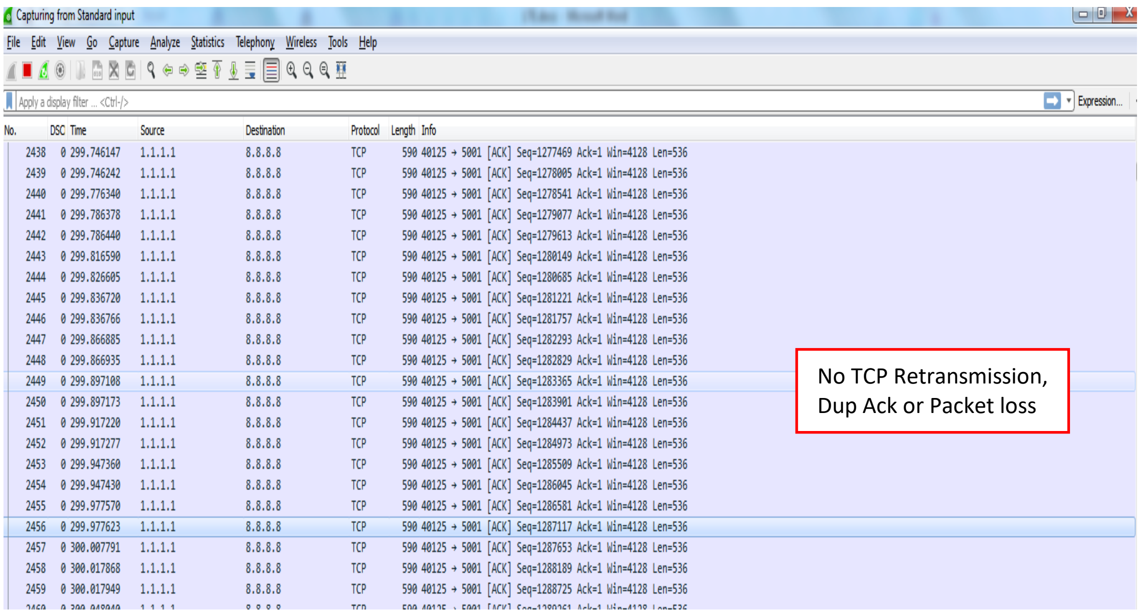


Figure32. TCP Capture of Live Traffic, MNO2

Table 11 Detailed Information about the Behaviour of the Packets in the Second Scenario

Bearer Type	**Priority**	**Example**	**Packet Delay**	**Marking Used in the Emulation**	**Packets in Emulation**	**Drops % with # LBO**
GBR	2	Voice Call	100 ms	DCSP EF	12000000 bps	0 %
GBR	4	Video Call	150 ms	DSCP AF41	1500000 bps	0 %
GBR	3	Online Gaming (Real Time)	150 ms	DSCP AF31	2000000 bps	0 %
Non-GBR	6	TCP based services e.g. email, chat, ftp	300ms	DSCP Af21	700000 bps	0 %
Non-GBR	7	Interactive gaming	100 ms	DSCP Af11	170000 bps	0 %

Table 11 contains information related to the QoS for live traffic (voice (RT), Video (RTVI) & data (specifically GBR and non-GBR)); it can be seen that the values are satisfactory. After comparing the results obtained in the first and second scenarios, it can be concluded that the performance is much better with the VNE agreement between MNO1 and MNO2 (second scenario).

4.3.3 Third Emulation Scenario

This scenario considers that MNO1 has an LBO Agreement. The traffic for LBO was generated with QoS AF22 like an FTP (downloading service), such that traffic from an end user that downloads a specific file or application will go through the LBO, and the rest of the traffic such as voice, gaming, among others, will go through the LTE Core Network of MNO1.

Back in Figure 21, the device named Gateway with LBO in the VNE with LBO agreement cloud, will monitor the traffic and make a decision based upon its nature: if the traffic is due to downloading (marked as AF22 in the Emulation) it will pass through the LBO cloud, and any other traffic will go through MNO1. Figure 33 depicts the logic flowchart of the process.

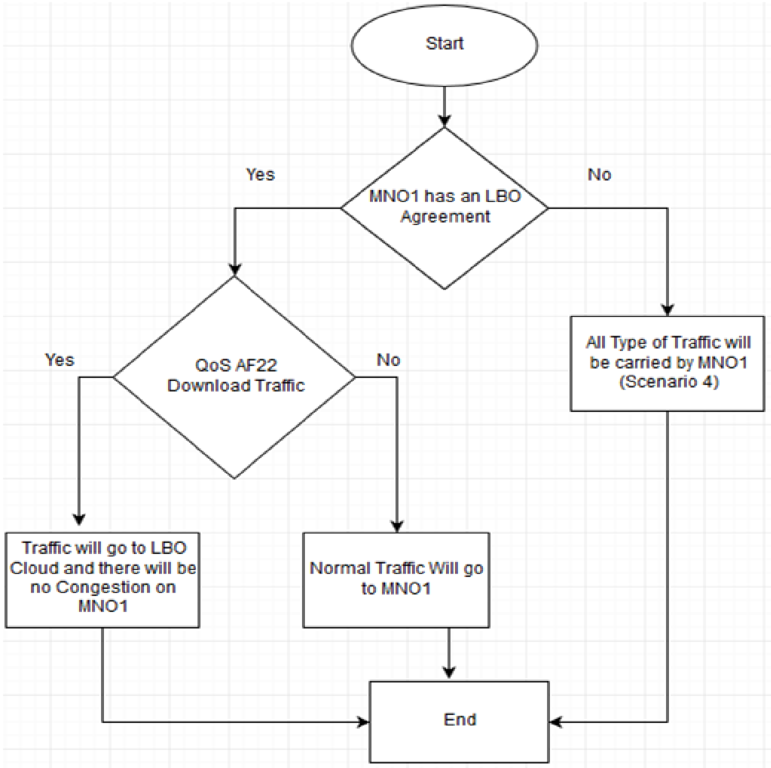


Figure 33. Logic Flowchart of the Third Scenario

Figure 34 displays a traffic capture that was done Via Wireshark on link e0/1 on the device named RB in the LBO cloud, showing the download traffic being carried over LBO. On the other hand, Figure 30 shows that no download traffic is carried over the LTE Core Network of MNO (traffic capture Via Wireshark on link e0/0 on device Gateway-With-LBO in the VNE with LBO agreement cloud. This small traffic (as indicated by the Packet/s Scale) is not real traffic and is called a keep alive traffic in order to let the protocols communicate between each other.

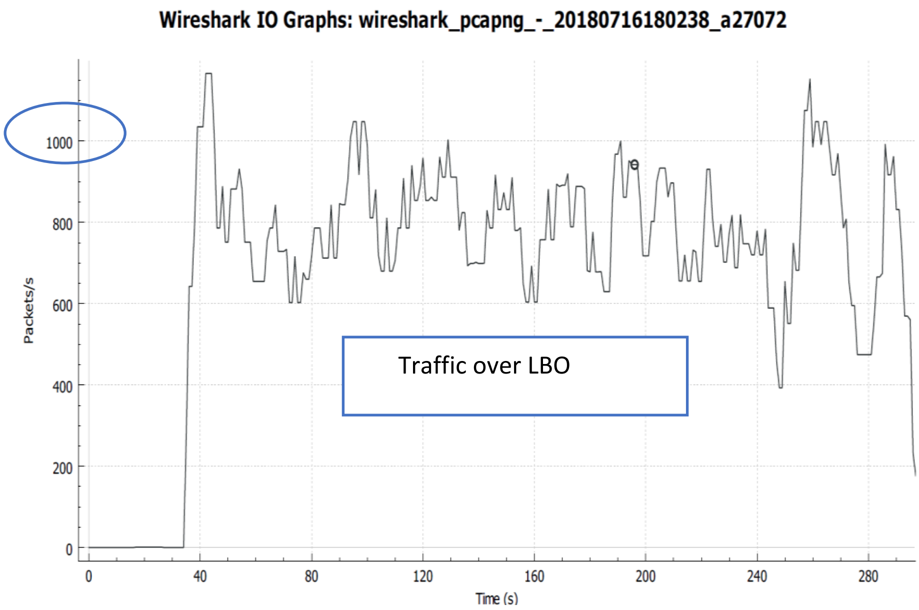


Figure 34. Download Traffic carried over LBO.

Figure 34 shows the Download Traffic for users (which is marked with QoS Value AF22). The variation in the graph indicates the utilisation and the packet exchange of the application itself (FTP) between sender (Internet) and receiver (user).

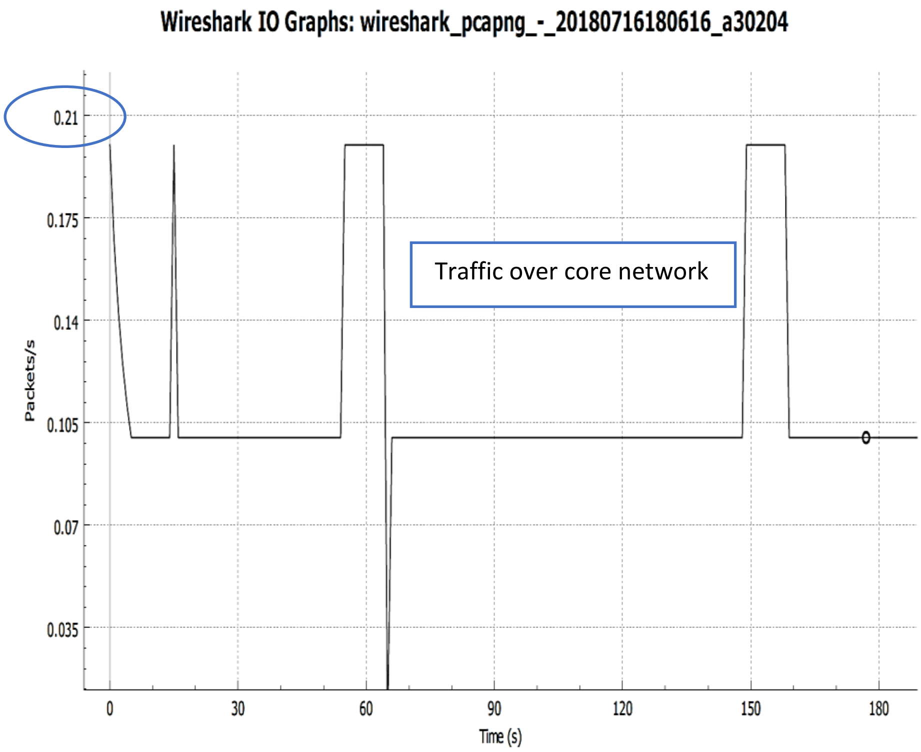


Figure 35. Traffic Carried Over the LTE Core Network of MNO1

The few packets in Figure 35 are known as keep alive packets and are used by the protocols to exchange messages among the network devices and may be referred to as signalling packets.

This scenario confirms the importance of the LBO agreement. The LBO carried a lot of download traffic and protected the LTE Core Network of MNO1 from congestion and carrying any excess traffic.

4.3.4 Fourth Emulation Scenario

Figure 36 represents the logic flowchart of this scenario, in which the MNO1 has no LBO Agreement. This scenario was emulated by shutting the device named RB in The LBO cloud, thus the LTE Core Network of MNO1 will carry all traffic including Downloads (AF22).

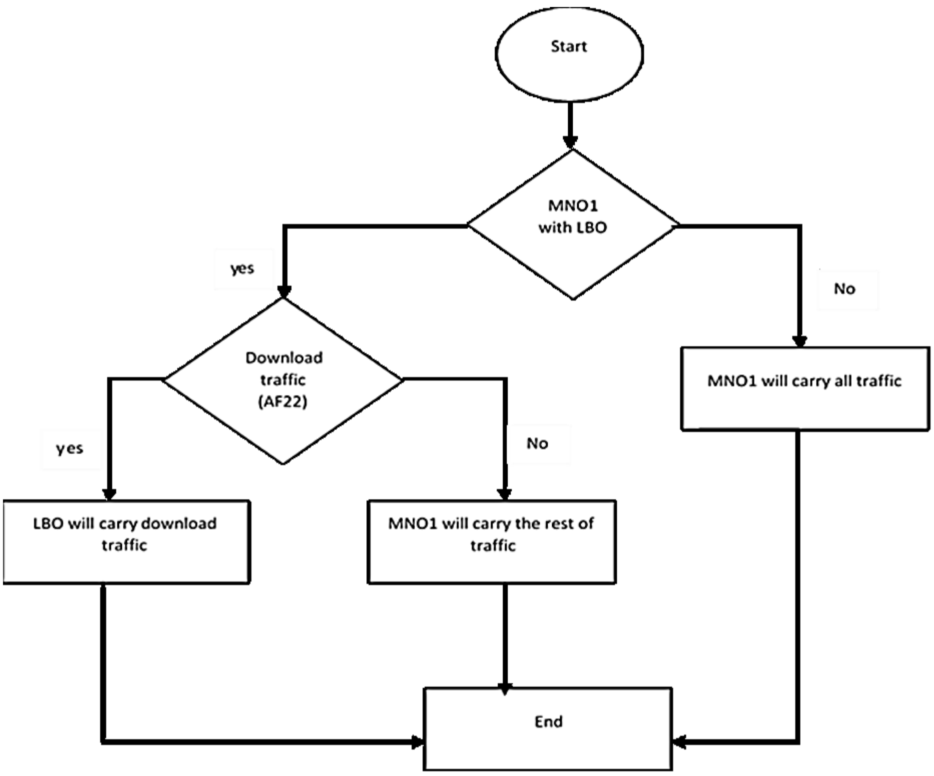


Figure 36. Logic Flowchart of The Fourth Scenario

Download traffic (marked as AF22) was generated during the emulation, and Figure 37 shows that such traffic was carried over the LTE Core Network (capture was made on interface e0/0), while Figure 38 shows that no traffic is carried on the LBO cloud (capture was made on Gateway-with-LBO device on interface e1/1).

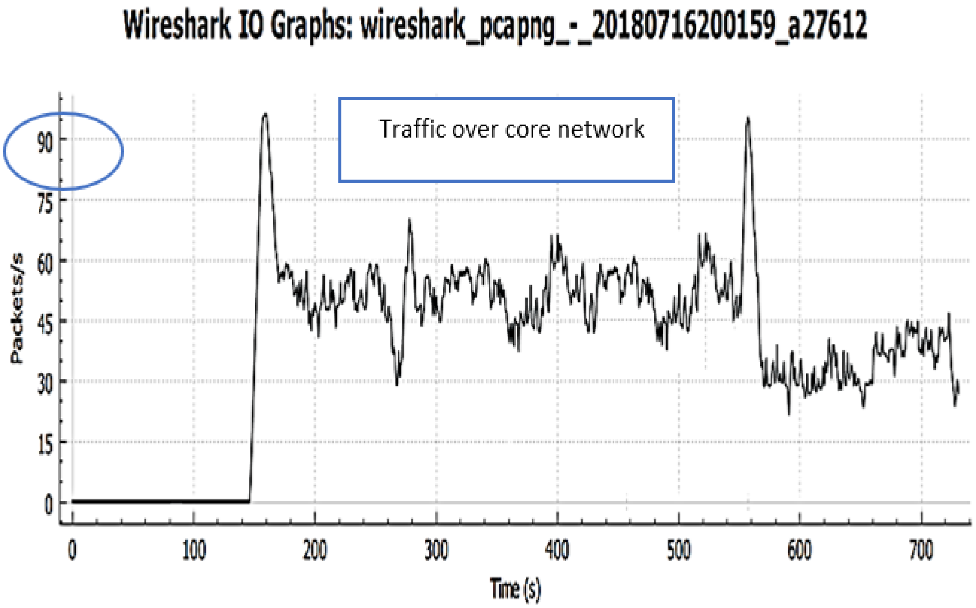


Figure 37. Download Traffic Carried Over the LTE Core Network of MNO1 In the Fourth Scenario

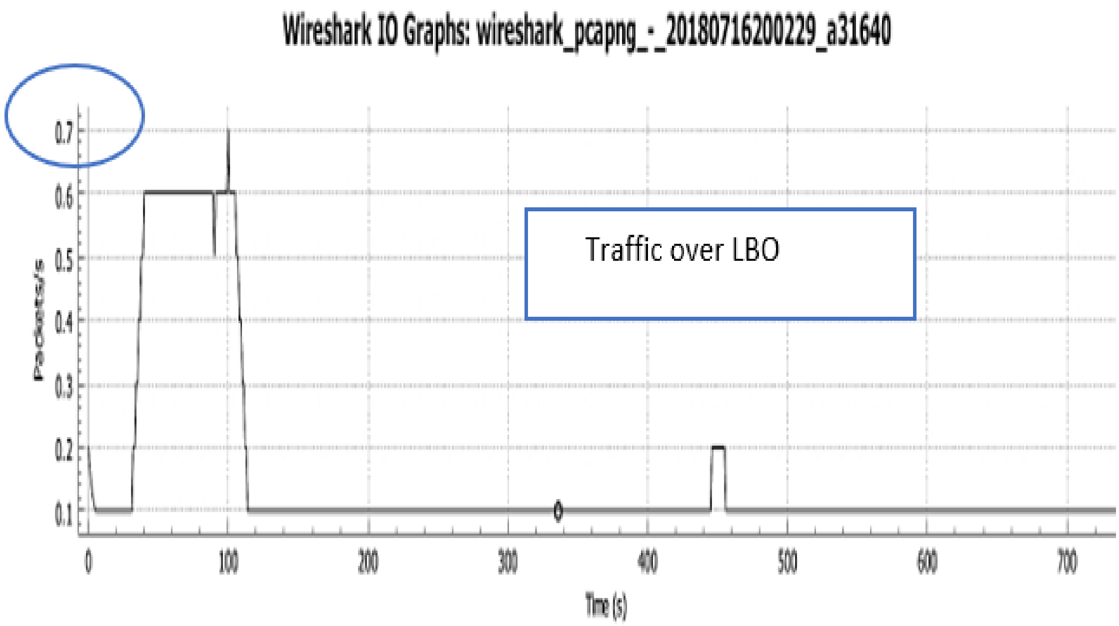


Figure 38. Download Traffic Carried over LBO in the Fourth Scenario

The variation in the graph emulates the use of the packet exchange of the application (FTP) between sender (Internet) and receiver (user), which is explained as signalling packets, similar to the previous scenario.

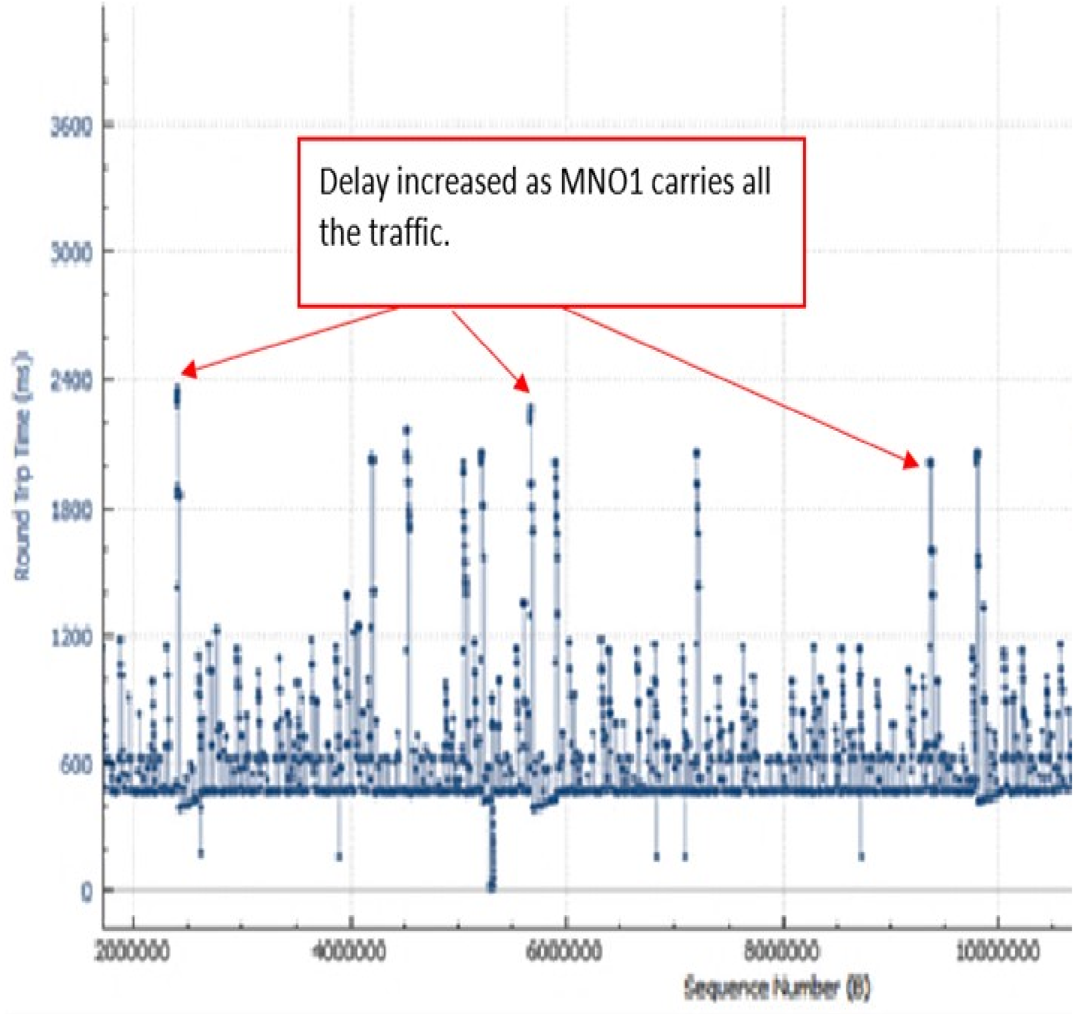


Figure 39. All Traffic is carried over the LTE Core Network of MNO1 in the Fourth Scenario.

Figure 39 shows that the delay increased over the MNO1 core network, as it carries now all traffic, download or normal traffic, as there is no LBO. This scenario clearly confirms the conclusion obtained in the previous scenario about the importance of the LBO agreement. Without it, all download traffic is carried over the core network.

4.3.5 Fifth Emulation Scenario

This emulation scenario considers that MNO2 does not have LBO, but it uses LBO by means of an agreement with MNO1 to carry the download traffic and pass it through the LBO Cloud. Figure 40 illustrates the logic flowchart of this scenario.

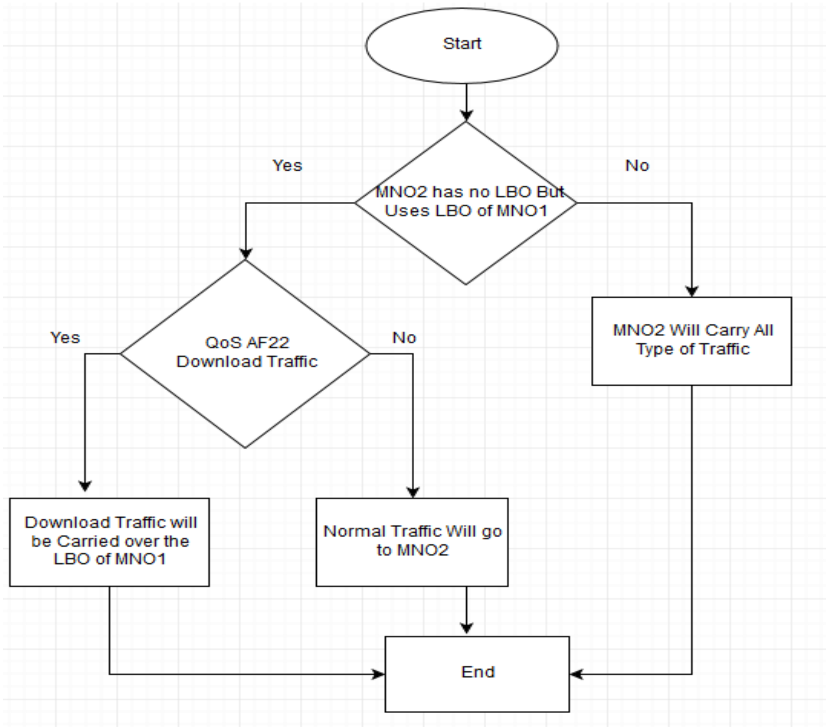
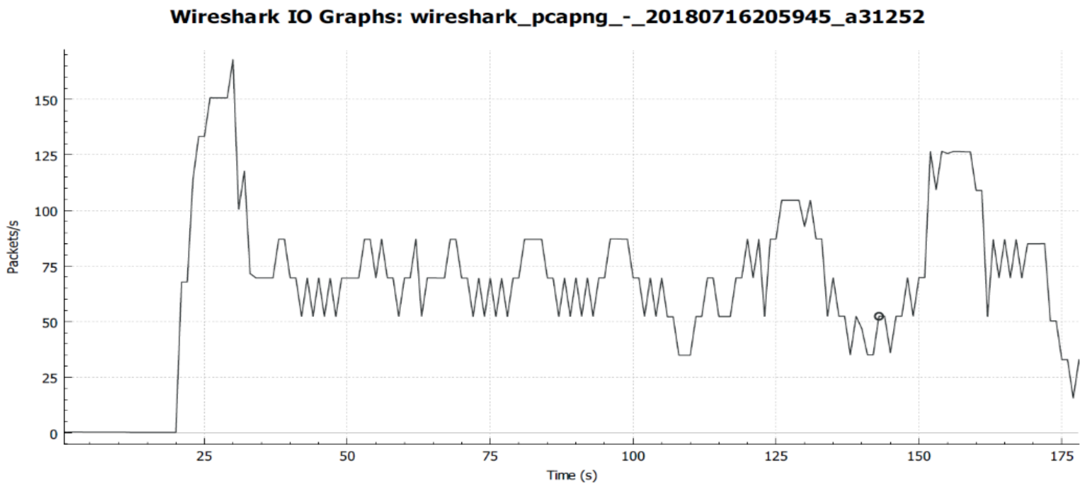


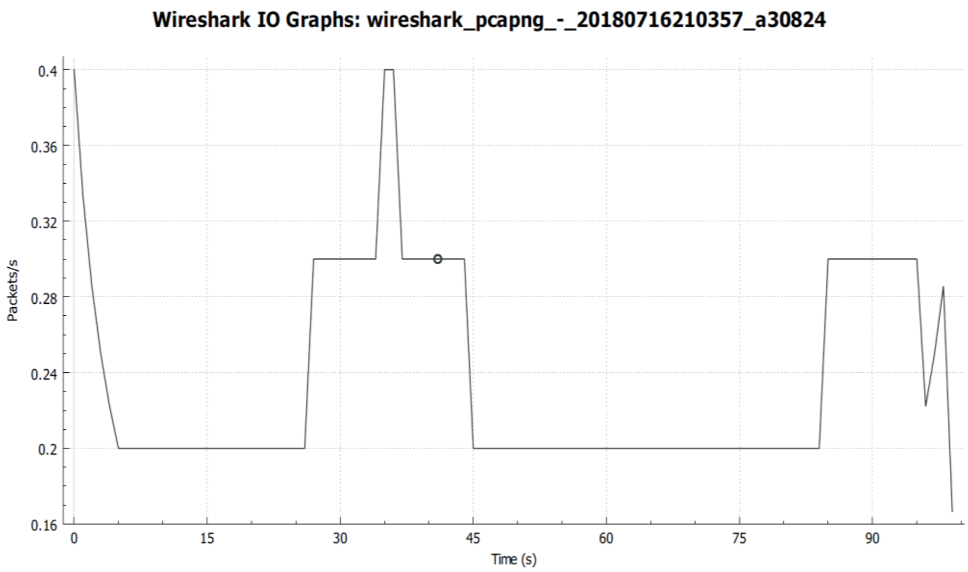
Figure 40. Logic Flowchart of the Fifth Scenario

In this scenario a download traffic (marked as AF22) was generated. The result is represented in Figure 41(a), which shows that this download traffic was carried over device ISR4K48 in the LTE Core Network of MNO2 up to 175 sec, when the traffic was forwarded to the core network of MNO1 and then to the LBO ISR4k53 in the LBO Cloud. Figure 21 illustrates the above-mentioned information.

On the other hand, Figure 41(b) illustrates the subsequent significant decrease in download traffic carried by interface e0/1 in device ISR4K48 in the LTE Core Network of MNO2; note that this small traffic (as indicated by the Packet/s Scale) corresponds to the keep alive traffic. This scenario validates the importance of the LBO, even when it is implemented by means of an agreement with another service provider. Figure 41(b) also shows that there is no more download traffic carried over interface e0/1 on ASR4k48, as all of download traffic will go to interface e1/0 on ASR4k48 and then to the LBO through MNO1 cloud (See Figure 21).



(a)



(b)

Figure 41. Download Traffic Carried Over Device ASR4k48 in Core Network of MNO2 in The Fifth Scenario: (a) Before Forwarding to The LBO Of MNO1, (b) After Forwarding to The LBO Of MNO1

4.4 Evaluation of the proposed framework

In this section the proposed framework is evaluated by comparing with [93]. In particular, advantages and disadvantages of both approaches are stated in terms of the following aspects:

- Type of virtualisation: Compared with the OnDemand virtualisation in this work, full virtualisation in [93] has the drawback of being more complex.
- Domain of virtualisation: Single domain refers to dealing with 1 MNO. When many MVNOs are present, the dedicated spectrum can be owned by one operator and managed by the hypervisor, since the MVNOs do not own additional spectrum (for illustration, see section (4.1)). This research deals with multiple domain, which means that more physical operators are involved and managed by the VNE taking into consideration the relationship between the SPs and InPs, thus the capacity here is larger than the first approach.
- Level of virtualisation: In [93] the air interface level on the eNodeBs includes only virtualisation. On the other hand, in this work the air interface level uses MATLAB and the core network level uses the EVE emulation program. The disadvantage here is that numerical results have been given in the air interface level, rather than providing mathematical results.
- Proposed framework: In [93] the proposed framework consists of two algorithms, one for the implementation of Load Balancing Techniques among the MVNOs and the second manages spectrum

sharing. The framework presented here consists of two algorithms, in addition to the LBO techniques agreements, one for single domain and another for the multi domain, taking into consideration the call admission, LBO and VNE agreements among the MNOs and resource availability as well (see Figure 18). The novelty here is the LBO, since the overhead traffic can be avoided from the core network and act as an alternative path, rather than reject or block the process for resource utilisation.

- Performance: In general both works have shown enhancement regarding sharing and virtualisation from different perspective, but the advantage of the approach proposed here is that it can relatively take more traffic, since it deals with more than one MNO and maintains QoS of the provided service to the end users, by implementing the LBO together with the virtualisation techniques.

Chapter 5

5 Future Research Work and Conclusion

5.1 Future Research.

A variety of services such as, e-banking, e-commerce, social media, video streaming, gamming, among others, are delivered over the Internet every day, being part of what is called the cyber society. In this capacity, the technology is moving toward the IoT (Internet of Things), which refers to utilising the internet by being mobile anywhere, by anything (ICT Device) and at any time, leveraging from internet coverage and usage. Obviously, the Internet developing technologies need to cope with the growth in data utilisation in a cost-effective way, while maintaining QoS.

As mentioned, is this work, mobile networks have evolved from legacy to IP based networks. As a consequence of to this evolution, the demand on the network infrastructure will increase as the technology of IoT keeps developing. In addition, IoT automotive devices and applications have been used in many areas such as health sector, tracking, monitoring and industrial, where it has become a fundamental part in life. It is expected that there will be about 25 billion devices in operation by 2020, all connected to the internet including fixed and mobile devices. The interesting issue is that M2M communication has exceeded human interaction in data generation by 80 percent [77].

In order to be a successful technology and keep growing, the IoT requires a lot of support from wireless technology, specifically, from mobile networks.

Now, it comes down to the MNOs responsibility to continuously develop their network architecture, to handle the large volume of data, while simultaneously fulfilling QoS issues. However, multiple generations of mobile networks still coexist today, because many end-users may not steadily upgrade their devices. Therefore, MNOs have to always support managing multiple network technologies over long-time of period. In this case the revenue growth may not be recovered due to CAPEX and OPEX. The second challenge is the high demand on expanding network capacity. As described earlier in this work, many approaches and methods have been proposed to expand the capacity of mobile networks. Nevertheless, all face challenges such as, high cost, implementation complexity and interoperability issues among different technologies. Moreover, the exponential growth of traffic demand still cannot be handled, even with a continuous infrastructure improvement. Other actions have been executed by many governances, and regulations have further impacted in the profitability of MNOs [77].

As discussed, avoiding excessive traffic and latency in a mobile network is one of the main goals of any MNO. Next generation 5G mobile networks will lead the MNOs toward this goal, by promoting the interaction of various technologies and techniques in its platform, some of which are the following [77]:

- Software-Defined Networks (SDN): by joining IT and Cloud notions into one software platform and implemented in the mobile core networks functionality

- Agnostic network access: to reduce latency and network backhaul, by accessing subscriber information.

- Mobile edge computing: The end-user will benefit from this technique, because all network applications, service applications and subscriber profiles will be available at the edge of the mobile network, which leads to less latency in general.
- 5G network slicing: all end-user IoT devices are serviced by separate and virtualised core networks, which results in less congestion.

5G networks are designed to attain the following objectives [77]:

- Supporting many connected devices: This will result from the deployment and evolution of IoT technology.
- Achieving extreme low latency: 5G networks have to support real-time communications with the IoT devices, e.g. medical equipment due to its critical conditions.
- Efficiently utilising the spectrum: it is necessary to develop new techniques and frameworks to increase the spectrum utilisation.
- Bandwidth and data rate increase: this is considered as one of the high priority issues, since high data rate has always been a major challenge for MNOs.
- Providing ubiquitous connectivity: Multiple radio technologies will coexist, and the future network should support ubiquitous computing among the technologies.

5.2 Conclusion

This work proposed a novel collaborative framework that combines Call Admission Control (CAC), network virtualisation and Local Break Out (LBO), to avoid congestion and guarantee quality of service in mobile networks.

The first step in developing the strategy was building a mathematical model, to determine the average number of resource blocks required for voice and data calls, and the blocking probability of both types of calls. The average numbers of physical resource blocks (PRB) required for voice and data calls were computed using an LTE simulator. Then, these average numbers are used to allocate the available PRBs, in order to satisfy the requirements both in the normal and virtualisation scenarios. In addition, a Poisson Distribution Simulator (PDS) is used to compute the voice and data calls blocking probabilities in the virtualisation scenario.

Then, a first simulation stage was carried out in which MATLAB was used to run a numerical example for verifying the mathematical model and the proposed framework. The simulations were conducted with the Simple and Distributed CAC algorithms. With the simple CAC, both voice and data call blocking probabilities increased compared to the distributed CAC. This was expected because the resources in the simple CAC are not available in a combined manner. In addition, a difference of about one order of magnitude existed between the data call blocking probability in the normal and the virtualised case with distributed CAC. A similar difference was obtained between the voice call blocking probabilities in both cases. Regarding bandwidth utilisation, it was 10% greater for the virtualised case. Based on these results which were according to the expectations, it can be said that the mathematical model implemented in MATLAB was verified satisfactorily. This model constitutes the support of the applications that were used to emulate real traffic situations in LTE core networks in different scenarios.

Afterward, the proposed collaborative framework was developed to offload the MNO network by means of virtualisation and an LBO Strategy. In particular, the Virtual Network Enabler worked in an automated way to switch excess traffic from one MNO to the other, if the former was congested. This switching was performed based on predefined thresholds for congestion and delay & vice versa between both MNOs. The OnDemand sharing among operators via virtualisation helps them take advantage of offloading mechanisms available with one (or few) operators. In addition, the LBO partially directs the traffic to other types of networks when required.

The proposed strategy was evaluated in an emulation stage, in which commercial applications such as Emulated Virtual Environment (EVE) and Wireshark, were used for emulating the traffic in the network under five different scenarios. The main purpose was to emulate real traffic in each scenario, to validate the performance of the proposed offloading and LBO strategies.

The first two scenarios corresponded to cases without and with Virtual Network Enabler agreement between the MNOs. Results show that the performance was significantly better with the Virtual Network Enabler agreement between MNO1 and MNO2 (second scenario). The third considered the case where the MNO1 has an LBO Agreement, while the fourth scenario was the opposite. The results confirm the importance of the LBO since it carried a lot of download traffic and protected the LTE Core Network of MNO1 from congestion and carrying excess traffic. At last, the fifth emulation scenario considered that the MNO2 does not have LBO, but it uses LBO by means of an agreement with MNO1 to carry the download traffic and pass it through the LBO cloud. This scenario corroborated the importance of the LBO, even if it is implemented by means of an agreement with another service provider. Hence, as expected, the performance is superior with the Virtual Network Enabler. In addition, the importance of the LBO strategy was confirmed, even if it is implemented through a different operator. In general, the proposed strategy resulted in a more efficient use of the spectrum, and reduced probability of failure of the system. The advantages of the proposed framework can be summarized as dealing with a heterogeneous environment and offloading of data traffic.

Future research work should be focused on developing optimal/robust algorithms for fulfilling the significant increase in the demand, expected from upcoming technologies such as the fifth generation (5G) of mobile networks. The former plans to incorporate vehicles and appliances to the Internet, thus increasing the demand, while the latter is promising milestones such as maintaining a speed of 1 Gb/sec for the greatest possible number of users.

Generally, 5G and beyond will support a set of various technologies and communication types as a consensus. Therefore, inovating strategies are urgently required to cope with this plethora. In addition, air-interfaces have to support heterogeneous and diverse applications which are imposing a set of very challenging requirements, such as hundreds of Mbs, satisfactory end-user throughput, low latency and high reliability.

References

[1] Ericsson. (2016). *Ericsson Mobility Report: On the pulse of the networked society.* Retrieved from: https://www.ericsson.com/en/ericsson-technology-review/archive/2016.

[2] Ericsson (2015). *Ericsson Mobility Report: On the pulse of the networked society.* Retrieved from: https://www.ericsson.com/en/ericsson-technology-review/archive/2015.

[3] Maallawi, R., Agoulmine, N., Radier, B., & Meriem, T. B. (2015). A comprehensive survey on offload techniques and management in wireless access and core networks", *IEEE Communications Surveys & Tutorials,* 17(3), 1582-1604.

[4] Wibowo, F. X. A., Bangun, A. A. P., Kurniawan, A., & Hendrawan, H. (2011). Multimedia broadcast multicast service over single frequency network (MBSFN) in LTE based femtocell. In *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics (ICEEI): 17-19 July 2011*, Bandung, Indonesia (pp. 1-5). doi: 10.1109/ICEEI.2011.6021623.

[5] Awobuluyi, O., Nightingale, J., Wang, Q., & Alcaraz-Calero, J. M. (2015). Video quality in 5G networks: context aware QoE management in the SDN control plane. In *Proceedings of the 2015 IEEE International Conference on Computer and Information Technology, Ubiquitous Computing and Communications, Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*: *26-28 October 2015*, Liverpool, UK (pp. 1657-1662). doi: 10.1109/CIT.2015.7378753.

[6] Ericsson. (2015). *Ericsson Technology Review*. Retrieved from: https://www.ericsson.com/en/ericsson-technology-review.

[7] Albinali, S., & Alani, O. (2014). *Collaborative mobile network operators framework for efficient spectrum utilization in LTE-A*. Paper presented in *The 15th Annual PostGraduate Symposium on the Convergence of Telecommunication, Networking and Broadcasting:* Liverpool, United Kingdom.

[8] Hossain, E., & Hasan, M. (2015). 5G mobile: Key enabling technologies and research challenges. *IEEE Instrumentation & Measurement Magazine,* 18(3), 11-21. doi: 10.1109/MIM.2015.7108393.

[9] Bizanis, N., & Kuipers, F. A. (2016). SDN and virtualisation solutions for the Internet of Things: A survey. *IEEE Access,* 4, 5591-5606. doi: 10.1109/ACCESS.2016.2607786.

[10] Liang, C., & Yu, F. R. (2015). Wireless network virtualisation: A survey, some research issues and challenges. *IEEE Communications Surveys & Tutorials,* 17(1), 358-380. doi: 10.1109/COMST.2014.2352118.

[11] Ericsson. (2017). *Ericsson Mobility Report: On the pulse of the networked society.* Retrieved from: https://www.ericsson.com/assets/local/mobility-report/documents/2017/emr-interim-february-2017.pdf.

[12] Ghosh, S., & Konar, A. (2013). An Overview of Call Admission Control in Mobile Cellular Networks. In *Call Admission Control in Mobile Cellular Networks*. *Studies in Computational Intelligence*, vol. 437 (pp. 1-62). Berlin Heidelberg: Springer.

[13] Barnes, S. D., & Maharaj, B. T. (2013). A comparison of spectrum occupancy in the South African 900 MHz GSM mobile bands. In *Proceedings of the 2013 AFRICON Conference: 9-12 September 2013,* Pointe-Aux-Piments, Mauritius (pp. 1-5). doi: 10.1109/AFRCON.2013.6757597.

[14] Dezhang, C., Jingjing, Y., Jida, W., Hao, T., & Ming, H. (2012). Spectrum occupancy analysis based on radio monitoring network. In *Proceedings of the 2012 1st IEEE International Conference on Communications in China (ICCC): 15-17 August 2012,* Beijing, China (pp. 739-744). doi: 10.1109/ICCChina.2012.6356981.

[15] Peng, L., Juncheng, J., Qian, Z., & Hamdi, M. (2010). Cooperation among wireless service providers: opportunity, challenge, and solution [Dynamic Spectrum Management]. *IEEE Wireless Communications,* 17(4), 55-61. doi: 10.1109/MWC.2010.5547922.

[16] Zaki, Y. (2012). *Future Mobile Communications: LTE optimization and mobile network virtualisation vol. 1*. Berlin: Springer Vieweg.

[17] Bhat, P., Nagata, S., Campoy, L., Berberana, I., Derham, T., Guangyi, L., …, Yang, J. (2012) LTE-advanced: an operator perspective. *IEEE Communications Magazine,* 50(2), 104-114. doi: 10.1109/MCOM.2012.6146489.

[18] Kokku, R., Mahindra, R., Zhang, H., & Rangarajan, S. (2013). CellSlice: Mobile wireless resource slicing for active RAN sharing. In Proceedings of the *2013 Fifth International Conference on Communication Systems and Networks (COMSNETS): 7-10 January 2013,* Bangalore, India (pp. 1-10). doi: 10.1109/COMSNETS.2013.6465548.

[19] Kalil, M., Shami, A., & Ye, Y. (2014). Wireless resources virtualization in LTE systems. In *Proceedings of the 2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS): 27 April-May 2, 2014*, Toronto, Canada (pp. 363-368). doi: 10.1109/INFCOMW.2014.6849259.

[20] Wang, X., Krishnamurthy, P., & Tipper, D. (2015). A collaborative spectrum-sharing framework for LTE virtualisation. In *Proceedings of the 2015 IEEE Conference on Collaboration and Internet Computing (CIC)*: *27-30 October 2015*, Hangzhou, China (pp. 260-269). doi: 10.1109/CIC.2015.41.

[21] Liang, C., & Yu, F. R. (2015). Mobile virtual network admission control and resource allocation for wireless network virtualisation: A robust optimization approach. In *Proceedings of the 2015 IEEE Global Communications Conference (GLOBECOM)*: *6-10 December 2015*, San Diego, USA (pp. 1-6). doi: 10.1109/GLOCOM.2015.7417760.

[22] Chochlidakis, G. & Friderikos, V. (2015). Robust virtual network embedding for mobile networks. In *Proceedings of the 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*: *30* August-2 September 2015, Hong Kong, China (pp. 1867-1871). doi: 10.1109/PIMRC.2015.7343603.

[23] Gao, L., Li, P., Pan, Z., Liu, N., & You, X. (2016). Virtualisation framework and VCG based resource block allocation scheme for LTE virtualisation. In *Proceedings of the 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring): 15-18 May 2016,* Nanjing, China (pp. 1-6). doi: 10.1109/VTCSpring.2016.7504529.

[24] Zhang, Y., Zhao, L., Lopez-Perez, D., & Chen, K. C. (2016). Energy-efficient virtual resource allocation in OFDMA systems. In *Proceedings of the 2016 IEEE Global Communications Conference (GLOBECOM): 4-8 December 2016*, Washington, USA (pp. 1-6). doi: 10.1109/GLOCOM.2016.7841810.

[25] Fu, F. & Kozat, U. C. (2013). Stochastic game for wireless network virtualisation. *IEEE/ACM Transactions on Networking,* 21(1), 84-97. doi: 10.1109/TNET.2012.2190419.

[26] Liu, X., Li, M., Song , M., Tian, Z., & Jianxun, F. (2016). Wireless virtual network embedding based on spectrum sharing allocation. In *Proceedings of the 2016 11th International Conference on Computer Science & Education (ICCSE)*: *23-25 August 2016*, Nagoya, Japan (pp. 670-675). doi: 10.1109/ICCSE.2016.7581660.

[27] Yang, M., Li, Y., Jin, D., Yuan, J., Su, L., & Zeng, L. (2013). Opportunistic spectrum sharing based resource allocation for wireles virtualisation. In *Proceedings of the 2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*: *3-5 July 2013*, Taichung, Taiwan (pp. 51-58). doi: 10.1109/IMIS.2013.97.

[28] Liu, B. & Tian, H. (2013). A bankruptcy game-based resource allocation approach among virtual mobile operators. *IEEE Communications Letters,* 17(7), 1420-1423. doi: 10.1109/LCOMM.2013.052013.130959.

[29] Zhu, K. & Hossain, E. (2016). Virtualisation of 5G mobile networks as a hierarchical combinatorial auction. *IEEE Transactions on Mobile Computing*, 15(10), 2640-2654. doi: 10.1109/TMC.2015.2506578.

[30] Samdanis, K., Taleb, T., & Schmid, S. (2012). Traffic offload enhancements for eUTRAN. IEEE Communications Surveys & Tutorials, 14(3), 884-896. doi: 10.1109/SURV.2011.072711.00168.

[31] Falowo, O. E. (2010). Joint call admission control algorithm for reducing call blocking/dropping probability in heterogeneous wireless networks supporting multihoming. In *Proceedings of the 2010 IEEE GLOBECOM Workshops (GC Wkshps): 6-10 December 2010*, Miami, USA (pp. 611-615).

[32] Chau, T. C., Wong, K. Y. M., & Bo, L. (2006). Optimal call admission control with QoS guarantee in a voice/data integrated mobile network. *IEEE Transactions on Wireless Communications,* 5(5), 1133-1141. doi: 10.1109/TWC.2006.1633366.

[33] Benslimane, A., Assi, C., Hossain, E., & Vuran, M.C. (2009). Special issue of "Computer Communications" on Cognitive Radio and Dynamic Spectrum Sharing Systems. *Computer Communications*, 32, 1903-1904. doi: 10.1016/j.comcom.2009.10.004.

[34] Junfeng, X., Hu, R. Q., Yi, Q., Lei, G., & Bo, W. (2013). Expanding LTE network spectrum with cognitive radios: From concept to implementation. *IEEE Wireless Communications*, 20(2), 12-19. doi: 10.1109/MWC.2013.6507389.

[35] Montazerolghaem, A., Yaghmaee, M. H., Leon-Garcia, A., Naghibzadeh, M., & Tashtarian, F. (2016). A Load-Balanced Call Admission Controller for IMS Cloud Computing. *IEEE Transactions on Network and Service Management, 13*(4), 806-822. doi: 10.1109/TNSM.2016.2572161

[36] Aglianò, S., Ashjaei, M., Behnam, M., & Bello, L. L. (2018, 9-10 May 2018). *Resource management and control in virtualized SDN networks.* Paper presented at the 2018 Real-Time and Embedded Systems and Technologies (RTEST).

[37] Ramantas, K., Antonopoulos, A., Kartsakli, E., Mekikis, P., Vardakas, J., & Verikoukis, C. (2018, 1-5 July 2018). *A C-RAN Based 5G Platform with a Fully Virtualized, SDN Controlled Optical/Wireless Fronthaul.* Paper presented at the 2018 20th International Conference on Transparent Optical Networks (ICTON).

[38] Ghaderi, M. & Boutaba, R. (2006). Call admission control in mobile networks: a comprehensive survey. *Wireless communications and mobile computing*, 6(1), 69-93. doi: 10.1002/wcm.246.

[39] ElNashar, A., El-saidny, M., & Sherif. M. (2014). Design, deployment and performance of 4G-LTE networks: A practical approach. Chichester: Wiley.

[40] Dahlman, E. (2008). *3G evolution: HSPA and LTE for mobile broadband*: Academic Press.

[41] Tzeng, S. S. (2009). Call admission control policies in mobile wireless networks with spectrum renting. *Computer Communications,* 32(18), 1905–1913. doi: 10.1016/j.comcom.2009.07.017.

[42] 3GPP. (2011). *Feasibility Study for Evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN)*. (25.912 Release 7) Retrieved from: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=1341

[43] Tzeng, S. S. & Hsu, Y. C. (2010). Channel reservation in mobile wireless networks with spectrum leasing. In *Proceedings of 19th International Conference on Computer Communications and Networks (ICCCN): 2-5 August 2010*, Zurich, Switzerland (pp. 1-7). doi: 10.1109/ICCCN.2010.5560096.

[44] Elnashar, A., El-saidny, M., Elnashar, A., El-saidny, M., & Sherif, M. (2014). LTE network architecture and protocols. In *Design, Deployment and Performance of 4G-LTE Networks* (pp 1-45).New Jersey: Jhon Wiley & Sons, Ltd.

[45] Dahlman, E., Parkvall, S., & Skold, J. (2011). *4G: LTE/LTE-advanced for mobile broadband*. Amsterdam: Academic Press.

[46] Triggs, R. (2015). LTE Advanced Pro and the road to 5G explained. Retrieved from:https://www.androidauthority.com/lte-advanced-pro-and-the-road-to-5g-explained-684095/.

[47] Keysight. (2016). *LTE Physical Layer Overview*. Retrieved from http://rfmw.em.keysight.com/wireless/helpfiles/89600b/webhelp/subsystems/lte/content/lte_overview.htmxx

[48] 3GPP. (2017). *Quality of Service (QoS) concept and architecture* (TS 23.107. Release2017). Retrieved from: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=783

[49] Keysight. (2016). *LTE Physical Layer Overview*. Retrieved from http://rfmw.em.keysight.com/wireless/helpfiles/89600b/webhelp/subsystems/lte/content/lte_overview.htmxx

[50] Flore, D. (2017) LTE evolution and 5G. *Qualcomm Technologies Inc. 3GPP.* Retrieved from: http://www.3gpp.org/ftp/Information/presentations/Presentations_2017/2017_02_flore_LTE%20evolution%20and%205G.pdf.xx

[51] International Telecommunication Union. (2016). New technology trends in wireless. Retrieved from: https://www.itu.int/en/ITU-D/.../2016/.../01-PTA-Keynote.pdf.xx

[52] Zaki, Y., Zhao, L., Goerg, C., & Timm-Giel, A. (2011). LTE mobile network virtualization. *Mobile Networks and Applications*, 16(4), 424-432. doi: 10.1007/s11036-011-0321-7.

[53] Brunner, M., Abramowicz, H., Niebert, N., & Correia, L. (2010). 4WARD: a European perspective towards the future internet. *IEICE transactions on communications,* 93-B (3), 442-445. doi: 10.1587/transcom. E93.B.442.

[54] N. LambdaRail. (2008). *VINI: Virtual Network Infrastructure*. Retrieved from: http://www.startap.net/translight/pages/applications/2008/VINI.html.

[55] Matsubara, D., Egawa, T., Nishinaga, N., Kafle, V., Shin, M., & Galis, L. (2013). Toward future networks: A viewpoint from ITU-T. *IEEE Communications Magazine*, 51(3), 112-118. doi: 10.1109/MCOM.2013.6476874.

[56] Bataa, O., Young-il, K., Dorj, E., Oyunbileg, B., Gonchigsumlaa, K., Chuluu, O., & Orosoo T. (2012). Novel design of eMBMS based on femtocell. In *Proceedings of the 2012 14th International Conference on Advanced Communication Technology (ICACT): 19-22 February 2012,* PyeongChang, South Korea (pp. 377-382).

[57] Paul, U., Subramanian, A. P., Buddhikot, M. M., & Das, S. R. (2011). Understanding traffic dynamics in mobile data networks. In *Proceedings of the 2011 IEEE INFOCOM: 10-15 April 2011*. Shanghai, China (pp. 882-890). doi: 10.1109/INFCOM.2011.5935313.

[58] Namba, S., Warabino, T., & Kaneko, S. (2012). BBU-RRH switching schemes for centralized RAN. In *Proceedings of the 2012 7th International ICST Conference on Communications and Networking in China (CHINACOM): 8-10 August 2012*, Kun Ming, China (pp. 762-766). doi: 10.1109/ChinaCom.2012.6417586.

[59] Srinivasan, D., Dey, J., Kumar, S., & Mukherjee, R. (2012). Data offload approaches for mobile operators improving network efficiency and strengthening Quality of Service. Retrieved from: http://www.wipro.com/documents/resource-center/Data_Offload_Approaches_for_Mobile_Operators.pdf.

[60] Wu, G., Li, Q., Hu, R. Q., & Qian, Y. (2013). Overview of heterogeneous networks. In *Heterogeneous Mobile Networks* (pp. 1-25). Chichester: John Wiley & Sons Ltd.

[61] Zhu, K. & Hossain, E. (2016). Virtualisation of 5G mobile networks as a hierarchical combinatorial auction. *IEEE Transactions on Mobile Computing*, 15(10), 2640-2654. doi: 10.1109/TMC.2015.2506578

[62] Samdanis, K., Taleb, T., & Schmid, S. (2012). Traffic offload enhancements for eUTRAN. *IEEE Communications Surveys & Tutorials,* 14(3), 884-896. doi: 10.1109/SURV.2011.072711.00168.

[63] Maallawi, R., Agoulmine, N., Radier, B., & Meriem, T. B. (2015). A Comprehensive Survey on Offload Techniques and Management in Wireless Access and Core Networks. *IEEE Communications Surveys & Tutorials, 17*(3), 1582-1604. doi: 10.1109/COMST.2014.2373356

[64] Niyato, D. & Hossain, E. (2005). Call admission control for QoS provisioning in 4G wireless networks: issues and approaches. *IEEE Network*, 19(5), 5-11. doi: 10.1109/MNET.2005.1509946.

[65] Ahmed, M. H. (2005). Call admission control in wireless networks: a comprehensive survey. *IEEE Communications Surveys & Tutorials*, 7(1), 49-68. doi: 10.1109/COMST.2005.1423334.

[66] Wang J., & Fu, X. (2012). Challenges of CAC in heterogeneous wireless cognitive networks. *Physics Procedia,* 25, 2218-2224.

[67] Ye, Y., Vora, N., Sairanen, J., & Sahasrabudhe, M. (2012). Enabling local breakout from eNB in LTE networks. In *Proceedings of the 2012 IEEE International Conference on Communications (ICC): 10-15 June* 2012, Ottawa, Canada (pp. 6982-6986). doi: 10.1109/ICC.2012.6364840.

[68] Maallawi, R., Agoulmine, N., Radier, B., & Meriem, T. B. (2015). A Comprehensive Survey on Offload Techniques and Management in Wireless Access and Core Networks. *IEEE Communications Surveys & Tutorials, 17*(3), 1582-1604. doi: 10.1109/COMST.2014.2373356

[69] Cox, C. (2012). *An Introduction to LTE: LTE, LTE-advanced, SAE and 4G mobile communications*. Chichester: Wiley.

[70] Dahlman, E., Parkvall, S., & Skold, J. (2011). *4G: LTE/LTE-advanced for mobile broadband*. Amsterdam: Academic Press.

[71] Da Silva, M. R. M., Correia, A. M., Dinis, R., Souto, N., & Silva, J. C. (2012). *Transmission techniques for 4G systems*. Boca Raton: CRC Press Llc.

[72] Falowo, O. E. (2010). Joint call admission control algorithm for reducing call blocking/dropping probability in heterogeneous wireless networks supporting multihoming. In *Proceedings of the 2010 IEEE GLOBECOM Workshops (GC Wkshps): 6-10 December 2010*, Miami, USA (pp. 611-615).

[73] Chau, T. C., Wong, K. Y. M., & Bo, L. (2006). Optimal call admission control with QoS guarantee in a voice/data integrated mobile network. *IEEE Transactions on Wireless Communications,* 5(5), 1133-1141. doi: 10.1109/TWC.2006.1633366.

[74] Benslimane, A., Assi, C., Hossain, E., & Vuran, M.C. (2009). Special issue of "Computer Communications" on Cognitive Radio and Dynamic Spectrum Sharing Systems. *Computer Communications*, 32, 1903-1904. doi: 10.1016/j.comcom.2009.10.004.

[75] Junfeng, X., Hu, R. Q., Yi, Q., Lei, G., & Bo, W. (2013). Expanding LTE network spectrum with cognitive radios: From concept to implementation. *IEEE Wireless Communications*, 20(2), 12-19. doi: 10.1109/MWC.2013.6507389.

[76] Wibowo, F., Bangun, P., Kurniawan, A., & Hendrawan, A. (2011). Transmission scheme mechanism in femtocell based eMBMS: A review. In Proceedings of the 2011 6th International Conference on Telecommunication Systems, Services, and Applications (TSSA): 20-21 October 2011, Bali, Indonesia (pp. 154-158), DOI: 10.1109/TSSA.2011.6095425.

[77] Zhang, Y. (2018). Network Function Virtualization: Concepts and Applicability in 5G Networks. USA: Wiley-IEEE Press.

[78] Hyytiä, E., & Virtamo, J. (2007). Random waypoint mobility model in mobile networks. Wireless Networks, 13(2), 177–188. doi: 10.1007/s11276-006-4600-3.

[79] Iversen, V. (2001). Teletraffic Engineering Handbook, ITU-D SG 2/16 & ITC, Draft 2001-06-20. Retrieved from: https://www.itu.int/dms_pub/itu-d/opb/stg/D-STG-SG02.16.1-2001-PDF-E.pdf.

[80] Roopesh Kumar Polaganga. "Self-Similarity And Modeling Of Lte / Lte-A Data Traffic". Master Thesis. University of Texas in Arlington. TX,USA.2015.

[81] Dmitri A. Moltchanov. "Basics of traffic modeling II". Available in: http://www.cs.tut./kurssit/ELT-53606/

[82]] Pablo Ameigeiras, Yuanye Wang, Jorge Navarro-Ortiz, Preben E Mogensen and Juan M Lopez-Soler "Traffic models impact on OFDMA scheduling design". Ameigeiras et al. EURASIP Journal on Wireless Communications and Networking 2012, 2012:61. http://jwcn.eurasipjournals.com/content/2012/1/61*

[83] Li, Y., Qian, M., Jin, M., Hui, P., Wang, Z., & Chen, S. (2014). Multiple mobile data offloading through disruption tolerant networks. IEEE Transactions on mobile computing, 13(7), 1579-1596. doi: 10.1109/TMC.2013.61.

[84] https://www.eve-ng.net/

[85] https://www.wireshark.org/

[86] https://www.cisco.com/c/en/us/products/collateral/routers/4000-series-integrated-services-routers-isr/data_sheet-c78-732542.html

[87] https://www.cisco.com/c/en/us/products/wireless/asr-5000-series/index.html

[88] https://www.cisco.com/c/dam/global/en_ae/assets/expo2011/saudiarabia/pdfs/lte-design-and-deployment-strategies-zeljko-savic.pdf

[89] https://www.cisco.com/c/en/us/td/docs/wireless/asr_5000/20/SecGW/b_20-SecGW-Admin/b_20-SecGW-Admin_chapter_01.pdf

[90] http://www.pillarcommunications.com/sw/swchannel/images/ProductCatalog/ProductPage/File/data_sheet_c78-6062236.pdf

[91] https://newsroom.cisco.com/press-release content?type= webcontent&articleId =1878822

[92] https://accedian.com/enterprises/blog/network-packet-loss-retransmissions-and-duplicate-acknowledgements/

[93] M. Li, L. Zhao, X. Li, X. Li, Y. Zaki, A. Timm-Giel, et al., "Investigation of Network Virtualization and Load Balancing Techniques in LTE Networks," in 2012 IEEE 75th Vehicular Technology Conference (VTC Spring), 2012, pp. 1-5.

APPENDIX A: MATLAB Code for LTE Simulator

The MATLAB code for LTE Simulator is shown below in parts:

A. RWP model (Main)

```
% Get output from RWP mobility model
for ii = 1 : N
   test_Execute;
   s_mobility_Opr(ii) = s_mobility;
end
```

B. Hypervisor time points interpolation code:

```
% Hypervisor loop
for ii = 1 : N
   for jj = 1 : UE
      mm = 1;
      for kk = 2 : length(s_mobility_Opr(ii).VS_NODE(jj).V_TIME)
         while (mm < s_mobility_Opr(ii).VS_NODE(jj).V_TIME(kk))

            % Time index
            s_mobility_Opr_Intp(ii).VS_NODE(jj).V_TIME(mm) = mm;

            % For Position
            t_1 = s_mobility_Opr(ii).VS_NODE(jj).V_POSITION_X(kk-1);
            t_2 = s_mobility_Opr(ii).VS_NODE(jj).V_POSITION_X(kk);
            t_3 = s_mobility_Opr(ii).VS_NODE(jj).V_POSITION_Y(kk-1);
            t_4 = s_mobility_Opr(ii).VS_NODE(jj).V_POSITION_Y(kk);
            t_5 = s_mobility_Opr(ii).VS_NODE(jj).V_TIME(kk-1);
            t_6 = s_mobility_Opr(ii).VS_NODE(jj).V_TIME(kk);

            s_mobility_Opr_Intp(ii).VS_NODE(jj).V_POSITION_X(mm) = ...
               ((mm-t_5)*t_2 + (t_6-mm)*t_1)/(t_6-t_5);

            s_mobility_Opr_Intp(ii).VS_NODE(jj).V_POSITION_Y(mm) = ...
               ((mm-t_5)*t_4 + (t_6-mm)*t_3)/(t_6-t_5);

            s_mobility_Opr_Intp(ii).VS_NODE(jj).V_SPEED_X(mm) = ...
               s_mobility_Opr(ii).VS_NODE(jj).V_SPEED_X(kk-1);

            s_mobility_Opr_Intp(ii).VS_NODE(jj).V_SPEED_Y(mm) = ...
```

```
                s_mobility_Opr(ii).VS_NODE(jj).V_SPEED_Y(kk-1);

            mm = mm + HypRes;
          end
        end
      end
    end
```

C. SNR calculation code for all operators:

```
Ptx = 0; % dBm per PRB
PRB_BW = 180e-3; % in MHz
SNR = zeros(N, UE, simTime-1);
CQI = zeros(N, UE, simTime-1);
PRB = zeros(N, UE, simTime-1);
est_BW = zeros(N, simTime-1);

% Channel
Chan = Rayleighchan(TS, FD, TAU, PDB);
Chan.ResetBeforeFiltering = 1;

for ii = 1 : N

  % Initialization of Log normally distributed RV
  S_prev = ld_sd * randn(UE, 1);

  for kk = 1 : simTime-1

    % Scheduling matrix
    sch_mtx = zeros(V_fr, NRB(ii));

    for jj = 1 : UE

      % Distance from eNB
      R_km = ...
(s_mobility_Opr_Intp(ii).VS_NODE(jj).V_POSITION_X(kk)/1000)^2 + ...
        (s_mobility_Opr_Intp(ii).VS_NODE(jj).V_POSITION_Y(kk)/1000)^2;

      % SNR (only based on Path Loss currently)
      SNR(ii, jj, kk) = Ptx - (128.1 + 18.8 * log10(R_km)) - ...
        (-174 + 10*log10(BW(1, ii)*1e6));

      % Shadowing and slow fading model for updating SNR
      speed = sqrt(...
        (s_mobility_Opr_Intp(ii).VS_NODE(jj).V_SPEED_X(kk))^2 + ...
        (s_mobility_Opr_Intp(ii).VS_NODE(jj).V_SPEED_Y(kk))^2);
      delta = speed * HypRes;
```

```
S_prev(jj, 1) = S_prev(jj, 1) * exp(-delta/Xc) + ...
   sqrt((delta^2) * (1 - exp(-2*delta/Xc))) * randn();
SNR(ii, jj, kk) = SNR(ii, jj, kk) + S_prev(jj, 1);

% Fast fading (Jakes) model
% Specify Rayleigh fading frequency selective and fast fading channel
FF_v = Fast_Fading(Chan, NRB(ii), fc, BW(ii));

% Choose best available PRB (scheduling)
[FF_v_sorted, Idx] = sort(FF_v, 'descend');

% Assign PRB's till all data bits in a voice sample are acc.
UE_assigned = 0;
data_bits_t = 0;
while (UE_assigned == 0)

   flag = 0;
   for ll = 1 : NRB(ii)
      for mm = 1 : V_fr
         if (sch_mtx(mm, Idx(ll)) == 0)

            % Update SNR with Fast Fading value
            SNR_FF = SNR(ii, jj, kk) + ...
               10*log10(FF_v(Idx(ll)));

            % Assigned
            flag = 1;
            break;
         end
      end
      if (flag == 1)
         break;
      end
   end

   % CQI from SNR based on
   % (www.sharetechnote.com/Handbook_LTE_CQI.html)
   CQI(ii, jj, kk) = CQI_from_SNR(SNR_FF);

   % Load bits
   data_bits_t = data_bits_t + ...
      data_bits_from_CQI(CQI(ii, jj, kk));

   % Update scheduling matrix
   sch_mtx(mm, Idx(ll)) = 1;

   % Update
```

```
            if (data_bits_t > dbpvs)
                UE_assigned = 1;
            end
        end
    end

    % Total Bandwidth across UE's for an operator
    est_BW(ii, kk) = sum(sum(sch_mtx)) * PRB_BW / V_fr;

    disp(kk);
  end
end
```

APPENDIX B: MATLAB Code for PDS Simulator

The gen_all_calls.m code is shown below:

```
function [core_ntwk_util, ofld_srvr_util, Pr_Bl_video_ave, Pr_Bl_voice_ave] =
gen_all_calls(Video_ofld_srvr_PRB)

% Define parameters
define_params;

% Generate voice and video calls
[Pr_Bl_voice, Pr_Bl_video, core_ntwk_BW, ofld_srvr_BW] = ...
    gen_calls(arr_iat_voice, mean_len_voice, ...
    arr_iat_video, mean_len_video, Video_ofld_srvr_PRB);

% Averages
core_ntwk_util                                                            =
mean(transpose(core_ntwk_BW(:,1:ceil((1/resln)*simTime)))))./(PRB*RB_BW);
ofld_srvr_util                                                            =
mean(transpose(ofld_srvr_BW(:,1:ceil((1/resln)*simTime)))))./(Video_ofld_srvr_PRB*R
B_BW);

Pr_Bl_voice_ave = mean(Pr_Bl_voice);
Pr_Bl_video_ave = mean(Pr_Bl_video);
```

The gen_calls.m code is shown below:

```
function [Pr_Bl_voice, Pr_Bl_video, core_ntwk_BW, ...
    ofld_srvr_BW] = gen_calls(arr_iat_voice, lambda_l_voice, ...
    arr_iat_video, lambda_l_video, Video_ofld_srvr_PRB)

% Define parameters
define_params;

% Initialize
num_RBs = zeros(N, 1000*mean_len_video + ceil((1/resln)*simTime));
ofld_srvr_RB = zeros(N, 1000*mean_len_video + ceil((1/resln)*simTime));
core_ntwk_RB = zeros(N, 1000*mean_len_video + ceil((1/resln)*simTime));
incoming_voice_calls = zeros(N, 1);
incoming_video_calls = zeros(N, 1);
rejected_voice_calls = zeros(N, 1);
rejected_video_calls = zeros(N, 1);
```

```
% Initialize
next_call_time_array(1:N, 1) = ceil((1/resln)*exprnd(arr_iat_voice));
if (video_single_user == 1)
   next_call_time_array(N+1:2*N, 1) = ...
      [ceil((1/resln)*exprnd(arr_iat_video(1))); simTime*(1/resln)*ones(N-1, 1)];
else
   next_call_time_array(N+1:2*N, 1) = ceil((1/resln)*exprnd(arr_iat_video(1), N,
1));
end
[Val, Idx] = min(next_call_time_array);

% Random nos. arrays
kk = ones(N, 1); mm = ones(N, 1);
rnd_voice_end = ceil((1/resln)*exprnd(lambda_l_voice,
ceil(50*simTime/min(arr_iat_voice)), 1));

size_var_voice = size(ceil((1/resln)*exprnd(arr_iat_voice(1),
ceil(50*simTime/min(arr_iat_voice)), 1)), 1);
size_var_video = size(ceil((1/resln)*exprnd(arr_iat_video(1),
ceil(50*simTime/min(arr_iat_video)), 1)), 1);
rnd_voice_iat = zeros(size_var_voice, N);
rnd_video_iat = zeros(size_var_video, N);

for ii = 1 : N
   rnd_video_iat(:, ii) = ceil((1/resln)*exprnd(arr_iat_video(ii),
ceil(50*simTime/min(arr_iat_video)), 1));
   rnd_voice_iat(:, ii) = ceil((1/resln)*exprnd(arr_iat_voice(ii),
ceil(50*simTime/min(arr_iat_voice)), 1));
end
rnd_video_end = ceil((1/resln)*exprnd(lambda_l_video,
ceil(50*simTime/min(arr_iat_video))+N, 1));

while (Val < ((1/resln)*simTime))

   % Next call is voice call
   Opr = mod(Idx-1,N)+1;
   if (Idx <= N)

      % Incoming call update
      incoming_voice_calls(Opr, 1) = incoming_voice_calls(Opr, 1) + 1;

      % Check - accept/reject
      [AI_RB_array] = accept_reject_voice_calls_virt(num_RBs, Val, Opr);

      % Voice call end time
      voice_call_end_time = Val + rnd_voice_end(kk);
```

```
    % CNR BW
    core_ntwk_RB(Opr, Val : voice_call_end_time-1) = ...
        core_ntwk_RB(Opr, Val : voice_call_end_time-1) ...
        + delta_ave_voice;

    if (sum(AI_RB_array ~= 0) ~= 0)

        for bb = 1:N

            % AIR BW
            num_RBs(bb, Val : voice_call_end_time-1) = ...
                num_RBs(bb, Val : voice_call_end_time-1) + ...
                AI_RB_array(bb);
        end
    else

        % Total video calls - Reject (only keep for a count in P_Bl)
        rejected_voice_calls(Opr, 1) = rejected_voice_calls(Opr, 1) + 1;
    end

    % Next voice call
    next_call_time_array(Idx, 1) = next_call_time_array(Idx, 1) + ...
        rnd_voice_iat(kk(Opr), Opr);
    kk(Opr) = kk(Opr) + 1;

else

    % Incoming call update
    incoming_video_calls(Opr, 1) = incoming_video_calls(Opr, 1) + 1;

    % Check - accept/reject
    [CN_Opr, Ofld_Opr, AI_RB_array] = ...
        accept_reject_video_calls_virt(num_RBs, ...
        ofld_srvr_RB, Val, Opr, Video_ofld_srvr_PRB);

    % Video call end time
    video_call_end_time = Val + rnd_video_end(sum(mm));

    % Ofld
    if (Ofld_Opr ~= 0)

        % Offload server bandwidth
        ofld_srvr_RB(Ofld_Opr, Val : video_call_end_time-1) = ...
            ofld_srvr_RB(Ofld_Opr, Val : video_call_end_time-1) ...
            + delta_ave_video;
    end
```

```
        if (CN_Opr ~= 0)

            % CNR BW
            core_ntwk_RB(CN_Opr, Val : video_call_end_time-1) = ...
                core_ntwk_RB(CN_Opr, Val : video_call_end_time-1) ...
                + delta_ave_video;
        end

        if (sum(AI_RB_array ~= 0) ~= 0)

            for bb = 1:N

                % AIR BW
                num_RBs(bb, Val : video_call_end_time-1) = ...
                    num_RBs(bb, Val : video_call_end_time-1) + ...
                    AI_RB_array(bb);
            end
        else

            % Total video calls - Reject (only keep for a count in P_Bl)
            rejected_video_calls(Opr, 1) = rejected_video_calls(Opr, 1) + 1;
        end

        % Next video call
        next_call_time_array(Idx, 1) = next_call_time_array(Idx, 1) + ...
            rnd_video_iat(mm(Opr), Opr);
        mm(Opr) = mm(Opr) + 1;
    end

    % Update
    [Val, Idx] = min(next_call_time_array);
end

% P(Bl) (+0.01 is a Hack to avoid NaN)
Pr_Bl_voice = min(1, rejected_voice_calls./(incoming_voice_calls+0.01));
Pr_Bl_video = min(1, rejected_video_calls./(incoming_video_calls+0.01));

% Ofld and CN BW
ofld_srvr_BW = ofld_srvr_RB * RB_BW;
core_ntwk_BW = core_ntwk_RB * RB_BW;
```

The accept_reject_video_calls.m code is shown below (CAC algorithm is embedded):

```
function [CN_Opr, Ofld_Opr, AI_RB_array] = ...
   accept_reject_video_calls_virt(num_RBs, ofld_srvr_RB, Val, Opr, ...
   Video_ofld_srvr_PRB)

% Define parameters
define_params;
AI_RB_array = zeros(1, N);
AI_RB_array_temp = zeros(1, N);
Ofld_Opr = 0;
CN_Opr = 0;

% Define loaded RBs
loaded_RBs_ave = num_RBs(:, Val);

% Define loaded Offload server
loaded_RBs_ofld_servr_ave = ofld_srvr_RB(:, Val);

% Resources available from Offloading
avlbl_ofld_virt = Video_ofld_srvr_PRB' - loaded_RBs_ofld_servr_ave;
[max_AOS_Val, max_AOS_idx] = max(avlbl_ofld_virt);

% AI Resources available from virt.
CSCV_RB = ((alpha_par .* PRB) - loaded_RBs_ave') .* gamma_video;
CSCV_RB(Opr) = ((alpha_par(Opr) * PRB(Opr)) - loaded_RBs_ave(Opr));

% Sort resources with VNE
[CSCV_RB_sorted, CSCV_RB_sorted_idx] = sort(CSCV_RB, 'descend');

% Total resources for VNE
total_CSCV_RB_VNE = 0;

% Initialize the 'Remaining To Be Alotted RBs'
rem_tba_RB = delta_ave_video;

for ii = 1:N

   alotted_RB = min(rem_tba_RB, CSCV_RB(CSCV_RB_sorted_idx(ii)));
   total_CSCV_RB_VNE = total_CSCV_RB_VNE + alotted_RB;

   AI_RB_array_temp(CSCV_RB_sorted_idx(ii)) = alotted_RB;

   if (total_CSCV_RB_VNE == delta_ave_video)
      break;
   end
   rem_tba_RB = delta_ave_video - total_CSCV_RB_VNE;
end
```

```
% Which Opr the call should go to
if ((alpha_par(Opr)*PRB(Opr) - loaded_RBs_ave(Opr)) >= delta_ave_video)

   % Call assigned to HO - to (HO_usage_b4_ofld)% HO resources occupation
   CN_Opr = Opr;
   AI_RB_array(Opr) = delta_ave_video;

elseif (total_CSCV_RB_VNE == delta_ave_video)

   % Offload
   if (max_AOS_Val >= delta_ave_video)

      % Offload available, AIR MVNE resources available
      % Virtualized Offload
      Ofld_Opr = max_AOS_idx;

   else

      % Offload not available, AIR MVNE resources available
      CN_Opr = Opr;
   end

   % Air Interface Resources
   AI_RB_array = AI_RB_array_temp;

end
```

The accept_reject_voice_calls.m code is shown below (CAC algorithm is embedded):

```
function [AI_RB_array] = accept_reject_voice_calls_virt(num_RBs, Val, Opr)

% Define parameters
define_params;
AI_RB_array = zeros(1, N);
AI_RB_array_temp = zeros(1, N);

% Define loaded RBs
loaded_RBs_ave = num_RBs(:, Val);

% AI Resources available from virt.
CSCV_RB = ((alpha_par .* PRB) - loaded_RBs_ave') .* gamma_voice;
CSCV_RB(Opr) = ((alpha_par(Opr) * PRB(Opr)) - loaded_RBs_ave(Opr));

% Sort resources with VNE
[CSCV_RB_sorted, CSCV_RB_sorted_idx] = sort(CSCV_RB, 'descend');
```

```
% Total resources for VNE
total_CSCV_RB_VNE = 0;

% Initialize the 'Remaining To Be Alotted RBs'
rem_tba_RB = delta_ave_voice;

for ii = 1:N

   alotted_RB = min(rem_tba_RB, CSCV_RB(CSCV_RB_sorted_idx(ii)));
   total_CSCV_RB_VNE = total_CSCV_RB_VNE + alotted_RB;

   AI_RB_array_temp(CSCV_RB_sorted_idx(ii)) = alotted_RB;

   if (total_CSCV_RB_VNE == delta_ave_voice)
      break;
   end
   rem_tba_RB = delta_ave_voice - total_CSCV_RB_VNE;
end

% Which Opr the call should go to
if ((alpha_par(Opr)*PRB(Opr) - loaded_RBs_ave(Opr)) >= delta_ave_voice)

   % Call assigned to HO
   AI_RB_array(Opr) = delta_ave_voice;

elseif (total_CSCV_RB_VNE == delta_ave_voice)

   AI_RB_array = AI_RB_array_temp;

end
```

The define_params.m code is shown below:

```
% Common Parameters
simTime = 50000; % Total simulation time
resln = 0.1; % Resolution of time used for generating calls (seconds)

N = 5; % no. of operators
alpha_par = [1 1 1 1 1]; % alpha parameter (for HO calls)
PRB = [75 50 25 25 50]; % Air interface resource blocks per operator

% RB BW
RB_BW = 200e3;
voice_rate = 15e3;
video_rate = 356e3;
```

```
% HO left capacity before Offloading
HO_left_cap_b4_ofld = [0.5 0.5 0.5 0.5 0.5];

% Voice parameters
gamma_voice = [0.9 0.9 0.9 0.9 0.9]; % assumed values
delta_ave_voice = 0.12; % With 29 dBm Tx Power, 0.3282 with 0 dBm
arr_iat_voice = [2/3 1 2 2 1]; % Voice Call Inter arrival time
mean_len_voice = 50; % Mean voice call length

% Video parameters
gamma_video = [1 1 1 1 1]; % is 1 for NRT (video) application
delta_ave_video = 2.5; % With 29 dBm Tx Power, 10.6733 with 0 dBm
arr_iat_video = [2 3 6 6 3]; % Video Call Inter arrival time
mean_len_video = 50; % Mean video call length
video_single_user = 0; % Single user only if = 1
```

APPENDIX C: Sample of Commands For Emulation Configuration

```
Router#sh int Ethernet0/0 | i drops
  Input queue: 0/75/0/0 (size/max/drops/flushes); Total output drops: 240
     12 unknown protocol drops  < note that there are 12 packets for protocols communication
```

Command Line Proof & statistics :

```
Router#sh policy-map interface fastEthernet 0/0 output
 FastEthernet0/0

  Service-policy output: COS-OUT-SHAPED-Gi8.590

    Class-map: class-default (match-any)
      1561 packets, 2242923 bytes
      5 minute offered rate 35000 bps, drop rate 0 bps
      Match: any
      Traffic Shaping
           Target/Average   Byte   Sustain   Excess    Interval  Increment
             Rate           Limit  bits/int  bits/int  (ms)      (bytes)
          9296000/9296000   58100  232400    232400    25        29050

        Adapt  Queue     Packets   Bytes     Packets   Bytes     Shaping
        Active Depth                         Delayed   Active
        -      0         1561      2242923   0         0         no

      Service-policy : COS-OUT-Gi8.590

        Class-map: DSCP-OUT-RT (match-any)
          712 packets, 1077968 bytes
          5 minute offered rate 1200000 bps, drop rate 0 bps
          Match:  dscp ef (46)
            712 packets, 1077968 bytes
            5 minute rate 12000000 bps
          Queueing
            Strict Priority
            Output Queue: Conversation 264
            Bandwidth 2000 (kbps) Burst 50000 (Bytes)
            (pkts matched/bytes matched) 0/0
            (total drops/bytes drops) 0/0
          police:
              cir 5840000 bps, bc 730000 bytes, be 730000 bytes
            conformed 712 packets, 1077968 bytes; actions:
              set-dscp-transmit ef
```

```
        exceeded 0 packets, 0 bytes; actions:
          drop
        violated 0 packets, 0 bytes; actions:
          drop
        conformed 20000 bps, exceed 0 bps, violate 0 bps

    Class-map: DSCP-OUT-RTVI (match-any)
      764 packets, 1156696 bytes
      5 minute offered rate 1500000 bps, drop rate 0 bps
      Match:  dscp af41 (34)
        764 packets, 1156696 bytes
        5 minute rate 1500000 bps
      Queueing
        Output Queue: Conversation 265
        Bandwidth remaining 7 (%)Max Threshold 64 (packets)
        (pkts matched/bytes matched) 0/0
    (depth/total drops/no-buffer drops) 0/0/0
      police:
          cir 5840000 bps, bc 730000 bytes, be 730000 bytes
        conformed 764 packets, 1156696 bytes; actions:
          transmit
        exceeded 0 packets, 0 bytes; actions:
          drop
        violated 0 packets, 0 bytes; actions:
          drop
        conformed 20000 bps, exceed 0 bps, violate 0 bps

    Class-map: DSCP-OUT-D1 (match-any)
      529 packets, 60306 bytes
      5 minute offered rate 2000000 bps, drop rate 0 bps
      Match:  dscp af31 (26)
        529 packets, 60306 bytes
        5 minute rate 2000000 bps
      Match:  dscp af32 (28)
        0 packets, 0 bytes
        5 minute rate 0 bps
      Queueing
        Output Queue: Conversation 266
Bandwidth remaining 54 (%)Max Threshold 64 (packets)
        (pkts matched/bytes matched) 0/0
    (depth/total drops/no-buffer drops) 20/0/0

    Class-map: DSCP-OUT-D2 (match-any)
      1292 packets, 1194488 bytes
      5 minute offered rate 700000 bps, drop rate 0 bps
      Match:  dscp af21 (18)
        1292 packets, 1194488 bytes
```

5 minute rate 700000 bps
Match: dscp af22 (20)
0 packets, 0 bytes
5 minute rate 0 bps
Queueing
Output Queue: Conversation 267
Bandwidth remaining 25 (%)Max Threshold 64 (packets)
(pkts matched/bytes matched) 0/0
(depth/total drops/no-buffer drops) 0/0/0

Class-map: DSCP-OUT-D3 (match-any)
0 packets, 0 bytes
5 minute offered rate 170000 bps, drop rate 0 bps
Match: dscp af11 (10)
0 packets, 0 bytes
5 minute rate 0 bps
Match: dscp af12 (12)
0 packets, 0 bytes
5 minute rate 0 bps
Queueing
Output Queue: Conversation 268
Bandwidth remaining 8 (%)Max Threshold 64 (packets)
(pkts matched/bytes matched) 0/0
(depth/total drops/no-buffer drops) 26/0/0

Class-map: class-default (match-any)
859 packets, 1149022 bytes
5 minute offered rate 30000 bps, drop rate 0 bps

Configuration commands

Some Configuration for essential device :
Gateway-with-LBO device #sh run
Router#sh run
Building configuration...

Current configuration : 2526 bytes
!! Last configuration change at 13:55:39 EET Mon Jul 16 2018
!version 15.4
service timestamps debug datetime msec
service timestamps log datetime msec
no service password-encryption
!hostname Router
!boot-start-marker
boot-end-marker
!!
enable password cisco
!

```
no aaa new-model
clock timezone EET 2 0
mmi polling-interval 60
no mmi auto-configure
no mmi pvc
mmi snmp-timeout 180
!
ip cef
no ipv6 cef
!
multilink bundle-name authenticated
!!
redundancy
!
no cdp run
!
track 100 ip sla 100
!
!
interface Ethernet0/0
 ip address 6.6.6.1 255.255.255.0
 load-interval 30
!
interface Ethernet0/1
 no ip address
 shutdown
!
interface Ethernet0/2
 no ip address
 shutdown
!
interface Ethernet0/3
 ip address 3.3.3.1 255.255.255.0
 load-interval 30
!
interface Ethernet1/0
 ip address 2.2.2.2 255.255.255.0
 load-interval 30
!
interface Ethernet1/1
 ip address 40.40.40.1 255.255.255.0
!
interface Ethernet1/2
 no ip address
 shutdown
!
interface Ethernet1/3
```

```
 no ip address
 shutdown
!
ip forward-protocol nd
!
!
no ip http server
no ip http secure-server
ip route 0.0.0.0 0.0.0.0 Ethernet0/0 6.6.6.2
ip route 0.0.0.0 0.0.0.0 Ethernet0/3 3.3.3.2 250
ip route 1.1.1.0 255.255.255.0 2.2.2.1
!
ip access-list extended LBO
 permit icmp host 1.1.1.1 host 8.8.8.8 dscp af22
!
ip sla 100
 icmp-echo 40.40.40.2
!
route-map LBO permit 10
 match ip address LBO
 set ip next-hop verify-availability 40.40.40.2 1 track 100
 set ip next-hop 40.40.40.2
!
!control-plane
!!
line con 0
 logging synchronous
line aux 0
line vty 0 4
 exec-timeout 0 0
 password cisco
 login
 transport input telnet ssh
!
event manager applet int-monitor223
 event interface name Ethernet0/0 parameter transmit_rate_bps entry-op ge entry-val 380000 entry-type value poll-interval 1 maxrun 140
 action 1.0 cli command "enable"
 action 2.0 cli command "config t"
 action 3.0 cli command "no ip route 0.0.0.0 0.0.0.0 Ethernet0/0 6.6.6.2"
 action 4.0 cli command "no ip route 0.0.0.0 0.0.0.0 Ethernet0/3 3.3.3.2 250"
 action 5.0 cli command "ip route 0.0.0.0 0.0.0.0 Ethernet0/0 6.6.6.2 250 "
 action 6.0 cli command "ip route 0.0.0.0 0.0.0.0 Ethernet0/3 3.3.3.2"
 action 7.0 wait 120
 action 8.0 cli command "no ip route 0.0.0.0 0.0.0.0 Ethernet0/0 6.6.6.2 250"
 action 9.0 cli command "no ip route 0.0.0.0 0.0.0.0 Ethernet0/3 3.3.3.2"
 action 9.1 cli command "ip route 0.0.0.0 0.0.0.0 Ethernet0/0 6.6.6.2"
```

```
 action 9.2 cli command "ip route 0.0.0.0 0.0.0.0 Ethernet0/3 3.3.3.2 250"
!
end

Switch5 #sh run
Building configuration...

Current configuration : 2839 bytes
!
! Last configuration change at 15:12:50 UTC Sat Jul 14 2018
!
version 15.2
service timestamps debug datetime msec
service timestamps log datetime msec
no service password-encryption
service compress-config
!
hostname Switch
!
boot-start-marker
boot-end-marker
!
!no aaa new-model
!
!ip cef
no ipv6 cef
!
spanning-tree mode rapid-pvst
spanning-tree extend system-id
!
vlan internal allocation policy ascending
!
!interface GigabitEthernet0/1
 media-type rj45
 negotiation auto
!
interface GigabitEthernet0/3
 media-type rj45
 negotiation auto
!
interface GigabitEthernet0/2
 no switchport
 ip address 1.1.1.2 255.255.255.0
 negotiation auto
!
interface GigabitEthernet0/0
 no switchport
```

```
 ip address 2.2.2.1 255.255.255.0
 negotiation auto
!
ip forward-protocol nd
!
no ip http server
no ip http secure-server
!
ip route 0.0.0.0 0.0.0.0 GigabitEthernet0/0 2.2.2.2
!
!control-plane
!
banner exec ^C
**************************************************************************
* IOSv is strictly limited to use for evaluation, demonstration and IOS  *
* education. IOSv is provided as-is and is not supported by Cisco's      *
* Technical Advisory Center. Any use or disclosure, in whole or in part, *
* of the IOSv Software or Documentation to any third party for any       *
* purposes is expressly prohibited except as otherwise authorized by     *
* Cisco in writing.                                                      *
**************************************************************************^C
banner incoming ^C
**************************************************************************
* IOSv is strictly limited to use for evaluation, demonstration and IOS  *
* education. IOSv is provided as-is and is not supported by Cisco's      *
* Technical Advisory Center. Any use or disclosure, in whole or in part, *
* of the IOSv Software or Documentation to any third party for any       *
* purposes is expressly prohibited except as otherwise authorized by     *
* Cisco in writing.                                                      *
**************************************************************************^C
banner login ^C
**************************************************************************
* IOSv is strictly limited to use for evaluation, demonstration and IOS  *
* education. IOSv is provided as-is and is not supported by Cisco's      *
* Technical Advisory Center. Any use or disclosure, in whole or in part, *
* of the IOSv Software or Documentation to any third party for any       *
* purposes is expressly prohibited except as otherwise authorized by     *
* Cisco in writing.                                                      *
**************************************************************************^C
!
line con 0
```

```
line aux 0
line vty 0 4
 login
!
!
end


Router#sh run
Building configuration...

Current configuration : 1120 bytes
!
! Last configuration change at 19:08:02 EET Sat Jul 14 2018
!
version 15.4
service timestamps debug datetime msec
service timestamps log datetime msec
no service password-encryption
!
hostname Router
!
boot-start-marker
boot-end-marker
!
!
!
no aaa new-model
clock timezone EET 2 0
mmi polling-interval 60
no mmi auto-configure
no mmi pvc
mmi snmp-timeout 180
!
!ip vrf test
!
ip cef
no ipv6 cef
!
multilink bundle-name authenticated
!
!
redundancy
!
!!
interface Ethernet0/0
 ip address 30.30.30.1 255.255.255.0
```

```
!
interface Ethernet0/1
 ip address 10.10.10.2 255.255.255.0
!
interface Ethernet0/2
 no ip address
 shutdown
!
interface Ethernet0/3
 no ip address
 shutdown
!
!
router eigrp 500
!
router isis
!
router isis 12
!
router bgp 500
 bgp log-neighbor-changes
!
ip forward-protocol nd
!
!
no ip http server
no ip http secure-server
ip route 1.1.1.0 255.255.255.0 10.10.10.1
ip route 8.8.8.0 255.255.255.0 30.30.30.2
!
control-plane
!
line con 0
 logging synchronous
line aux 0
line vty 0 4
 login
 transport input none
!

End
```

R19

```
R19#sh run
Building configuration...

Current configuration : 1447 bytes
```

```
!
! Last configuration change at 12:01:15 EET Mon Jul 16 2018
!
version 15.4
service timestamps debug datetime msec
service timestamps log datetime msec
no service password-encryption
!
hostname Router
!
boot-start-marker
boot-end-marker
!
no aaa new-model
clock timezone EET 2 0
mmi polling-interval 60
no mmi auto-configure
no mmi pvc
mmi snmp-timeout 180
!
ip cef
no ipv6 cef
!
multilink bundle-name authenticated
!
redundancy
!
interface Ethernet0/0
 ip address 14.14.14.2 255.255.255.0
 ip policy route-map LBO
!
interface Ethernet0/1
 ip address 13.13.13.1 255.255.255.0
!
interface Ethernet0/2
 no ip address
 shutdown
!
interface Ethernet0/3
 no ip address
 shutdown
!
interface Ethernet1/0
 ip address 200.200.200.1 255.255.255.0
!
interface Ethernet1/1
 no ip address
```

```
 shutdown
!
interface Ethernet1/2
 no ip address
 shutdown
!
interface Ethernet1/3
 no ip address
 shutdown
!
ip forward-protocol nd
!
!
no ip http server
no ip http secure-server
ip route 1.1.1.0 255.255.255.0 14.14.14.1
ip route 5.5.5.0 255.255.255.0 14.14.14.1
ip route 8.8.8.0 255.255.255.0 13.13.13.2
!
ip access-list extended LBO
 permit icmp host 5.5.5.1 host 8.8.8.8 dscp af22
!
!
route-map LBO permit 10
 match ip address LBO
 set ip next-hop 200.200.200.2
!
control-plane
!
line con 0
 logging synchronous
line aux 0
line vty 0 4
 login
 transport input none
!!
end
```

APPENDIX D: Published Paper

Paper title: “Collaborative Mobile Network Operators Framework for Efficient Spectrum Utilization in LTE-A”

Authors:

Subah Al-Binali

Dr. Omar Alani

The paper was accepted, introduced and published in the proceedings of PGNET2014 in Liverpool from 23-24/6/2014.