Technologies and Architecture School

Sciences and Information Technology Department

Improved Planning and Resource Management in Next Generation Green Mobile Communication Networks

Luís Carlos Barruncho dos Santos Gonçalves

Thesis specially presented for the fulfillment of the degree of

Doctor in Information Science and Technology

Jury:

Doctor Pedro Ramos, PhD, Associate Professor, ISCTE – Instituto Universitário de Lisboa

Doctor António José Castelo Branco Rodrigues, PhD, Assistant Professor, Instituto Superior Técnico

Doctor Fernando José da Silva Velez, PhD, Coordinator Professor, Universidade da Beira Interior

Doctor José Alberto Gouveia Fonseca, PhD, Associate Professor, Universidade de Aveiro

Doctor Américo Manuel Carapeto Correia, PhD, Full Professor, ISCTE – Instituto Universitário de Lisboa

July, 2020

Abstract

In upcoming years, mobile communication networks will experience a disruptive reinventing process through the deployment of post 5th Generation (5G) mobile networks. Profound impacts are expected on network planning processes, maintenance and operations, on mobile services, subscribers with major changes in their data consumption and generation behaviours, as well as on devices itself, with a myriad of different equipment communicating over such networks. Post 5G will be characterized by a profound transformation of several aspects: processes, technology, economic, social, but also environmental aspects, with energy efficiency and carbon neutrality playing an important role. It will represent a network of networks: where different types of access networks will coexist, an increasing diversity of devices of different nature, massive cloud computing utilization and subscribers with unprecedented data-consuming behaviours. All at greater throughput and quality of service, as unseen in previous generations.

The present research work uses 5G new radio (NR) latest release as baseline for developing the research activities, with future networks post 5G NR in focus. Two approaches were followed: i) method re-engineering, to propose new mechanisms and overcome existing or predictably existing limitations and ii) concept design and innovation, to propose and present innovative methods or mechanisms to enhance and improve the design, planning, operation, maintenance and optimization of 5G networks. Four main research areas were addressed, focusing on optimization and enhancement of 5G NR future networks, the usage of edge virtualized functions, subscriber's behavior towards the generation of data and a carbon sequestering model aiming to achieve carbon neutrality. Several contributions have been made and demonstrated, either through models of methodologies that will, on each of the research areas, provide significant improvements and enhancements from the planning phase to the operational phase, always focusing on optimizing resource management. All the contributions are retro compatible with 5G NR and can also be applied to what starts being foreseen as future mobile networks. From the subscriber's perspective and the ultimate goal of providing the best quality of experience possible, still considering the mobile network operator's (MNO) perspective, the different proposed or developed approaches resulted in optimization methods for the numerous problems identified throughout the work. Overall, all of such contributed individually but aggregately as a whole to improve and enhance globally future mobile networks. Therefore, an answer to the main question was provided: how to further optimize a next-generation network - developed with optimization in mind - making it even more efficient while, simultaneously, becoming neutral concerning carbon emissions. The developed model for MNOs which aimed to achieve carbon neutrality through CO_2 sequestration together with the subscriber's behaviour model - topics still not deeply focused nowadays – are two of the main contributions of this thesis and of utmost importance for post-5G networks.

Keywords: 5G, traffic offload, traffic steering, carbon neutrality, optimization, efficiency, planning, subscriber behaviour, advanced analytics.

Resumo

Nos próximos anos espera-se que as redes de comunicações móveis se reinventem para lá da 5ª Geração (5G), com impactos profundos ao nível da forma como são planeadas, mantidas e operacionalizadas, ao nível do comportamento dos subscritores de serviços móveis, e através de uma miríade de dispositivos a comunicar através das mesmas. Estas redes serão profundamente transformadoras em termos tecnológicos, económicos, sociais, mas também ambientais, sendo a eficiência energética e a neutralidade carbónica aspetos que sofrem uma profunda melhoria. Paradoxalmente, numa rede em que coexistirão diferentes tipos de redes de acesso, mais dispositivos, utilização massiva de sistema de computação em nuvem, e subscritores com comportamentos de consumo de serviços inéditos nas gerações anteriores. O trabalho desenvolvido utiliza como base a *release* mais recente das redes 5G NR (New Radio), sendo o principal *focus* as redes pós-5G. Foi adotada uma abordagem de "reengenharia de métodos" (com o objetivo de propor mecanismos para resolver limitações existentes ou previsíveis) e de "inovação e design de conceitos", em que são apresentadas técnicas e metodologias inovadoras, com o principal objetivo de contribuir para um desenho e operação otimizadas desta geração de redes celulares.

Quatro grandes áreas de investigação foram endereçadas, contribuindo individualmente para um todo: melhorias e otimização generalizada de redes pós-5G, a utilização de virtualização de funções de rede, a análise comportamental dos subscritores no respeitante à geração e consumo de tráfego e finalmente, um modelo de sequestro de carbono com o objetivo de compensar as emissões produzidas por esse tipo de redes que se prevê ser massiva, almejando atingir a neutralidade carbónica. Como resultado deste trabalho, foram feitas e demonstradas várias contribuições, através de modelos ou metodologias, representando em cada área de investigação melhorias e otimizações, que, todas contribuindo para o mesmo objetivo, tiveram em consideração a retro compatibilidade e aplicabilidade ao que se prevê que sejam as futuras redes pós 5G.

Focando sempre na perspetiva do subscritor da melhor experiência possível, mas também no lado do operador de serviço móvel – que pretende otimizar as suas redes, reduzir custos e maximizar o nível de qualidade de serviço prestado - as diferentes abordagens que foram desenvolvidas ou propostas, tiveram como resultado a resolução ou otimização dos diferentes problemas identificados, contribuindo de forma agregada para a melhoria do sistema no seu todo, respondendo à questão principal de como otimizar ainda mais uma rede desenvolvida para ser extremamente eficiente, tornando-a, simultaneamente, neutra em termos de emissões de carbono. Das principais contribuições deste trabalho relevam-se precisamente o modelo de compensação das emissões de CO_2, com vista à neutralidade carbónica e um modelo de análise comportamental dos subscritores, dois temas ainda pouco explorados e extremamente importantes em contexto de redes futuras pós-5G.

Palavras-Chave: 5G, neutralidade de carbono, *DenseNets*, direcionamento de tráfego, otimização, eficiência, planeamento, comportamento subscritores, analítica avançada.

Acknowledgements

As I write these words, I realize now that every PhD is a lifetime challenge for any person who embraces it, not solely an action that starts and simply finishes after a couple of years. As usually said, a long walk all it takes is a small step, and this small step was taken in back in 2012. And suddenly, where an empty circle existed, a small step in the centre appeared, following its way to the edge of the circle in order to continue a never-ending journey of work, research and construction of this thesis, which in the end, is only the first of several future walks.

As Lucretius, the Roman philosopher once said, *ex nihilo nihil fit*, meaning that nothing comes from nothing, being a father, family head and fully employed, to embrace such a challenge especially in *quasi* part time, seemed like a daunting task at the time. But no one ever walks alone, and I was very fortunate to have two Professors that did not let fear win me over and were able to transfer onto me all their enthusiasm about the journey ahead. For that, I am and will always be very thankful to Professor Francisco Cercas and Professor Américo Correia. They were, in fact the catalysts of this journey. I still remember and cherish the day when Professor Américo Correia introduced me to my supervisors. His words of incentive were constant and admirable. At that point, I knew I had taken the first step in the middle of the PhD circle. Two different supervisors, that paved my way towards the perimeter of the circle and allowed me to start my endless journey.

There are not enough words to express my gratitude to my supervisors, Professors Nuno Souto and Pedro Sebastião. They became two people that have endless and tirelessly walked with me down the PhD path, always available, motivating, alongside with their full knowledge and invaluable advices. We shared afternoons, evenings, meals, ideas, thoughts, and were present whenever needed. Aside being excellent professionals with profound knowledge of all subjects, both revealed kindness in sharing and providing their valuable knowledge and experience. Today, I can say that both experts held my hand through all the intellectual and technical tumultuous processes of bringing light to this thesis. I am truly and deeply thankful to both, without whom I would not have achieve this final stage.

During this whole process, which was becoming progressively more difficult due to additional responsibilities and being promoted on the hierarchy path in my full-time job,

as Head of Cybersecurity, IT Risk and Compliance, both supervisors never let me fall in discourage. Harsh times eventually fell over me and I was forced to interrupt my journey for a year and slowed down the rate at which the research work was conducted. Family needs as well as increased responsibilities at work left almost no time to pursue the objectives on time. Time was slipping through my days, but words of encouragement continued to stream down from my supervisors. Harsh times were, simultaneously, admirable times thankfully to both and also to Professor Américo Correia.

Along my path I encountered other people to whom I wish to thank for. Starting with Professors Luis Botelho and Octavian Postolache, who provided helpful suggestions, recommendations and valuable advices. Also, I would like Professors Ana Almeida and Elsa Cardoso for being such fantastic supporters of the work that was being created and for being such inestimable persons alongside this project, embracing me as one of their own on the fulfilling process of teaching Data Science, which eventually led to improve even more the current work.

A special word for the enormous support provided by BALU, which I believe shined light several times over some of the harsh and most difficult cloudy moments, always with an encouraging word and relentlessly believing that this work would come to a successful ending. BALU is a fantastic person who I thank very much for being tireless and unconditionally by my side, injecting knowledge but most importantly, courage to endure. Finally, some deep words of gratitude to my family. To my parents, Rosário Gonsalves and Carlos Gonsalves, and my uncle Mário Barruncho, who were always there by my side, encouraging me to continue even in the most difficult times. I saved, naturally, the most important people of my life for last: my daughter Bianca, which was born almost when this journey started and my son Rafael, who was born one year after I was already pursuing this challenge. Thank you both for all those hours you gave me to conclude my work, although you never really understood why Daddy was not playing with you, preferring to sit by the computer. Hopefully you will understand one day. To the love of my life, Cláudia, my beautiful wife, who encouraged me the first time, believed and was there for me every day. This work has been possible only through their sacrifice and encouragement and therefore to all of them is dedicated.

Contents

List of Tables

Chapter II – Articles

2.1 Article nr. #1

2.2 Article nr. #2

2.3 Article nr. #3

2.4 Article nr. #4

List of Figures

Chapter I – Introduction

Chapter II – Articles

2.1 Article nr. #1

2.2 Article nr. #2

2.3 Article nr. #3

2.4 Article nr. #4

Chapter III – Conclusions

List of Acronyms

2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
5GC	5G Core
5G gNB	5G Next Generation NodeB
5G NR	5G New Radio
ANDSF	Access Network Discovery and Selection Function
ARPU	Average Revenue per User
CDN	Content Delivery Network
CN	Core Network
CO_2	Carbon Dioxide
C-RAN	Cloud RAN
CSS	Carbon Sequestration and Storage
D2D	Device to Device
EC	Edge Computing
EE	Energy Efficiency
eMBB	Enhanced Mobile Broadband
HO	Hand-Over
HS	Hotspot
HS2	Hotspot 2.0
IoT	Internet of Things
KYC	Know Your Customer
LTE	Long Term Evolution
M2M	Machine to Machine
mMTC	Massive Machine Type Communications
MNO	Mobile Network Operator
MR	Multi-RAT
NF	Network Function
NFV	NF Virtualization
NR	New Radio
NR-U	NR Unlicensed
OPEX	Operational Expenditure
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
SDN	Software Defined Network
SIB	System Information Block
UAV	Unmanned Aerial Vehicle
UE	User Equipment
URLLC	Ultra Reliable and Low Latency Communications
WIFI	Wireless Fidelity
WIFI6	WIFI version 6

Page intentionally left blank

Chapter I – Introduction

This chapter describes the drivers that formed the starting point of this work. By focusing on the foundations and reasons for conducting this work, the main research questions and goals are presented.

Afterwards, the focus shifts to the background of the domains and subject areas addressed. At this point the work gains some diversity, which contributes to its overall completeness. Especially, it was taken into consideration that, although 5th Generation (5G) cellular networks are the working baseline and future networks beyond 5G new radio (5G NR) the main research objective, setting the path forward, a comprehensive approach to the research problems had to consider previous versions, like Long Term Evolution (LTE). As it will be seen throughout the work, several of the proposed techniques were quite innovative, according to peer review.

It should be noted that this is a multi-disciplinary research work, focusing on subjects that are very relevant and timely. However, some of the subsections are not completely exhaustive, in order to avoid repeating the background sections of the articles that were published, which are fully copied from each journal to this document. Each publication date depends on each journals' internal processes and methodologies, meaning that some dates do not totally reflect the chronological order of this work.

Additional journal articles that are under peer review have also been fully copied to this document, and a specific section exists to list all the international work that was developed and presented at several conferences.

In order to detail the research questions, but especially the different contributions that resulted from this work, a section is included for such presentation. The aim is to help clarify even further those aspects, despite being referred in each of the different articles.

The last section of Chapter I addresses the aspects related to the document organization, which is of utmost importance for easy readability and understanding of the research flow, considering that this document adopts a thesis by articles structure.

1.1. Motivation and Goals

With the forthcoming advent of pervasive communications, seamless and global connectivity is crucial. Communications invisibility must be a reality, meaning that subscribers should not feel the difference of being connected, either indoor, outdoor, over cellular or other access technologies[1]. That is one of the main objectives of 5G systems. Data communications within mobility should therefore be at its highest level of transparency. Always on data connections must provide "everywhere, anytime" real data transparent communications, not only to handheld terminals, but also to any autonomous computing entity with communications capabilities. The access to data communications should be simple for embedded devices and, in some cases, global overall signal coverage should be mandatory (*e.g.*, embedded health monitoring devices). This kind of pervasive communications will provide the ability of context and location-based services and applications, highly integrated user and machine interaction and overall transparent usage of data connections. Systems and networks beyond 5G NR, will be even more highly disruptive when compared with former cellular network generations [2]. This high level of disruption will result in unprecedented innovation, where all sorts of devices will coexist, different radio base stations, networks within networks and different radio access technologies: a myriad of new concepts, methods and processes that, as never before, need to be properly managed and optimized [3]. This was the biggest motivation: how to further optimize a network which has been carefully designed to be optimal as much as possible.

In order to pursue such motivation, several questions needed to be raised: how can all those radio access technologies coexist and provide transparent and seamless connectivity to the massive number of subscribers that are expected to service such networks? And, moreover, how to properly coordinate all these interactions? How can the advent of de-materialized cloud systems leverage 5G core (5GC) and radio access network (RAN) optimizations? And, at the end, is there a way to improve existing cellular generations and deployments in order to allow a smoother migration to future networks, passing through 5G NR, without changing existing standards? Yet, a very important aspect, which role will the subscribers' behavior play in such complex and dense systems? Will behavior regarding data generation drive network changes or planning processes? Is there any relationship between introducing new services, cease existing ones and the behavior of the subscriber?

Can it be modeled in a subscriber-centric perspective? Is it possible to use Data Science methods like advanced clustering to improve overall 5G NR planning and resource management processes, therefore paving the way to future networks giving birth to the concept of intelligent cellular networks? 5G NR considers concepts like network function (NF) virtualization: is there any function that can be virtualized, setting the scenario for NF virtualization (NFV) maximization in future networks, leveraging and further enhancing the described goals? These are some of the main questions that this research work presented answers to. In fact, these are the research questions of this work. Yet, additionally and becoming progressively more important, another research question had to be raised: in such dense networks, with massive number of subscribers and radio base stations, where different radio access technologies co-exist, is it possible to keep such intensive resource-hungry processing and simultaneous be environmentally friendly? Is there a way for mobile network operators to develop programs in order to reduce its carbon dioxide emissions into the atmosphere, or the carbon footprint of the whole life cycle, from smartphone manufacturing all the way to servicing? Can MNOs become carbon neutral by offsetting their carbon footprint? Can 5G NR but especially, future networks be environment friendly? Within this thesis answers to all of these questions have been provided and detailed in Chapter II, but listed, for simplicity in Section 1.4.

Figure 1 summarizes the flow of activities that were undertaken, and which will be presented in the following subsections.

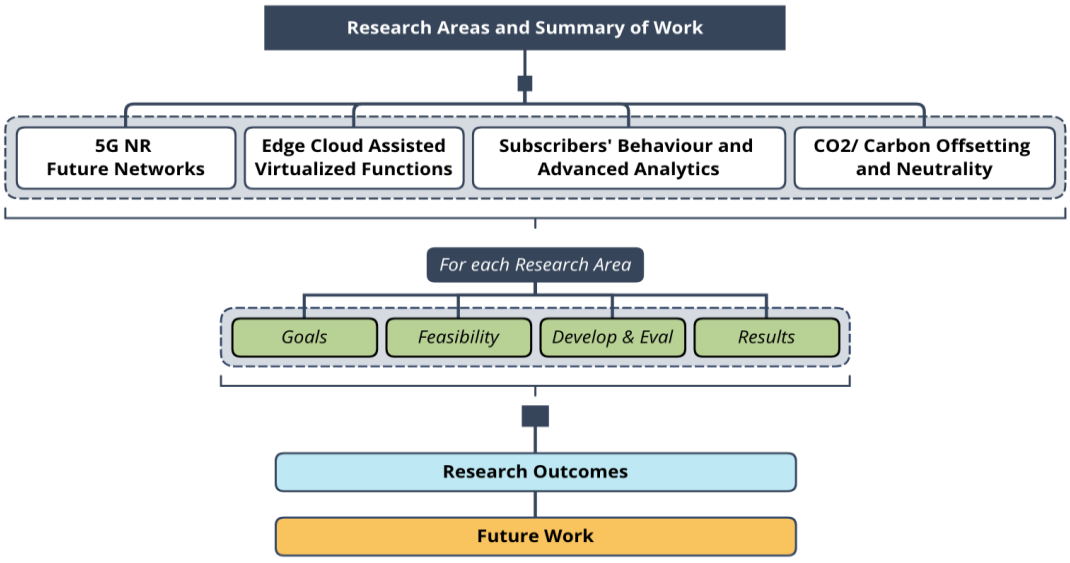


Figure 1. Summary of activities.

There are four main areas of analysis in the presented work, as it will be shown in the following sections: 5G NR future networks, focusing on optimization and enhancement of the overall planning and resource management; the usage of edge virtualized functions, which will be massively adopted in post-5G NR networks, allowing each function after disaggregation, to be individually optimized, with performance gains, leading to better services and applications which ultimately can lower overall costs; the subscribers analytics which, as will be described in this thesis and shown in the published work, is one of the most promising aspects for future research and development. Regarding this particular topic, the objective was to develop a model that, based on subscriber's behavior towards the generation of data, would evaluate its impact on overall network performance and resource management. Finally, a fourth main research area was addressed, as depicted in Figure 1, which focused on carbon neutrality, considering that 5G NR and future mobile networks are expected to increase power consumption as never before. From that perspective, the objective was to find a method that would allow for compensating that expectation, by reducing overall power consumption but, most importantly, allow for offsetting the carbon footprint of the MNO and, eventually, achieve carbon neutrality. To note that the four research areas should not be understood as siloed research, but all as a contributing part to the whole focus of the work which was the development of several new methodologies to overall improve planning and resource management in next generation mobile networks, while keeping them as green as possible.

For each of the referred research areas, goals were defined, their feasibility was analyzed, and models were developed and evaluated. This led to several results in each of the research areas, leading ultimately to research artifacts, which are the multiple international publications, listed in sub-section 1.3 and all the scientific contributions listed in subsection 1.4. Finally, a second research outcome adding to the artifacts is the list of future subjects that can be further investigated, as presented in Section 3.3.

As a summary, starting with motivation and perceived goals, the work passed onto defining four research areas, all integrated among itself contributing to the major objective of the work, as demonstrated in this thesis. From that, research artifacts and future work was generated as main outcomes, finalizing a structured approach to the whole work.

1.2. Background

In this section the most relevant background for the research questions focused on this thesis is presented. Each sub-section will focus on a specific subject that triggered the research activity and, overall, lead to the aggregated results presented in this document. First, a quick background is provided on 4th Generation (4G) and 5G NR networks highlighting the respective challenges and research questions that were addressed from a network planning perspective. Secondly, the focus is not the future mobile cellular network itself, but the way that several RATs can be integrated, resulting in a multi-RAT (MR) scenario, aiming to seamlessly improve the subscriber's QoE. The third subsection focuses on the problem of energy consumption. It provides a background on Carbon Sequestration and Storage (CSS) techniques and carbon neutrality, which has been having great attention lately. Finally, a subsection is presented focusing on the background related to subscribers' behaviour analytics, which has been minimal. It addresses how behaviour analytics has been focused on as a research problem.

1.2.1. From 2G to 5G and Beyond

It has been a couple of years since the second generation (2G) of mobile cellular networks gave its place to the third generation (3G), characterized by the first fully data-driven applications and services. From 3G to 4G or LTE, mobile cellular systems became even more centric on the subscriber, on its ability to generate or demand for high traffic volumes, either on uplink or downlink and services and applications evolved fueled by rapid advances related to end user devices' technology. 4G witnessed a boom in traffic need and, in reality, became a bridge generation to 5G and beyond, where massive disruption is expected with the past generations. 4G has paved the way to a smoother introduction of 5G, due to a progressive implementation of several methods and techniques recently put in place, especially with LTE advanced pro.

Nevertheless, 5G / 5G NR is the most advanced generation of cellular networks that have been designed up so far. Its main purpose, besides being designed with optimization in mind, is to address the progressively higher demand for massive connectivity, on several different RATs, with a multitude of devices, ranging from smartphones to sensors and machines [4][5]. Especially, as stated previously, 5G NR will enable ultra-reliable

communications and real time services, supporting critical services, enabling a wide range of remote services due to its very low latency. First 5G specifications became available in third generation partnership program's (3GPP) Release 15 and Release 16 with additional enhancements already published [6][7]. At start the main objective was simple: to provide subscribers with services that would give them the experience of extreme broadband access at their hands, through their mobile devices. The latest release enhanced the standard very relevantly, including additional features to support enhanced mobile broadband (eMBB), massive machine type communications (mMTC), e.g., internet of things (IoT) and ultra-reliable and low latency communications (URLCC).

5G NR has been developed, therefore, in order to target and meet the requirements of highly mobile and always connected data driven subscribers. To leverage that, it is notable the set of technological advances, methodologies and paradigms that have been developed, resulting in a clear disruption with previous cellular generations [8]. Although the aim of this work is not to focus specifically on 5G NR but on several enhancements, that, when aggregated, aim to optimize it even further, it is unavoidable to refer the principal 5G NR aspects that this work was focused on. First, to note the fact that this generation is the first to shift the focus from solely subscriber-centric applications to device/machine applications, with both coexisting for the first time. Other aspects like end to end network slicing are paramount in order to optimize the whole network operation and also resource management [9][10]. By considering the existence of software defined network (SDN) and also NFV the ingredients are met in order to guarantee highly efficient dense, heterogeneous and optimized beyond 5G networks, considering, for instance, the existence of base stations located in unmanned aerial vehicles (UAVs) [11][12]. But, also very important, such methodologies do not only enable high levels of optimization, but also tend to decrease overall costs to MNOs. As the cost factor is something that is considered several times on this thesis and an important aspect for also optimizing MNO's business models, it must also be noted and referred.

Additionally, another aspect which is enabled by 5G NR is the usage of dematerialized environments, supporting the disaggregation offered by NFV [13]-[16]. If there are several advantages of applying NFV and disaggregation in the core network (CN), it is also very interesting to note that such environments can also be applied at the RAN edge. As so,

Cloud-RAN (C-RAN) will set the path to additional flexibility, on RAN edge, allowing flexible and scalable network and radio resource management, for 5G NR and beyond networks [17]. By disaggregating the 5G next generation NodeB (5G gNB) in different elements, called functions, the majority can be virtualized, allowing the incorporation of different technologies, rising the difficulty on interoperability level, but lowering the overall cost and improving performance and reliability [18]. It is also particularly important for some of the aspects that this research work focuses on: densifying networks with different RATs [19][20], in order to enable higher network capacity, by aggregating different links, multihoming connections, intelligently steering traffic to the best quality RAT at any given time, reduce overall interference on the network and also, and probably, the most important part from the subscriber's perspective, allowing inter-RAT handover (HO) which are seamless and transparent at milliseconds or sub millisecond rates. Also, although not explored in this work, C-RAN can also play an important role in the transition process from older generations (e.g., LTE) to intermediate architectures (e.g., non-standalone 5G NR) and finally to 5G NR (standalone) and beyond. Another perspective, which is slightly focused on this work, is that C-RAN will also become an enabler for co-existence of different RATs in unlicensed spectrum, which is one of the most relevant aspects [21].

Finally, and because this work also focused on that part, edge computing (EC), which is now being focused on as an enabler for enhancing 5G NR's performance even further, brings higher relevance the paradigm of splitting the system into several computing and performance edges, from the device/IoT edge, to the Core edge, through as referred the optimized RAN edge. Combining cloud environments, NFV, content delivery networks (CDN) and having access/peering points at each edge will for sure be a game changer to 5G NR beyond networks.

Other concepts which have been attracting focus is the usage of data science methods for enhancing overall network operation [22]-[24]. Supervised and non-supervised learning, advanced clustering, behavior analytics and artificial intelligence, to name a few, will play a unique role in optimizing the whole 5G NR network, transforming it slowly into a partial autonomous system, the so called intelligent 5G.

Nevertheless, all of this simply shows that 5G NR has been developed in order to be optimized and has been optimized since the first release version and it will continue to. As so, that is the reason we use current 5G NR as a baseline to focus on future networks beyond 5G, where further enhancements and optimizations are expected to be developed in order to fully meet the challenges ahead.

1.2.2. Radio Access Technology Coexistence

Considering data growth as the major driver, it is now clear that several RATs will need to coexist in 5G NR but with enhanced densification in beyond 5G networks, over licensed and also unlicensed spectrum and evolving to MR [19][20]. As referred, 5G NR will be a network of networks, and MR can be a conjunction of 5G NR with LTE (non-standalone), LTE and wireless fidelity (WIFI) or, one of the main focus of this work, 5G NR with WIFI6 [25][26]. Densified 5G gNB deployments are crucial to increase cellular capacity, fulfill demand and provide the highest throughput levels for advanced services over cellular connections. WIFI, on the other hand, is the most suitable technology for "best effort" class of services and applications over unlicensed spectrum. Nevertheless, WIFI has come a long way and it features several mechanisms today that increased the reliability and quality of service of the system, making it suitable for latency and quality critical services.

Also, from 5G NR and future perspectives, existing developments regarding the usage of 5G NR over unlicensed spectrum (NR-U) will be crucial to enable additional benefits of co-existence of both RATs over the same unlicensed spectrum [27][28].

Integrating both technologies will enable seamless, transparent and pervasive communication scenarios, either outdoor or indoor locations, dimming to a higher extent the frontiers between both environments. In fact, release 15 and 16 already consider the existence of 5G NR deployed in conjunction with other RATs. In this work Certified Passpoint Hotspot (HS) 2.0 (HS2) and WIFI6 have been considered in the research work of this thesis, as presented in detail on Chapter II. WIFI6 (802.11ax) being the sixth generation of WIFI has come a long way since its previous generations [29]. It has been developed in order to provide increased performance and throughput, lower latency but, especially, increased device density [30] - [33]. This particular aspect is the main factor that explains why it has been pointed as the best RAT to complement 5G NR: the support

of massive end user devices as expected in 5G NR networks [34]. Both shall be able to unprecedently increase overall network performance, optimize radio resource management and provide higher throughput and lower latency, supporting a whole new set of services and applications that will appear.

Those particular improvements, will allow mission critical IoT devices and applications to start servicing - supported by both RATs - enabling new use-cases in manufacturing, healthcare, energy, remote sensing and many other industries [35]. Both will enable extreme low latency, unlocking a myriad of advanced applications with virtual reality and augmented reality being key enablers for industries such as healthcare, education and hospitality [36][37].

In such heterogeneous networks, where these two RATs will co-exist and densify in future networks, achieving a seamless level of integration for both technologies is not an easy task, especially when one starts focusing on inter-RAT HO. For this matter, it is crucial that HOs are performed in the most seamless and optimized manner as possible [38]. Also, offloading traffic from 5G NR to WIFI6 should be seamless and transparent for the subscribers, especially when real time, high data rate services are being used or mission critical services are being supported [39]. As so, the HO process and its triggers have become as important as never before. Contextual HOs should be performed considering the network conditions of both RATs, and might be triggered by the end user device, by the access network discovery and selection function (ANDSF) within the 5GC or as a result of combining both triggers. The main concern is to avoid performing inter-RAT HOs to lower quality connections, due to uncoordinated or poorly measured network quality, and also to avoid the ping pong effect where inter-RAT HOs are performed back and forth, completely ruining the end user's quality of experience but also, creating additional negative impact on surrounding devices, communication channels and the network as a whole [40]. These scenarios become extreme when one considers devices located at cell edges or associated to highly mobile subscribers, where the rate of needed HOs can rise relevantly. This is another aspect where EC techniques can greatly contribute to improve performance in future networks beyond 5G, which is also a research subject in this work, by deciding to place a virtualized ANDSF function in network edge [41]. Interestingly enough, although not the focus of this current work, optimization may be further increased

by carefully choosing how to place the functions into the mobile edge [42].

1.2.3. Carbon Neutrality

With 5G NR being expected already to become highly densified and heterogenous, future networks will increase even more such densification, creating additional challenges despite several energy efficiency gains and optimizations [43]: the number of smartphones, radio base stations, machines and other devices communicating will require unprecedented levels of system capacity [44][45]. In order to support such capacity increase, MNOs will see their environmental footprint become higher. Thus, MNOs would also take into consideration their environmental impact on greenhouse gas emissions and change its operational models in order to become greener [46]. Energy efficiency methods have been proposed almost on all system edges, from device to core and also RAN but such are not enough to eliminate all emissions [47] - [49]. As so, MNOs need to focus on achieving carbon neutrality, meaning that their operation should be performed in order to compensate for its carbon footprint, through carbon offset techniques.

This work focuses on carbon offsetting through biotic sequestration, precisely to avoid the pitfall of some proposed techniques, that try to address this reduction by the development and deployment of additional technology [50] - [53], which, *de per si*, contributes to increasing carbon footprint, thus eliminating any advantages that they were designed for.

The two approaches which are presented in the current thesis, address precisely the reduction of the overall carbon footprint, without contributing further to its increase. The proposed methodology is to use biotic sequestration, which an MNO can develop through its own program or contribute to existing for-nonprofit organizations created specifically with that in mind [53][54].

Although several other techniques to capture carbon exist, as it is deeply discussed on Chapter II, such do not provide the level of carbon offsetting that the proposed techniques do. Because it is our belief that this subject has not been explored enough (on the carbon offsetting part) this work aimed to reduce that gap with the proposed methods, with relevant contributions. Nevertheless, there is still large space for additional research in this subject, which we also aim to pursue in future work, as referred in Chapter III.

1.2.4. Subscriber-Centric Behavior Analytics

All the factors focused on the last subsections contribute to enhance the overall 5G NR and beyond networks' quality of service [55], but one other aspect will become increasingly important and centric on future networks: quality of experience (QoE), which is a very different concept [56]. Related to human perception of each mobile service, QoE is one of the most important concepts to consider in mobile networks nowadays and will progressively increase its importance in future networks, especially considering the different types of new services and applications, as well as previously referred, the density of users servicing on a MR context. Having the ability to measure or predict the perceived QoE, which may be very subjective, is one important challenge. Related to perceived QoE is the behavior of mobile network subscribers, which relates directly with business aspects for MNOs. Thus, it is of utmost importance to be able to characterize subscribers into behavioral segments, in order to quantitatively measure the impact of each segment on the network (regarding traffic generation capacity, for instance) and on the overall MNO's business model, from a perspective of introducing new services or discontinuing existing ones, and also by considering churn rates. From this perspective we believe that the ability to characterize and predict subscribers' behavior is fundamental and should be taken into consideration in 5G NR and beyond cellular network planning, operation and radio resource management activities. This subject is one of the most innovative contributions of this thesis, and a subject that has never really been explored as far as this research work goes. Especially considering subscriber behavior within beyond 5G NR networks, with very stringent quality, performance and latency requirements. It is, therefore, crucial to be able to tightly couple the subscriber's behavior to the impact that it has on the whole network capacity, regarding traffic generation but also from a business model perspective its impact on churn rates, which is something of utmost economic importance to MNOs [57][58].

As an example of how subscribers' impact can be very relevant, let us assume, as shown in the previous subsection, that an MNO successfully develops a carbon dioxide (CO_2) offset program, achieving carbon neutrality. In sum, it becomes the "greener" MNO in a certain market. Just by becoming that, the sole knowledge that the MNO is the "first green" of its kind, it can drive other subscribers from other MNOs – whose behavior is more

sensitive to carbon neutrality and environmental friendliness – to churn, leaving their "not so greener" MNO and becoming new clients. This will cause an immediate impact over the business model of the former MNO, but also can shift and unbalance traffic and capacity equilibrium, leading the MNO to rapidly change its network, in order to avoid service degradation and, maybe, unbalance also the carbon footprint, ceasing its carbon neutral operation. With a simple example, it is demonstrated the importance of knowing your customer for MNOs.

By having the knowledge about subscribers' behavior, cellular planning as well as service introduction and portfolio can be tailored and adapted, best suiting the users and increasing the average return per user. From a subscriber perspective, a better user-centric service will be provided thus increasing the quality of experience and overall satisfaction with the whole cellular service.

These factors will contribute to modelling user centric traffic generation capabilities and, in the end, enable a whole new approach of mobile network planning, operation and evolution, considering all of the above-mentioned aspects and business modeling. Subscriber behavior analysis was demonstrated to be a matter of utmost importance. It was demonstrated that the positive implications for MNOs are very relevant, when advanced analytics is performed: subscribers can be clustered into groups, and sometimes new groups appear from within those already analyzed, allowing for adaptations from network to service and application domains. In this thesis, the developed model for subscriber behavior analysis has shown that it is possible to plan subscriber cluster-centric network deployments, service offering and, at the same time, if behavior changes, that information can be automatically available for the MNO to act, either on planning or on resource management levels.

A whole new dimension of heterogeneous network devices and deployment types will characterize 5G and beyond networks. From that perspective, where "machines" will co-exist with subscribers, from a traffic generation and consumption perspective, the developed behavior model can be extended to internet of things (IoT), device to device (D2D) and machine to machine (M2M) communications, where behavior parameters are not human anymore, but as long as each element represents a certain traffic generation pattern, there is no need to distinguish between human or machine behavior towards data

production or consumption. From a traffic generation pattern perspective, it does not really matter "what" creates the traffic: either a human operated UE or an automatic device. Thus, one of the relevant advantages of the model is that it was generically developed, allowing it to be extended and support such "non-human" analysis within ultra-dense mobile networks post 5G NR.

At the same time, subscribers are increasing the amount of traffic they generate, creating serious capacity challenges for upcoming mobile networks. Demographic, economic and cultural are just examples of characteristics that can influence user behavior. All are considered as factors that can influence the behavior of subscribers, especially from a QoE perspective, which resulted in this thesis in the development of a user behavioral model, based on user segmentation and traffic generation abilities. As an example, based on subscribers' behavioral data, prediction methods can be applied and, for instance, better churn rate analytics can be performed, individually and per cluster, which represents, in the end, one of the most important challenges that MNOs will increasingly face in the future.

This work has taken this research an extra mile and bridged data analytics methods coming from Data Science area of expertise. As such, advanced clustering was performed over the existing data about subscribers, and it was shown that several advantages can result from bridging cellular network planning with advanced analytics. We believe that the road has been paved in order to further explore additional techniques, by applying supervised and non-supervised learning, as well as artificial intelligence, as future work, which is already being prepared.

1.3. Contributions

During the course of developing this research work, several questions have been answered by contributions resulting from all the published materials. Despite the current document focusing mainly on the already mentioned journal articles and its contributions, for the sake of completeness, no distinction will be made between the sources of contributions. Instead, all contributions along the research work are presented in this section, especially those coming from work developed and published in conferences, while all other contributions from accepted journal papers will be focused in greater detail on each subsection of Chapter II. Nevertheless, and once again, to avoid repetition, those contributions will be briefly presented.

The contributions resulting from the whole developed research are:

1. A method which allowed evaluating the best type of network deployment for 5G networks. The developed work concluded that for 5G deployments, small cells and femtocells as well as heterogeneous networks are the best solution to:
    a. Cope with massive 5G and beyond networks' traffic requirements;
    b. Reduce the overall carbon footprint of 5G and beyond networks;
    c. Minimize overall network costs.

2. A totally novel mechanism to address poor cell edge performance, which is expected to unprecedently rise in densified 5G NR deployments, but most relevantly in post-5G networks (the rise of DenseNets). The proposed approach offloads subscribers at cell edge users to existing WIFI HS2 access points (AP), based on radio link quality assessment. The results showed that:
    a. By steering cell edge user equipment's (UEs) in prioritized way, no service degradation is imposed to other UEs currently servicing closer to the 5G gNB;
    b. Poorly served UEs at cell edge will have its quality of service (QoS) improved by steering to and servicing in a better radio access technology (RAT);
    c. Negative impacts from cell-edge subscribers on overall 5G radio resource management is reduced.

3. An intelligent mechanism for RAT selection and traffic offloading, applied to MNOs that provide both cellular and broadband RAT, which is applicable to 4G, 5G NR and beyond with WIFI HS2.0 or WIFI6. This methodology is UE initiated and differs from others substantially by further using system information block (SIB) signaling, without any changes to standardized protocols, as many others have proposed in the literature. This means that the proposed mechanism allows for immediate implementation and adoption as no changes are required to existing standards. The result of this contribution is an intelligence-driven inter-RAT UE initiated steering mechanism. The results showed that:
    a. This approach can be used immediately by any MNO that deploys both RATs, without protocol changes to the network;
    b. Efficient RAT steering can be obtained at minimal cost and overall network

impact;

c. RAT steering may be performed by the UE but on a more intelligent fashion, considering several performance and real time awareness network data;
d. Enhanced global heterogeneous network capacity is increased and improved, as well as optimized resource management;
e. Overall increase of QoS and subscribers' QoE can be obtained.

4. Considering the expected densification of post 5G NR cellular networks, especially with heterogeneous 5G gNBs, a method to evaluate the feasibility of achieving Carbon Neutrality was developed, based on biotic CO_2 sequestration methods. The results of the proposed CO_2 sequestration method shown that:
    a. It is feasible for an MNO to achieve carbon neutrality by totally compensating for the greenhouse gas emissions of their network infrastructures;
    b. For the heterogeneous 5G NR deployments that resulted from the first contribution as the best deployment type, by using low energy powered 5G gNBs (e.g., Femtocells) carbon neutrality is achievable at negligible additional expenditure;
    c. MNOs can engage greener and environmentally friendly network operations by developing CO_2 offset programs using the proposed methodology.

5. In order to further enhance radio resource management, the overall network planning process, and reduce the overall CO_2 emissions, a user behavioral impact model was designed, shifting the focus from network centric to subscriber centric analytics, following the expected trend for post 5G NR networks. This work was, at the time, the first one to develop such a model and associate it with the overall planning and operation process of a cellular network, especially for 5G NR. The developed model aims to evaluate how subscribers' behavior can impact overall 5G NR network capacity. The model was developed, and four different user segments were proposed, as well as corresponding behavioral characteristics, concerning data centric traffic generation capabilities. The model was then evaluated using real traffic data. The main contribution is a traffic model that allows quantifying the impact of user behavior towards the generation of mobile data traffic over 5G NR cellular networks as a

baseline. The results have shown:

a. That traffic consumption patterns, and user segments have different impacts on overall network traffic and should be accounted for;
b. Subscriber's behavior is of utmost usefulness for MNOs to plan the insertion of new services or the removal of existing ones;
c. That the proposed model allows certainty and risk analytics, enabling MNOs to identify most stable user segments to new services or changes in existing ones;
d. The possibility of deploying prioritization mechanisms for subscribers based on their behavioral characteristics and service consumption habits, increasing overall average return per user (ARPU) but also decreasing the probability of churn rate.

6. Considering the special context of 5G NR as a baseline, an advanced subscriber behavior-centric clustering model was proposed. In order to enable advanced analytics from Data Science field of research to be applied on highly dense and complex cellular networks, a new model was developed, which resulted in identifying new subscriber clusters that are usually not considered but are increasingly important and avid data traffic consumers and producers. The results of the work have shown that:
    a. The new identified clusters represent a highly relevant group of subscribers that should be properly addressed and considered by MNOs;
    b. Advanced analytics can be applied to overall MNO business model optimization;
    c. Radio resource management optimization can be highly improved and overall operation can benefit from techniques like machine learning or artificial intelligence, clearly paving the way to future intelligent 5G NR networks.
7. The next contribution is a cloud assisted traffic steering and offloading method. Considering advanced methods for 5G NR like NFV and EC, two models were developed considering post 5G NR networks' challenges, in order to virtualize the ANDSF function and bring RAT steering decisions to initiate from the core edge, with metrics and performance analytics being performed on both device edge and also, on

a coordinated fashion, on the core edge. The classical datacenter ANDSF deployment method was compared with the two-proposed cloud-based virtualized ANDSF functions and it was shown that:

   a. The proposed methods, due to their elasticity, scalability and overall flexibility can provide enhanced performance to traffic steering decisions;
   b. Overall cost associated with resource management, inter-RAT HOs and associated mechanisms is reduced very relevantly;
   c. The proposed methods are the best future proof approach, as it considers EC which is still being widely discussed as well as WIFI6 with enhanced quality of service and integration with post-5G NR.

8. The final contribution of this research work focuses once again on carbon offsetting: this time it was considered the existence of intelligent mechanisms in 5G NR cellular networks, namely energy efficiency (EE) methods. In this advanced context, in which it is considered that the most advanced EE methods are deployed, and the focus is not solely the infrastructure but also the other 5G NR components, like the UEs, a two-tier carbon offset methodology is proposed, combining both EE and CO_2 sequestration. Overall, the results have shown that:
   a. Global CO_2 emissions can be dramatically reduced on all network edges by first applying EE methods and complementing with CO_2 sequestration methods;
   b. Carbon offset and neutrality can be achieved much faster at lower investment and operating costs, with the added value of overcoming some of the limitations of biotic sequestration;
   c. Greener operation can be maintained on all network edges, from the device to the CN edge, for higher periods of time;
   d. Overall network running costs are relevantly reduced and resource management becomes highly optimized.

All of the above contribute to the overall research objective of this work: enabling advanced but at the same time smoother 5G NR introduction. This is promoted through an analytical

perspective, focusing on subscribers' behavior and their traffic consumption patterns, the usage of EC and cloud environments, with virtualized NFs, improved inter-RAT mechanisms for seamless service continuity from the end-user perspective, especially on mobility scenarios, enhanced radio resource management and cost optimization, but also leveraging the application of data science methodologies, e.g., machine learning, classification or clustering, leading to more optimized resource allocation, capacity planning, but especially, also allowing to predict changes and future trends regarding subscribers' actions and behaviors. Such will enable analytical cellular planning methodologies, addressing expected complexities for intelligence-driven 5G NR networks. Finally, all of these aspects are properly aligned with environmental efficiency, through the application of the proposed carbon offset methodologies, leading to carbon neutrality on daily operations and finally, to greener next generation, highly densified cellular networks.

1.4. List of Publications

The work related with this thesis resulted in the publication of the following articles:

1. Gonçalves, L. C., Sebastião, P., Souto, N. and Correia, A., One Step Greener: Reducing 5G and Beyond Networks' Carbon Footprint by 2-Tiering Energy Efficiency with CO_2 Offsetting *MDPI Electronics* 2020, *9*, 464.
2. Gonçalves, L.C.; Sebastião, P.; Souto, N.; Correia, A., Extending 5G Capacity Planning Through Advanced Subscriber Behavior-Centric Clustering. *MDPI Electronics* 2019, *8*, 1385.
3. Gonçalves, L. C., Sebastião, P., Souto, N. and Correia, A., On the impact of user segmentation and behavior analysis over traffic generation in beyond 4G networks. Trans. Emerging Tel. Tech., 28: e2933. doi: 10.1002/ett.2933, 2017
4. L.C. Gonçalves, P. Sebastião, A. Correia, N.S. Souto, 5G Mobile Challenges: A Feasibility Study on Achieving Carbon Neutrality, IEEE International Conf. on Telecommunications - ICT, Thessaloniki, Greece, Vol. 1, pp. 1 - 8, May 2016.
5. L.C. Gonçalves, P. Sebastião, N.S. Souto, A. Correia, Addressing Cell Edge Performance by Extending ANDSF and Inter-RAT UE Steering, IEEE International Symp. on Wireless Communication Systems - ISWCS, Barcelona, Spain, Vol. 1 , pp. 1 - 3 , August 2014.
6. L.C. Gonçalves, P. Sebastião, N.S. Souto, A. Correia, Network Aware Traffic Steering and Selection In Heterogeneous Wi-Fi/LTE-A Networks, IEEE European Conf. on Networks and

Communications - EUCNC, Bologna, Italy, Vol. 1, pp. 1 - 2 , June 2014.

7. L.C. Gonçalves, P. Sebastião, N.S. Souto, A. Correia, Subscriber Group Behavioral Analysis for Data-Centric Service Consumption Beyond LTE-Advanced, IEEE Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace, Aalborg, Denmark, Vol. 1, pp. 1 - 5, May 2014.

8. L.C. Gonçalves, F. V. Vaz, A. Correia, P. Sebastião, Femtocell deployment in LTE-A networks: a sustainability, economical and capacity analysis, IEEE International Symp. on Personal, Indoor and Mobile Radio Communications - PIMRC, London, United Kingdom, Vol. 1, pp. 3443 - 3447, September 2013.

9. F. V. Vaz, L.C. Gonçalves, P. Sebastião, A. Correia, Economic and Environmental Comparative Analysis on Macro-Femtocell deployments in LTE-A, IEEE Wireless Communications, Vehicular Technology, Information Theory and Aerospace, Atlantic City, New Jersey, United States, Vol. 1 , pp. 1 - 5 , June 2013.

10. L.C. Gonçalves, P. Sebastião, F. V. Vaz, A. Correia, Technical, Financial and Environmental assessment of femtocell deployments in LTE-A networks, Conf. on Telecommunications - ConfTele, Castelo Banco, Portugal, Vol. 1, pp. 169 - 173, May 2013.

The following journal articles have been submitted and are being reviewed:

- Gonçalves, L. C., Sebastião, P., Souto, N. and Correia, A., Cloud-Assisted Multi-RAT Steering in 5G/Beyond and Wi-Fi6 DenseNets, (Submitted, under peer review in *MDPI Applied Sciences Journal*)

Additionally, the following Book chapter has been accepted and is in the process of being finalized for submission to EIT's *Intelligent Wireless Communications Book.* To this chapter contributes the majority of the methods and work presented in this thesis, with special emphasis on subscribers' behaviour analysis and advanced network resource management based or assisted by cloud and EC.

- Gonçalves, L. C., Sebastião, P., Souto, N. and Correia, A., Cardoso, E., Applications of Artificial Intelligence for 5G Advanced Resource Management (being finalized)

1.5. Thesis Organization

As previously referred, this document is presented on a thesis by articles format. The main reason behind choosing this possibility is that this kind of approach allows the author to dedicate additional effort in the writing and publication of high-quality scientific papers, as presented in Section 1.4 and on the following chapters [59].

The main part of the thesis, which is Chapter II, is composed of four journal articles: three of them accepted and published already, available online and each with its respective volume number attributed. One additional journal publication is under final revisions by the reviewers and have been included in the same chapter, due to its relevancy to the contributions of this work. Regarding published work, other scientific papers have been developed and presented at different international conferences. Nevertheless, all have been properly submitted to peer-review processes before acceptance for publication. In this work and in order to follow the thesis by articles structure, we have chosen to list them and focus primarily on journal articles, which still remains, nowadays, one of the best forms for communicating important scientific results [60].

As previously referred, the presented research work focused on four main subjects: first the objective was to perform gap analysis regarding current planning methodologies for both 3G and 4G and find the gaps to a more advanced planning process in 5G NR and beyond networks. The next subjects focused on enhancing the overall results from the first, and how to overcome its limitations regarding environmental aspects like greenhouse gas emissions. Each subject is coupled and tightly related to the previous one, making the whole work a continuous flow of research topics and contributions all adding up to the final result. With that perspective in mind, the objective was to make each of the main subjects publishable in peer reviewed journals. Quality instead of quantity was chosen relating to journal published works, but extensive quality work was also done and presented in conferences, part of it receiving several high-quality reviews, by the reviewers as well as by the experts attending the conferences. Nevertheless, as referred, the work was prepared in order to achieve at least, one publication of each topic on conference proceedings and one publication of each major topic on international journals. Regarding journals, this was achieved for two main subjects, with two additional ones still being under peer reviews.

To make sure that every aspect was properly presented in this thesis, it was chosen to retain each article's full version under the form of sub-chapters. That resulted in several referencing systems: one per each of the fully included papers and another one for this thesis document, referring to Chapters I and III. For each article presented in Chapter II its own structure is provided, in the specific journal structure. Its own references are provided also, exactly as published or required after revisions and proof reading. Also, regarding pagination, and in order for this thesis to be completely clear and structured as best as possible, we have chosen to slightly change each of the original articles page numbering. This way, each paper is introduced in this thesis in a smooth way, with all numbering being consistent, resulting also, in a consistent overall index.

Each of the presented articles in Chapter II, as referred, should be considered as an independent document, a work on its own, each with its own section, chapters, abstracts, references, pagination and format, according to the journal's requirements. Nevertheless, it should be retained that still, each and every one of the research articles contribute to the major thesis' goals and objectives. This approach leads to the need of creating a final chapter, developed in an umbrella-style, concluding and gathering all the works together, discussing in a macroscopic perspective the overall results of this thesis. As so, this thesis is organized as follows:

- Chapter II presents the main research activities and the main results, in the form of four journal articles. It is formed by four different sections and each represents an individual article.
- Chapter III focuses on bringing everything together with a through discussion around the several results that were attained as well as to what degree the contributions have helped to mitigate the initial research questions and problems.
- Final remarks and future work following this research are focused on the final part of Chapter III.

Page intentionally left blank

Chapter II – Articles

2.1. Article nr. #1

This article presents the first approach to the subscriber behavior modelling problem. The goal was to develop an impact model that, based on four different clusters of subscribers – each cluster with its own characteristics – would allow evaluating the traffic generation capabilities of each group. By this time, no advanced clustering was performed because the aim was solely to address its impact from network planning and capacity perspectives. This article set the path to subscriber centric concepts and analytics

The main contribution to the present thesis was the traffic impact model itself, but also the concepts lying behind it, namely that by properly characterizing subscriber's behavior, several benefits and improvements could be achieved on overall network resource management and service quality. Also, a very relevant contribution was accomplished by introducing, for the first time four particular user segments representing typical subscribers of mobile network services.

Article details:

- Title: On the impact of user segmentation and behavior analysis over traffic generation in beyond 4G networks;
- Date: February 2016 (made available in 2017);
- Journal: Transactions Emerging Telecommunications Technologies;
- Scimago/Scopus Journal Ranking: Quartile 2;
- Publisher: Wiley.

TRANSACTIONS ON EMERGING TELECOMMUNICATIONS TECHNOLOGIES
Trans. Emerging Tel. Tech. 2015; 28:e2933
Published online 17 March 2017 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/ett.2933

RESEARCH ARTICLE

On the impact of user segmentation and behaviour analysis over traffic generation in beyond 4G networks

Luís Carlos Gonçalves*, Pedro Sebastião, Nuno Souto and Américo Correia

ISCTE-Instituto Universitário de Lisboa / Instituto de Telecomunicações, Lisboa, Portugal

ABSTRACT

In this paper, a survey of the mobile data traffic growth over the last decade is performed, showing historically how data consumption patterns have evolved. Based on the results from the survey and the most important factors for unprecedented traffic growth from last years, a set of user-centric services and scenarios are identified as the most prevalent by 2020. Based on service characteristics and subscribers' behaviour and consumption regarding traffic generation, four user segments are derived. All these factors allow introducing an impact model that characterises user segment's behaviour over post-4G cellular networks, namely, its consistency and stability towards traffic generation, data consumption and service usage, providing mobile network operators with relevant information to predict user segment dispersion, stability and risk. Real and estimated market data are used to test the impact model, and most relevant results are shown. Overall, the proposed impact model aims to bring together technological, sociological and also economical perspectives into a single analytical framework, meeting both mobile network operators and subscriber's expectations of high quality of service with continuous cost decrease. Copyright © 2015 John Wiley & Sons, Ltd.

*** Correspondence**

Luís Carlos Gonçalves, ISCTE-Instituto Universitário de Lisboa / Instituto de Telecomunicações, Lisboa, Portugal.

E-mail: lcbsg@iscte-iul.pt

Received 29 September 2014; Revised 26 January 2015; Accepted 2 February 2015

1. INTRODUCTION

Cellular telephone users' behaviour has changed notably in the last decade. Nowadays, users do not rely on equipment mostly for voice communications as they used to [1–6]. A fact is that cellular telephones are globally being replaced by smartphones [2, 3]. These are advanced feature handsets giving users additional services and applications beyond simple voice calls, becoming personal agendas, entertainment systems and most recently, virtual wallets. Smartphones, more than ever, are becoming part of users' lives not only as a handset that allows seamless communication but also as a daily life manager. This not only enabled new ways of communication for mobile users but also posed new challenges to mobile network operators (MNOs) [6–9].

Mobile network operators, as more individuals are using their telephones for data services (communication, education, recreation and leisure), must assure that their cellular networks are up to the challenge, delivering these services with high performance, quality of service (QoS) and adequate traffic capacity.

However, smartphones are not the only devices that play an important role in mobile traffic generation nowadays [10–12]. Tablet computing has become increasingly popular and could easily overwhelm MNO's capacity with the usage of real-time video streaming [12–14]. This service stands out because of its intrinsic nature of requiring large amounts of data: mobile subscribers using a tablet will watch higher resolution videos due to the screen size, which they would not if using a smartphone—a low resolution video would be enough considering the screen size—creating even more data traffic load on mobile networks. Furthermore, users expect video streams to be displayed in real time, thus creating several continuous streams of data traffic on the mobile network. Several research and market studies [5, 7, 13–19] compared the amount of data traffic across multiple mobile operators and found that as networks become faster, the percentage of video crossing them greatly increases.

It is imperative that MNOs are prepared for increasingly bandwidth demand not only from a technological standpoint but also from market perspectives. MNOs should be able to profile their subscribers' behaviour and predict data consumption evolutions on their networks. Therefore, it is of utmost importance for an MNO to

- profile their subscribers into user segments relating them with the type of services they use;
- gather statistical information on data traffic generation and future trending, as well as new future services;
- predict and score the impact of user segments on their networks when deploying new technologies or increasing the availability of existing ones;
- be able to adapt its service and value offers to the customers and their data usage profile.

This work starts by presenting a historical evolutional study about data traffic consumption indicators and patterns over the last years. It also focuses on existing mobile data traffic forecasts and predicts new data traffic characteristics for 2020. Some of these aspects have recently been addressed by the METIS project, by focusing on scenarios, services, applications and technology by 2020, the same period [20, 21] we consider in our work but from different perspectives. Our predicted traffic characteristics framework enables introducing four user segments and a set of data-centric services, mirroring what we expect to be the mobile data market in the early next decade.

Finally, based on those user segments and their characteristics, a behavioural impact model is proposed and tested with real data estimates from existing market studies, allowing the quantification of the impact of each user segment's traffic usage patterns and behaviour on mobile networks. Unlike previous impact analyses, like the economical and ecological impact studies performed within the EARTH project [22], in our approach, we provide a mathematical model framework considering sub- scribers' behaviour reflected into traffic generation beyond 4G networks. A recent work [23] has bridged services, applications and traffic with some behavioural features but taking into account physiological aspects of human to device interaction, which is different from our perspective, as we focus not on user interaction with handheld devices and potential traffic generation but on subscribers' ability of generating traffic, when grouped in similar characteristics user segments.

The paper is organised as follows. Section 2 presents both historical and future evolution on data traffic generation and consumption from several perspectives: mobile data subscriptions and devices, applications and services and standardisation. Section 3 focuses on mobile data characterisation. Section 4 identifies a set of user-centric services and scenarios that we believe to be the most prevalent by 2020, responsible for generating the highest amounts of data traffic. In order to better profile sub- scribers, Section 5 introduces a set of user segments, derived from subscribers' behaviour towards data generation and consumption. Those user segments will then be used as a part of a behavioural impact model that we introduce in Section 6. This model is evaluated with real traffic data from market studies, and results are presented in Section 7. Finally, Section 7 summarises the whole work reporting the most significant conclusions.

2. DATA TRAFFIC EVOLUTION

This section provides an overview of the factors that have contributed the most for mobile traffic increase. Section 2.1 focuses on the evolution of mobile data subscriptions and Section 2.2 on the evolution of mobile Internet-enabled devices and their usage increase. Section 2.3 focuses on mobile applications and always on services and their contribution to mobile traffic generation. Finally, Section 2.4 addresses the evolution of technological standards and the way they enabled the unprecedented increase of mobile traffic.

2.1. Mobile data subscription

By the end of 2007, there were 300 million Internet subscribers in the world, through fixed broadband subscriptions with their Internet service providers. In 5 years span time, that number more than doubled reaching 620 million. Estimations predict that by 2014, the global number of fixed broadband subscriptions will rise to 940 million [1]. World universal mobile telecommunication system (UMTS) and high-speed packet access (HSPA) sub- scribers have doubled between 2008 and 2012. Estimations predict that by the end of 2014, mobile subscriptions in both technologies will rise up to 2.2 billion, surpassing the number of fixed broadband subscriptions [2]. The fact that HSPA has been a natural evolution of already- deployed UMTS networks was the facilitator factor of global subscription growth of mobile broadband. Based on existing studies and market data from 2010 to 2016, the corresponding trend was derived and expanded enabling expected HSPA subscriptions growth estimations up to 2020 [1, 5–8, 16, 17]. The results are depicted in Figure 1. These results from the performed estimations are modelled by (1) and (2), showing that UMTS subscriptions are expected to stall from 2014 onwards, while HSPA technologies will increase almost exponentially. Two equations were derived from the estimations, by regression-based analysis of market data [1, 5–8, 16, 17], modelling the number of subscribers for HSPA, S_{HSPA}, and for UMTS, S_{UMTS}, on a given current year Y_C, relative to a given start year, Y_S:

$$S_{HSPA} = [T_{HSPA} \cdot (Y_C - Y_S + 1)^{G_{HSPA}}] \cdot 10^6 \quad (1)$$

$$S_{UMTS} = \lceil T_{UMTS} \cdot (Y_C - Y_S + 1)^{G_{UMTS}} \rceil \cdot 10^6 \quad (2)$$

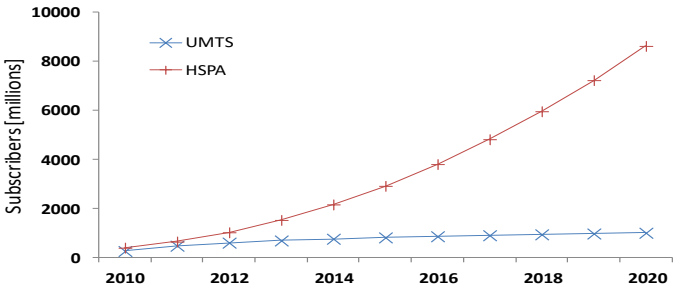


Figure 1. Estimated global subscribers' growth for UMTS and HSPA in 2010-2020 period.

The aforementioned expressions additionally consider growth indexes, G_{HSPA}, and G_{UMTS} as well as two technology factors, T_{HSPA} and T_{UMTS} respectively corresponding to HSPA and UMTS technologies. In this case and from Figure 1, $Y_S = 2010, G_{HSPA} = 1.35$, $G_{UMTS} = 0.16, T_{HSPA} = 282.5$ andT_{UMTS}=370.5. The value G_{HSPA} versus G_{UMTS} reflects a higher exponential characteristic regarding the number of HSPA subscribers when compared to UMTS, which estimations predict that will stall.

Three years from its debut, long term evolution (LTE) is estimated to have a total of 746 million subscribers, equivalent to 10 percent of the approximately 7.3 billion total mobile broadband subscribers worldwide [3].

2.2. Mobile Internet-enabled devices

The impact of wireless devices on mobile data traffic is unprecedented. Because of several technological advances in the mobile telephone area, the landscape of mobile data has dramatically changed when looking back a few years. Handset technology also evolved, with more capable equipment and higher battery capacity. This led to a new class of handsets: more powerful handsets (nowadays known as *smartphones*), tablet devices, *netbooks* and laptops equipped with mobile broadband dongles and most recently, cellular-enabled gaming consoles. Mobile devices are expected to continue benefiting from major break- throughs to occur in the next decade becoming increasingly complex and intelligent devices. This means that in the future, not only handheld devices will be connected but also other networked devices. We expect an unprecedented increase of the number of novel connectivity-enabled devices within cellular networks for the period 2015–2020, for example, Smart Televisions.

A total of 12.95 million wireless-enabled laptops, notebooks or dongles units were purchased by subscribers, bringing the total number of data-capable devices on mobile networks to 264.5 million in June 2010. All these aspects together have made cellular mobile Internet sur- pass the number of connected computers to the Internet [4]. In the period 2010–2015, worldwide share of smartphone in global mobile shipments will rise from nearly 25 per cent in 2010 to 56 per cent in 2015, representing a massive smartphone penetration rate across the globe [5]. Using the same trending and prediction methods presented

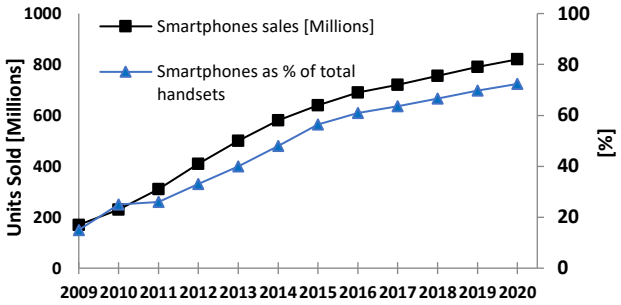


Figure 2. Smartphone sales and as percentage of world handsets.

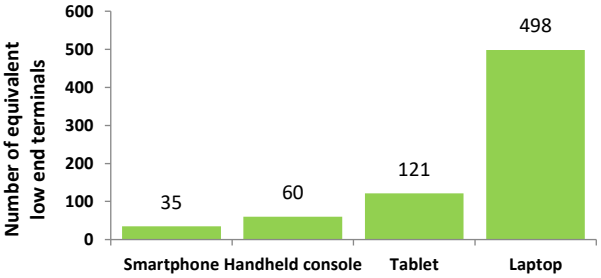


Figure 3. Traffic generation capacity relative to low end telephones (adapted from [7]).

in the previous section, our estimations show that smartphone sales will continue to grow and that by 2020, smart- phones will account for 72 per cent of global handsets, as depicted in Figure 2 [2–5, 16]. In respect to the amount of traffic, these innovative devices must be seriously considered by mobile operators, as a single smartphone or tablet is capable of generating large quantities of traffic. Figure 3 presents the equivalent monthly traffic generated by advanced mobile Internet-based devices compared with basic feature telephones. Mobile traffic generated by hand- held gaming consoles with mobile connectivity is expected to grow considerably in the 2012–2020 period. This will be highly driven by online social gaming, which is already massively adopted over the Internet and will begin shifting from fixed to mobile access.

2.3. Mobile applications and always on services

The introduction of laptops and high-end mobile handsets onto mobile cellular networks is a key driver of traffic generation, because they offer content and applications not supported by the previous generations of mobile devices. The types of mobile services used by sub- scribers are intimately related to the performance of the handheld devices they use. High-end handheld devices drive the appearance of more complex and demanding services and applications with video streaming leading traffic generation. Nowadays, existing subscriber communities generate, distribute and consume content, creating high amounts of traffic, whether in downlink (DL) or uplink (UL) directions [24]. New behaviours emerge, with subscribers starting to maximise the use of uplink connections to generate content, whether via video or audio recordings and photography uploads. Most popular applications and services will be entertainment (e.g. games, music, food, travel and sports) as well as information and daily tasks services related (e.g. maps and navigation, weather, news and banking). Figure 4 shows the estimated traffic share per service for 2016 [7]. From it, the conclusion can be drawn that traffic is switching from mobile web and data access to mobile video streaming, a real-time service with high demand of QoS. It is also shown that by 2016, video streaming is expected to account for 70 per cent of all generated mobile data traffic. This indicates that mobile user's behaviour is changing from non-real time to real-time services usage. This behaviour constitutes an enormous challenge for mobile operators, as their networks should withstand this kind of traffic demand. Data usage is also strongly influenced by the availability of applications. Application stores like Google's Play or Apple's App Store, inter alia, have changed the paradigm of available applications [12]. Both stores have already surpassed 15 billion application downloads, representing respectful traffic amounts from two perspectives: the traffic generated by download of the application itself and the traffic that the application generates while being executed. More recently, cloud technology has driven additional traffic generation, considering that many applications and services are available that way. Recent concepts like software as a service and cloud gaming are becoming extremely popular among individuals

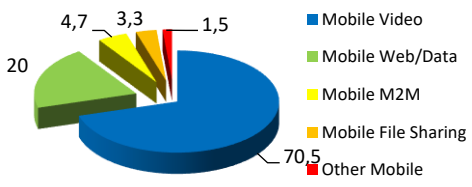


Figure 4. Expected traffic share per service [25].

and enterprises, relying solely on a robust, high quality and always on mobile connection to the Internet [25]. Cloud services and applications can also be used to overcome handset technological limitations, boosting its ability to access data content beyond its design [7].

All these aspects sustain stating that the paradigm has shifted: users are becoming themselves persistently connected and generating traffic almost continuously.

2.4. Standardisation

Standards have been defining the way technological advances cope with traffic growth. By 1999, the 3rd generation partnership programme (3GPP) Release 99 (Rel-99) defined UMTS specifications enabling more spectrum efficiency and better performing voice and data services [26]. Two years later, Rel-4 introduced call and bearer separation into the core network [27]. Rel-5 introduced high-speed downlink packet access [28], and Rel-6 added further enhancements including high-speed uplink packet access [29]. Both HSPA technologies enabled exponential growth of capacity and traffic demands and allowed mobile broadband Internet access to have enormous growth. Rel-7 defined evolved HSPA or HSPAC [30]. Rel-8 specifications defined enhancements to HSPAC technology and introduced LTE [31]. Rel-9 included additional enhancements for HSPAC but was focused on LTE enhancements [32]. We highlight Rel-9 due to the introduction of enhancements supporting home NodeB (hNodeB) and evolved NodeB (eNodeB), mostly known as femtocells. Release 10 defined a new set of enhancements towards LTE advanced (LTE-A), focusing on self- organising networks, heterogeneous networks, hNodeB and eNodeB enhancements as well as machine to machine (M2M) communications [33]. Release 11 stabilised its work on core networks by 2012 and radio access protocols and RAN performance during 2013 [34]. Release 12 is currently being finished and due to be completed by March 2015 [35]. At the same time, Release 13 is already under- way, under feature studies and is expected to be completed by March 2016 [36]. Figure 5 shows the specifications timeline from Release 4 onwards.

Overall, the releases introduced several radio resource optimisation techniques, resulting in both uplink and downlink connections with increased throughput while progressively reducing latency[†], a service quality indicator directly related to user's experience. Figure 6 presents the evolution on UL and DL data rates and the corresponding latency. From Figure 6, one should note that, for example, LTE latency is one-seventh of HSPA and one- third of HSPAC latencies, and that LTE-A nominal latency is expected to decrease to 5 ms.

[†]Latency being defined as the time data takes to transverse the network, end-to-end.

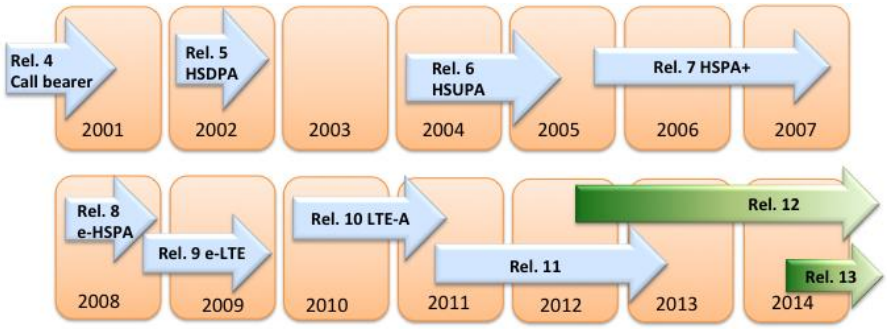


Figure 5. 3GPP specifications timeline.

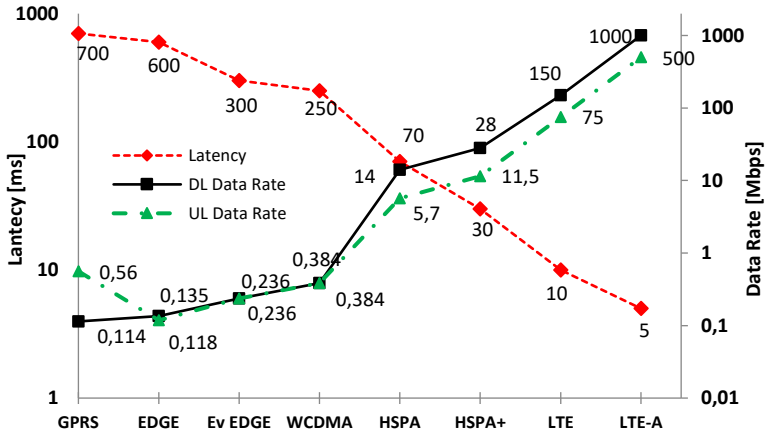


Figure 6. Latency and typical UL/DL data rates for cellular technologies.

As users began experiencing services with progressively lower levels of latency, which resulted in better service experience, they adapted their service usage to the capacity of the communications channel. This resulted in a behavioural shift from non-real time to real-time service usage. Whereas this constitutes unprecedented QoS for end users, it also means that the number of subscriptions will grow even more, and generated traffic will increase accordingly, thereby creating new capacity concerns for mobile operators. Latency has decreased rapidly the last years and will keep that tendency with high pace. We believe that at this rate, by 2015–2016, latency will drop below 1 ms, as depicted in Figure 7, allowing unprecedented user experiences and high-quality services.

Based on the several technologies' latencies, we have estimated latency per year, L_{Year}, by extrapolation, resulting in the graphic of Figure 7. The equation that models technology latency is given in (3) and considers a latency factor L_f, and yearly decrease constant L_d, in a given current year Y_C, relative to a given start year, Y_S:

$$L_{Year[ms]} = L_f \cdot e^{-L_d \cdot (Y_C - Y_S + 1)} \qquad (3)$$

In this case and from Figure 7, we find $L_f = 2579{,}6$, L_d =-0,753 and $Y_S = 1998$. The evolution of 3GPP's standards has become faster and stronger, providing significant new capabilities and features over the years. These features provide mobile

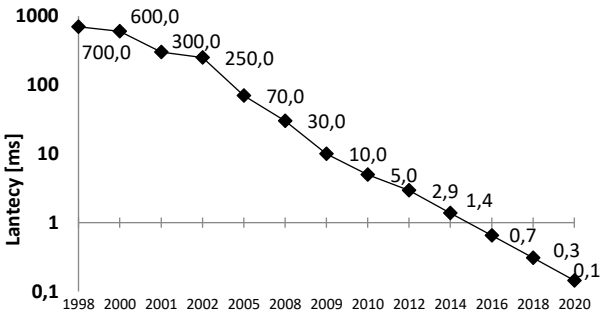


Figure 7. Latency decrease forecast.

operators with solutions for meeting fast growing wireless data usage demands of subscribers. We, therefore, believe that standards development rates have been partially driven by worldwide mobile data traffic explosion as never before.

3. TRAFFIC CHARACTERISATION

Mobile traffic characterization is a very important aspect for MNOs to account for. A historical perspective of mobile traffic evolution is of utmost importance for future planning of mobile networks' capacity. This section aims to characterize mobile traffic and data generation. Sections 3.1 and 3.2 present two different perspectives we consider the most representative: data quantification and data physical origin, respectively.

3.1. Mobile traffic quantification

By 2015, the mobile data traffic footprint of a single subscriber could be 450 times what it was 10 years earlier in 2005 [5]. Forecasts state that by 2020, Asia will represent 34.3 per cent of the total world mobile traffic, Europe 22 per cent and Americas 21.4 per cent. Figures 8 and 9 present the world mobile traffic evolution projected for the 2010–2020 decade. In the 2015–2020 period, world daily mobile traffic is expected to grow 150 per cent [5].

By 2015, wired connected devices are expected to account for 46 per cent of global IP traffic, whereas mobile devices will generate 54 per cent of global IP traffic using cellular connections [5]. From an MNO perspective, it is important to be able to predict the total monthly traffic that their subscribers will generate in the future. This gives MNOs the tools to adapt its network capacity but most importantly, adequate their service offerings in order to raise customers' satisfaction and reduce churn rates.

Another important indicator is the estimated traffic per mobile subscription, that is, per subscriber. This factor is extremely important for average return per user (ARPU) calculations. From several market studies and estimations up to 2016 [1–8, 16, 17], it is possible to perform regression-based analysis on the expected evolution on traffic generation per subscriber for the period up to 2016 and extrapolate to 2020. The results are presented in Figure 10.

The resulting equation that models world monthly estimated traffic per subscriber, $U_{MT_{Year}}$, is given in (4) considering a traffic factor T_f, and a monthly traffic constant T_i, in a given current year Y_C, relative to a given start year, Y_S:

$$U_{MT_{Year}[GB]} = T_f \cdot (Y_C - Y_S - 1)^{T_i} \qquad (4)$$

In this case, the results show that $T_f = 3{,}778$, $T_i = 1{,}4389$ and $Y_S = 2010$.

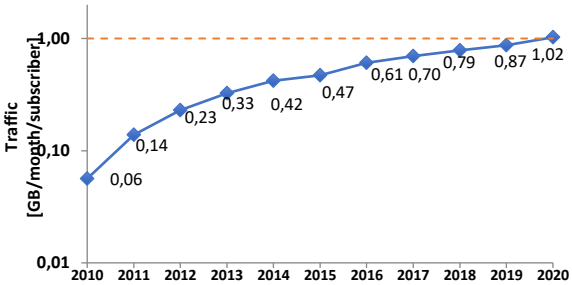


Figure 8. World estimated monthly traffic per subscriber.

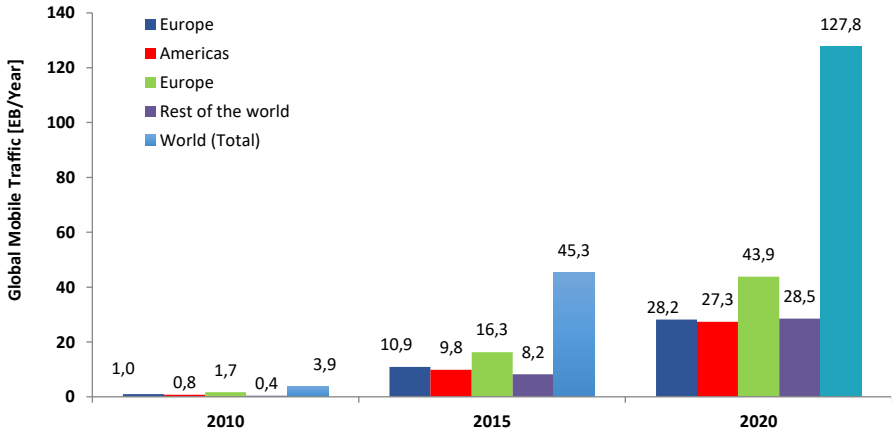


Figure 9. Estimated traffic per World Region (adapted from [7]).

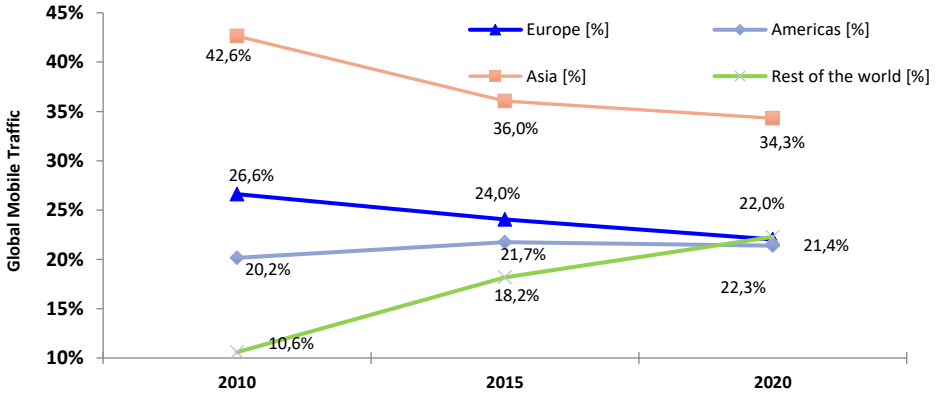


Figure 10. Global mobile traffic percentage per World Region.(adapted from [7]).

3.2. Mobile traffic origin

The major part of this traffic growth originates from indoors where cellular mobile networks are not as efficient as they are outdoors [22]. In order to achieve improved indoor radio capacity and coverage additional macrocell, base station deployment is not a solution, neither financially nor technically. This means that indoor users will remain unable to fully enjoy data capacity provided by newer standards and marketed by mobile operators. This problem is especially relevant when considering third generation and fourth generation systems, with higher frequency bands, more prone to signal attenuation. 3GPP introduced a heterogeneous mobile network environment, where outdoor macrocells coexist with indoor smaller cells, known as femtocells [37]. This concept is considered the solution for providing indoor capacity and coverage and meeting the demand for indoor data traffic, while allowing traffic offloading through existing fixed broad- band access. Other widely debated solutions exist, namely, relays but are not within the scope of the current work. Several femtocell architectures have been proposed, as well as use cases and services, addressing its challenges, advantages and disadvantages [22–24, 26, 37–39]. Additional research has been conducted on business models to allow femtocell deployment to take place maximising the average return per user while minimising capital and operational costs [capital expenditure (CAPEX) and operational expenditure (OPEX), respectively], for mobile operators, for both cases where the mobile operator or the subscribers own the femtocell [37, 40–43]. Either from a technical or a business perspective, it is of utmost importance for any MNO to have deep statistical knowledge about their subscribers' service profile and data traffic behaviour. Those statistics enable mobile operators to define new product strategies and new business models and also find what is the best period to deploy network enhancements.

4. SERVICES AND SCENARIOS

The services identified in this section are capable of generating medium to high amounts of mobile data traffic in every data transaction, either in an always on or deferred fashion. The services were chosen following user-centric behaviours. This kind of behaviour is defined by mobile data subscribers moving through different places over time, using a wide range of devices and data services and applications, adapted from [44]. Our focus is the users and their actions while using mobile data services, that is, their data traffic behaviour. We do not consider the existence of classical voice services by 2020. Instead, voice over IP data connections is assumed considering that this ser- vice is actually one of the main trends nowadays [11]. The following subsections will present the data-centric services we believe to be the most used around 2020, as well as a set of assumptions about services and technologies. Both will enable the definition of the user behavioral segments presented in Section 5.

TABLE 1. MAPPING OF DATA CENTRIC SERVICES TO QCI'S.

Data Centric Services	QCI Number								
	GBR				Non GBR				
	1	2	3	4	5	6	7	8	9
S_0						S		P	
S_1			S				S		
S_2		S		S		S	S	P	
S_3	S								
S_4		S				S	P	P	
S_5						S		P	
S_6						S		P	
S_7					S				

4.1. Data-centric services

We consider the following services to be the most prevalent in the 2016–2020 period [8, 24, 25, 45, 46]:

- s_0 – Mobile Social Networking (chat and e-mail);
- s_1 – Mobile Social Gaming (including real time cloud gaming and non real-time interactive gaming);
- s_2 – Video Streaming (including television, live broadcast);
- s_3 – Voice over data (VoIP, conversational voice);
- s_4 – Peer-to-Peer communications (including conversational video call);
- s_5 – Mobile Web Browsing;
- s_6 – Mobile commerce and banking;
- s_7 – M2M communications‡.

These services can be mapped in accordance with LTE/LTE-A's QoS class identifier (QCI) [47]. As an example, we have considered guaranteed bit rate and non-guaranteed bit rate for all services that apply, resulting in the mapping presented in Table I. Regarding premium and non-premium services and users, our mapping follows the standard, considering the existence of QCIs for premium services and premium subscribers. Premium and standard (non-premium) QCIs considered are mapped with an S and a P, respectively.

4.2. Service and technology assumptions

Derived from the existing information and trending from the last years and also considering technological pre- dictions for the near future, the following assumptions are made:

- Handheld devices and mobile Internet connections become more affordable. Data connections will be more reliable, with higher QoS, namely, lower latency

‡We consider Consumer Telematics, Vending/Kiosk/ATM, Digital Signage and mobile health the most used signalling-based applications for the consumer behavioural segments identified in this work.

and higher data rates.

- Peak data rates of 1Gbps on the downlink and 500Mbps on the uplink;
- User mobility will increase and users expect the services they use to be available anywhere and anytime;
- User experience perception similar to fixed broadband access;
- Cognitive Radio techniques are already deployed, allowing increased spectrum efficiency and channel capacity [48];
- hNodeB and eNodeB (femtocells) are common within indoor households and outdoors [49, 50]
- Cellular traffic rerouting using fixed broadband access exists;
- New femtocell deployment has no complexity associated (zero touch installation), low cost and energy consumption [51];
- Service and applications markets will keep growing at high pace, as well as mobile telephone and smartphone technologies;
- Mobile network operators have adapted their business models to content based pricing;
- Services, applications and handheld devices hardware components are optimized towards battery capacity maximisation.
- Mobile application data security, integrity and privacy is a reality;
- Mobile and fixed convergence is a reality, with single sign-on for all systems;
- Mobile connection enabled health monitoring devices will be a reality;
- Device usage and handling complexity will decrease and user friendliness will increase.
- Sensors will increasingly be embedded in more devices, systems, and infrastructures. Intelligent housing systems will exist based on M2M communications, universally available [52].
- Online social interaction will increase and users will want to personalize applications, services and control information usage;
- Transparent billing and easy payment online systems are effective and highly secure;
- Cloud services are widely adopted.

These assumptions form the basis to identify user segments and respective data behaviour identified in the next section.

5. USER BEHAVIOURAL SEGMENTS

Scenario subjects are human individuals or groups of individuals, all subscribers of a set of mobile services in accordance to their behaviour. A group of subjects from the same kind are seen as a consumer segment. For the services identified, we consider four different consumer segments. The first two segments were improved from [9] and two additional segments are introduced. Based on data-centric services and the assumption from Section 4, we summarise their usage behaviour in the near future. Figure 11 summarises the main characteristics of the four segments.

5.1. Moklofs

Moklof stands for mobile kid with many friends. This segment contains young individuals, characterised by early adoption behaviour towards technologies. These mobile subscribers are strongly focused on mobile entertainment and messaging services, as well as online social communities, which are already part of their lives, as they are always connected and available. Online gaming and social networking, video and music streaming are very popular services as well as always on cloud services. They are very sensitive to content and interactivity, thus making them mass adopters of new applications and services. This segment is known to generate high amounts of data traffic on cellular networks, mainly because of the use of streaming services and applications. Nevertheless, economic restrictions keep this kind of subscribers from adopting premium services, that is, mostly mapped to non-premium QCIs.

5.2. Yupplots

Yupplots stands for young urban people/parents with lack of time. This is a more mature segment than the Moklofs, characterised by an efficiency behaviour towards technology due to the lack of free time. Yupplots are more focused than Moklofs, preferring voice calls and telepresence applications that will bring them closer to family and friends as well as applications that will make them more efficient in their daily lives. Yupplots are characterised by using mobile commerce and banking, while they are at work, commuting or at home. Security, geographical and location awareness as well as remote surveillance and control are services these individuals are fond of allowing them to keep track of their infants or households during the day. Yupplots are not early adopters of services and applications. Technology usage must not be time-consuming, allowing them more time for social activities. Mobile work is always present: Yupplots are not entirely telecommuters as they do not tend to have nomad lives. Yupplots are fond of automatically performed tasks and automatically fed information. M2M communications are highly valued by this segment.

5.3. Supmuts

Supmut stands for senior urban people with much time. They are considered the grandfathers of Moklofs. The senior concept comes not only from age but also from technological stands.

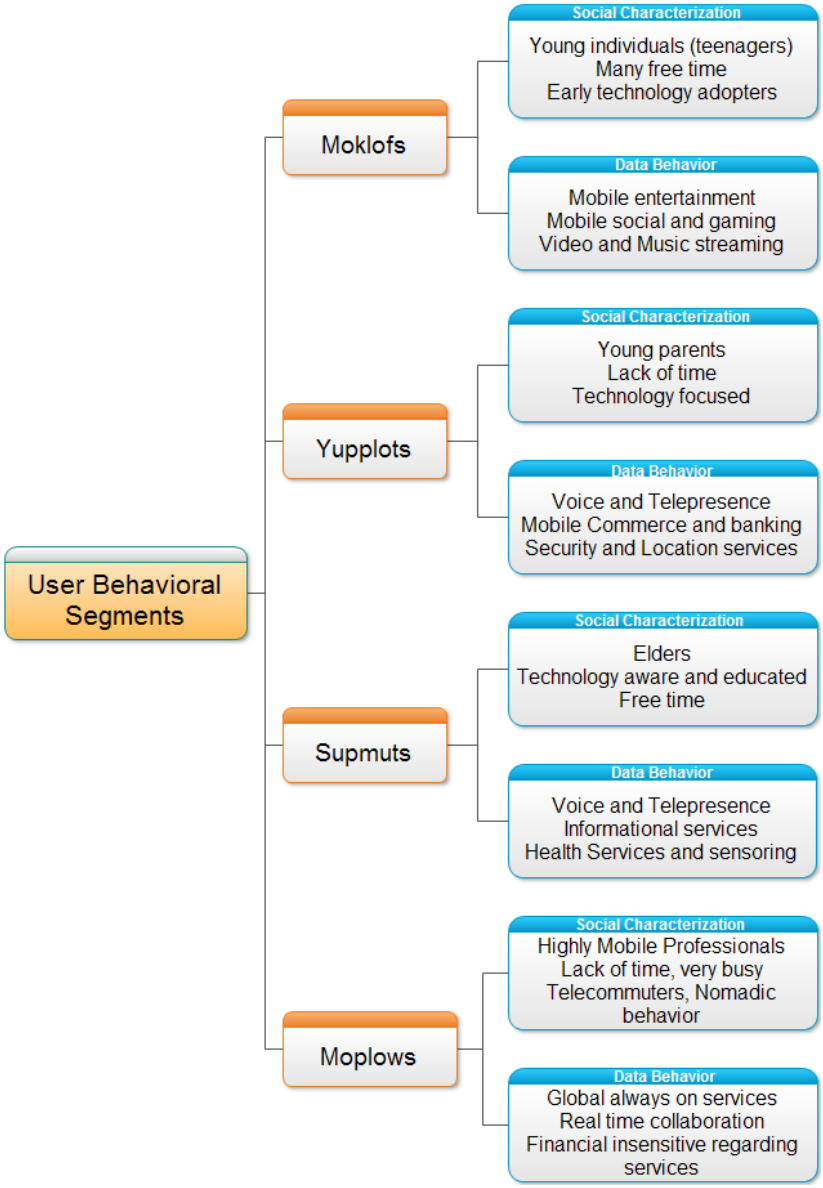


Figure 11. User behavioural segments.

Supmuts are a technologically aware and educated user segment as they followed several technologies from start all up to maturity and are much focused technology friendly individuals. *Supmuts* have free time for pleasure living and active life styling, doing what they were not allowed to when they were *Yupplots*. *Supmuts* are very sensitive to social networking services, with particular interest in video calls. Travelling and online informational services are much appreciated. One-to-one and one-to-many communication is very important to keep in touch with family and friends, with video being the predominant mean of communication. Health-focused services and applications are very important for this segment, and automatic body function sensoring is well appreciated. Videoconferencing is of vital importance because it allows *Supmuts* to be in touch with their doctors whenever they need to.

5.4. Moplows

Moplow stands for mobile professionals with much work. They are a very busy user segment, always travelling around the world, characterised by high mobility and enormous need for high-quality, robust, always available mobile services. This is a financial insensitive segment regarding technology services and applications. They require utmost quality products with reliability being the top requirement. They need high-quality up-to-date and real-time mobile information services as well as mobile commerce and payments. Mobility is the key, and *Moplows* need fast, high-capacity data connections to their home or work offices to synchronise data. These data connections may be automatic, without any human intervention, allowing *Moplows* to maximise time efficiency. As they are always travelling, personal communications with

family and friends are very important. Because financial restrictions do not apply, they have top of the line, powerful mobile equipment and demand high-quality services from mobile networks. Service-based, this kind of sub- scribers is more prone to adopt premium services, that is, mapped to premium QCIs. They are frequent users of video conferencing with family, friends and work. Global network coverage, seamless communication and high- quality and data rate services are essential for this segment. Moplows are the real telecommuters with high nomadic behaviour.

6. USER BEHAVIOUR: IMPACT MODEL

In this section, we propose an impact model that quantitatively measures the impact of user segments' data traffic generation. This method considers the user segments defined in Section 5 as well as the data-centric services from Section 4, considered the most prevalent ones in the future. In addition, and to be as generic and close to reality as possible, the method allows integrating existing real-world statistical information that can be extracted from any source. We start by defining a probability distribution function for all six services listed in Table II, extracted from [4, 7]: user behavioural segments.

$$p_s(s) = \sum_{i=0}^{I} p_i \delta(s - i) \tag{5}$$

where i is the service number, $p_j$is the usage probability of the i^{th} service and I representing the upper limit of the services index range, which for the services from Table II corresponds to 5.
The most prevalent services identified in Section IV-A form a **T** column vector representing the traffic share of each S service for a given Y year.

$$\mathbf{T_Y} = \begin{bmatrix} T_{s_0} \\ \vdots \\ T_{s_k} \end{bmatrix}, \tag{6}$$

where
$s_0 = Mobile\ Social\ Networking$
$s_1 = Mobile\ Social\ Gaming$
$s_2 = Video\ Streaming$
$s_3 = Voice\ over\ Data$
$s_4 = P2P\ Communications$
$s_5 = Mobile\ Web\ Browsing$
$s_6 = Mobile\ Commerce\ and\ Banking$
$s_7 = M2M\ Communications$

Considering that some services from Section IV-A might integrate a single service category from Table II, we define a generic probability distribution function given by the following:

$$p_{s_s}(s_s) = \sum_{k=0}^{K} p_k \delta(s_s - k) \tag{7}$$

where p_k is the usage probability of each of the k^{th} service and K representing the upper limit of the services index range which, for the services considered as most

TABLE 2. Monthly traffic share [4,7].

Service share [%]	Video	Data	M2M	File sharing	Gaming	VoIP
2016	70.5	20	4.7	3.3	1.1	0.3
2020	72	17.5	7.0	2.2	1.0	0.3

M2M, machine to machine; VoIP, voice over data.

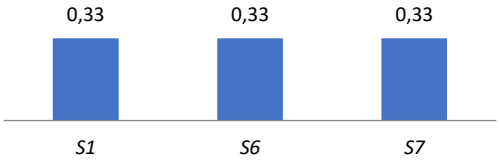


Figure 12. Service distribution for data category [percent].

prevalent in Section 4,1, corresponds to 7. Additionally, in our case, we consider services s_1, s_6 and s_7 mapped to Data category from Table II, with equal distributions (Figure 12), which was chosen for simplicity, but any other distribution for these three services can be used since the general model supports this configuration.
In order to associate services to user segments, we defined a probability distribution function per segment for all services as

$$p_{u_s|s_s}(u_s|s_s) = \sum_{l=1}^{L} p_{l|s_s} \delta(u_s - l) \tag{8}$$

where L represents the upper limit of the user segments, which is 4, and $p_{l|s_s}$ is the probability of each s_s service's traffic being generated by each user segment l. We further admit that the percentage of service usage per each user segment is split into N_l levels, given by:

$$N_l = \frac{100}{f_r} \tag{9}$$

where $f_r$is a resolution factor. For the current work, $f_r = 10$, the percentage of service usage per user segment will be quantified in multiples of 10 per cent in order to allow enough granularity on lower penetration rate services. Other values from existing studies, however, can be used. From data source [7][4], services and the respective traffic values were extracted, in TB/month:

$$\mathbf{T_{Y=2016}} = \begin{bmatrix} 6.7 \\ 1.1 \\ 70.5 \\ 0.3 \\ 3.3 \\ 6.7 \\ 6.7 \\ 4.7 \end{bmatrix} \quad \text{and} \quad \mathbf{T_{Y=2020}} = \begin{bmatrix} 5.8 \\ 1 \\ 72 \\ 0.3 \\ 2.2 \\ 5.8 \\ 5.8 \\ 7.0 \end{bmatrix} \tag{10}$$

A probability matrix is defined representing the probability of each segment using one of the T_{s_j}services, for L user segments:

$$\mathbf{P} = \begin{bmatrix} P_{(U=1,S=0)} & \cdots & P_{(U=1,S=K)} \\ \vdots & \ddots & \vdots \\ P_{(U=L,S=0)} & \cdots & P_{(U=L,S=K)} \end{bmatrix} \tag{11}$$

where,

$$\sum_{U=1}^{L} P(U,S) = 1, \qquad U = 1, \ldots L\ ; S = 0, \ldots K \tag{12}$$

In this case study, we assume the following probability distribution of the services per user segments, based on the expected traffic consumption patterns for each service [7][4]:

$$\mathbf{P} = \begin{bmatrix} 0.8 & 0.9 & 0.5 & 0.1 & 0.2 & 0.2 & 0.02 & 0.08 \\ 0.1 & 0.02 & 0.2 & 0.2 & 0.24 & 0.2 & 0.4 & 0.2 \\ 0.02 & 0.02 & 0.1 & 0.3 & 0.16 & 0.2 & 0.08 & 0.5 \\ 0.08 & 0.06 & 0.4 & 0.4 & 0.4 & 0.4 & 0.5 & 0.22 \end{bmatrix} \quad (13)$$

The first column, referring to service s_0, means that from the total traffic for that service, user segment U_1 is responsible for generating 80% of it, user segment U_2 is responsible for 10% and so on. Finally, traffic PDF is introduced using

$$p_T(T) = \sum_{l=1,k=0}^{L,K} p_{l|s_s} \delta(u_s - l) \cdot p_k \delta(s_s - k) \quad (14)$$

and compound traffic is obtained with (10) and (13), in the following way:

$$\mathbf{T_C} = \mathbf{P} \cdot \mathbf{T_Y} = \begin{bmatrix} T_{C_{(U=1)}} \\ \vdots \\ T_{C_{(U=L)}} \end{bmatrix} \quad (15)$$

7. RESULTS

The proposed impact model is generic enough to be applied to different market data. As referred, extracting real data market study data from [4, 7] allows testing the proposed model. Figure 12 shows the expected results for traffic per user segment by 2016. It can be seen that video streaming, s_2, is the service that generates most traffic among all user segments, as it is the one with higher penetration rate from the extracted estimations. From Figure 13, one can observe that traffic generation wise, all segments present similar generation capacity, focused on video streaming, and disperse according to their behaviour on other services consumption. The model shows that by 2016, *Moplows* and *Yupplots* segments are expected to account for the highest impact level regarding data traffic generation. These results are consistent when considering the fact that both segments have the highest penetration rates of high impact services, for example, video streaming and P2P communications. When comparing both user segments, *Moplows* present a higher penetration rate of mobile social networking due to their nomadic behaviour, which is categorised as the second service with the highest impact from data [7].

Interesting is the fact that *Moklof's* impact is very close to *Yupplots*. This happens because, although *Moklofs* have higher usage of real-time services than *Yupplots*, the latter compensate with high usage of M2M communications, mobile banking and commerce services. The results also show that *Supmuts* are expected to have high data traffic impact due to their intrinsic technology awareness. This derives directly from the behaviour of each segment regarding service usage levels.

One important aspect to note is the way segmentation is performed. The level of segmentation can represent an important factor when MNOs' capacity planning process is underway. If no user segmentation exists, that is, no traffic behaviours considered, by not having data on traffic consumption behaviour patterns, over-dimensioning might occur. Following over-dimensioning, increased CAPEX and OPEX will occur leading to profit decrease. If user segments are considered, a deeper analysis into traffic characteristics is possible, per service as depicted in Figure 13, for instance, allowing MNOs' deeper knowledge and more accurate prevision, enabling more realistic capacity planning. Thus, capacity over-dimensioning can be reduced.

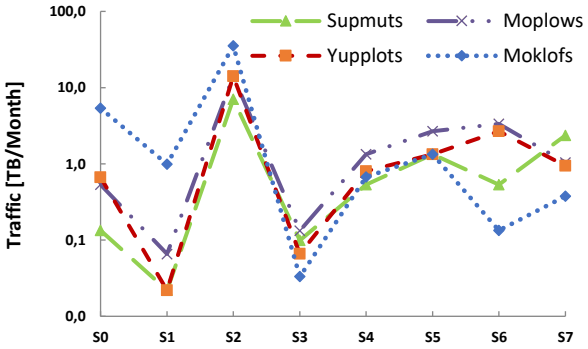


Figure 13. Expected traffic per user segment by 2016.

Another relevant aspect about user segments is that it enables MNOs to adopt prioritisation schemes based on that information, through the assignment of different levels of importance according to the behaviour of the users belonging to each segment. As an example, Moplow is one segment that can lead to ARPU maximisation, just because its users have higher economic capacity when compared with the other user segments. Thus, the MNO could choose to prioritise capacity allocation based on that information, maximising its income and the users' QoS. As referred in Section 4, the existence of premium and non-premium QCIs itself indicates the possibility of having differentiated services on the same network. One of the most interesting ways an MNO can implement differentiated services is if it knows the characteristics of their users, that is, their behaviour. By performing user segmentation according to known behaviour, the MNO should be able to perform differentiated services, thus adapting its offer, increasing the overall QoS for the users and reducing costs. The decision variable could be as simple as premium services and only two user segments: one user segment would be formed by all users who do not require premium services and all the others would form a second user group. This simple split would allow the MNO to better tailor its service offering, while reducing indirect costs of a non-optimised

capacity planning process: profit loss due to increased CAPEX and OPEX.

Segmentation depth can be higher, depending on the level of information that the MNO has about their users. If detailed data about user behaviour exists, allowing the MNO to split users into several behavioural segments, we expect traffic over-dimensioning to decrease on average. Taking into account user behaviour, capacity planning can be accomplished with a higher degree of certainty and more exact than the system without the users' behaviour.

Figure 14 shows the traffic each service generates by 2016 in a transversal way among all user segments. Once again, it is visible that video streaming and mobile social gaming are expected to be the services that generate the most and the least traffic, respectively. Mobile social gaming is expected to generate less traffic due to the real-time nature of the service, which requires low latency levels, below 1 ms, that technology does not support yet, confirming what is pointed out in [23]. Mobile Web browsing is the service that represents the most uniform behaviour among the different user segments.

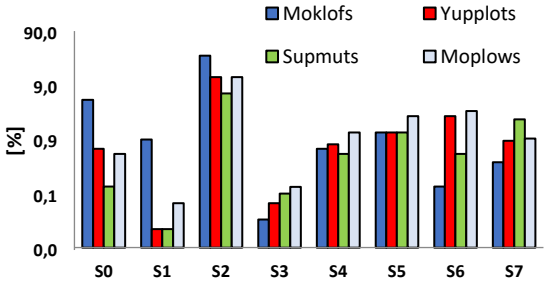


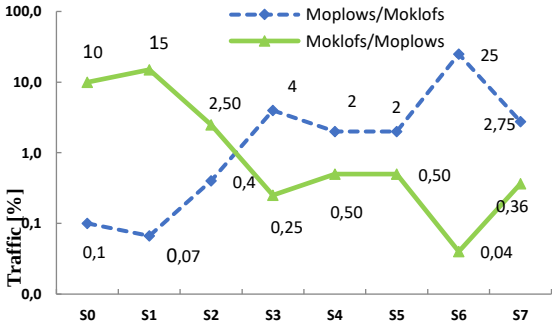
Figure 14. User segment penetration per service by 2016.

The proposed impact model is generic enough, allowing direct comparison between user segments in order to quantify existing differences and traffic generation capabilities, thus overall network impact. Doing so, Figure 15 shows the traffic generation capabilities relationship between Moplows and Moklofs, the user segments that generate the most traffic.

It can be seen that the biggest differences in traffic follow the behavior or both segments: Moklofs generate 10 and 15 times more traffic than Moplows in mobile social networking and gaming, respectively, with a clear inversion in traffic relationship when considering other services where Moplows are more focused on than Moklofs, for example, mobile commerce and banking, where the latter are expected to generate 25 times more traffic than the former.

One can even state that Moklofs' prevalence is focused on services s_0 to s_2, with s_3 being the service where both user segments generate almost the same amount of data (Moplows generate 400 per cent more traffic compared with Moklofs) and also the turning point, considering that Moplows become prevalent regarding traffic generation on all the following services.

Another analysis can be made to compare the user segment that generates the most traffic (Moklofs) with the second and last user segments. Figure 16 shows such comparison. If one wants to look deeper into direct comparison between user segments, it is interesting to notice that between *Moklofs* and *Supmuts*, mobile social networking and mobile social gaming are the two services where the gap is bigger. M2M communications and mobile commerce and banking are the two services

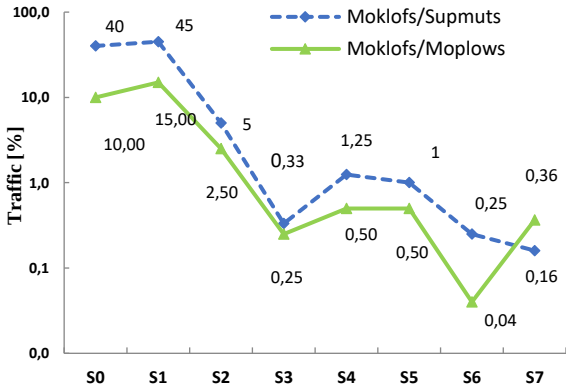


where the two user segments become closer to each other when traffic generation is concerned.

Figure 15. Traffic generation relationship between the two top user segments.

Figure 16. Traffic comparison between *Moklofs* and *Moplows* and *Moklofs* and Supmuts by 2016.

Moklofs are expected to generate 40 and 45 times more traffic than Supmuts on those two services, respectively. Additional comparison can be made between all the user segments due to the generic nature of the impact framework presented. For instance, when comparing the two segments that are expected to generate the most traffic, it can be seen that the gap between them is not as big as expected. For instance, the biggest difference between



Moklofs and Supmuts exists on mobile social gaming, with the latter expected to generate 15 times more traffic than the former. Conversely, the former generates 25 times more mobile commerce and banking traffic than the latter. The presented analysis can be extended to 2020 prediction data, as presented in Table II or any

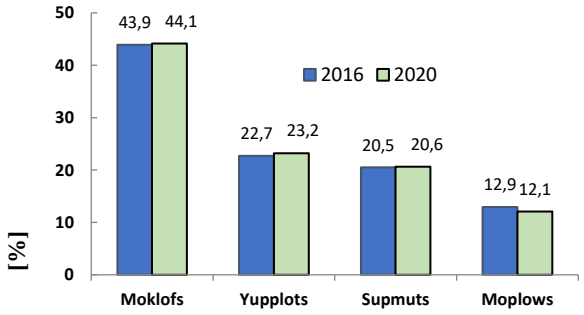


Figure 17. User segments' share of global monthly traffic.

As expected, Moklofs and Moplows are the two user segments that are expected to generate more traffic, thus creating bigger impact on the network.

Yupplots and Supmuts are expected to increase their impact level mainly because penetration rates of mobile commerce and banking as well as M2M communications are estimated to rise, with relevant monthly traffic increase. On the other hand, Supmuts are very sensitive to M2M communications—particularly mobile health monitoring applications—that explain the increase of its impact level. Moplows are the only segments that are expected to generate less traffic by 2020 (□6.2 per cent), particularly because of the usage probability decrease of file sharing [7].

Also, it is important to consider the possibility of measuring traffic spread or dispersion, by comparing the traffic generation probability of each service per user segment. Figure 18 presents traffic dispersion pattern per user segment, when considering total traffic generation. Moklofs are the user segments in which traffic quantification is more dispersed, as well as Supmuts, representing the two user segments for which average return per user is more dispersed, traffic wise. On the other hand, Yupplots and Moplows are very focused on one service, being the two where traffic generation dispersion is smaller.

Service wise, Figure 19 shows the results of traffic spread for user segments. Of all the traffics that each user segment generates, it can be seen that Moklofs have the least spread, with high levels of consistency and stability around the video streaming service. On the other hand, Moplows and Yupplots seem to be the most uncertain user segments, with higher levels of traffic spreading through the services.

This is also a very important aspect because associated with consistency is the certainty concept, meaning that a more consistent user segment (regarding its traffic generation through the usage of several services) is also more certain to maintain its behaviour, thus representing less risk to an MNO.

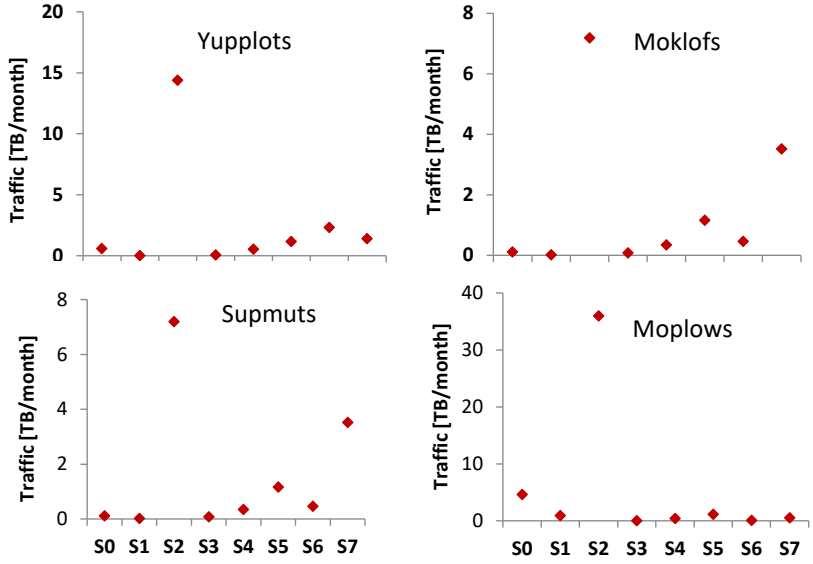


Figure 18. Total traffic spread patterns per user segment.

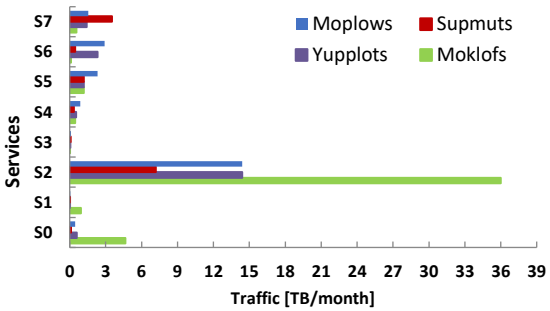


Figure 19. Service traffic spread per user segment.

It also shows that central tendency, as the main service, is video streaming for all user segments. Thus, MNOs can directly observe that resource allocation will mostly focus on this service. Conversely, user segments with higher traffic spread are more likely to be unstable, changing their traffic generation behaviour representing less certainty for an MNO, thus higher risk. This analysis is based on a measure of spread or dispersion, but if one would like to focus on central tendency measurements, the presented impact model framework supports it. The model can also be applied from an economical perspective when this kind of data exist, for example, annual financial reports from MNOs.

We identify two degrees for integration of economics into the model: on the user segment level and service levels. On the user segment side, user segments may be constructed based on economic factors, which influence the evolution of traffic patterns and user behaviour. This allows MNOs to develop multilevel analysis and optimise their network infrastructure based on demographic, period of day (busy hour and positioning of users during a day) and economic factors. From a service level perspective, keeping the user segments fixed, the MNO can consider service and application pricing extracted from internal databases and derive indicators like average ARPU or return of investment. Finally, both degrees of integration can be explored, with user segment and services and applications tailored accordingly to economic aspects. Customer driven and innovation CAPEX are two examples of indicators that can be output from the model when feeding it considering social–behavioural, technological, economic and ecological data.

8. CONCLUSION

The main factors contributing to unprecedented growth of mobile data traffic were presented. A set of data-centric services were identified as the most prevalent in the 2016–2020 period, based on existing studies and several technological and behavioural assumptions. Those services were mapped to traffic consumption estimations extracted from existing market studies. Two adapted and two new user segments were introduced and profiled according to subscribers' behaviour towards mobile data traffic generation and used services. The number of user segments and their characteristics can be tailored depending on behaviour similarities or differences, demographic and economic data as well as traffic consumption patterns that MNOs might possess. Nevertheless, it should be considered that high group segmentation could result in core network's performance degradation (e.g. additional delay and processing needs). However, we believe that detailed group segmentation results in more efficient subscriber behaviour matching, and that future cellular communication systems' resource allocation and management should consider subscriber behaviour and respective segments, thus improving subscribers' QoS and general satisfaction. From an MNO perspective, this can lead to a churn rate reduction and ARPU increase especially when considering optimisation techniques for packetisation, which allow an increased overhead reduction [53]. With traffic and user segment characterisation, an impact model was derived, in order to account for social, technological and economical aspects of traffic generation and consumption into a single analytical framework. The model was developed to be as generic as possible, allowing incorporating data from market studies or from existing MNOs' databases. The introduced model was applied to existing and estimated market study data for 2016 up to 2020, respectively, in order to estimate user segment impact, and several results were drawn. User segments' consistency and stability analysis was performed based on the measure of user traffic spread over the considered services. It was also discussed that just by considering user behavioural information regarding traffic-type generation, and grouping users into two different groups, capacity planning over-dimensioning could be reduced, and segment-based service differentiation and resource allocation prioritisation can be performed.

Conceptual analysis like certainty and risk per user segment can be extracted from the results, allowing to identify the most stable ones (those who present most certainty in their traffic generation behaviour) and, consequently, less risky to sudden traffic behaviour change, allowing an MNO to adapt its service offering. MNOs can highly bene- fit from knowing the network impact of each user segment, in order to adapt their business and service provisioning as close as possible to subscriber's characteristics and behaviour. Although it is not the scope of this work, the proposed model also supports economic analysis, providing that such data are available. By having all subscribers' segments well-characterised, their impact quantified and service offering adapted to their behaviour as a result of applying the impact model, operational and capital expenditure analysis can be performed, allowing MNOs to reduce those expenditures, maximise the return of investment, as well as bring subscribers additional advantages as customised services and cost reduction.

ACKNOWLEDGEMENTS

Trans. Emerging Tel. Tech. 2017; 28:e2933 © 2015 John Wiley & Sons, Ltd. DOI: 10.1002/ett

This work was partially supported by the FCT-Fundação para a Ciência e Tecnologia (projects ADIN-Advanced PHY/MAC Design for Infrastructure-less Networks PTDC/EEI-TEL/2990/2012, NEUF – LTE Advanced Enhancements using Femtocells PTDC/EEA TEL/ 120666/2010 and PEst-OE/EEI/LA0008/2013).

REFERENCES

1. IDATE. *Consulting and Research: DigiWorld Year-Book 2011*, Technical Publication. IDATE: Montpellier, France, 2011.
2. GSA. *Global Mobile Suppliers Association: Global Mobile Suppliers Association—GSM/3G Market/Technology Update*, Report. GSA, Global mobile Suppliers Association, Sawbridgeworth, UK, 2012.
3. Global mobile Suppliers Association: Status of the LTE Ecosystem, Report, 2012.
4. Signorini E. Mobile Broadband Connected Future: from Billions of People to Billions of Things, Executive Report. Yankee Group Research, Inc., Boston, Massachusetts, United States, Oct. 2011.
5. UMTS Forum. Recognising the Promise of Mobile BroadBand. UMTS Forum: London, UK, 2010.
6. UMTS forum. Mobile traffic forecasts 2010–2020, Report no 44, UMTS Forum, London, UK, 2011.
7. Cisco Systems Inc. Global Mobile Data Traffic Fore- cast Update 2011–2016. Cisco Systems, Inc.: San Jose, California, United States, 2012, 2011–2016.
8. Signorini E. Mobile Broadband Connected Future: From Billions of People to Billions of Things, Executive Presentation. Yankee Group Research, Inc., London, UK, 2011.
9. Lindgren M, Jedbratt J, Svensson E. Beyond Mobile: People, Communications and Marketing in a Mobilized World, Palgrave Macmillan, 2002.
10. 4G Americas: The Evolution of HSPA- The 3GPP Standards Progress for Fast Mobile Broadband Using HSPA C, White Paper, 2011.
11. Koekkoek H. Distimo Publication Full Year 2011, Technical Publication. Distimo: Utrecht, The Netherlands, 2012.
12. Distimo Inc. Distimo Publication Full Year 2011. Distimo: Utrecht, The Netherlands, 2012.
13. IDATE. The Digital World' s challenges. DigiWorld Yearbook 2010. IDATE: Montpellier, France, May 2010.
14. IDATE. The challenges of the Digital World. DigiWorld Yearbook 2011. IDATE: Montpellier, France, Jun. 2011.
15. 4G Americas. Mobile Broadband Evolution: 3GPPRelease 10 and Beyond—HSPA C, SAE/LTE and LTE Advanced, White Paper. 4G Americas: Bellevue, Washington, United States of America, 2011.
16. Cisco Systems Inc. Cisco Visual Networking Index: Forecast and Methodology, 2011–2015. Cisco Systems, Inc.: San Jose, California, United States, 2011, 2011–2015.
17. UMTS Forum. Mobile Traffic Forecast. UMTS Forum: London, UK, 2011, 2010–2020.
18. IDATE. The challenges of the Digital World, DigiWorld Yearbook 2012. IDATE: Montpellier, France,2012.
19. OFCOM. International Communications Market Report 2011. OFCOM: London, UK, 2011.
20. Osseiran A. et al. Scenarios for 5G mobile and wireless communications: the vision of the METIS project. IEEE Communications Magazine 2014; 52: 26–35.
21. Monserrat J. et al. Rethinking the mobile and wireless network architecture: the METIS research into 5G. In IEEE European Conference on Networks and Communications, Bologna, Italy, 2014; 1–5.
22. Energy aware radio and network technologies. EARTH, Economic and Ecological Impact of ICT, Deliverable D2.1, 2011.
23. Fettweis GP. The tactile internet: applications and challenges. IEEE Vehicular Technology Magazine 2014; 9: 64–70.
24. IDATE. Mobile Payments and Mobile Money 2010.
25. CORDIS European Comission. Research in Future Cloud Computing, Expert Group Report, 2012.
26. 3GPP. "*Overview of 3GPP Rel. 1999"*, V0.1.1, 2010.
27. 27. 3GPP. "Overview of 3GPP Rel. 4", V1.1.2, 2010.
28. 3GPP. "Overview of 3GPP Rel. 5", V0.1.1, 2010.
29. 3GPP. "Overview of 3GPP Rel. 6", V0.1.1, 2010.
30. 3GPP. "Overview of 3GPP Rel. 7", V0.9.6, 2012.
31. 3GPP. "Overview of 3GPP Rel. 8", V0.2.8, 2012.
32. 3GPP. "Overview of 3GPP Rel. 9", V0.2.7, 2012.
33. 3GPP. "Overview of 3GPP Rel. 10", V0.1.6, 2012.

34. 3GPP. “Overview of 3GPP Rel. 11”, V0.1.2, 2012.

35. 3GPP. “Overview of 3GPP Rel. 12”, V0.0.5, 2012.

36. 3GPP. “Overview of 3GPP Rel. 13”, V0.0.6, 2014.

37. Markendahl J, Mäkitalo O, Werding J. Analysis of Cost Structure and business model options for wireless access provisioning using femtocell solutions. In 19th European Regional ITS Conference, 2008.

38. Tyrrel A, Zdarsky F, Mino E, Lopez M. Use cases, enablers and requirements for evolved femtocells. In IEEE 73rd Vehicular Technology Conference, Budapest, Hungary, 2011; 1–5.

39. 3GPP TS 25.467. “UTRAN architecture for 3G home node B (HNB)”, V10.5.0, 2012.

40. Lin P , Zhang J, Chen Y , Zhang Q. Macro-femto heterogeneous network deployment and management: from business models to technical solutions. IEEE Wireless Communications Magazine 2011; 18: 64–70.

41. Chandrasekhar V , Andrews G. Femtocell networks: a survey. IEEE Communication Magazine 2008; 46: 59–69.

42. Shetty N, Parekh S, Walrand J. Economics of femtocells. In IEEE Global Telecommunications Conference, Honolulu, Hawaii, 2009; 1–6.

43. Vezin J, Giupponi L, Tyrrell A, Mino E, Miroslaw M. A femtocell business model: the BeFEMTO view. In Future Network and MobileSummit 2011 Conference, Warsaw, Poland, 2011; 15–17.

44. Tafazolli R, Sasse A. Technologies for the Wireless Future: Wireless World Research Forum (WWRF), Vol. 2. John Wiley & Sons: London, 2006.

45. Distimo Inc. Distimo Publication August 2012, Aug. 2012.

46. BeFemto. Description of baseline reference systems, use cases, requirements, evaluation and impact on business model, Deliverable 2.1.

47. 3GPP TS 23.203. “Policy and charging control architec- ture”, V13.2.0, 2014.

48. Alti A, Laborie S, Phillipe R. Dynamic semantic-based adaptation of multimedia documents. Transactions on Emerging Telecommunications Technologies 2014; 25: 239–258. DOI: 10.1002/ett.2677

49. 3GPP TS 22.220. Service requirements for home node B (HNB) and home eNode B (HeNB), V11.5.0, 2012.

50. 3GPP TR 23.830. Architecture aspects of home nodeB and home eNodeB, V9.0.0, 2009.

51. Morosi S, Piunti P, Re E. Sleep mode management in cellular networks: a traffic based technique enabling energy saving. Transactions on Emerging Telecommunications Technologies 2013; 24: 331–341. DOI: 10.1002/ett.2621

52. Machine-to-machine: an emerging communication paradigm. Transactions on Emerging Telecommunica- tions Technologies 2012; 24(5): 494–495.

53. Torres CP, Fitzek F, Lucani E. Network coding is the 5G key enabling technology: effects and strategies to manage heterogeneous packet lengths. Transactions on Emerging Telecommunications Technologies 2015; 26: 46–55. DOI: 10.1002/ett.2899

2.2. Article nr. #2

This article presents an advanced method for performing behavior analysis over 5G NR dense and heterogeneous networks, by the application of advanced clustering techniques.

The main contribution to the present thesis was to uncover the existence of two new clusters of subscribers that contribute very relevantly to unbalance the amount of traffic that can be generated as a result of their behavior. In terms of Business Intelligence but, especially 5G NR cellular network planning and resource management, such gap constitutes an important aspect to be addressed. Other relevant contributions were made, namely related to paving the way to apply advanced data science techniques to the information that MNOs possess regarding its subscribers' behavior towards data generation. Regarding network, capacity, and resource management planning processes, it is the first time that advanced clustering and such groups have been considered in the context of 5G NR optimal planning and resource management. It is also demonstrated that there are extensive advantages for both operators and subscribers by performing advanced subscriber clustering and analytics.

Article details:

- Title: Extending 5G Capacity Planning Through Advanced Subscriber Behavior-Centric Clustering;
- Date: November 2019;
- Journal: Electronics;
- Scimago/Scopus Journal Ranking: Quartile 1;
- Publisher: MDPI.

electronics

MDPI

Article

Extending 5G Capacity Planning Through Advanced Subscriber Behavior-Centric Clustering

Luís Carlos Gonçalves [1,2,*], Pedro Sebastião [1,2], Nuno Souto [1,2] and Américo Correia [1,2]

[1] Technology and Information Science Department, ISCTE-Instituto Universitário de Lisboa, Av. Forças Armadas, Lisboa 1649-026, Portugal; pedro.sebastiao@iscte-iul.pt (P.S.); Nuno.Souto@lx.it.pt (N.S.); americo.correia@iscte-iul.pt (A.C.)

[2] Radio Systems Group, Instituto de Telecomunicações, Av. Forças Armadas 1649-026, Lisboa, Portugal

* Correspondence: lcbsg@iscte-iul.pt; Tel.: +351-213-130-991

Received: 25 October 2019; Accepted: 14 November 2019; Published: 21 November 2019

Abstract: This work focuses on providing enhanced capacity planning and resource management for 5G networks through bridging data science concepts with usual network planning processes. For this purpose, we propose using a subscriber-centric clustering approach, based on subscribers' behavior, leading to the concept of intelligent 5G networks, ultimately resulting in relevant advantages and improvements to the cellular planning process. Such advanced data-science-related techniques provide powerful insights into subscribers' characteristics that can be extremely useful for mobile network operators. We demonstrate the advantages of using such techniques, focusing on the particular case of subscribers' behavior, which has not yet been the subject of relevant studies. In this sense, we extend previously developed work, contributing further by showing that by applying advanced clustering, two new behavioral clusters appear, whose traffic generation and capacity demand profiles are very relevant for network planning and resource management and, therefore, should be taken into account by mobile network operators. As far as we are aware, for network, capacity, and resource management planning processes, it is the first time that such groups have been considered. We also contribute by demonstrating that there are extensive advantages for both operators and subscribers by performing advanced subscriber clustering and analytics.

Keywords: 5G; advanced clustering; behavior modelling; capacity planning; intelligent 5G; subscriber centricity; subscriber clusters; resource management

1. Introduction

Smartphones and tablets have become very convenient end user devices that can replace several other devices, providing a multitude of multimedia

functionalities that are no longer limited to specific people, occupations, or social status. On the other hand, the pervasiveness of smartphones and tablets have transformed them almost into children's toys, with small children using them mainly to watch videos. The unprecedented availability of new services, new data rates, and applications, with the introduction of 5G, implies that network operators must be prepared and plan their networks according to expected capacity demand. Nevertheless, mobile network operators (MNOs) might not have the chance to exhaustively test the introduction of new services and applications along with any eventual capacity exhaustion that might happen. Therefore, especially considering densities that are expected both on network and subscriber planes, it is complex to analyze both planes' behavior against the capacity that needs to be guaranteed. 5G's most prevalent "use-cases" are enhanced mobile broadband (eMBB), ultrareliable and low-latency communications (URLLC), and massive machine type communications (mMTC); this work focuses on eMBB.

Thus, it is of utmost importance that 5G network planning processes start using advanced analytics and focus on subscriber's behavior towards traffic generation. However, 5G ultra dense networks have other challenges such as low latency requirements, meaning that not only subscribers' behavior must be considered but also non-human devices, such as the Internet of things (IoT) [1]. This is where advanced subscriber clustering comes in, which is the basis for defining the research problem in this work—how can capacity planning and even network planning be supported by using knowledge about subscribers? Also, will such knowledge be an enabler in better helping the planning and resource management processes in 5G, as well as in providing deeper insights into subscribers' traffic consumption habits, resulting in overall service improvement and better radio resource management? This work focuses on human behavior analytics and its impact on mobile network capacity.

Nowadays, advanced data science techniques are becoming mainstream and data analytics has never before been such a focus. Therefore, new data is being generated every day, and mobile networks are not an exception. With 5G and beyond, data from subscribers that characterizes them will increase as never before. Thus, there is a clear advantage of using that data and applying it to develop subscriber-centric clustering on the basis of behavior that will impact very positively on both capacity network planning process and resource management.

Furthermore, from another perspective, it will be shown that there are new types of subscribers using mobile networks, and that potential new clusters can appear when previously not expected. As an example, in this work, we show that in comparison with [2], simply by having additional data that was not available at

the time, changes must be made to the clustering set that characterizes subscribers. Today, subscribers aged from 0 to 12 years old are increasingly using tablets and smartphones to access high capacity demanding services such as video streaming and online cooperative gaming [3,4]. Additionally, up until recently, the majority of data on subscribers' behavior mainly focused on ages starting at 18 years, as this is considered the legal age to respond to surveys, which are the typical instruments used to measure adherence to cellular services [5–9]. We, on the other hand, focus on very young age groups, the same area as sociology and psychology studies are currently focusing upon, looking for behavior and mental disorders due to excessive handheld devices usage. In our case, we look at such data from an MNO perspective.

In this paper, our main contribution is twofold: first, we demonstrate the advantage of considering subscriber-centric clustering based on behavior both in terms of capacity network planning process and resource management. Our second main contribution is the definition and characterization of new subscriber clusters comprising subscribers aged from 0 to 12 years old, whereas the majority of studies mainly focus on ages starting at 18 years old. The objective of the current study, extending from our previous study, is how it can be demonstrated that advanced clustering based on behavior can, in fact, be a very useful tool for MNOs and how also, due to behavioral changes, former approaches need to be re-fit, making this an ongoing process. As a result of our objectives, it is shown that parents that have children and let them use their smart devices represent a pattern change in traffic consumption and, therefore, must be considered for analysis. This represents a new cluster of subscribers, and thus new traffic generation capabilities, and as so allows us to present our approach to quantify the impact of new subscriber cluster on network traffic generation and performance, as well the challenges that this represents for MNOs.

2. Subscriber-Centric Clustering

We start by defining the concept of market, which, in cellular networks, consists of existing and potential buyers/consumers of mobile products and services. It is beyond the scope of this work to focus on products (e.g., type of mobile end user device), as our model intends to focus on behavioral aspects of service adoption and its corresponding impacts over cellular networks and capacity planning. It is also beyond this work to explore the several clustering techniques from data science fields of expertise.

The concept of cellular market is defined as a set of mobile services that subscribers are willing to buy in order to satisfy individual needs. Such action can be satisfied through an exchange relationship with the MNO. Such needs can be

different and vary according to demographic and behavioral aspects, thus requiring advanced analysis techniques such as subscriber segmentation in order to group them and best tailor and fit the service offering. In that perspective, we propose that advanced segmentation techniques should be applied to subscribers, not only on a post-factum perspective (sell the best service to the subscribers that most treasure it) but on a pre-factum perspective, which will use such information in order to enable different, more optimized, and intelligent cellular planning approaches based on deep knowledge about the subscribers.

Advanced segmentation techniques will go beyond the usage of simple demographic data—which is currently the class of data that is most used—such as age and income. It can even be applied to service characterization but also extended to a higher abstraction level of non-human devices, as long as its traffic generation pattern is known [10]. Advanced segmentation systems will allow subscriber-centric clustering and provide at least two main advantages: giving MNOs the ability to enhance the quality of service among subscribers that are much more sensitive to it, and also to increase, for example, the average return per user (ARPU) associated with each of the subscriber clusters. Figure 1 presents the subscriber-centric clustering process proposed in this work, which will be further detailed.

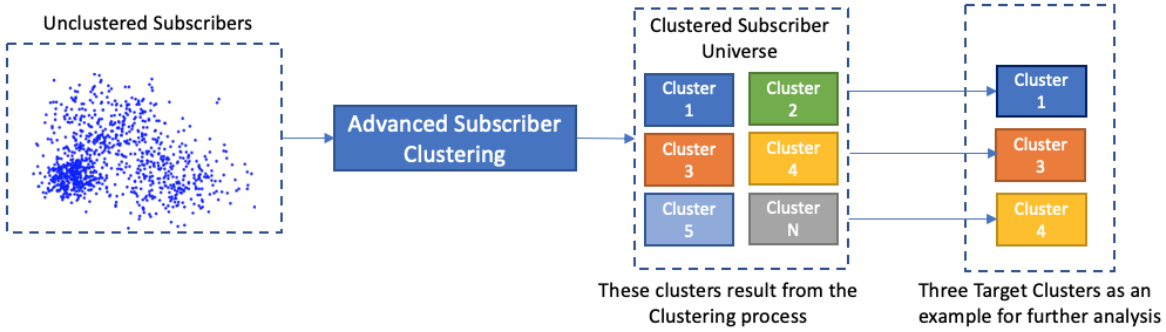


Figure 1. Subscriber-centric clustering process.

2.1. Subscriber Clustering

Subscriber segmentation or clustering can be referred as the process of splitting the subscriber base, aiming to provide deeper insights into the landscape of the customer market, as seen in Figure 1. Such a process reveals subscribers' characteristics that will enable grouping them into clusters that have one or more service or behavior in common. Such processes are known as clustering algorithms. By applying such advanced techniques, the MNO has the ability to proper analyze each group and tailor their planning process or service offering according to the characteristics that mostly define the cluster, as it was performed

by [11] using temporal analysis. The process of advanced clustering is essential and of utmost importance—if clustering is not well performed, the proper definition of a cluster of subscribers is unlikely to be successful, thus misleading the MNO during the analysis and planning processes. This is a process that has not been substantially explored in this area in question, even though there is some previous work [12,13] that can be found about techniques to cluster subscribers. As mentioned, the scope of this work is not to propose a novel clustering technique, but to demonstrate how behavioral clustering can be of utmost importance for MNOs, enabling the application of advanced analytics.

2.2. Subscriber Cluster

A subscriber cluster is a subset or a segment that results from the process of subscriber clustering. It is a sub-group or cluster of subscribers that is identified as the result of exercising advanced segmentation techniques and that has similar needs, service, and purchase behavior. In order for an MNO to successfully consider each cluster into its planning process, the cluster needs to have as many characteristics as possible, and, also, from a post-factum perspective, needs to be reachable and accessible as much as possible.

After applying the processes of advanced clustering, each MNO will have a set of clusters that will be fed into the whole network capacity planning process and resource management. In order for the MNO to properly develop its planning process, the clusters need to be characterized as best as possible, especially in terms of size and characteristics, allowing variables such as profitability to become part of the pre-factum process. After having assessed the substantiality and measurability capacities of each cluster, as well as other characteristics that will be further presented on the current work, MNOs will have clusters that are not totally distinguishable, and behavior overlapping is expected to exist with high probability.

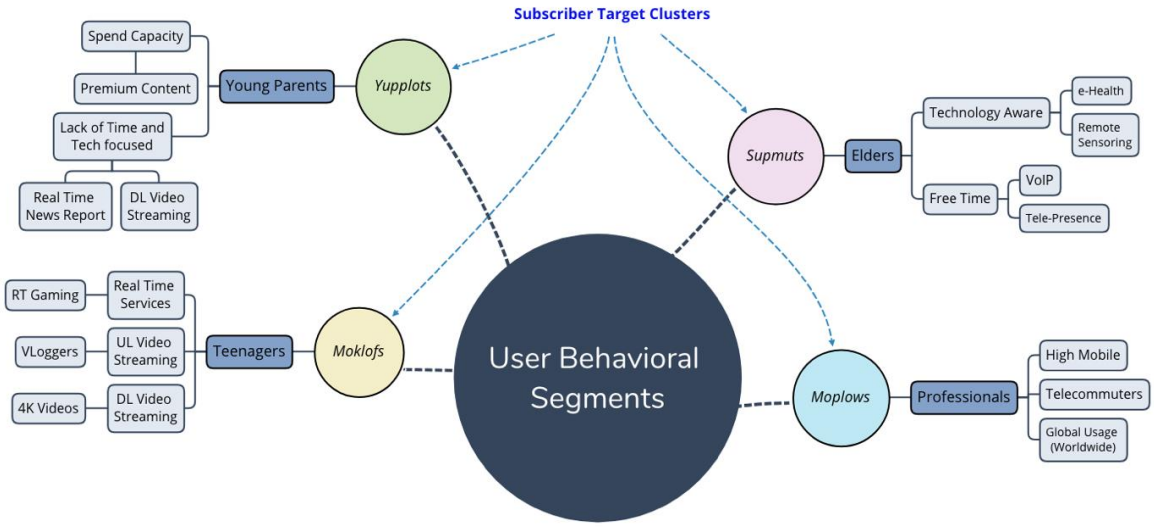


Figure 2. User behavioral clusters and resumed characteristics: Real Time (RT), Downlink (DL), Uplink (UL), Voice over IP (VoIP).

All resulting clusters should allow the MNO to distinguish them apart [13], allowing it to respond and plan the aspects of the whole network differently. Considering the work previously presented in [2], the concept of subscriber segments relates to user behavioral segments terminology, as depicted in Figure 2. These may constitute the main four subscriber clusters this work will extend upon, considering the specifics of 5G and advanced behavior-centric clustering.

2.3. Subscribers Target Cluster

A subscriber target cluster is one of the several clusters that the MNO identifies as being prone to the development of specific planning. For example, a target cluster where video streaming is the most used service will require special attention from the MNO in terms of capacity and network planning, as a consequence. After having applied advanced clustering techniques to their customer base and therefore having a set of clusters identified, the MNO must consider each one of them as a target cluster.

Also, from a post-factum perspective, such a group may become a target group to introduce additional or related services to in order to increase ARPU or, from a quality of experience perspective, should become a target cluster to focus on in order to maximize quality of experience (QoE) and likely reduce the churn probability among such a group's subscribers.

Another interesting example is the introduction of new services. How can an MNO define the best part of its network to deploy the first 5G testbed? Also, which subscribers would be more willing to immediately adhere to the new services?

This is an example of how advanced clustering and extracting deep knowledge about customer behavior and characteristics can help an MNO plan to deploy and manage its network. Potential target clusters that can be focused on are represented in Figure 3, showing where the advanced analytics can extract more information from existing unknowns.

2.4. Expected Benefits of Subscriber-Centric Clustering

From both technical and financial perspectives, by applying the referred technique, a set of benefits can be expected. Subscriber-centric clustering allows more optimized and less costly planning processes for an MNO and overall resource management optimization. This can lead to global reduction of capital expenditure (CAPEX)—investment—and, in particular, operational expenditure (OPEX)—maintenance and support—by having the ability—with the flexibility of 5G and beyond networks—to adapt the network to the changing patterns of the different clusters of subscribers and introduce advanced services due to enhanced resource management capabilities [14]. The advantages of applying subscriber clustering for financial gain for an MNO can be observed in several reports. Some focus solely on the advantages of service personalization, which is a direct result of applying advanced clustering techniques to subscribers [15].

From service planning and resource management based on subscribers' behavior, which is this work's focus, subscriber-centric clustering allows MNOs to properly plan their network according to deep knowledge from their subscriber's base and corresponding behavior towards data and service consumption, overall leading to increased capacity requirements. The MNO can focus not only on the technical behavioral aspects of the clusters but also, from an economic perspective, can drive successful campaigns in order to increase the average return per user (ARPU) in several groups, reduce the churn rate in others and optimize sales channels, and, especially, focus on the clusters that contribute most to either planning capacity or profitability.

Thus, MNOs can focus not simply on overall behavior of the subscribers, but also on their future profitability, thus better focusing their resources on the most profitable clusters that, in conjunction with the ability to use software-defined networks (SDN), will for sure contribute to the maximization of resource management at the same time [16].

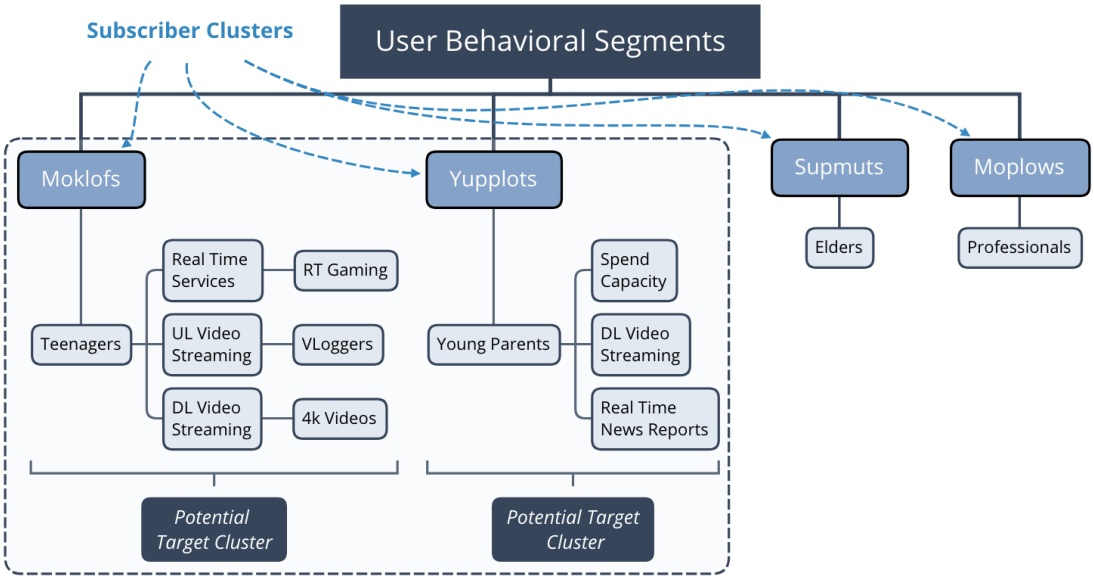


Figure 3. Potential target clusters to focus upon.

Another expected benefit is related to the frequency of such analysis. If performed continuously, or at least very frequently, it enables MNOs to have a fresh look at their customers' behavior towards service consumption and overall network capacity and adequacy. This leads to an improved understanding of the needs and wants of the subscribers, allowing for identifying lack of network capacity, which in turn enables the concept of intelligent network to rise, adequately complemented with performance and risk key indicators, leading to better overall QoE, quality of service (QoS), and customer satisfaction. From a service perspective, subscriber-centric clustering also allows MNOs to better fit their service offering to the several user groups, considering their consumption patterns and most used services. By doing so, subscribers will gain the notion that the service provided by their MNO network fully matches their needs and, interestingly, such notion can lead the MNO to competitive advantage scenarios compared with others. Some examples of advantages were briefly discussed, but others can surely be derived from subscriber-centric clustering, which can become the distinguishing factor between visionary MNOs in a highly mutable customer market. The following section will focus on some of the parameters that can be considered to perform cluster characterization, which is of utmost importance to properly define the subscribers' clusters.

3. Cluster-Centric Characterization

As referred to in the previous section, there are several variables that can be considered in order to perform clustering among subscribers. The most important

part, however, is that the clusters should be most representative of their subscribers as possible. The ability to split the subscriber base into homogenous clusters each with different sets of behaviors, needs, and desires is fundamental, especially in the network planning phase [17]. The most common methods found in the literature include demographic, psychographic, transactional, and behavioral segmentation. Despite the focus of this work being behavior, we briefly describe some of the parameters that can be evaluated in each one of these methods.

Geographic segmentation is based on geographical characteristics such as, for instance, place of birth, nationality, home address, and region of interest. Broadly, every geographical parameter associated with the subscriber can be used. Another interesting aspect is to use geo-textual data to feed certain services to subscribers [18]. Demographic clustering can be performed by looking at subscriber's information such as age, gender, marital status, and also financial income [12,13,15].

Psychographic clustering uses parameters that are subjective and focused on subscribers' attitudes and interests. This kind of clustering can be performed together with classical clustering methods, such as demographic, in order to complement the analysis with deeper insights from subscribers' perspectives.

Transactional clustering usually takes into consideration parameters such as the number of purchased products, financial volume, and number of items purchased, as well as time of the day the purchase occurs. From a cellular planning perspective, such parameters are very important because they will contribute to the identification of one or more clusters of subscribers that buy high-end mobile devices—which have the ability to generate more traffic—or that have more than on device, also contributing to higher levels of traffic capacity needs [4].

Behavioral clustering is the main focus of this work, which is described in greater detail in the following section and that will further enable the disclosure of subscriber clusters that have not been addressed from a cellular planning perspective and that will be characterized by high capacity demand.

4. Behavioral Clustering

4G has greatly changed users' behavior towards the usage of mobile services. 5G and beyond networks are expected to produce even bigger changes in subscribers' behaviors. Behavioral clustering focuses on parameters that can characterize the actions and behavior of subscribers. It is a set of clustering variables with a high degree of uncertainty, especially if the used parameters are not properly fit or the subscribers have erroneous actions that do not allow for pattern extraction. Examples of such parameters are benefits sought from a service

or set of services, sense of happiness that a service can provide (e.g., using social networks massively), quality of experience, and spend ability, among others. Behavioral variables are becoming increasingly important as subscribers' behavior is changing due the usage of mobile networks. This leads to very subjective variables that can be used to cluster yet, at the same time, new service models are enabled over cellular networks (e.g., real-time YouTubers, continuously uploading high amounts of data). Interesting concepts such as revenue per experience thus start to become a reality, and quality of experience becomes more prevalent than quality of service, which can lead to another subset of parameters focusing around the likeliness of a subscriber to re-use certain services in the future. Spend ability is a very interesting parameter that can be used in order for MNOs to prepare service tailoring and introduction, as well as plan a higher capacity area, on the basis of the expected spend capacity of the identified cluster.

Additionally, the ability of knowing which subscribers are more willing to spend on network capacity, a better service or set of services, as well as subscribers that exist, previously unknown to have such ability, illustrates the power of using behavior as a special clustering as opposed to just using the more traditional ones such as simple demographic data.

Behavioral clustering is one of the most powerful clustering techniques due specifically to its ability not only to characterize each subscriber's behavior but also, when additional data science mechanisms are used, to predict subscribers' future behavior. Although behavior prediction is out of the scope of this work, it is fair to state that behavioral clustering is a superset of clustering that can use any of the traditional clustering methods' variables, take them all into consideration, but also having the added value of considering other intrinsic subjective parameters in a combined fashion. All will contribute to the development behavioral models and clusters, allowing MNOs to potentially satisfy all subscribers in the end, as well as developing more adequate intelligent cellular networks.

5. New Subscriber Behavioral Clusters

As mentioned previously, new subscriber behavior clusters can be derived from potential target clusters, as presented below. Such new reality will result in adaptations of the impact model from previous work, shown in Subsection 5.2.

5.1. New Behavioral Clusters

The introduced subscriber cluster is derived from demonstrated behavior of children towards the usage of cellular networks and end user devices. A focus on ages between 0 and 12 years old is precisely a range of ages that is usually not

considered within the field of network capacity planning. Therefore, it is a 0–12 year-old behavior-based cluster, which is additionally split into three sub-clusters, as presented in Table 1, on the basis of different observed behavior.

We define toddlers as babies aged 0–2, preschoolers between 3 and 5 years old, and tweens between 8 to 12 years old. We felt the need to split these last two clusters into younger and older individuals due to some overlapping behavior, which is perfectly expectable. Such clustering seems the most appropriate considering several studies [2–9]. It is not the focus of this section to cover all subscriber's segments, but to focus solely on those which we believe are the new user behavioral segments that must be considered, as presented in Table 1, which will be further explored in the next section.

Table 3. Age clusters considered in the current work.

Cluster	Sub-Cluster	Ages
Toddlers	-	0–2
Preschoolers	Younger	3–4
	Older	4–5
Tweens	Younger	6–8
	Older	8–12

Figure 4 clearly shows that smartphone ownership has been rising among preschoolers [3]. Pew Research Center has shown that service consumption changes according to age, with Facebook and YouTube being the most prevalent services among adults over smartphones, but, on the other hand, young adults (between 18 to 24 years old) are clearly shifting to a set of platforms depending on their behavioral characteristics [4]. Another interesting factor that directly characterizes the behavior towards usage of services is the amount of time that each subscriber spends daily on online platforms. In [4], it was shown that in a group of children aged 3 to 5, almost 85% had and used smartphones to access media. A study focused on mobile usage among young children showed that many parents in the United Kingdom are worried about not effectively monitoring their young children's usage of handheld devices, with extensive time being spent on such activities, mainly for video watching and educational applications. All of these are high capacity-demanding mobile services that MNOs should be aware of, especially when, as shown in the study, watching online videos is one of the main activities of young children [19]. In the United States, it can be seen in Table 2 that there is a clear behavior towards accessing social networks several times a day, on a clear demonstration that such applications and services are becoming part of subscriber's life. This is a clear indicator that, when compared with the past,

the behavior has changed and that mobile usage is becoming more prevalent and real-time based in order to sustain such habits and needs.

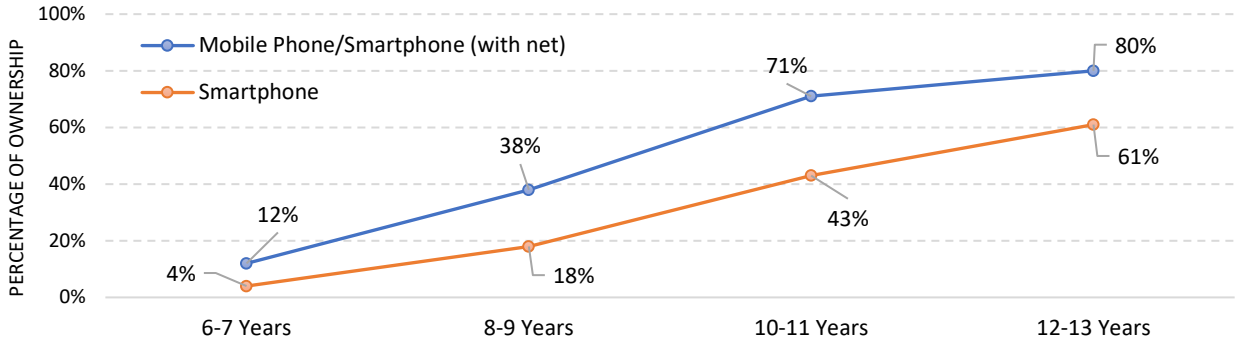


Figure 4. Smartphone ownership split among tween sub-groups.

The majority of households have a TV, smartphones, and tablets, as depicted in Figure 5a.

Table 4. Percentage of U.S. adults that use each site.

Usage Frequency	Facebook	Snapchat	Instagram	YouTube	Twitter
Several times a day	51%	49%	38%	29%	26%
Once a day	23%	15%	22%	17%	20%
Less often	26%	36%	40%	54%	54%

This clearly shows a tendency of three major devices for screen visualization. It was shown that 96.6% of children used mobile devices, in some cases right after turning one year old [19]. However, most interestingly from a behavioral perspective, is the reason behind such usage at a young age—most parents confessed that they gave smartphones to their children in order to calm them down (65%), which is increasingly becoming a common reason, especially in restaurants to keep children quiet. A total of 29% of parents even assumed that smartphones should be given to children at bedtime in order, once again, to calm them down (Figure 5b) [20].

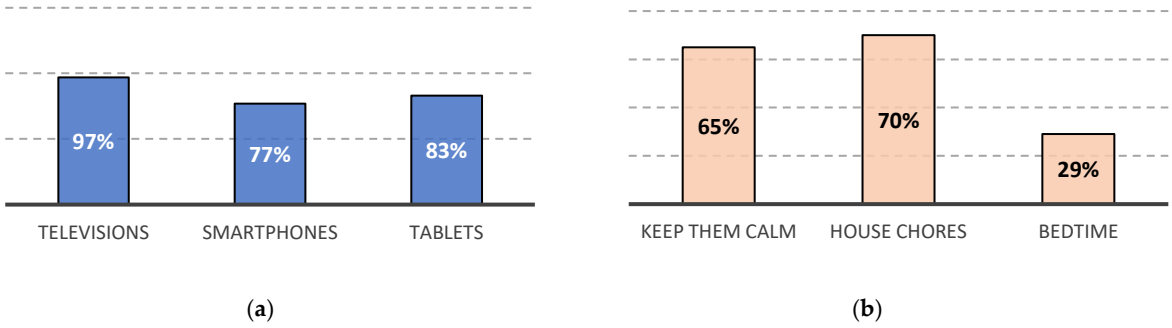


Figure 5. (**a**) Household media (%); (**b**) smartphone usage for three class of ages (%).

An indirect behavior of parents regarding their children results in a segment of the population, in this case very young children, starting to have access to devices and beginning to generate relevant amounts of traffic (mainly based on video or real-time gaming). Another very relevant aspect is that this group of children started using mobile media before the age of one year old. By the age of two, most children used a device daily and spent relevant screen time on television and mobile devices. Most three- and four-year-old used devices without any help, in a completely autonomous way. By four years old, most of them had their own device [17,19].

In another perspective, reinforcing the statement that these new behavioral clusters must be taken into account is the fact that children between one and nine years old tend to use smartphones repeatedly without separating from the device [6]. This, once again, constitutes a great source of traffic generation, and capacity demands rise, as usually these users consume streaming services, mainly YouTube videos, on high resolution through their smartphones [21]. It was also shown that for children under six years old, attending kindergarten and daily care centers, 80% used smartphones and the most preferred usage was precisely watching children's videos. Once again, the main reason for children using smartphones was to keep them calm or for house chores, as depicted in Figure 5b adapted from [19]. However, these are not the only reasons—to amuse them mainly while eating out, which is complemented by meal times at home or during long trips, is also a strong reason [6,7]. In [8], it is shown that for trips higher than 1.5 h, a continuous stream of video or online gaming will be generating a great amount of network traffic.

A recent work has narrowed down the services to only three main groups used by tweens: contacting friends, entertainment, and visiting websites [21]. A recent survey from AT&T [22] shows that end user device usage from younger American segments has increased exponentially since 2016—84% of preschoolers and 96% of tweens have their own internet-connected devices, smartphones, tablets, computer, or gaming system, and preschoolers have sole access to their own equipment.

Naturally, these results vary from context and demographic area, but it is not the aim of this work to focus on those differences. Instead, our aim is to show that, in fact, there are user clusters that did not exist in the past and that should now be considered for mobile network planning and capacity quantification.

From a parental perspective, it can be affirmed that parents use their smartphones as a distraction tool or as a reward to their children, once again, revealing high usage of smartphone by children [19,23]. Parents in their twenties, thirties, and above assume the initiative of giving smartphones to their children.

Parents in their thirties were the most common. A 2018 report shows that toddlers and young preschoolers use their parents' smartphone to access their content, whereas older preschoolers and tweens have their own phone. When considering tablets, 19% of toddlers and young preschoolers have their own tablet, as well as 47% of older preschoolers [8,24], as depicted on Figure 6. The report also shows that toddlers might use their parents' smartphones, but more than that, some might have their own tablet, which typically, for the same content and especially considering the screen size, downloads videos with higher quality, meaning additional traffic generation.

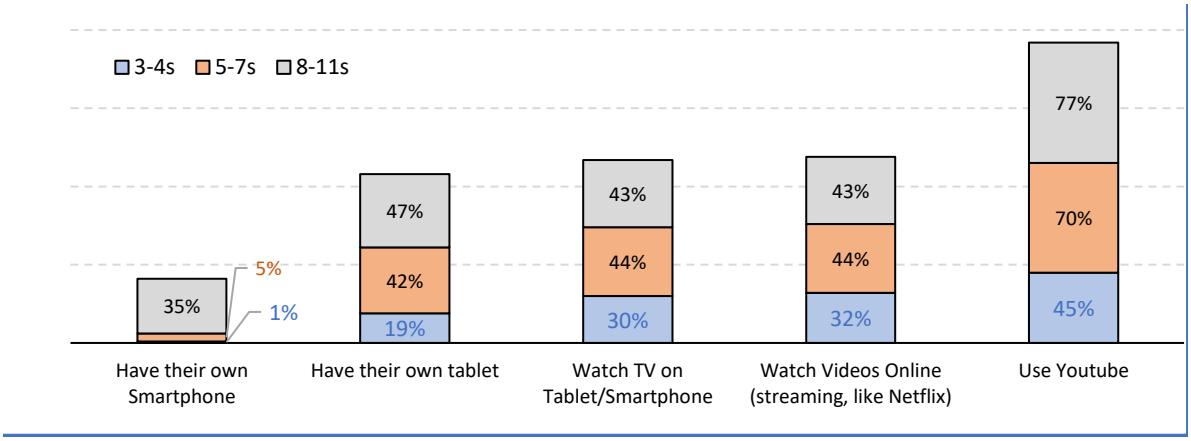


Figure 6. Device ownership and usage habits for the three classes of ages.

Another relevant fact that contributes even more to premature access to smart devices by younger children is the increasing belief from parents that mediation practices (e.g., content monitoring and time usage restrictions) may minimize negative effects and actually contribute beneficially to enhance or accelerate children's learning processes [7,17,21,25]. Such a belief will be reflected as additional challenges for MNOs in terms of capacity and QoS. High traffic-demanding applications remain the main usage of smartphones, ranging from watching TV, to streaming videos (e.g., Netflix), to YouTube. Older preschoolers and tweens are the most demanding, with YouTube usage above 70% on their hand-held devices, mostly for watching videos and listening to music through video clips (Figure 6).

These results also show that YouTube is becoming the viewing platform of choice, with rising popularity particularly among tweens, who are increasingly becoming content generators (*Vloggers*, YouTubers). In a long timeframe comparison between 2011 and 2017 and also another between 2014 and 2019, regarding screen usage among toddlers and young preschoolers, it can be seen that TV screen time has been decreasing, and that smartphone usage has risen

drastically [9,26]. Computer and gaming console usage has also been dropping, clearly supporting the idea that, progressively, the smartphone is replacing those devices (Figure 7).

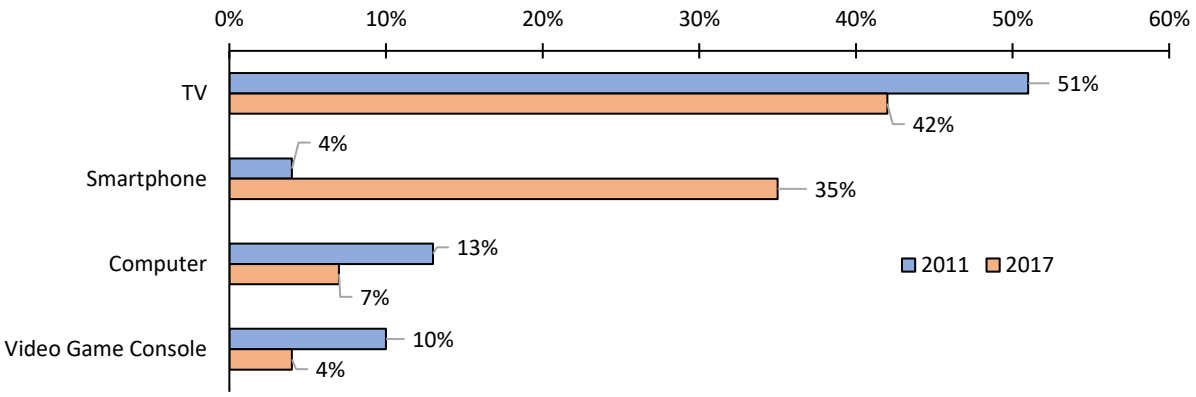


Figure 7. Share of screen time for ages 0–8 years old (%).

As demonstrated in the previous section, there is a new behavior pattern that will force young parents cluster to be split according to whether they have children or not. Thus, a new behavior cluster appears from Yupplots (young parents with lack of time), accommodating those parents that have young children, either toddlers or preschoolers, allowing them to have access to devices, and thus representing a behavioral change. This is the first result of applying advanced clustering analytics merged to cellular concepts. Figure 8 shows the result of the proposed extension, with two clusters that became target clusters and, by further using advanced clustering analytics, result in two additional clusters. In the end, there are six subscriber clusters, two of them becoming the target clusters from the perspective of this work (tweens and young parents with children), clearly showing the advantage of performing advanced clustering analytics over existing data. From an impact perspective, both new clusters will contribute differently. The tweens cluster will represent increased capacity need, up to now, only considering ages from 13 onwards. This is considered as a new behavioral cluster because, as was shown in previous sections, the majority of children at age six already own their own device [17,19].

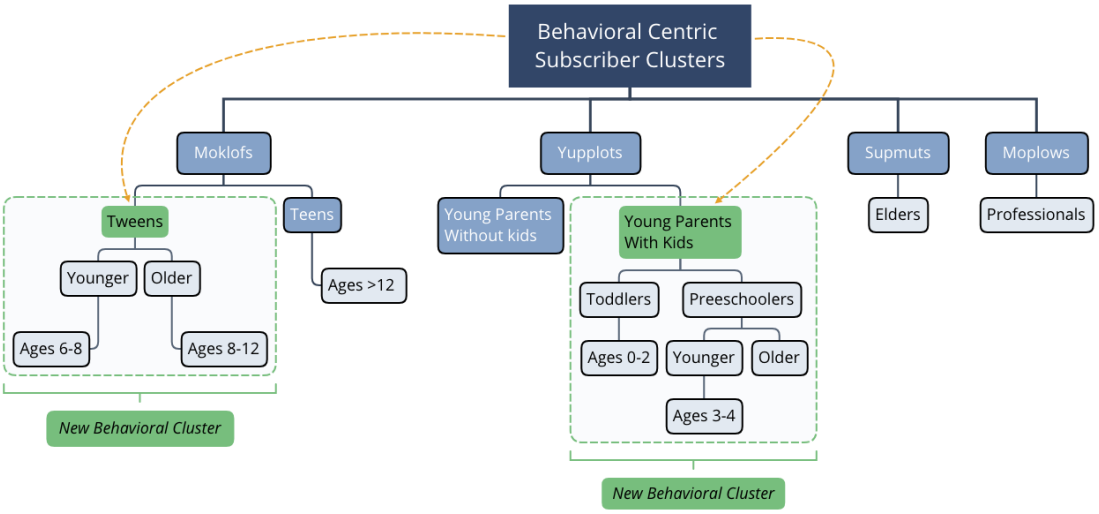


Figure 8. New behavioral-centric subscriber clusters. Moklofs: mobile kids with lots of friends, Yupplots: young parents with lack of time, Supmuts: senior urban people with much time, Moplows: mobile professionals with lots of work.

5.2. Adaptation of Subscriber-Centric Impact Model

The changes to context and, especially, changes to behavioral clusters, with two additional ones to be considered, must be properly considered in the impact model presented in [2]. Thus, we present the major changes to the model, which will attest its flexibility, which was one of the major aspects when it was developed. Setting the baseline of services, we consider the most prevalent services for 2020–2022 [24,27], as previously presented in Table 3.

Those services form a **T** vector representing the traffic share of each *S* service.

$$\boldsymbol{T} = \begin{bmatrix} T_{s_0} \\ \vdots \\ T_{s_k} \end{bmatrix}, \quad \text{where} \quad \begin{aligned} s_0 &= Mobile\ social\ networking \\ s_1 &= Mobile\ social\ gaming \\ s_2 &= Video\ streaming \\ s_3 &= Voiceover\ data \\ s_4 &= P2P\ communications \\ s_5 &= Mobile\ web\ browsing \\ s_6 &= Mobile\ commerce\ and\ banking \\ s_7 &= M2M\ communications \end{aligned}$$

Table 5. Services considered (un-aggregated).

Service
Video streaming
Mobile social networking
Mobile social gaming
Voiceover data
Peer-to-peer communications
Mobile web browsing
Mobile commerce and banking
Machine-to-machine (M2M) communications

In the elements of the vector, k represents the k^{th} service, which ranges from 0 to 7. Such services are then categorized and some of them aggregated considering their similarities, leaving video streaming as the most prevalent, which we, in this work will assume differently—in the previous work we considered eight services, but in this case we will only consider video and non-video services, as represented in Table 4. Services were treated individually in the previous work, however, in this work, in order to only have two classes (video and non-video), service aggregation had to be performed, as represented in Table 4 by X. Such assumption and simplification aim to understand the impact of the new behavioral clusters that mainly have video streaming service usage behavior, which is the main point to demonstrate in this work. This means that, in this case, k ranges only from 0 to 1. Each of the initial four behavioral clusters were then mapped according to their contribution to the total traffic generated by each of those five classes.

Table 6. Service aggregation.

Service	Previous Work k = [0:7]	This Work	
		Mobile Video traffic (k = 0)	Non-Video Traffic (k = 1)
Video streaming	X	X	-
Mobile social networking	X	-	X
Mobile social gaming	X	-	X
Voiceover data	X	-	X
Peer-to-peer communications	X	-	X
Mobile web browsing	X	-	X
Mobile commerce and banking	X	-	X
Machine-to-machine (M2M) communications	X	-	X

Two periods in time were considered in the previous work, regarding expected total traffic generated by each service: $Y = 2016$ and $Y = 2020$ and compared the contribution of each behavioral cluster both in 2016 and 2020. Traffic followed a probability distribution function per behavioral cluster as a function of services usage $p_{s_s}(s_s) = \overset{K+1}{k=0} p_{k-}(s_s - k)$, where p_k the usage probability of each of the k^{th} service and K is the total number of considered services, S_s. In this case, we have narrowed down K to only two service classes. In order to evaluate if the new clusters have impact on overall traffic capacity and demand, we begin by presenting a 6 year evolution of mobile traffic, starting in 2017 and forecasting up to 2022, as depicted in Figure 9 [24,27]. We focus solely on two service categories, only to show the impact of the four previous clusters versus the six new clusters

resulting from the advanced behavior clustering process. Those two categories are represented in Table 4. It can be seen from Figure 9 that mobile video traffic is the service that has been increasing the most and is expected to continue, which is why for simplification we used only two categories.

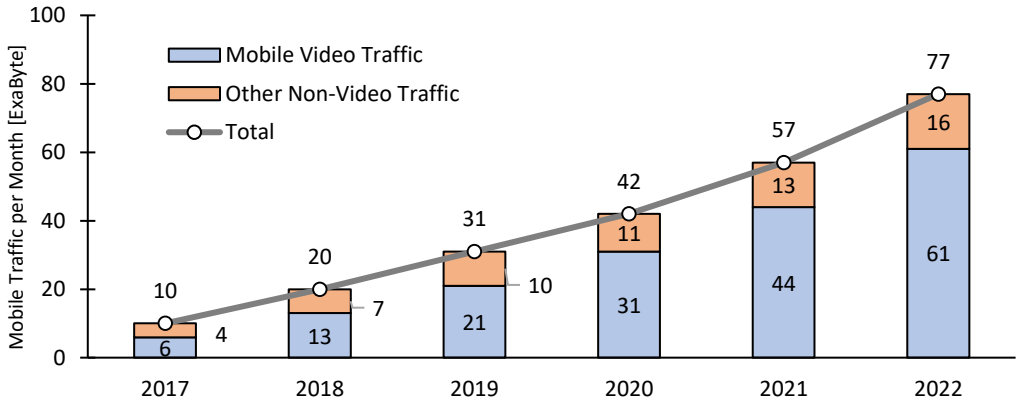


Figure 9. Mobile traffic per month (video and non-video).

A probability matrix was defined representing the probability of each segment using one of the T_{s_j} services, considering that, in this case, the number of user segments L has increased from four to six.

$$\mathbf{P} = \begin{bmatrix} P_{(U=1,S=0)} & - & P_{(U=1,S=M)} \\ - & - & - \\ P_{(U=L,S=0)} & - & P_{(U=L,S=M)} \end{bmatrix}, \text{ where U represents a cluster.}$$

The sum of each cluster impact applied to its corresponding service should equal one, according to $\sum_{U=1}^{L} P(U,S) = 1$, with $U = 1, L; S = 0, M$. In this case, *max(L) = 6* and *max(M) = 1*. We then assumed the following probability distribution of the services per user segments, with the additional two clusters derived from the existing ones:

The assumptions made are the following when comparing tweens with Moklofs (mobile kids with lots of friends):

- Tweens use less than 50% of social networking;
- Tweens spend more than 5% of the time playing games;
- Tweens spend more than 30% of the time watching mobile videos;
- Tweens are trend less in terms of using voice over IP (VoIP), mobile commerce and banking, as well as machine-to-machine (M2M) services.

Regarding Yupplots without children and with children, the following assumptions were made:

- The usage of mobile video increases twofold when children use the device (main usage);
- All other services practically remain the same;

- Video streaming is the predominant change.

Such assumptions are supported by existing studies referred to throughout this work. Nevertheless, the impact model, which is not the subject of this work, is flexible enough to accommodate any change in behavior, as detailed in [2].

Finally, with the existing probability distribution of the services and using the traffic per month values from Figure 9, resulting in a traffic vector $\boldsymbol{T}$, one can calculate the overall impact (amount of generated compound traffic, $\boldsymbol{T_C}$), resulting in a quantification of overall video traffic and non-video traffic generation per behavioral cluster.

$$\mathbf{T_C} = \begin{bmatrix} T_{Y(Video)} \\ - \\ T_{Y(Non-Video)} \end{bmatrix}, \text{and each cluster derives from:}$$

$$T_{Y\,\text{Video}} = \sum_{u=1}^{L} P_u(S_2) \sum_{-y=2017}^{Y} T_y \quad \text{and} \quad T_{Y\,\text{Non-Video}} = \left(\sum_{u=1,s=0}^{L,K} P_u(S) \sum_{-y=2017}^{Y} T_y \right) - T_{Y\,\text{Video}}$$

where in this case y = 2017 and Y = 2022. This way, it is possible to calculate the total traffic (video and non-video) that each of the clusters generate based on their behavioral characteristics, as presented in the next section.

6. Results

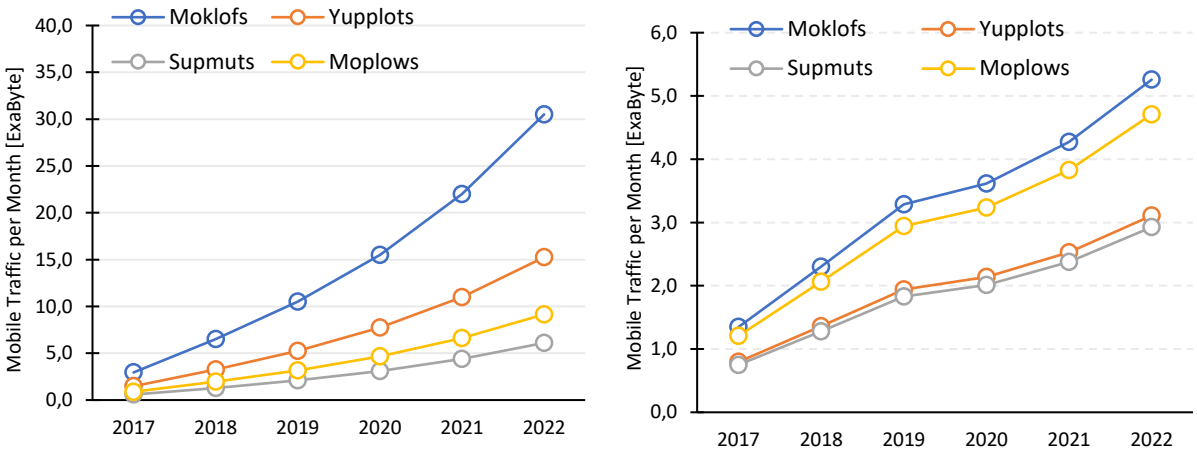
This section presents the results from applying both advanced behavior clustering on the clusters derived as explained in the last section, as well as the impact that such new clusters have from traffic generation's behavioral perspective. As seen earlier, on the basis of the information presented, two new behavioral clusters should be expected to extend from the initial four. Moklofs (mobile kids with lots of friends) [2] were split, and instead of considering solely teenagers, tweens were also considered, in accordance to Table 1. This section focuses only upon these two additional clusters versus the original ones, as the objective is to show the advantages of performing advanced clustering, which, in our case, applies only to two of the original four clusters. Yupplots (young parents with lack of time) were also split into two different behavioral clusters.

The Yupplots cluster without children maintained the impact levels as previously discussed [2], but for the new cluster it was necessary to break down the capacity into two: directly generated, that is, by the parents, and indirectly generated, that is, by children, whether the child is a toddler or a preschooler.

Considering that such service is the one that relates more to the behavior of the new clusters, it appears to be adequate to use it in order to evaluate the impact

of considering both a new cluster and splitting the Yupplots cluster. Figure 10a presents the breakdown of evolution of traffic for the four original clusters in what concerns mobile video traffic when applying the cluster characteristics to the data from Figure 9.

It can be seen that Moklofs and Yupplots are the behavioral clusters with the highest level of traffic usage. Figure 10b shows exactly the same breakdown but for the second category of services (non-video traffic).



(**a**) (**b**)

Figure 10. Mobile traffic per month for each of the original four clusters: (**a**) video and (**b**) non-video.

Figure 11 presents the video and non-video traffic generated per month for each of the four original behavior clusters. It can be seen that by 2017 video traffic was not that much higher than non-video traffic. Nevertheless, as the time goes by, and especially considering 5G, the expectation is that by 2022 mobile video accounts for 79% of all traffic per month and non-video will account for 21%. By 2017, video traffic represented 59% of the whole mobile traffic generated per month and 41% represented the sum of all remaining services.

Up to this point, it has been shown that mobile video is the prevalent service on all four behavioral clusters, as it was concluded in [2]. It was shown also that such traffic will increase significantly over the next few years, especially fueled by 5G and behavioral changes on subscribers. From that perspective, we have applied the impact methodology developed in [2] to two new clusters resulting from this work (refer also to Figure 10):

- The Moklofs cluster was split into two behavioral segments, one which represents teenagers in general and the other a new cluster focused solely on tweens.

- The Yupplots cluster was split into two behavioral segments, one which consists of young parents without children, and a new one, which mimics the behavior of young parents that have children and whose traffic consumption behavior changes accordingly.

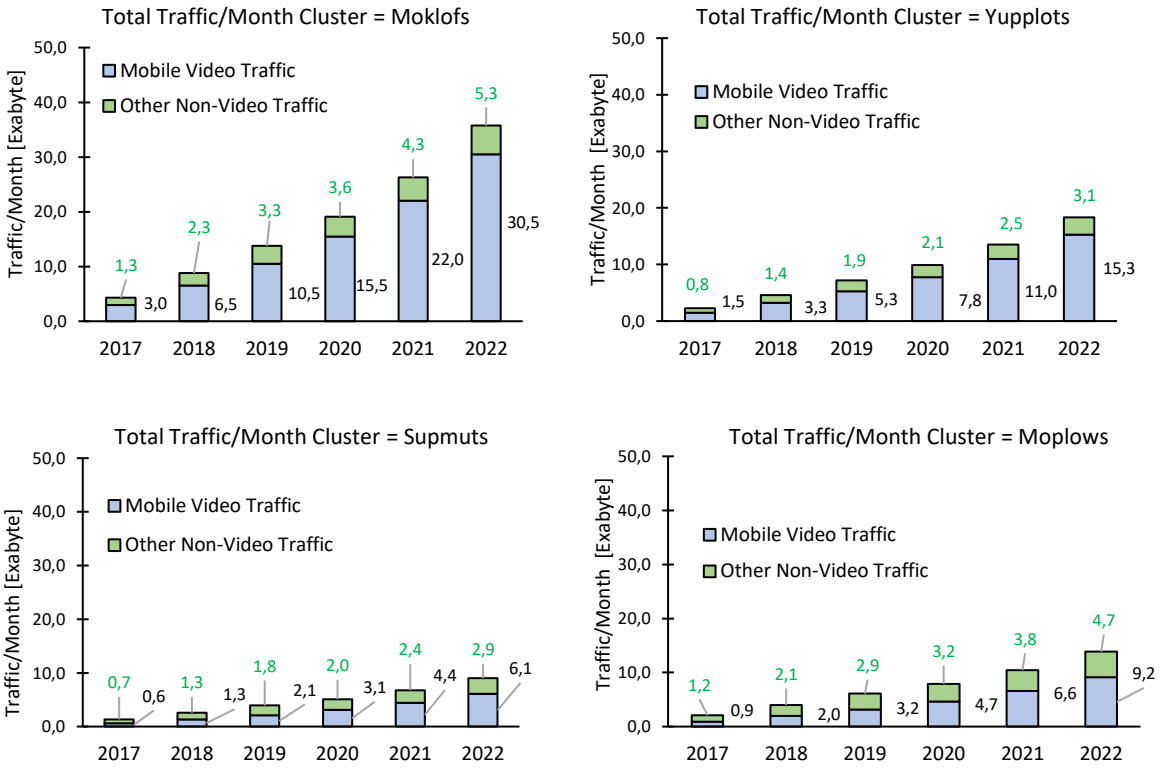


Figure 11. Traffic generation breakdown per month and per behavioral cluster, over 6 years.

Figure 12 shows the results in terms of traffic generation when considering the new clusters by 2022. It can be seen that there was an increase in traffic generation for the parents that have children, considering that there was a behavior change as previously presented in Section 5, leading to additional usage of video streaming. Regarding tweens, the new cluster, it can be seen that, at such ages, video consumption is higher, whereas social networking (within other services) does not represent much of a traffic increase. Thus, by 2022, tweens are expected to generate 148% more traffic than teens, and Yupplots with children using their devices will represent a traffic increase of 186% compared to normal Yupplots' behavioral patterns.

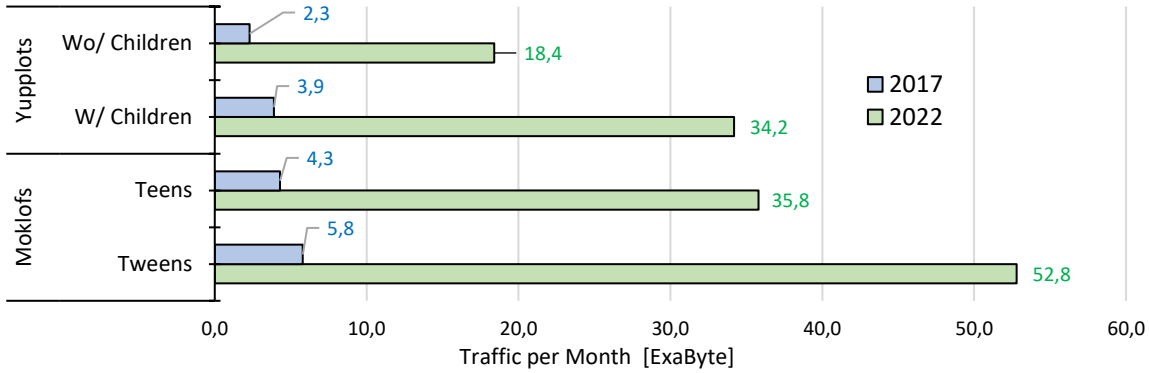


Figure 12. Traffic per month: impact of new clusters in traffic generation by 2022 versus 2017 in terms of mobile video traffic.

Figure 12 presents the traffic per month by 2017. It can be seen that, curiously, Yupplots behavior changes due to child usage of smart devices for video streaming, which brings this cluster closer to a behavior typical of teens.

In fact, this aspect is also shown in Figure 13. This is one of the conclusions that can be drawn after applying advanced clustering when the two initial clusters have been divided and looked at more closely.

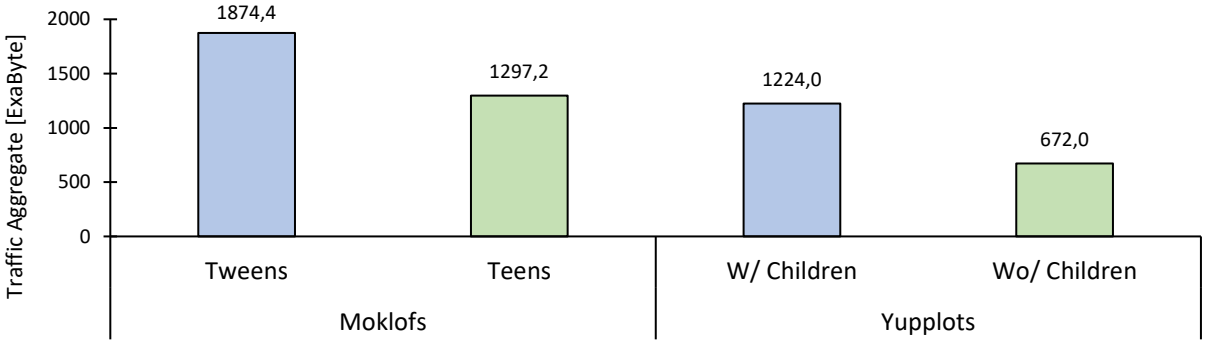


Figure 13. Traffic aggregate for new clustering in a 5-year time span.

Nevertheless, by 2017, traffic generated by tweens represented 135% more than that generated by teens. As for Yupplots, when children are given smart devices, this cluster behavior transforms and increases traffic consumption by 172% compared with normal behavior. Figure 14 presents an interesting result. It can be seen that by 2022, according to the estimates, if an MNO would consider solely teens and Yupplots as simple clusters, the total amount of expected traffic would be nearly 1970 Exabyte. Nevertheless, the tweens cluster and also the behavior changes of Yupplots when children start to use their devices must be considered. In such cases, without the detailed analytics and advanced clustering, the MNO would expect an aggregate traffic of 1970 Exabyte, disregarding and

overlooking an additional 3100 Exabyte of approximate generated traffic. On the other hand, by performing the proposed advanced clustering, an MNO should have the notion that there are two additional specific clusters and thus it should plan its network to sustain a total of 5067 Exabyte in a time span of 5 years. Thus, a MNO who was able to perform advanced clustering, such as the proposed approach versus one that would stand with only four clusters, would know that in 5 years' time, span traffic increase due to new behavior clusters would represent an increase of 157% and could therein prepare its network planning and resource management accordingly, representing a competitive advantage over others.

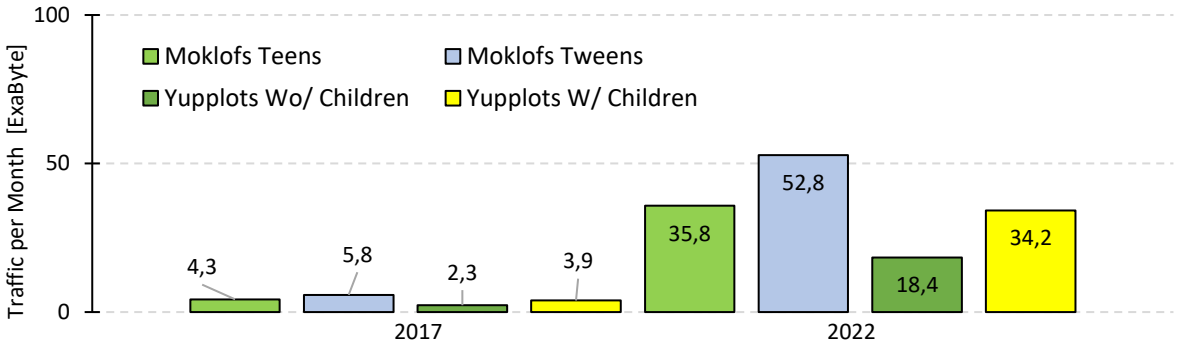


Figure 14. Traffic per month 2017 versus 2022 for new clusters.

This example clearly shows the advantage of bridging advanced subscriber analytics, and Figure 15 presents the comparison of traffic per month generated by each behavioral cluster when considering 2017 versus 2022. It can be seen that by 2017, a more uniform traffic amount was present among clusters, mainly due to the fact that mobile video streaming was more paired with all other services. However, when considering 2022, the behavioral change of subscriber clusters indicate that traffic per month is expected to grow unprecedentedly, especially within the clusters that mostly use mobile video as the main service, that is, Tweens, Teens, and Yupplots with children, which represent a behavior change that bring them closer to teen traffic patterns. Figure 15 presents the average monthly traffic per month for all four segments that resulted from splitting the original two.

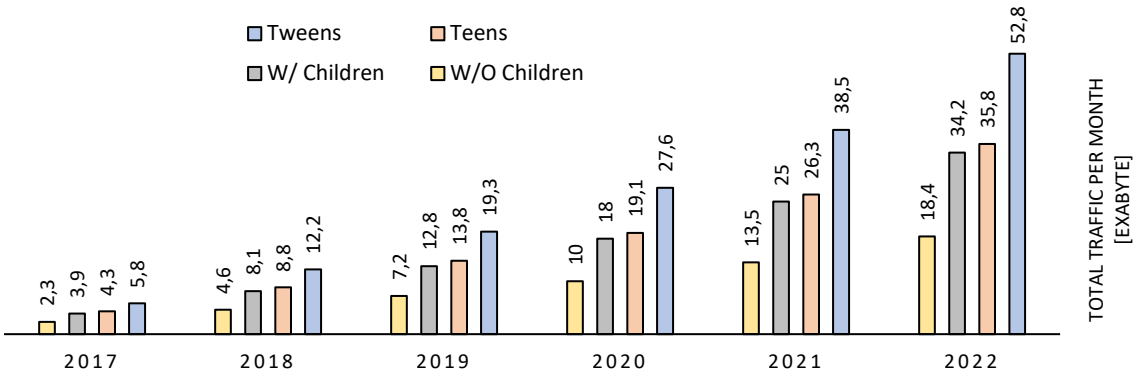


Figure 15. Total traffic per month per year for new behavioral clusters.

As previously mentioned, the results of applying the advanced clustering techniques referenced in the first sections of this work lead to extend the existing four segments from [2] to a total of six segments. Figure 16 shows the overall traffic impact that each of the six behavioral clusters represent. It can be seen that the two new segments resulting from this work are part of the top three segments with highest traffic generation. As so, segments very relevant to traffic generation are now considered after advanced clustering has been performed.

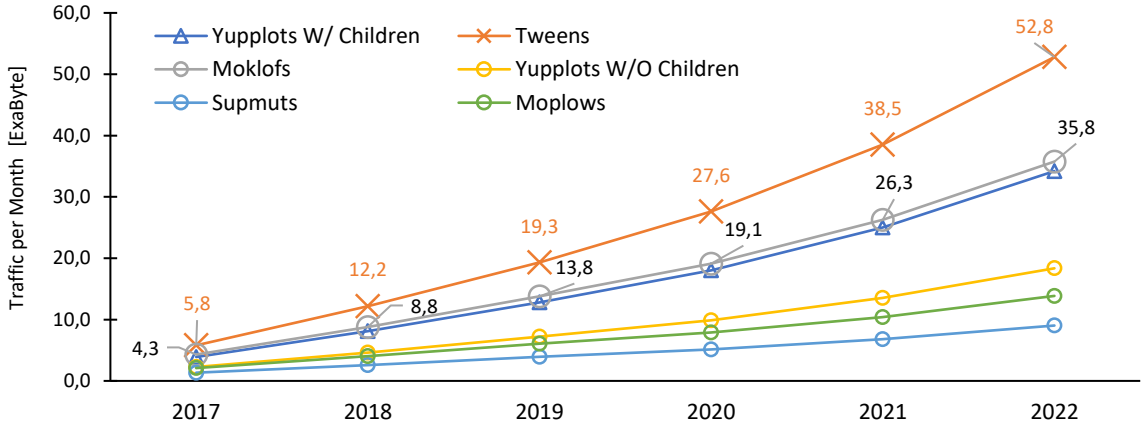


Figure 16. Traffic per month for all six clusters.

Finally, in a direct comparison between the traffic per month generated by the four clusters and six clusters, Figure 17 shows that the new clusters resulting from advanced clustering should not be ignored, that is, the cluster formed by tweens and the one by the behavior of parents when letting their children use smart devices. Their sum, as depicted, from 2017 to 2019 represents almost the same traffic generation as the four initial clusters. However, from 2020 onwards, these two segments will surpass the traffic generated by the original four clusters and, therefore, should be considered and focused upon in depth. As previously

mentioned, the end user device will play a very crucial role in future networks, not only 5G, but also beyond, becoming the main source of traffic generation and consumption.

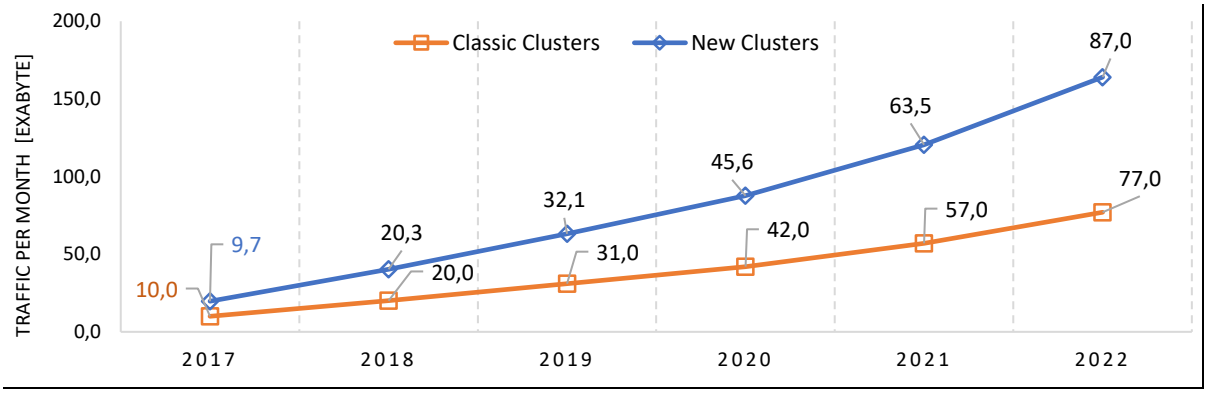


Figure 17. Traffic per month considering four (classic) and two (new) clusters.

With higher cellular network capacity, higher data rates, and more capable and performant smartphones, we can imagine that the majority of services will be provided by the handheld device, thus representing an unprecedented traffic generation in cellular networks. It is, therefore, of utmost importance to further develop advanced analytics over subscriber data, with important conclusions potentially appearing as the ones presented in this paper—the need to consider a new cluster of subscribers and also the discovery of a behavior change that was only possible to identify by using such techniques.

7. Conclusions

In this paper, our main contribution is twofold: first, we demonstrated the advantage of applying subscriber-centric clustering based on behavior, both to capacity network planning process and resource management. We also focused on the advantages of applying several types of clustering.

Next, our second main contribution was the definition and characterization of a new subscriber cluster comprising subscribers aged from 0 to 12 years old—extending previous work clusters—who are increasingly using tablets and smartphones to access high capacity-demanding services such as video streaming and online cooperative gaming. The other part of the second contribution was that by using different clustering based on behavior, we could demonstrate that one cluster needed to be split into two different clusters, precisely due to behavioral change—parents that have children and let them use their smart devices represent

a pattern change in traffic consumption and, therefore, must be considered for analysis.

We showed that tweens and older preschoolers are avid consumers of high data rate services over smartphones, and thus should be considered as a relevant subscriber cluster from an MNO perspective. This comes as normal nowadays. Nevertheless, additionally, in terms of capacity planning, it was demonstrated that toddlers and young preschoolers constitute an indirect cluster of high demand subscribers, hidden by the fact that parents' smartphones are mainly used, although some have their own devices at very young ages. Nevertheless, for a cluster of parents that are not identified as having data-hungry behaviors, the fact that having children may unbalance the clustering process results in a different cluster. This shows clearly one aspect that is of utmost importance for MNOs—knowing your customer. In fact, one of the major pitfalls in this approach will materialize if MNOs do not have proper and sufficient quality information or data from their subscribers. As we have shown, and such can be seen as proof of the importance of quality data, simply by overlooking the fact that young parents might share their devices with their children makes a large difference and, thus, a new behavioral class appears, one that was being overlooked. Herein, proper data and information are paramount, especially when considering applying methodologies such as data mining, or even artificial intelligence (e.g., artificial neural networks) to detect and predict behavioral anomalies and, possibly, new clusters of analysis. If data are not sufficient or properly treated, MNO's approaches to network and capacity planning might not be realistic. As such, although it was not the objective of this work to address data quality, in order for results to be reliable and fit for feeding to the impact model, we minimized that issue by using proper and valid data from known sources, as referenced throughout the work. Nevertheless, MNOs should take into consideration that proper data preparation and handling should be performed in order to have a workable, valid, and quality dataset.

It was shown that there are relevant advantages of having the subscribers segmented into clusters using behavioral parameters as variables—it allows for a more effective process of cellular planning while giving the ability, especially in 5G dense networks for MNOs, to adapt their capacity and planning strategies more frequently, as well as enhance resource management. As an example, just by knowing that there is a cluster of subscribers who have been parents lately might be enough to expect an increase in capacity demand.

We also contributed by focusing on age groups that usually are not considered for subjects regarding cellular capacity and planning. Typically, only subjects older than 18 years of age are considered because they are old enough to be

allowed in a survey, which is the most common process used to gather behavioral information. Although teenagers are often also considered through online surveys, there is a gap, which we have discussed, ranging from very young ages to pre-teenagers. The reality shows that global overall cellular service consumption and, thereafter, capacity demand starts at very young ages, sometimes under one year old. Finally, we contributed to the extension of previous work by introducing two clusters that have not been focused upon previously from cellular planning and resource management perspectives, also introducing such new cluster groups to the impact scoring methodology used before in order to evaluate overall network impact over capacity towards traffic generation [2]. The results showed that tweens should definitely be considered, especially by 2022, in a high-capacity driven network, and they also uncovered a behavior change in one of the Yupplots clusters due to the fact that children are given their parents' smart devices, indirectly changing traffic consumption patterns. Clearly it was shown that, when giving their children smart devices, the Yupplots' traffic pattern changes, becoming closer to the traffic and behavioral pattern of teens.

Our work showed that subscriber clustering should be considered one of the most important analysis vectors when considering network and capacity planning, especially in high-density and high capacity such as 5G and beyond. This kind of technique is of utmost importance to MNOs, and we have argued that advanced clustering and analytics should the thoroughly performed prior to any service introduction, also on the network planning phase, as well as during full production, in order to maximize resource management, according to the ever-changing behavior of subscribers. The present work can be applied generically to any user in the world, on a national level or cross-border, depending, once again, on the existence of data about the subscribers. The approach is not dependent on any specific geography or demographic information. As long as there is a national, European or any other kind of dataset regarding subscribers, behavioral features might be extracted, and MNOs can start applying clustering techniques and advanced analytics in order to evaluate possible advantages of such approaches.

This is one of the actual limitations of the current work, which is in having a proper and recent dataset to work with. Future work will focus on finding the right dataset, which will allow taking this approach one step further by enabling applying mechanisms such as data mining, artificial intelligence, and deep learning to the new behavioral clusters in order to enhance overall resource management and capacity planning, as well as end users' quality of experience.

Author Contributions: Conceptualization, L.G., P.S., N.S., and A.C.; data curation, P.S.; formal analysis, L.G. and N.S.; project administration, N.S. and A.C.; supervision, P.S., N.S., and A.C.; validation, L.G., P.S., N.S., and A.C.; writing—original draft, L.G.; writing—review and editing, L.G.

Funding: This work was funded by Fundação para a Ciência e a Tecnologia / Ministério Ciência Tecnologia e Ensino Superior through national funds and when applicable co-funded by Fundo Europeu de Desenvolvimento Regional (FEDER) – PT2020 partnership agreement under the project UID/EEA/50008/2019.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Siddiqi, M.A.; Yu, H.; Joung, J. 5G Ultra-Reliable Low-Latency Communication Implementation Challenges and Operational Issues with IoT Devices. *Electronics* **2019**, *8*, 981.
2. Goncalves, L.; Souto, N.; Sebastiao, P.; Correira A. On the Impact of User Segmentation and Behavior Analysis over Traffic Generation in Beyond 4G Networks. *Trans. Emerg. Telecommun. Technol.* **2014**, *28*. doi:10.1002/ett.2933.
3. Statista. *Smartphone Ownership among Children and Teenagers by Age Group in Germany in 2016*; 2019.
4. Pew Research Center. *Social Media Use in 2018, Report*; Pew Research Center: Washington, DC, USA, 2018.
5. Kim, S.J.; Cho, S.M.; Lim, K.Y. The effects of high exposure to smartphone from ages 3 to 5 years on children's behaviors. In Proceedings of the 25th European Congress of Psychiatry/European Psychiatry, Florence, Italy, 1–4 April 2017; Volume 41S.
6. Cho, K.-S.; Lee, J.-M. Influence of smartphone addiction proneness of young children on problematic behaviors and emotional intelligence: Mediating self-assessment effects of parents using smartphones. *Comput. Hum. Behav.* **2017**, *66*, 303–331
7. Genc, Z. Parents' perceptions about the mobile technology use of preschool aged children. *Procedia-Soc. Behav. Sci.* **2014**, *146*, 55–60.
8. OFCOM. *Children and Parents: Media Use and Attitudes*; Report; OFCOM: London, UK, 2019.
9. Common Sense Media, Survey on Screen Share Time. Available online: https://www.commonsensemedia.org/sites/default/files/uploads/research/2019-census-8-to-18-full-report-updated.pdf (Last accessed on 6th Nov. 2019).
10. Parimalam, T.; Sundaram, K.M. Efficient Clustering Techniques for Web Services Clustering. In Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, India, 14–16 December 2017; pp. 1–4.
11. K. M. Shabana and J. Wilson, A novel method for automatic discovery, annotation and interactive visualization of prominent clusters in mobile subscriber datasets, *IEEE 9th International Conference on Research Challenges in Information Science (RCIS)*, **2015**, pp. 127-132.
12. Deng, H.; Qi, Y.; Liu, J.; Yang, J. Analysis of mobile subscribers' behavior pattern based on non-negative matrix factorization. In Proceedings of the IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC), Beijing, China, 23–25 September 2016; pp. 180–185.
13. K. M. Shabana, J. Wilson and S. Chaudhury, A Multi-view Non-parametric Clustering Approach to Mobile Subscriber Segmentation, IEEE 18th Conference on Business Informatics (CBI), 2016, pp. 173-181.
14. Khan, U.A.; Lee, S.S. Three-Dimensional Resource Allocation in D2D-Based V2V Communication. *Electronics* **2019**, *8*, 962.

15. Segment Research. *The 2017 State of Personalization Report*; Technical Report; Segment Research: Beijing, China, 2017.
16. Wang, M.; Karakoc, N.; Ferrari, L.; Shantharama, P.; Thyagaturu, A.S.; Reisslein, M.; Scaglione, A. A Multi-Layer Multi-Timescale Network Utility Maximization Framework for the SDN-Based LayBack Architecture Enabling Wireless Backhaul Resource Sharing. *Electronics* **2019**, *8*, 937.
17. Ochoa, W. *Mobile Screen Technologies and Parents of Young Children: Investigating Diverse Parents' Attitudes, Beliefs, and Their Interactions with Children*; UC Irvine: Irvine, CA, USA, 2019.
18. Chen, L.; Shang, S.; Zheng, K.; Kalnis, P. Cluster-Based Subscription Matching for Geo-Textual Data Streams. In Proceedings of the IEEE 35th International Conference on Data Engineering (ICDE), Macau, China, 8–11 April 2019; pp. 890–901.
19. OFCOM. *Life on the Small Screen: What Children are Watching and Why*; Report Annex 1; OFCOM: London, UK, January 2019.
20. Kabali, H.K.; Irigoyen, M.M.; Nunez-Davis, R.; Budacki, J.G.; Mohanty, S.H.; Leister, K.P.; Bonner, R.L. Exposure and Use of Mobile Media Devices by Young Children. *Official Journal of the American Academy of Pediatrics* **2019**.
21. Zaman, B.; Nouwen, M.; Vanattenhoven, J.; de Ferrerre, E.; Van Looy, J. A Qualitative Inquiry into the Contextualized Parental Mediation Practices of Young Children's Digital Media Use at Home. *J. Broadcast. Electron. Media* **2016**, *60*, 1–22.
22. AT&T: Screenready, Report 2018. Online: https://about.att.com/newsroom/2018/ScreenReady.html (Last accessed on 5 September 2019**)**.
23. Zaman, B.; Nouwen, M. *Parental Control: Advice for Parents, Researchers, and Industry*; EU Kids Online Report; EU Kids: London, UK, 2016.
24. Etaher, N.; Weir, G.R.S. Understanding children's mobile device usage. In Proceedings of the International Conference on Cybercrime and Computer Forensics, Vancouver, BC, Canada, 5 September 2016; pp. 1–7.
25. OFCOM. *Parental Mediation*; Report; OFCOM: London, UK, 2015.
26. OFCOM. *Children's Media Lives—Wave 5*; Report; OFCOM: London, UK, 2019.
27. Cisco. *Visual Networking Index*; Report; Cisco: San Jose, CA, USA, 2019.

2.3. Article nr. #3

This article presents two models of NF virtualization through the usage of cloud environments, in order to assist in the whole inter-RAT HO, for 5G NR and considering future WIFI6 deployments as a secondary RAT and post-5G future networks' challenges. The two models were developed considering a highly densified heterogeneous network (both on subscriber and RAN plane – Device/IoT Edge), and proposed to greatly improve existing deployments but, especially, future 5G NR network systems. The existence of massive IoT devices on device edge part of the network was also considered. NFV is a 5GC concept on the two models, fully aligned with the trend to virtualize several existing monolithic network functions, by disaggregating them, towards enhanced cloud 5GC and RAN.

The main contribution to the present thesis was the models themselves. By considering EC, it is shown that the models surpass current physical ANDSF classical datacenter deployments in performance, elasticity, cost and overall quality of the whole inter-RAT HO process, relevantly contributing to the whole optimization of densified 5G NR networks. The proposed models are fit for 5G NR standalone deployments as well as non-standalone and already have into consideration advanced concepts like device EC and full NFV and disaggregation.

Article details:

- Title: Cloud-Assisted Multi-RAT Steering in 5G/Beyond and Wi-Fi6 DenseNets;
- Date: December 2019;
- Journal: Applied Sciences;
- Scimago/Scopus Journal Ranking: Quartile 1;
- Publisher: MDPI (Still under Revision).

applied sciences

MDPI

Article

Cloud-Assisted Multi-RAT Steering in 5G/Beyond and Wi-Fi6 DenseNets

Luís Carlos Gonçalves[1,2,*], Pedro Sebastião[1,2], Nuno Souto[1,2] and Américo Correia[1,2]

[1] Technology and Information Science Department, ISCTE-Instituto Universitário de Lisboa, Av. Forças Armadas, Lisboa 1649-026, Portugal; pedro.sebastiao@iscte-iul.pt (P.S.); Nuno.Souto@iscte-iul.pt (N.S.); americo.correia@iscte-iul.pt (A.C.)

[2] Radio Systems Group, Instituto de Telecomunicações, Av. Forças Armadas 1649-026, Lisboa, Portugal

* Correspondence: lcbsg@iscte-iul.pt; Tel.: +351-213-130-991

Received: date; Accepted: date; Published: date

Abstract: In this work, a novel cloud assisted inter-radio access technology (RAT) steering mechanism for dense networks is introduced aiming to enable several improvements to both the mobile provider and the subscriber. Our proposed approach aims to offload calculations from the end user device. The proposed algorithm is not device triggered but controlled by a virtualized access network discovery and selection function (ANDSF) based on cloud environment, assisting the steering process and handling all complex computation in order to make complex but efficient decisions. These resources will receive real-time information and measurements from the user equipment (UE), as well as keep record of past measurements and experiences (*e.g.*, last ten Wi-Fi6 access points and their quality of service and existing or lacking traffic constraints) in order to develop a service quality-centric lifecycle for the UE and, in the end, for the subscriber itself. This way, battery power consumption on the handheld device is reduced, by having complex computations performed on cloud resources, *e.g.* quality of service (QoS) prediction. The proposed mechanisms and system models as well as the main contribution from this paper are valid for 5G backwards compatible with long term evolution (LTE) and also 5G new radio (NR), not backward compatible, and are designed to enable high performance network steering and traffic offloading within ultra-dense heterogeneous networks with millions of subscribers and connected devices with less energy consumption for user equipment's, thus distributing traffic load across different radio access in a very optimized and greener manner.

Keywords: 5G, ANDSF, Capacity planning, Cloud, DenseNet, Hybrid Cloud, Network Architecture, RAT, Resource Management, Traffic Offloading, Traffic Steering, Wi-Fi6

1. Introduction

5G NR, unlike previous generations of cellular networks, will become much more than a cellular radio definition. It has been designed to future integrate a broad platform of multiple RATs: over unlicensed, licensed and shared spectrum. Over unlicensed spectrum, Wi-Fi and its latest amendment Wi-Fi 6 (based on the IEEE's 802.11ax specifications), will play a relevant role in order to extend high quality, high speed, low latency wireless connectivity and services, especially indoors, compensating for poor cellular indoor coverage, where existing. Nowadays mobile devices are equipped with multiple radio interfaces, which support different RATs, including Wi-Fi and cellular. Current network designs, especially in 5G NR are becoming denser, from network element perspective, but also from traffic demand. On the network part, heterogeneous networks became mainstream, with macro base stations, pico and femtocells, but, in a co-existent perspective, Wi-Fi access points are also part of such equation, contributing to the densification of access points, generically. On the other hand, mobile network operators (MNOs) are continuously challenged by increasingly needs of capacity, in order to support high-data consumption behavior from subscribers, leading to several approaches regarding backhaul architectures [1]. As so, MNOs have been developing traffic offloading techniques, leveraging inter-RAT handovers, according to the RAT that, at each point, can provide better service quality [2][3].

When considering 5G NR – which is the focus on the current work – there are different services with different requirements which must be adequately addressed. For instance, enhanced mobile broadband (eMBB), ultrareliable and low-latency communications (URLLC), massive machine type communications (mMTC), have different requirements such as throughput, latency, bandwidth, but also the number of devices deployed on such networks. Therefore, in complex scenarios like these ones, Multiple RATs are considered in 5G NR but also in other technologies like Wi-Fi CERTIFIED 6™ (Wi-Fi6), in order to reach capacity and throughput requirements [4].

On the other hand, latency is one of the crucial factors that affect the QoS, especially when real-time, high bandwidth and interactive applications are used. Mobile cellular networks differ from the Internet or access networks in the fact that the latter does not guarantee QoS [5]. Internet is based on a best-effort service model, without QoS guarantee. So, by coupling such RAT, necessary to a cloud-assisted model, one would assume that service degradation would increase. While QoS methods over Internet connections do exist, in this work we focus on content delivery networks (CDN) which is an overlay technique to improve perceived quality for subscribers. As so, we assume that cloud infrastructures are provided of such transport mechanisms.

But, most importantly, the most advanced cloud service providers and infrastructures (e.g., Microsoft Azure, Amazon and Google Cloud) are fit with several direct links to Internet Exchange Points in several geographies, as well as additional links between their datacenters, which have higher quality than usual links. Therefore, Cloud infrastructures provide low-latency and very optimized data transfers, which will ultimately result in increased QoS for multimedia and interactive applications. Nowadays, cloud infrastructure is being used extensively and several critical real-time systems rely on such infrastructures exactly because of the very low levels of latency, overall quality of redundant connections and higher data throughput [**5**][6][7][8]. Therefore, cloud infrastructures can be used to enhance user experience in those applications that are expected to be most used in 5G NR and beyond networks: high-throughput, multimedia, interactive and real time, while reducing overall end-to-end delay. Also, very important, cloud infrastructures provide redundancy and fault-tolerance, in some cases enabling business continuity strategies to use cloud environments as disaster recovery and business continuity mechanisms. When processing capacity and performance is crucial, cloud environments appear as a very unique opportunity. Such infrastructures were built from the start with elasticity and processing capacity in mind. Cloud environments are being used today to replace or extend datacenters, to support services based in Internet of Things (IoT) services [8], mainly due to its elasticity capabilities and scaling performance and also with the most recent advances in Software Defined Networks (SDN), can also be applied to intelligent 5G NR networks for enhanced resource management [9][10][11]. From this perspective, cloud becomes a good choice when considering high computational and capacity needs, and a good enabler for network sharing [6][12][13][14][15][16][17]. Figure 1 presents the considered system model: a DenseNet both Indoor and Outdoor which at the same time is a HetNet with dual RAT technologies available: 5G NR and Wi-Fi6.

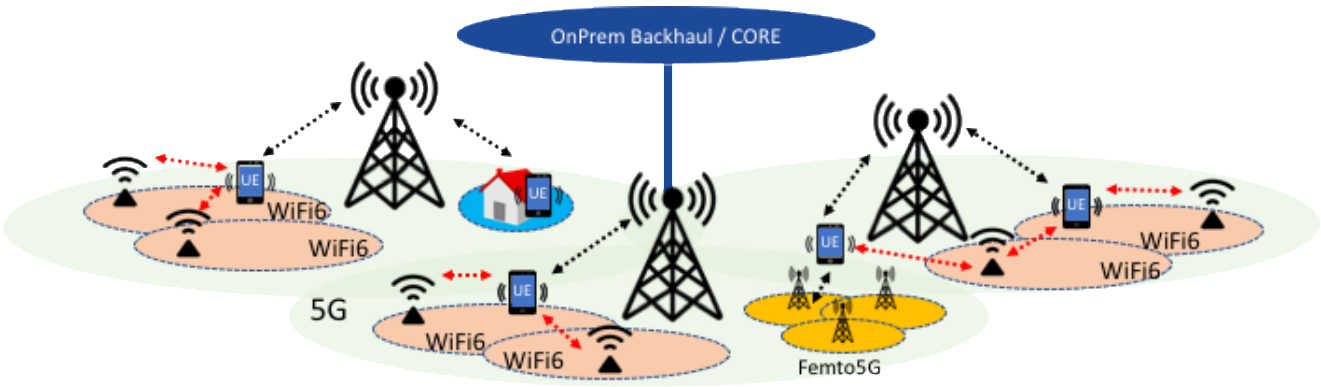


Figure 1. Considered model: dense HetNet indoor/outdoor (5G + Wi-Fi6).

Considering such advantages, in this paper we propose a fully cloud-based assisted traffic steering mechanism, controlled in the backhaul (and not by the UE) through high performance computation, low-latency links and high level of redundancy, overall contributing to a holistically enhanced service delivered to the subscriber, either over 5G NR or Wi-Fi6 radio access links. All based on a heterogeneous scenario of dual overlapped coverage of Wi-Fi6 and 5G NR networks, with macro and small cells as well as indoor femtocell deployments. Backhaul triggers and controls which Wi-Fi6 Access Point (AP) or 5G NR Base Stations the end user's device must steer and receive traffic from, depending on measured QoS criteria through cloud-based ANDSF virtualized function. The proposed mechanisms in this work are valid for next generation NodeB (gNB) providing NR user plane and control plane terminations towards the UE, for 5G NR access. Therefore, for the sake of simplicity all references onwards will be to radio base stations (RBSs). Nevertheless one should refer to Release 15 [18] and 16 [19] and aim to demonstrate that datacenter ANDSF servers should become cloud virtualized functions as soon as possible.

The following sections will briefly introduce Wi-Fi6 and focus on the ANDSF network virtualized function. The main contributions of this paper are presented under section IV, and the introduced cloud assisted model is presented in section 5. Finally, conclusions are drawn in Section 6.

2. Wi-Hi Hotspot 2.0 and Wi-Fi Certified 6

Wi-Fi Passpoint 2.0 [20] and Wi-Fi6 [21] are two solution that have been developed in order to enhance Wi-Fi to a level of quality that can enable seamless integration with cellular networks. Passpoint 2.0 has been enabling next generation Wi-Fi connectivity, providing enhanced capacity, coverage, and performance needed by subscribers, when using high-demanding applications. Although it is not the focus of this work, a brief explanation of the most relevant features for the presented model will be referred.

Wi-Fi6 brings additional capabilities which are very relevant from a DenseNet perspective. By considering the usage of Wi-Fi6 HetNet in this work, we expect an additional decrease of battery consumption in UEs, adding to the already lowered consumption provided by the cloud assisted steering mechanism proposed. Wi-Fi6 therefore, will increase even more the quality of use cases such as smart home and IoT uses within Dense HetNets [22]. The key of Wi-Fi6 include higher data rates and increased capacity and improved power efficiency. Considering the progressively more demanding behavior of subscribers, Wi-Fi6 is the perfect

combination to deploy along 5G NR and beyond, which, coupled with the proposed cloud-assisted mechanism will enhance the quality of services like streaming, ultra-high-definition movies, mission-critical business applications that require low latency and high throughput. Wi-Fi6 enhancements, based on pre-existing Passpoint 2.0 features, will support DenseNets over Wi-Fi RAT.

Wi-Fi6 features enhances orthogonal frequency division multiple access (OFDMA) in both uplink and downlink, which fits perfectly as a dual RAT for 5G NR and beyond networks. OFDMA will impact directly on increased radio efficiency thus lowering latency over ultra-demanding capacity scenarios. Additionally, enhancements to Multi-user multiple input, multiple output (MIMO) have been developed from the pre-existing in Passpoint 2.0. Such enhancements will increase the volume of data that can be transferred at a given time, which will, therefore, enable more subscribers to be served simultaneously on the same access point. Not exhaustively, battery life in UEs is expected to increase due to power consumption reduction through target wake time mechanisms and, finally, increased capacity and higher throughput at farther distances will also be enabled by beamforming technologies.

Wi-Fi6 enhancements enable it as a RAT that can co-exist with cellular networks, namely 5G NR, from a reliability standpoint. Quality of service has been a feature of Wi-Fi for long, replacing unreliable best effort mechanisms presented in the first technological releases of Wi-Fi. As so, Wi-Fi6 increases even more the reliability for services like e-Learning, online gaming, telepresence and telework, healthcare monitoring, but especially all low-latency and high throughput demanding services like video streaming both in downlink and uplink. It evolved from a best-effort RAT without QoS to a full QoS enabled RAT where, for instance, VoIP is one of the future main services, one of the most sensitive to delay and lack of network capacity. [23]. Wi-Fi6 also provides enhanced capabilities to support next generation highly capacity demanding environments like retail settings, stadiums, and transportation hubs, therefore, becoming a very promising second RAT to co-exist with 5G NR, especially for traffic offloading purposes on such environments, and in conjunction with ANDSF, the end user will not have to manually select and attach the device to WiFi6.

3. Access Network Discovery and Selection Function

ANDSF [23] is a discovery and selection algorithm defined by 3GPP, in order to allow User Equipment (UE) to discover and access non-3GPP RATs.
It is deployed on the Evolved Packet Core (EPC) as a standalone server, using standardized interface S14 to distribute selection information and policies to UEs, according to 3GPP specifications [25]-[29]. We assume that 5G NR Cores efficiently

provisioned [30] as a baseline to further enhance traffic flows and radio access selection and traffic offloading. Specifically, ANDSF provides UEs with three types of information sets, depending on each MNO configurations: Inter-System Mobility Policy (ISMP), Inter-System Routing Policy (ISRP) and Access Network Discovery Information (ANDI). ISMP addresses UEs capable of routing traffic over a single RAT at a time, while IRSP addresses UEs that can simultaneously route traffic over different RATs. ISMP provides policies for selecting which RAT to be used for routing all data traffic, while IRSP provides policies defining which traffic should be routed over what RAT. Finally, ANDI provides UEs with a list of access networks available in its vicinities. Such optimizes overall UE energy consumption and, when coupled with cloud-based services, such consumption can be further reduced [30], especially under a hotspot scenario [31]. By applying the right ANDSF policies, MNOs are able to start steering cellular users to Wi-Fi hotspots when load levels grow from low to medium and additional users when cellular load is high. The following subsections will present the three most important features that support highly reliable traffic offloading between RAT.

3.1. Selected IP Traffic Offload

Selected IP traffic offload (SIPTO), as depicted in Figure 2 is a traffic offloading technique. With SIPTO supported HetNets networks, all traffic generated in the UE that is destined to the Internet, is routed directly through existing broadband Internet access, without traversing the mobile core network. Considering that the majority of Internet access through cellular connections is generated indoors, this technology results in both cost reduction and traffic offloading from customer and MNO points of view, respectively [33][34].

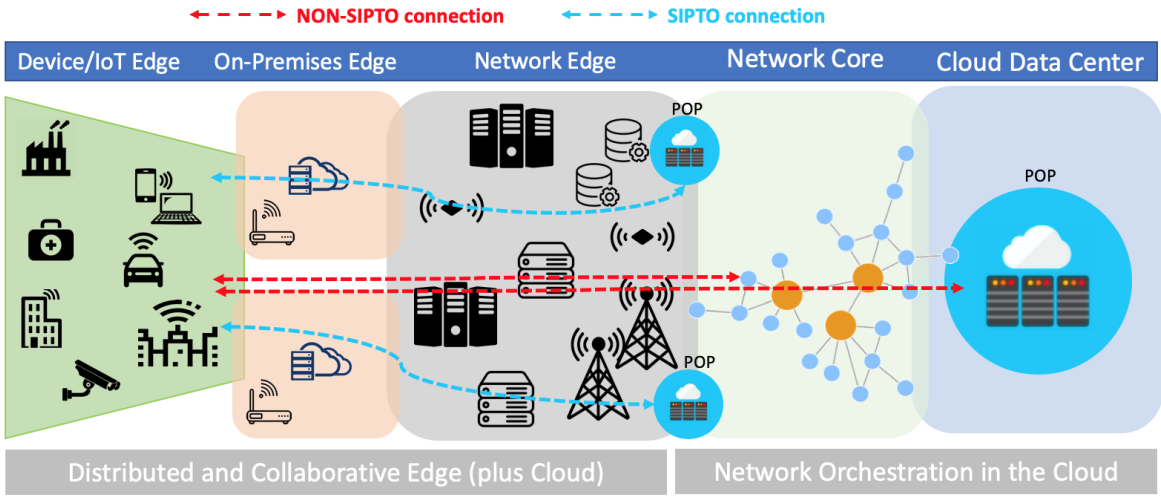


Figure 2. SIPTO for traffic offloading.

SIPTO can be performed or operate within the home premises, operating within the femtocell environment (On-Premise edge) or it can operate in the next edge, the Network edge, at Macro level. SIPTO can use Internet points of presence (POP) to traffic data directly to the Internet on edges, through the client on-premises edge or the network edge.

If such mechanism did not exist, traffic would be routed through all edges, into the network core and from there to the Internet. We define distributed and collaborative edge (DCE) as the Device/IoT, On-Premises and Network edges, and network orchestration edge (NOE) as the network core and cloud data center edge.

3.2. Local IP Access

Local IP access (LIPA), as depicted in Figure 3 is a traffic routing technique similar to SIPTO, particularly important on indoor network premises allowing for data offloading from cellular connections. When a UE supports LIPA, traffic destined to local IP networks (either residential or enterprise) is offloaded locally through the Wi-Fi transceiver and directly routed to those networks, without traversing the mobile network core, as would happen if a direct interface local IP networks would not exist. LIPA enables data offload locally to existing IP networks instead of routing it through the MNO's core network and back, as shown in Figure 2.

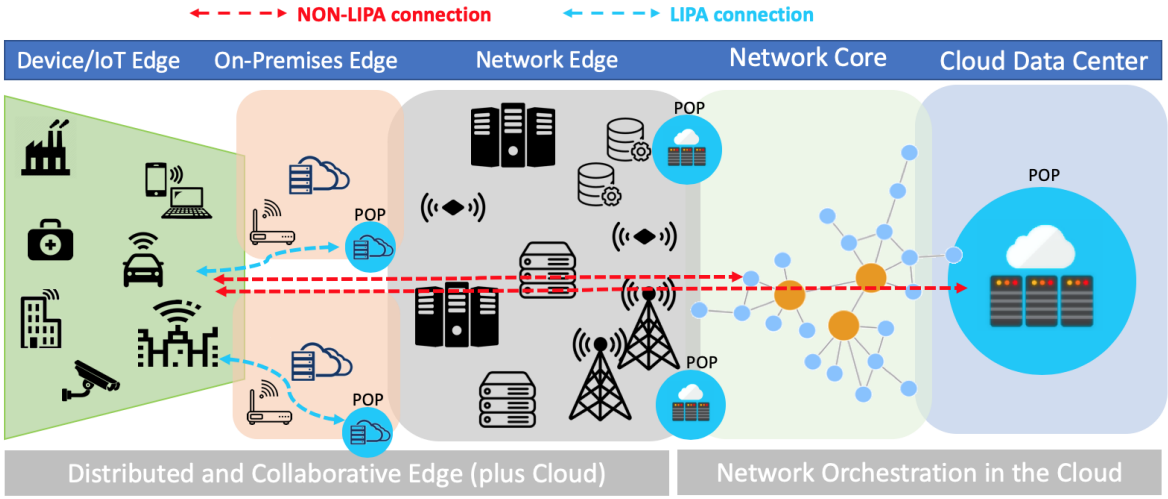


Figure 3. Local IP access for traffic offloading.

LIPA which is today an improved mechanism will be used to allow traffic to be routed directly from the UE to the ANDSF function through a dedicated channel rather than S14, reducing this way the signaling overhead over that channel [35][36]. In short, UE's information to and from ANDSF virtualized function will

be conveyed through S14 when the UE is connected to 5G NR gNBs and through LIPA when the UE is connected to Wi-Fi6.

3.3. 5G NR Edge Computing and Offloading

Based on LIPA and SIPTO, but from an edge computing perspective, with 5G NR enhanced architectures, optimal placement of offloading point will depend on the type of service. If one considers the concept of edge computing as the new reference architecture, as depicted in Figures 2, 3 and 4, ultra-low latency services with high bandwidth can be a reality.

The expected volume of data generated in 5G NR and beyond networks will force mobile network operators to have the possibility of scaling up, down or move computing resources and storage capacity accordingly to where it is needed, between edges. As such, increased virtualization of the core networks and the radio access networks will be unavoidable. By virtualizing or moving functions to the cloud, as such process becomes closer to the customer devices, enhancements in overall service quality will appear, especially from a latency point of view.

Thus, new service enablers will appear, and as depicted on Figure 4, Edge traffic offloading to ANDSF function can occur directly on the device edge and increasingly also on the customer premises edge. Such will enable very low latency round trip communication between end user devices and the ANDSF function to decide on RAT offloading, improving the overall process.

3.4. Transparent WiFi HotSpot Discovery

When considering HetNets from a mobility perspective, it is of utmost importance for the UEs to be able to transparently discover access points where to attach to as they become available while a user is moving. Additionally, the ability to distinguish the APs from other existing APs', ISPs' or domestic and enterprise's Wi-Fi hotspots is crucial.

In that sense, 3GPP introduced the access network discovery and selection function for Wi-Fi hotspot discovery. This feature consists of an operator-side component, in charge of storing operator policies towards transparent discovery and selection of Wi-Fi hotspots. Based on the location of the UE, this function maintains a list of which Wi-Fi hotspots are on the vicinities of the UE and are capable of receiving a transparent handover from cellular connections. Thus, ANDSF function enables the UE to distinguish the Wi-Fi hotspots from other service provider hotspots, guaranteeing seamless handovers within the HetNets. The usefulness of this feature is extent to a higher level through the possibility of deploying location-based policies, allowing a UE to select certain hotspots over

others on the same location based, e.g., on the kind of service being used or the capacity level of each hotspot [37]-[39].

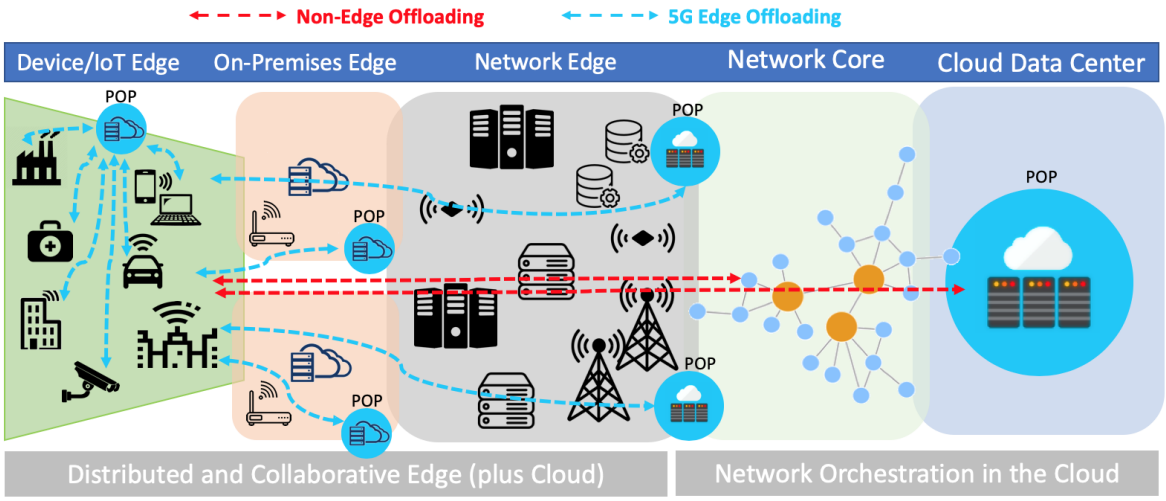


Figure 4. 5G NR Edge traffic offloading for ultra-low ANDSF communication and traffic offloading.

4. Main Contributions

The contributions of this work are three-fold and described in further detail below. The first contribution is related to the ANDSF function which we propose to be hosted in a cloud environment in two different ways: hybrid core and Cloud based. The second contribution is the deployment of a secondary channel to convey network-related information that will contribute to the steering decision. Finally, as a result of the first two contributions an enhanced cloud-assisted and network-initiated methodology is proposed for network steering between 5G NR and beyond and Wi-Fi6 networks.

4.1. Cloud-Based High Performance ANDSF Virtualized Function

Typically, the ANDSF is a physical server and is located within the core network. This greatly limits the overall performance of the whole steering process, when ultra-dense networks are considered, with millions of subscribers, several RATs and heterogeneous networks. Either the ANDSF is located on the core and steering decisions are left distributed to the UE – with all the disadvantages previously referred – or the ANDSF function is highly scalable, and capable of performing massive calculations in real time, while reducing latency in communicating the steering decision. That said, such capabilities can be supported by a cloud virtualized ANDSF function, which can be provided as an infrastructure, service or platform as a service (IaaS, SaaS or PaaS).

In reality, by virtualizing network functions, one is pursuing disaggregation of traditional network equipment, which got converted into functions. Such is the aim of virtualizing the ANDSF server into a function, which, in practice means building a different network, by separating tightly integrated systems into its individual components. The concept of disaggregation further allows, on a second stage, to recombine individually enhanced and optimized elements in a more efficient manner, which is proposed in this work.

By providing such flexibility in choosing which components are used, the component with the best attributes can be used. Those attributes can be cost, total cost of ownership, scalability, latency, *inter alia*. Another advantage of disaggregation is that each component can come from different providers, avoiding vendor lock-in, while allowing the introduction of new functions to the system, without having to reconstruct the entire system.

The main disadvantage of disaggregating is the additional complexity and the need for increased interoperability between virtualized functions which are not part of the same integrated system anymore, especially when such functions are plain and simple software (software defined networks - SND or software defined radio - SDR) running on a serverless approach. Specifically, the major challenges are how components interact and communicate with each other and how traffic flows are mapped through the different components. In order to achieve this, each virtualized function needs to provide an application programming interface (API) between disaggregated components.

By converting the ANDSF server into a virtualized function, we take a step further in the direction of edge computing and full network function virtualization. Such is what is proposed in two different models which are presented in Section 5 where they are compared with the classical physical ANDSF server located in the MNO datacenter's Core Network. The process of disaggregation has three main advantages:

1. Cost: the use of open-source hardware and software, but especially cloud environments, will lead to cost reductions also from a market competition perspective between vendors.

2. Feature Flexibility and Functionality: in the case of aggregated systems, adding features or functionalities must be performed according to a specific change cycle and depends on new hardware and software releases. On the other hand, upgrading the software on an integrated system often requires taking the system offline. In a disaggregated architecture, each virtualized

function can be added or changed, without impacting the other ones, as new software modules are developed. The best example of such advantage is the usage of cloud workloads and virtualized functions.

3. Ability to Scale: in a disaggregated architecture upgrades and enhancements happen only to those components and functions which are virtualized. All other remain untouched. If, e.g., a forwarding/user plane virtualized function needs to be scaled up due to capacity or bandwidth temporary or permanent scarcity, this can be accomplished without replacing or upgrading the control plane. On the other hand, if a virtualized function like ANDSF needs to be upgraded either in computing performance - more virtual central processing units (vCPUs) or mode memory – it can be performed on a disaggregated cloud instance where the virtual function lies, without any impact or change to the user control plane.

Such disaggregation and network function virtualization approach will contribute to reduce overall latency in steering calculations and decisions, while making available a totally scalable, elastic and also highly redundant server, taking advantage of native cloud architectures. Additionally, the ANDSF function can benefit from a CDN which are very mature and efficient mechanisms that several cloud providers use, reducing decision latency even more [5][12] [40][41]. By following this approach, the ANDSF function will no more be prone to become the single point of failure or the bottleneck in the whole inter-RAT steering process, becoming highly adequate to ultra-dense and highly performant heterogeneous networks like 5G NR plus Wi-Fi6. This method also minimizes the probability of a saturated ANDSF function providing steering decisions when the network conditions are already different and not the best anymore.

This method will relevantly increase overall efficiency of a core based ANDSF function. Two configurations are considered for the ANDSF function located on cloud environments: a Cloud-attached Hybrid Backhaul or a Cloud Detached Backhaul, as presented in Section 5 where they are evaluated on the three aspects upper referred: cost, feature flexibility/functionality and ability to scale.

4.2. Cloud Assisted Network-Initiated RAT Steering

Typically, RAT selection and data offloading is performed by two different methods: UE-side initiated [42]-[47] and network-side initiated [45]-[48]. When a UE which is already associated with one RAT finds itself under coverage of an additional one – under both RATs –, an evaluation process should be triggered which can result on a new association and inter-RAT handover, if better service

conditions exist. Such mechanisms were first introduced by 3GPP in its Release 12 specification and object of several enhancements afterwards [40].

Selfish Steering and Coordinated Fairness

UE-side initiated steering has been largely discussed and it is not the objective of this work to deeply focus. Positive aspects focus on the fact that devices and UEs on the device edge are in the best position to evaluate its service quality, according to the current service and application usage and also subscriber requirements. Also, it is taken into consideration that UEs are becoming more performant and advanced with higher computing capabilities, meaning that they are able to perform and conduct the necessary calculations in order to evaluate the best RAT to perform service on. We agree on those perspectives but, considering 5G NR and the 5G NR with ultra-dense heterogeneous networks, UEs will not have all the necessary capacity to perform such decisions alone. The environment in the vicinities will change very rapidly and for applications with latencies under 1ms as expected in 5G NR systems, the UE will tend to ping-pong (perform extensive handovers, especially if on a mobility scenario) among RATs. This will overload and increase congestion and queueing levels on ANDSF server instances, bringing down battery capacity thus reducing the UE availability. Each UE will tend to perform its own steering decisions in order to maximize their utility functions, without taking other UEs into consideration. This will lead to eventual performance impact on other UEs in the vicinity [50]. If one considers eventual malicious software or applications running on an UE, there is the possibility of providing false information and measurements to the ANDSF server, leading to a potential impact over the whole network and other UEs, which cause resource scarcity on the whole system if such malicious application is running on several UEs [51][52][53]. Finally, lack of a macroscopic awareness of networks and other UEs, as well as ANDSF server congestion levels, call for a centralized function to coordinate all measurements, network awareness and steering decisions, considering potential overall impact on the network.

For all those reasons, instead of a UE initiated selfish steering process, that can create performance drop or service disruption, we choose to focus on coordinated fairness in order to decide steering directions for UEs.

On the other hand, a network triggered RAT steering algorithm enables a more holistic vision of the whole network which can lead to better steering decisions and also enable advanced features like prioritization which ultimately can lead to increasing the ARPU. Taking this as an example, high-priority users will always have 5G NR as preferred network, while low-priority users can be steered through Wi-Fi6 or have a small portion of 5G NR capacity. This would allow increasing the

ARPU by partitioning the resources and RAT according to those subscribers that are willing to pay premium services for dedicated resources, while others use shared resources. The existence of a central coordinated and fairness oriented ANDSF controller function can periodically check all APs or RBSs for availability and resource management, from a centralized perspective. And it is up to this centralized controller to, within the Core, gain knowledge about the list or priority users and share resources considering that perspective, while not creating impact on the overall network. This is something that cannot be performed from a UE perspective, simply because it does not possess such information.

So, basically, the ANDSF virtual function can be fit with so many mechanisms in order to properly enhance and optimize resource management, while, at the same time, it keeps the whole network from suffering sudden performance drop or service disruption. If one considers congestion levels, the centralized and coordinated ANDSF function has the ability to take into account congestion levels in all networks and, to maximize UE's QoS, by feeding it a list of uncongested networks, to which it can steer traffic to. The UE itself, and especially if the decision algorithm is based only in signal strength measurements, may choose the RAT with the highest signal to noise ratio and steer traffic to that RAT which may already be congested, increasing the whole problem. The ANDSF centralized function, especially when located in cloud environment, will have the ability to perform all those calculations and autonomously scale up when additional computing performance is needed. Thus, congestion analysis is always performed, and all steering decisions sent to the UE will contribute to overall increase system performance.

Methodology

Taking into consideration the above aspects, we consider that the most adequate model is a mix of the two approaches. There still is UE-centric measurements in order to achieve their desired QoS, improve data throughput and reduce latency and attain its QoS requirements. We assume that the UE's behavior is enhanced, enabling always-best connectivity [54][55]. Nevertheless, we feel that for future proof and 5G NR networks, the ANDSF function should be enhanced with additional features as referred, conveying the UE the best network list for it to steer to. Our approach is depicted in Figure 5. This centralized function, on the backhaul will perform all necessary calculations in order to choose the best RAT at each giver time for each UE under dual coverage, taking into consideration the whole network and, especially, surrounding UEs. Such function would additionally benefit from a congestion watchdog process, that would enable it to perform load sharing or addition additional functions to the virtualized ANDSF

function, benefiting from the fact that it is cloud based, meaning that its processing capacity is scalable and elastic, and will not constitute a bottleneck when doing all the calculations for proper RAT steering. This is the point where the advantages of having a cloud virtualized ANDSF functions stand out. In a direct comparison with an ANDSF server located on an MNO's datacenter, in case of congestion due to 5G NR ultra-dense networks' UEs always trying to steer to the best RAT from their perspective, the cloud based ANDSF virtualized function has the ability to scale up automatically, continuously evaluate its queuing levels and congestion, and scale down when necessary. All is performed automatically, and also new ANDSF virtualized functions can be added to the existing ones, something that it is not possible on physical ANDSF servers locate on datacenters. In such a scenario, the same ANDSF server would face congestion, underperform and create impact on the overall system, eventually leading to service disruption or denial of service of the ANDSF function due to lack of additional resources. This is the main reason why we introduce the cloud virtualized ANDSF function.

In the proposed methodology, the virtualized ANDSF server will assist the network steering through a set of new mechanisms. As depicted in Figure 5, on the Device/IoT edge, it is up to the UE to perform the normal operations in this kind of mechanism, according to what was previously referred. The ANDSF virtual function is placed within cloud instance, on the cloud core edge and besides performing the normal ANDSF functions, it is also is enhanced with a congestion watchdog, that exploits the flexibility of cloud environments, in order to scale up, down or increase and decrease performance. The fact that it is a virtualized function also enables ANDSF to perform additional tasks in order to increase overall performance of the steering process. We consider, for instance, the ability to apply supervised and non-supervised learning, as well as artificial intelligence to perform advanced decisions based on the data that it receives from the device edge, but also from the results of continuous enumeration and calculation about the quality of available networks/RATs. Nevertheless, such is not the focus of our work, but solely a reference to a possible future development.

As so, in our work we consider this new virtualized ANDSF function quality of service and, therefore, in our system model the UE will focus solely on measuring quality indicators and conveying them to the Core edge, directly to the cloud ANDSF virtualized function, without performing any complex calculations and processing (which results also in higher battery life and better user experience in the device itself). We believe that this enhanced network function will lead to overall enhanced capacity and is additionally capable of higher performance levels due to the introduction of the cloud assisted processing, thus guaranteeing optimal

traffic steering and offloading, maximizing per subscriber system throughput and QoS.

4.3. Separate Data Channel

The proposed approach has another main benefit: reducing the overall signaling load over the S14 channel. By having the ANDSF function on the Cloud, independently of the cloud deployment model, the overall traffic of the S14 interface is reduced, due to direct communication of the UE with the ANDSF function. This is a very relevant aspect which will contribute greatly to reduce overall network load, but at the same time, to provide network measurements to the ANDSF function securely and with low latency associated. In the case of indoor located UEs, their information can be relayed directly to the cloud based ANDSF function, through the Internet connection that supports an indoor 5G NR femtocell and Wi-Fi6 access point.

Via backhaul network, cloud ANDSF virtualized function will receive real time data from end user devices, satisfying all subscribers, but on a side channel. After processing is complete, the cloud based ANDSF function(s) will send the results back to UEs. As main advantages, network traffic balance can be achieved and traffic peaks and S14 channel saturation can be avoided, as well as by taking advantage of the high availability characteristics of cloud servers. Transmission latency decrease would then be achieved regarding UEs, reducing overall response times compared with the traditional mechanisms.

5. Proposed Cloud-Assisted Models

As referred, we propose two different models to address the limitations of current ANDSF deployments, in order to prepare upcoming 5G and 5G NR. Aiming at improving overall network performance and resource management, we first present the two proposed ANDSF virtualized function models and finally compare them with the classical datacenter physical ANDSF currently in practice.

5.1. Architectural assumptions

We assume that 5G NR and Wi-Fi6 cells all belong to the same MNO. For the sake of clarity, we do not focus solely on current early 5G NR deployments. We consider the full 5G NR deployment, where the 5G NR Core (5G NRC) will still support evolved packet core (EPC), but only standalone deployments will exist. Additionally, we consider full availability of new capabilities described in Release 15 and Release 16, especially network slicing, urLCC and mMTC.

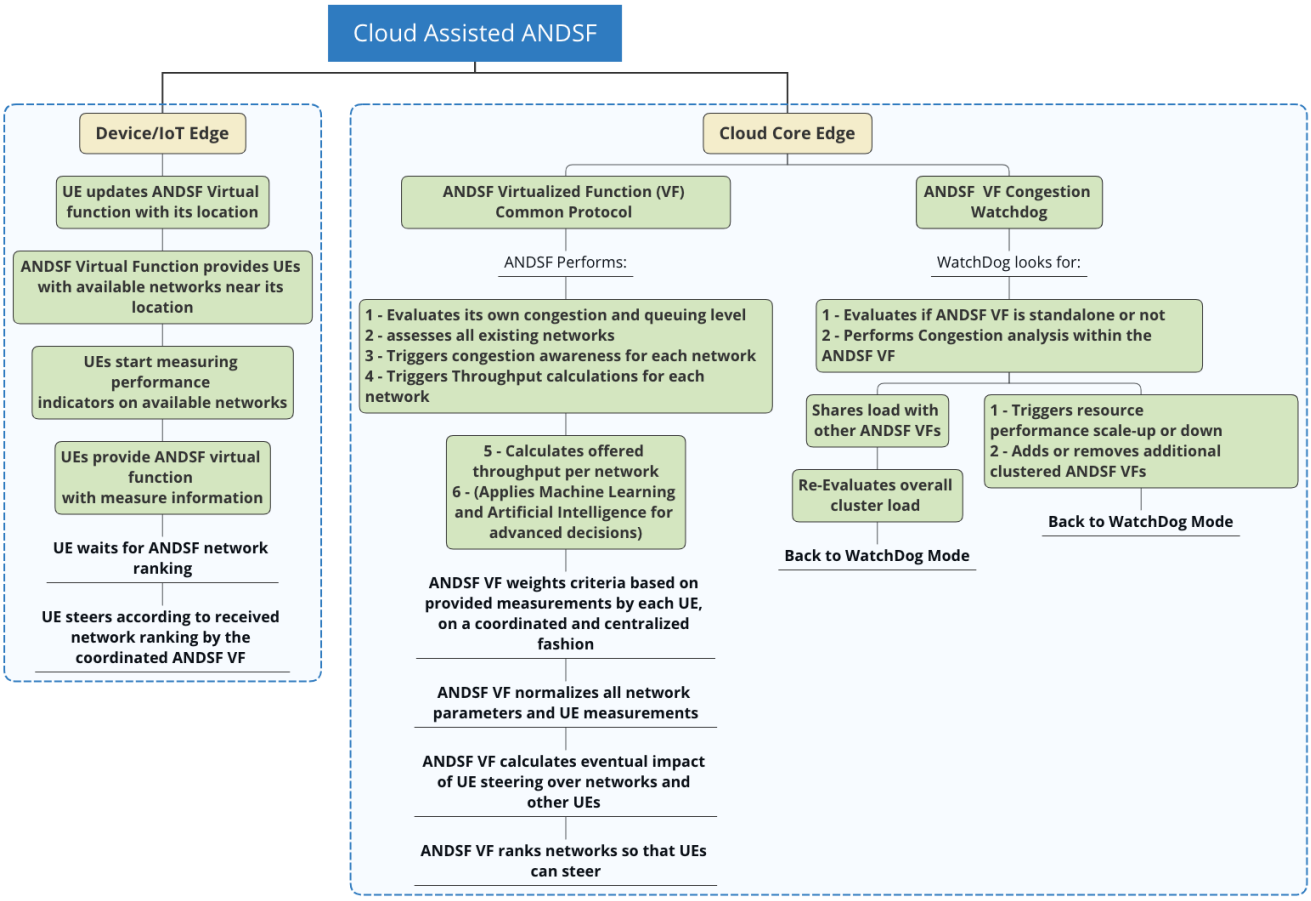


Figure 5. Cloud assisted ANDSF.

Without further detailing the architecture, as it is not the focus of this work, as referred previously, we consider the existence of matured network function virtualization (NFV) available, and SDR networks [53][56]. This will enable the main focal point of this work: virtualization and disaggregation of many of the RAN and mobile core functions and increased edge computing.

Therefore, we consider the Cloud-RAN approach as the basis to our proposed architectures, and, where considered, Split RAN splitting baseband functions into function blocks, which can be virtualized, and distributed in order to enhance spectral efficiency.

From the end device perspective, we consider that UEs are pre-configured by MNOs configuration regarding access to Wi-Fi hotspots and can be updated on the fly via ANDSF. In order to simplify the overall model, and because it is not the focus of the current work, we assume that both the gNBs and the APs backbone connections converge to a centralized controller through lossless links, i.e., no latency or signal/data degradation is assumed in the links. We also consider the existence of CDN that greatly reduces latency between the cloud based ANDSF

and the UEs. Femtocells that are located indoors will take advantage of their broadband connection, maintaining the advantages of contacting directly the cloud ANDSF function. As referred, we consider two different network deployment options: section 5.1 presents a Hybrid Datacenter approach, which can be more expensive, but more secure and reliable, due, especially to the fact that there needs to be direct links between the MNO Core and the Cloud provider datacenters. Section 5.2 presents a less expensive solution and with higher time-to-market, using a fully detached cloud model, with connections between the MNO Core and the Cloud supported over the Internet through CDN. Nevertheless, the two scenarios are presented and Figure 5 and 6 show the proposed architecture for each one.

5.2. Model 1: Cloud-Attached Hybrid MNO 5GC

The cloud-attached hybrid MNO backhaul is depicted in Figure 6. It consists of two main edges: i.) the DCE, which comprises the device/IoT, on-premises domestic and network edge, without any NFV considered and no cloud access and ii.) the introduced hybrid NOE, with two main components: the typical physical MNO datacenter with its on-prem architectural components and a cloud provider based extension to that datacenter, where part of the core functions reside for flexibility and scaling.

Both components are connected through a dedicated channel and NFV is considered in both parts of the hybrid network orchestration edge. Finally, the service edge, the Internet, is connected to the physical datacenter through two dedicated internet service provider links, for load balancing and high availability. In this proposed scenario, the ANDSF function is virtualized within the cloud data center. Such virtualization means that by being on a cloud-scale datacenter, the ANDSF function is already using distributed processing, is designed to scale up or down accordingly to contextual needs, and is also redundant, thus properly prepared to handle increasingly flows of data from the 5G NR distributed and collaborative edge.

Such flows and all communications and measurements originating from devices within that edge, (indoor via Broadband, Wi-Fi or 5G NR gNBs, or outdoor via Wi-Fi or 5G gNBs) are channeled by LIPA and conveyed through the network core (MNO Datacenter) and dedicated cloud links to the virtualized ANDSF function in the Cloud datacenter.

The MNO's leased dedicated communication circuits from its datacenter to the cloud provider datacenter, guarantee total data encryption but, especially, very low latency in data transmission.

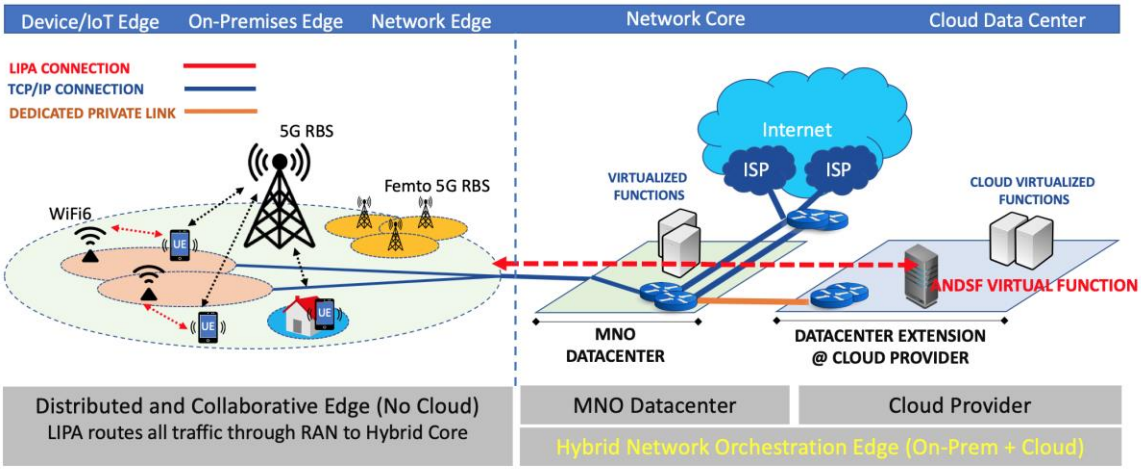


Figure 6. Cloud-attached hybrid MNO 5GC.

This way, no communication goes through the Internet, which is a best-effort solution not guaranteeing latency minimization as well as increased security on all data flows. Finally, it is up to the ANDSF virtualized function to perform all processing and issue the result to each device located on the DCE in order for traffic steering to occur. All processing occurs on the ANDSF virtual function due to its elastic computing capacity and performance, when compared with a classical physical server in a datacenter [57].

This model is a mid-term approach to 5GC, by enabling converting the ANDSF server into an NVF by adopting an attached private cloud environment.

5.3. Model 2: Cloud-Detached Hybrid MNO 5GC

The cloud detached hybrid 5GC is depicted in Figure 7. It consists of the same two edges as depicted on the previous model, the DCE and the NOE. The main difference is that, in this model, the cloud instance is detached from the MNO datacenter, meaning that there are no dedicated circuits between the datacenter components and the cloud virtual functions. The direct impact is that the assumption that the flows between both datacenter components do not have latency cannot be sustained anymore, as such connection is supported on the Internet, on a best effort paradigm. Thus, an additional latency must be considered between communicating devices from the DCE and the ANDSF virtualized function in the cloud.

Such flows and all communications and measurements originating from devices within that edge are routed by SIPTO directly through the MNO's Datacenter over the Internet links, to highly performant CDNs and finally to the virtualized ANDSF function in the Cloud datacenter. As assumed, CDNs relevantly reduce the additional latency.

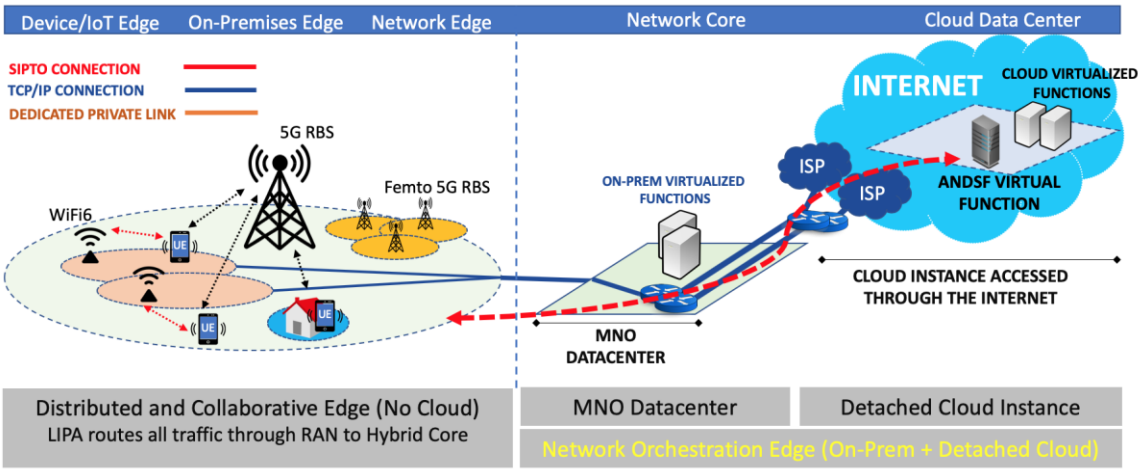


Figure 7. Cloud-detached hybrid MNO 5GC.

This architecture is not as secure as the previous one, especially considering data traversing the Internet, but is the closest to the full network function virtualization model, allowing for almost seamless and direct communication between all edges where NFV is present and the cloud instance. As an example, if one considers the POP and Internet access in the Device/IoT edge, all communicating devices may directly contact the virtualized ANDSF function through the Internet and receive the required information to better choose and steer traffic.

5.4. Discussion

In both scenarios, the overall RAT evaluation is performed, at first, with the UE reporting its cell ID or accurate GPS location information to ANDSF function. The ANDSF checks the map database of WLAN, finds out the nearest WLAN hotspot at user's location and delivers the WLAN network discovery information to the UE, which will then provide the ANDSF function with the necessary information for best RAT calculations. The ANDSF will perform such calculations and signal the UE which is the best RAT where it should steer to or, alternatively, indicate that no steering should be perform at the moment. All this process is repeated for all UEs in real time, weather located indoor or outdoor, and decisions are performed by the ANDSF function.

As referred in subsections 4.1, 5.1, 5.2 and 5.3, our approach is based on network function virtualization of the ANDSF function. By doing so, a step further in the direction of edge computing and full network function virtualization is taken. In this section comparison between the two proposed models and the classical one is presented, focusing mainly on the three principal advantages: cost, feature flexibility/functionality and ability to scale.

5.4.1. Total Cost of Ownership

Model 1 is the most expensive model for an MNO to support, from a capital expenditure point of view. The cost of dedicated circuits physically connecting the two components of the NOE largely surpasses the cost of Model 2, which only needs redundant Internet service provider links. Additionally, the trend in 5G NR will be to increasingly virtualize network functions, meaning that the operational cost of Model 1 will become higher due to the fact that part of the virtualized functions will tend to stay in the on-premises part of the NOE, instead of migrating as much as possible to the detached cloud environment, thus reducing considerably the operational costs.

We consider basic requirements for an ANDSF server as presented in Table 1.

Table 7. Minimum server requirements.

Service	Quantity (Physical)	Quantity (Virtual)
Number of Latest Generation Intel Cores	2 HexaCore/Server	2 vCPUs
RAM	8GB	8GB
Storage	1000GB HDD	1000GB SSD

For the total cost of ownership, we consider that the amortization period is 4 years (48 months) and that the overall service availability is 99.9% of the time, to have a comparison with the usual cloud offering. In this case, for guaranteeing 99.9% availability, a three-tier cluster is needed, meaning that three physical servers are needed.

We have designed such servers in cloud offerings from Amazon, Microsoft, SherWeb and HostGator and value difference among all offering is < 7%. For the physical server we have created the configurations through the online configuration pages from Dell and HP and the prices for the physical server differ <5%. All values consider manufacturer suggested retail price (MSRP), without any kind of discount taken into consideration. As so, directly comparing the cloud offering with the physical offering, Table 2 presents the prices and features available in both cases:

Table 8. Physical vs cloud comparisons.

Service/Feature	Physical	Cloud Virtualized
Average Monthly Cost per Server	~1700 € / Month	~ 490 € / Month
Elasticity	minimal	Very high
System availability	99,9%	99,999%
System refresh/amortization	48	Not Applicable
24 X 7 live support	N	Y
Free workload migration	N	Y
Enterprise class SAN all SSD disks	N	Y
Autonomy without local power	N	Y
Redundant Locations	N	Y

As it can be seen, the usage of Virtualized cloud functions not only reduces the monthly cost for the network function itself, but also introduces several advantages, especially when considering uptime and business continuity. In the case of lack of local electrical power, the cloud virtualized function will still be online. If the cloud datacenter loses energy, automatic failover exists to other datacenters located in other regions, thus, no service disruption exists. So, from the MNO's perspective, overall quality of service can be expected, at lower costs. Figure 8 presents the cost comparison between both monthly prices.

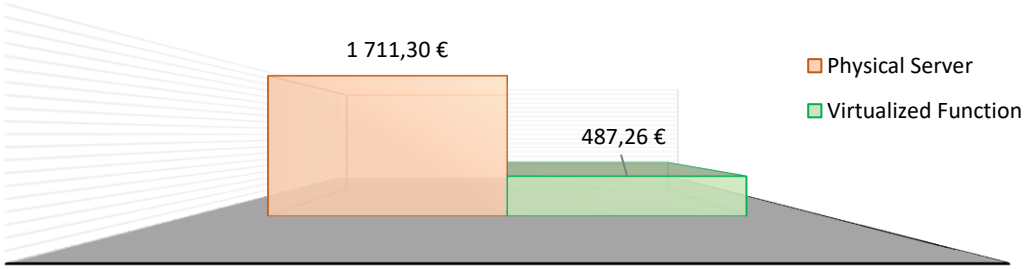


Figure 8. Average monthly cost comparison for physical and virtualized function.

Based on the above results, it can be seen that the average monthly saving is ~72%, meaning that, for the MNO, the virtualized function will represent a cost saving of 72% per virtualized function. But, as we considered 48-month period for server amortization, due to rapid technology change, it is of most interest to have an idea of the total cost of ownership in a 7-year time span. As so, based on the calculated values, Figure 9 shows a seven-year time span TCO, in order to allow for physical replacement to take place within the MNO's Core datacenter, and respective cost consideration.

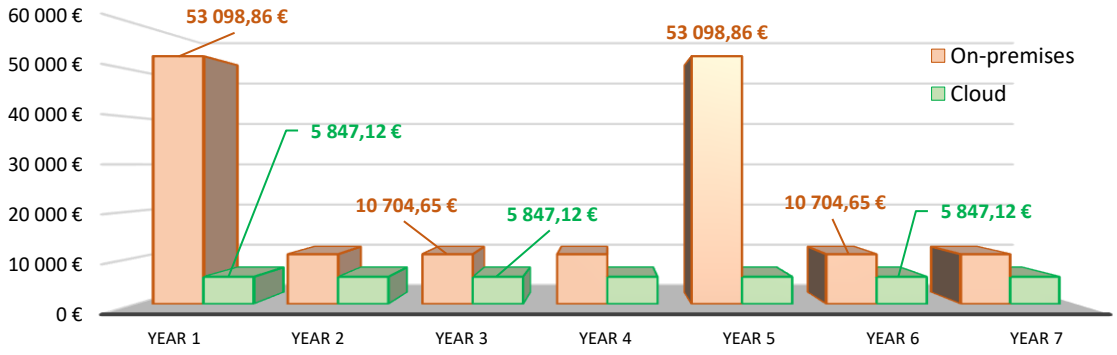


Figure 9. Average yearly cost comparison for physical and virtualized function.

The average saving in the period of 7 years is ~74%. Table 3 presents the yearly breakdown and savings.

Table 9. 7-Year yearly savings and total TCO saving

Year	Cost Physical [€]	Cost Cloud Virtualized [€]	Savings [€]	Savings [%]
Year 1	**53 098,86 €**	5 847,2 €	**47 251,73 €**	89 %
Year 2	10 704,65 €	5 847,2 €	4 857,52 €	45 %
Year 3	10 704,65 €	5 847,2 €	4 857,52 €	45 %
Year 4	10 704,65 €	5 847,2 €	4 857,52 €	45 %
Year 5	**53 098,86 €**	5 847,2 €	**47 251,73 €**	89 %
Year 6	10 704,65 €	5 847,2 €	4 857,52 €	45 %
Year 7	10 704,65 €	5 847,2 €	4 857,52 €	45 %
TCO:	**159 720,95 €**	**40 929,87 €**	**118 791,07 €**	**74 %**

As it can be clearly seen, from cost perspective, the most desirable scenario, and the way forward towards 5G NR and edge computing is the adoption of network function virtualization. As demonstrated, a virtualized ANDSF function based on the server requirements from Table 1 will cost the MNO around 5.900€ per year versus its physical counterpart which doubles the monthly price for the same performance and computing capacity. Table 4 presents a summary of overall monthly costs.

Table 10. Monthly cost summary

Physical Total Cost	
Refresh cycle first year	53 098, 86 €
Following Years	10 704, 65 €
Average Monthly cost (servers)	1 711, 30 €
Cloud Virtualized Total Cost (Monthly)	
Cloud Functions (servers)	487, 26 €
Average monthly savings in €	1 224, 04 €
Average monthly savings in %	72 %

From the total cost of ownership perspective, we have shown that the two proposed models are the ones that maximize the investment, minimizing the MNO's operational and capital expenditure, when compare with the classical physical server. Finally, from the end user perspective, the usage of cloud virtualized functions will contribute to increased quality of experience, as the probability of service disruption and downtime is greatly reduced, with the added benefit of, in the extreme case of cloud edge computing at the device edge, having the CDN with the possibility to reduce even further the latency by steering its traffic through different RATs.

5.4.2. Scale Flexibility and Performance

Scale flexibility while maintaining performance is one of the major advantages of the two proposed models when compared with the classical datacenter approach. The capability for scaling up resources in cloud virtualized environments is simply unrivaled, when compared with classic datacenter physical components.

Concerning performance, in this work we classify performance as the ability for the ANDSF virtual function to process as many inter-RAT handovers as possible, according to its own processing capacity. This will be tightly coupled with the ability to scale up in congestion situation. By congestion we mean ANDSF virtual function being saturated or close to saturation and not congestion on the distributed or collaborative edge. We define δ as the ANDSF function congestion rate. It depends directly on the processing capacity of the ANDSF function to properly perform the process presented in Figure 5 and process all inter-RAT handovers (HOs). A congestion factor of 0% means that there are no HOs issued from the DCE and that the ANDSF function is performing its usual measurements tasks. A congestion factor of 100% means that the ANDSF function is handling simultaneous HOs in the same amount as its maximum capacity.

$$\delta = \left(\sum_{i=1}^{N_{max_{HO}}} H_i\right) / N_{max_{HO}}$$

where $N_{max_{HO}}$ is the maximum number of simultaneous HOs that can be processed by the ANDSF function, H_i is the i^{th} HO request from device and $0 \leq \delta \leq 1$.

In order to avoid service interruption due to massive inter-RAT HO requests and saturation of ANDSF server, for both the physical server and the virtualized function from the proposed models, we consider a protection factor that will make sure that the ANDSF function will maintain service. We consider that, when the number of inter-RAT HO requests reach a certain percentage of the maximum processing capacity, the ANDSF function should cease processing additional requests, queuing them in a first in first out (FIFO) fashion. We call that the ANDSF function's protection factor, ρ_F that we assume to be equal to 70% of the whole processing capacity. Additionally, we consider that a triggering factor γ_T, is defined as a function of ρ_F the according to:

$$\gamma_T = \rho_F \cdot N_{max_{HO}} \text{ , where } \begin{Bmatrix} N_{max_{HO}} = 1000, \text{for physical ANDSF server} \\ N_{max_{HO}} = 1000, \text{for virtual ANDSF function} \end{Bmatrix}$$

Considering that the virtualized function is highly flexible in scaling up resources and capacity, which is not available to the physical ANDSF factor, we define the

following scenarios, which will clearly show the advantage of virtualizing the ANDSF function in what concerns overall performance. We present five different situations according to Table 5.

As simultaneous HOs keep reaching the ANDSF server / function, if it is "unprotected", meaning that ρ_F is not applied, in which case the server will continue processing requests without any resource protection. If the amount of simultaneous HOs suddenly spikes, the server might enter a denial of service state by resource exhaustion and interrupt the ANDSF service. If it is "protected", it means that ρ_F is applied and that the maximum number of simultaneous inter-RAT HOs to process is only 70% of the total capacity, meaning that ρ_F = 0.7 , leaving computing capacity for the server to keep servicing, without entering in a resource exhaustion mode

Table 11. ANDSF function congestion scenarios.

Scenario	Impact on function capacity
Unprotected (no γ_T) Physical Standalone (SA) ANDSF Server	Service Disruption
Unprotected (no γ_T) Physical ANDSF Server Cluster	Service Disruption
Protected Physical Standalone (SA) ANDSF Server	Queueing and FIFO
Protected Physical ANDSF Server Cluster	Queueing and FIFO
Protected Cloud Virtual ANDSF Function	Scale Up is performed. No service impact.

When 70% is achieved, γ_T, the ANDSF proposed function from Figure 5 triggers the congestion watchdog and it will either share load with other ANDSF servers, if in a cluster, or if standalone, it will cease processing new inter-RAT HOs. In the case of the cloud virtualized ANDSF function, it will try to scale up its resources automatically or, eventually, add a new ADSF function to the existing one. Figure 10 shows the result of such process, only for physical ANDSF servers.

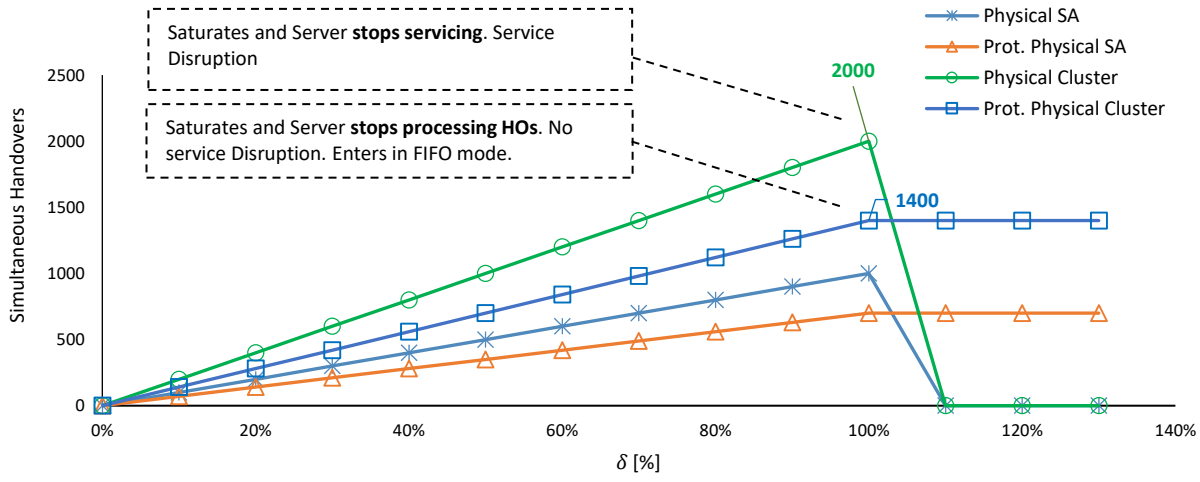


Figure 10. Physical ANDSF server congestion scenarios as a function of δ.

As it can be seen, all unprotected scenarios, either standalone or in a cluster, will reach its maximum capacity of simultaneous inter-RAT HOs (1000 for SA server and 2000 for two SA servers clustered, load balancing, with assumed doubled capacity) and simply have their server resources exhausted. Service is interrupted until the server is able to perform again. End users will experience extensive delay in inter-RAT especially if using low latency applications and if they are moving. This is the worst-case scenario, that clearly impacts the user's QoE, showing that a watchdog mechanism that constantly evaluates γ_T is needed. This is a feature that is introduced in this work due to the fact that with 5G NR massive inter-RAT HO requests are expected to be norm.

For the protected scenarios, it can be seen that when the standalone server or cluster reach their protection values (700 and 1400 respectively), γ_T is triggered and new HO requests are queued and accepted in a FIFO fashion. Servers do not get resource exhausted and no service interruption exists.

Finally, as HO requests are being processed by the ANDSF server, it is normal that saturation levels drop, more resources are freed on the server, and the whole process becomes cyclic.

When one considers the cloud virtual ANDSF function as proposed on the two models, a very different behavior occurs. As presented in Table 5, when the congestion watchdog trigger γ_T is reached, the ANDSF virtual function will make use of the scalability and elasticity of the cloud environment where it is running. We assume that, when the watchdog trigger happens, the server will automatically issue a scale-up request and perform necessary adjustments in order to scale its capacity 50% higher.

As shown in Figure 11, for the physical server the behavior is the same as referred before. The new and interesting part coming from the proposed models is that the cloud virtualized ANDSF function, because its disaggregated from other functions, when it reaches its capacity protection, the watchdog is triggered but instead of just making sure that resources are not exhausted, it scales up the number of vCPUs and RAM by 50%. We defined 50% just for the sake of demonstrating the different behavior, but it could scale up to any capacity. The other aspect, which greatly improves the overall quality of the whole process is that this can be performed automatically and can be adaptive: if 50% is not enough, the watchdog behavior we introduced for the ANDSF function will scale it further until the function has enough capacity to cope with the inter-RAT requests coming from the COE. Nevertheless, the ANDSF function watchdog can decide to double its capacity by instantiating a new cloud ANDSF function and perform load

balancing. This is the part that, as far as we are aware, is still difficult to perform on the UE end. Still, on the cloud environment, it is perfectly possible to do so as well as properly attach an artificial intelligence and machine learning function, developed specifically for ANDSF traffic analysis, to help enhance further the overall performance of the system.

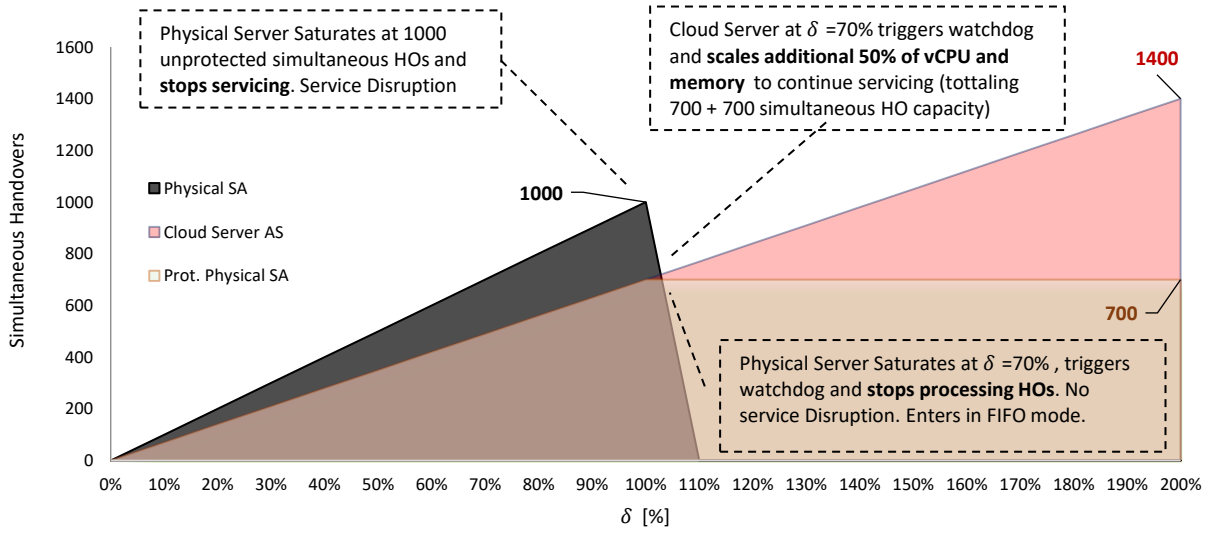


Figure 11. Physical vs cloud virtualized ANDSF server congestion scenarios as a function of δ.

All of these mechanisms in the end, will ensure that each of the Device/IoT Edge end user devices will be attached to the best RAT at any given instance, considering its vicinities and context. By being able to scale automatically and also adaptively, the cloud virtual ANDSF function will increase overall performance and quality of the inter-RAT HO mechanism, not only assuring that the best list of networks is sent to the Device/IoT Edge, but also that the end user will experience a seamless HO, without service interruption or degradation, thus increasing its overall QoE and, naturally, reducing the churning probability as a consequence. We have shown through the presented results that the proposed models maximize performance and scaling. In the following, we briefly focus on the cost of having the cloud virtualized ANDSF function's watchdog deciding on increasing function computing capacity or adding more virtual ANDSF functions.

Costs of scaling

From subsection 5.3.2 we have shown that the price of a single ANDSF virtualized function is approximately 490 € per month. However, as previously referred, it is important to know the cost of scaling up the server specifications, capacity and performance. We have considered in the previous analysis that, in

order to support more simultaneous RAT handovers, the ANDSF virtualized function would be able to scale up its vCPU and memory specifications by 50%. That means that the ANDSF server would scale up to 12GB of memory, 3vCPUs. Nevertheless, considering the cost difference in cloud environments, we also consider the scale up to double capacity, *i.e.*, 16GB RAM, 4vCPUs while keeping the same storage capacity. Figure 12 presents the monthly cost of each ANDSF virtualized function, after scaling up 50% and 100%.

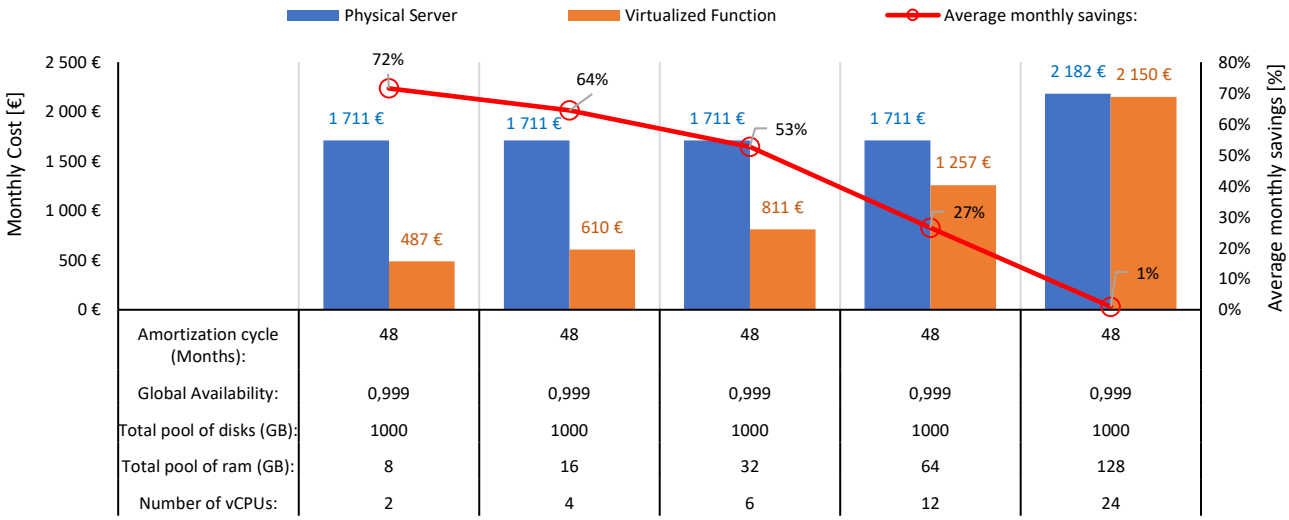


Figure 12. Average monthly cost comparison for virtualized function capacity scale-up.

The results show that a single cloud virtualized ANDSF function can scale up 12 times the number of vCPUs and 16 times the amount of RAM of its original specifications, and still represent monthly savings of 1%. It is, of course, demonstrated that from a scale and flexibility perspective, the virtualized function that we propose on both models, increases both.

Finally, from the perspective of scaling not the virtualized function itself, but other instances, per the cost of one physical ANDSF server, there can be three cloud virtualized ANDSF functions, meaning that there is the flexibility of scaling up in a factor of 1:3 and still obtain the same monthly costs that a single physical server represent. When scaling down is necessary, the virtualized function can perform those functions by itself, thus reducing capacity and cost on lower congestion periods.

All of these aspects allow us to conclude that the proposed models can substantially improve the overall efficiency of the Core Edge, through a cloud and network function virtualization architecture.

In short, the proposed cloud assisted system model can provide improved steering capabilities for ultra-high capacity demanding services running on top of 5G NR and Wi-Fi6 networks. Nevertheless, there are challenges in building such architectures, the most relevant related to latency minimization through the choice of a stable and highly reliable CDN network, as well as guaranteeing that the cloud provider maximizes availability and performance, in order not to transform the ANDSF function into a single point of failure. In this section we have shown through the results that network function virtualization applied to ANDSF network function will enable cost optimized, highly performant and flexible inter-RAT HO processes for dense 5G NR systems.

6. Conclusions

In this work, we proposed a cloud assisted ANDSF architecture, supported on network function virtualization and component disaggregation, as well as software defined networking, to specifically optimize the inter-RAT HO, steering process and traffic offloading in 5G NR ultra-dense networks, considering the co-existence of 5G NR and Wi-Fi6. Three major contributions were provided, and two different system models were also proposed. Both models can be deployed at any time, especially for the new 5G NR deployments, due to the fact that no protocol changes are required. This not only gives flexibility to MNOs, but especially, can have a profound impact on steering capabilities on ultra-dense networks, enabling unprecedented quality of service and experience to the subscriber. By employing such mechanisms, the cloud-assisted ANDSF-triggered steering process can increase its performance substantially, as well as reduce the overall battery consumption on the UEs, by changing the switching the majority of steering calculations and decisions to the Core Edge, where system-wide visibility exists, in opposition to possible distributed selfish process by each UE. The advantages of the proposed mechanism are several and from different nature and have been presented throughout the work. We believe that such architecture can provide relevant improvements in ultra-dense heterogeneous networks, both indoor and outdoor, overall providing both subscribers and MNOs with several advantages. All proposed architectures are possible today to deploy, which was also a concern of ours. Further work will focus on extending the existing results and explore other perspectives, namely how an artificial intelligence and non-supervised learning cloud virtualized function could be developed and attached to ANDSF in order to enhance even further the whole 5G NR system.

Author Contributions: Conceptualization, L.G., P.S., N.S., and A.C.; data curation, P.S.; formal analysis, L.G. and N.S.; project administration, N.S. and A.C.; supervision, P.S., N.S., and A.C.; validation, L.G., P.S., N.S., and A.C.; writing—original draft, L.G.; writing—review and editing, L.G.

Funding: This work was funded by Fundação para a Ciência e a Tecnologia / Ministério Ciência Tecnologia e Ensino Superior through national funds and when applicable co-funded by Fundo Europeu de Desenvolvimento Regional (FEDER) – PT2020 partnership agreement under the project UID/EEA/50008/2019.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. M. Jaber, M. A. Imran, R. Tafazolli and A. Tukmanov, "5G Backhaul Challenges and Emerging Research Directions: A Survey," in IEEE Access, vol. 4, pp. 1743-1766, 2016.
2. A. Maaref, J. Ma, M. Salem, H. Baligh, and K. Zarifi, "Device-centric radio access virtualization for 5g networks" in Globecom Workshops (GC Wkshps), Dec. 2014, pp. 887–893.
3. S.Andreev, M.Gerasimenko, O.Galinina,Y.Koucheryavy, N.Himayat, S.-P. Yeh, and S. Talwar, "Intelligent access network selection in con- verged multi-radio heterogeneous networks", IEEE Wireless Communications, vol. 21, Dec. 2014, pp. 86–96.
4. NGMN Alliance, "5G White Paper", Feb. 2015.
5. S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang and W. Wang, "A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications," in IEEE Access, vol. 5, pp. 6757-6779, 2017.
6. T. Um, G. M. Lee and H. Lee, "Trustworthiness management in sharing CDN infrastructure", International Conference on Information Networking (ICOIN), Chiang Mai, Thailand, 10-12 January 2018; pp. 73-75.
7. A. Sajid, H. Abbas and K. Saleem, "Cloud-Assisted IoT-Based SCADA Systems Security: A Review of the State of the Art and Future Challenges," in IEEE Access, vol. 4, pp. 1375-1384, 2016.
8. N. Mohamed, J. Al-Jaroodi, I. Jawhar, S. Lazarova-Molnar, and S. Mahmoud, ''SmartCityWare: A service-oriented middleware for cloud and fog enabled smart city services,'' IEEE Access, vol. 5, pp. 17576–17588, 2017.
9. W. Yu et al., "A Survey on the Edge Computing for the Internet of Things," in IEEE Access, vol. 6, pp. 6900-6919, 2018.
10. Wang, M.; Karakoc, N.; Ferrari, L.; Shantharama, P.; Thyagaturu, A.S.; Reisslein, M.; Scaglione, A. "A Multi-Layer Multi-Timescale Network Utility Maximization Framework for the SDN-Based LayBack Architecture Enabling Wireless Backhaul Resource Sharing". MDPI Electronics Journal, 2019, 8, 937.
11. H. I. Kobo, A. M. Abu-Mahfouz and G. P. Hancke, "A Survey on Software-Defined Wireless Sensor Networks: Challenges and Design Requirements," in IEEE Access, vol. 5, pp. 1872-1899, 2017.
12. Li Ling, Ma Xiaozhen and Huang Yulan, "CDN cloud: A novel scheme for combining CDN and cloud computing" Proceedings of 2013 2nd International Conference on Measurement, Information and Control, Harbin, China, 16–18 September 2013; pp. 687-690.
13. C. Yingying, S. Jain, V. K. Adhikari, Z.-L. Zhang, and K. Xu. "A first look at inter-data center traffic characteristics via yahoo! datasets." IEEE INFOCOM Proceedings, 2011, pp. 1620-1628. IEEE.
14. Yang, Song, Philipp Wieder, Ramin Yahyapour, Stojan Trajanovski, and Xiaoming Fu." Reliable Virtual

Machine Placement and Routing in Clouds" IEEE Transactions on Parallel and Distributed Systems 2017, vol.28, No. 10.

15. Makkes, Marc X., Ana-Maria Oprescu, Rudolf Strijkers, Cees de Laat, and Robert Meijer." Metro: low latency network paths with routers-on- demand." In Euro-Par 2013: 19th International Conference on European Conference on Parallel Processing, Aachen, Germany, 26-30 August 2013; pp. 333-342.
16. Cai, Chris X., Franck Le, Xin Sun, Geoffrey G. Xie, Hani Jamjoom, and Roy H. Campbell. "CRONets: Cloud-Routed Overlay Networks", IEEE 36th International Conference on Distributed Computing Systems (ICDCS), Nara, Japan, 27-30 June 2016; pp. 67-77.
17. Afraz, N.; Slyne, F.; Gill, H.; Ruffini, M. Evolution of Access Network Sharing and Its Role in 5G Networks. Appl. Sci. 2019, 9, 4566.
18. 3GPP TR 21.915, "Summary of Rel-15 Work Items" V15.0.0, Out. 2019.
19. 3GPP TR 21.916, "Summary of Rel-16 Work Items" V16.0.0, Sep. 2019.
20. Wi-Fi Hotspot 2.0 Specification Package, Passpoint Release 3, V3.1.0, May 2019. Online: https://www.wi-fi.org/downloads-registered-guest/Hotspot_2.0_Specification_Package_v3.1.zip/35974 (Last accessed on 5 December 2019).
21. Wi-Fi CERTIFIED 6™: "A new era in wireless connectivity", Wi-Fi Alliance, Sep.2019. Online: https://www.wi-fi.org/downloads-registered-guest/Wi-Fi_CERTIFIED_6_white_paper_20190912.pdf/35680 (Last accessed on 5 December 2019)
22. S. Saloni and A. Hegde, "WiFi-aware as a connectivity solution for IoT pairing IoT with WiFi aware technology: Enabling new proximity-based services" 2016 International Conference on Internet of Things and Applications (IOTA), Pune, India, 22-24 January 2016; pp. 137-142.
23. A. Mondal, C. Huang, J. Li, M. Jain and A. Kuzmanovic, "A Case for WiFi Relay: Improving VoIP Quality for WiFi Users" 2010 IEEE International Conference on Communications, Cape Town, Sout Africa, 23-27 May 2010;
24. 3GPP TS 24.312, "Access Network Discovery and Selection Function (ANDSF) Management Object (MO) (Rel. 15)", V15.0.0, Jun. 2018.
25. 3GPP TS 23.402, "Architecture enhancements for non-3GPP accesses (Rel. 16)" V16.0.0, Jun. 2019.
26. 3GPP TS 24.302, "Access to the 3GPP Evolved Packet Core (EPC) via non-3GPP access networks (Rel. 16)", V16.1.0, Jun. 2019.
27. IEEE 802.11u, "Interworking with External Networks", Feb. 2011.
28. 3GPP TS 22.278, "Service requirements for the Evolved Packet System (EPS) (Rel. 16)", V16.2.0, Jun. 2019
29. 3GPP TS 36.331, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control; (Rel. 15)", V15.6.0, Jun. 2019.
30. Li, X.; Guo, C.; Xu, J.; Gupta, L.; Jain, R. Towards Efficiently Provisioning 5G Core Network Slice Based on Resource and Topology Attributes. Appl. Sci. 2019, 9, 4361.
31. S. K. Goudos et al., "A Novel Design Approach for 5G Massive MIMO and NB-IoT Green Networks Using a Hybrid Jaya-Differential Evolution Algorithm," in IEEE Access, vol. 7, pp. 105687-105700, 2019.
32. Mitsolidou, C.; Vagionas, C.; Mesodiakaki, A.; Maniotis, P.; Kalfas, G.; G. H. Roeloffzen, C.; W. L. van Dijk, P.; M. Oldenbeuving, R.; Miliou, A.; Pleros, N. A 5G C-RAN Optical Fronthaul Architecture for Hotspot Areas Using OFDM-Based Analog IFoF Waveforms. Appl. Sci. 2019, 9, 4059.

33. W. F. Elsadek and M. N. Mikhail, "Inter-domain Mobility Management Using SDN for Residential/Enterprise Real Time Services," IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Vienna, Austria, 22–24 August 2016; pp. 43-50.

34. W. F. Elsadek and M. N. Mikhail, "IP mobility management using software defined networking: A review", IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 15-17 December 2017; pp. 76-81.

35. H. Cheon, S. Lee and J. Kim, "New LIPA/SIPTO offloading algorithm by network condition and application QoS requirement," 2015 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, 28-30 October 2015; pp. 191-196.

36. Ruiz, L.; Durán, R.J.; De Miguel, I.; Khodashenas, P.S.; Pedreño-Manresa, J.-J.; Merayo, N.; Aguado, J.C.; Pavón-Marino, P.; Siddiqui, S.; Mata, J.; Fernández, P.; Lorenzo, R.M.; Abril, E.J. A Genetic Algorithm for VNF Provisioning in NFV-Enabled Cloud/MEC RAN Architectures. Appl. Sci. 2018, 8, 2614.

37. C. Lai, H. Chang, M. Tsai and C. Lin, "Reducing Costs of LIPA Handover Through Bearer Reservation With Preemption", in IEEE Transactions on Vehicular Technology, July 2017, vol. 66, no. 7, pp. 6428-6438.

38. L. Gao, G. Iosifidis, J. Huang, L. Tassiulas and D. Li, "Bargaining-Based Mobile Data Offloading", in IEEE Journal on Selected Areas in Communications, June 2014, vol. 32, no. 6, pp. 1114-1125.

39. Wi-Fi CERTIFIED Passpoint®, Deployment Guidelines, Rev. 1.3 April 24, 2019.

40. M. Gerasimenko, N. Himayat, S.P. Yeh, S. Talwar, S. Andreev and Y. Koucheryavy, "Characterizing Performance of Load-aware Network Selection in Multi-radio (WiFi/LTE) Heterogeneous Networks", in GLOBECOM Workshop, Dec. 2013, pp. 397-402.

41. M. Klymash, O. Shpur, N. Peleh, O. Lavriv, R. Bak and O. Skybinskyi, "Increasing the Accessibility of Static Content using CDN Networks as PaaS" in IEEE 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), Polyana, Ukraine, 26 February – 2 March 2019; pp. 1-4.

42. F. Rebecchi, M. D. de Amorim, V. Conan, A. Passarella, R. Bruno and M. Conti, "Data Offloading Techniques in Cellular Networks: a Survey", IEEE Communications Surveys & Tutorials, Nov 2014, Vol. 17, No. 2, pp. 580-603.

43. K. Adachi, M. Li, P. H. Tan, Y. Zhou and S. Sun, "Q-Learning Based Intelligent Traffic Steering in Heterogeneous Network" in proc. of VTC Spring, pp. 1-5, May 2016.

44. M. El Helou, M. Ibrahim, S. Lahoud, K. Khawam, D. Mezher and B. Cousin, "A Network-assisted Approach for RAT Selection in Heterogeneous Cellular Networks", IEEE Journal on Selected Areas in Communications, Jun 2015, Vol. 33, No. 6, pp. 1055–1067.

45. Hwang, R.-H.; Peng, M.-C.; Cheng, K.-C. QoS-Guaranteed Radio Resource Management in LTE-A Co-Channel Networks with Dual Connectivity. Appl. Sci. 2019, 9, 3018.

46. B. H. Jung, N. Song and D. K. Sung, "A Network-assisted User-centric WiFi-Offloading Model for Maximizing Per-user Throughput in a Heterogeneous Network" IEEE Transactions on Vehicular Technology, Oct 2013., Vol. 63, Issue 99, pp. 1940 – 1945.

47. A. Roy and A. Karandikar "Optimal Radio Access Technology Selection Policy for LTE-WiFi Network" in International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), May 2015, pp. 291 - 298.

48. E. Khloussy, X. Gelabert and Y. Jiang, "A Revenue-Maximizing Scheme for Radio Access Technology Selection in Heterogeneous Wireless Networks with User Profile Differentiation" in Advances in

Communication Networking, Springer, 2013, pp. 66-77.

49. 3GPP TR 37.834 V12.0.0, “Study on WLAN/3GPP Radio Interworking”, Jan. 2015.
50. S. Su, B. Huang, C. Wang, C. Yeh and H. Wei, "Protocol Design and Game Theoretic Solutions for Device-to-Device Radio Resource Allocation," in IEEE Transactions on Vehicular Technology, vol. 66, no. 5, pp. 4271-4286, May 2017.
51. S.Adhikarla, M. S. Kang, and P. Tague, “Selfish Manipulation of Cooperative Cellular Communications via Channel Fabrication.” In Proceedings of ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec), Budapest, Hungary, 17 – 19 April 2013; pp.49-54.
52. H. Zhang, Y. Li, D. Jin and S. Chen, "Selfishness in device-to-device communication underlaying cellular networks," 2015 IEEE International Conference on Communication Workshop (ICCW), London, UK, 8-12 June 2015; pp. 675-679.
53. C. Gao, H. Zhang, X. Chen, Y. Li, D. Jin and S. Chen, "Impact of Selfishness in Device-to-Device Communication Underlying Cellular Networks," in IEEE Transactions on Vehicular Technology, vol. 66, no. 10, pp. 9338-9349, Oct. 2017.
54. Q. Nguyen-Vuong, N. Agoulmine, E. H. Cherkaoui and L. Toni, "Multicriteria Optimization of Access Selection to Improve the Quality of Experience in Heterogeneous Wireless Access Networks," in IEEE Transactions on Vehicular Technology, vol. 62, no. 4, pp. 1785-1800, May 2013.
55. M. Louta and P. Bellavista, "Bringing always best connectivity vision a step closer: challenges and perspectives," in IEEE Communications Magazine, vol. 51, no. 2, pp. 158-166, February 2013.
56. Tonini, F.; Khorsandi, B.M.; Bjornstad, S.; Veisllari, R.; Raffaelli, C. C-RAN Traffic Aggregation on Latency-Controlled Ethernet Links. Appl. Sci. 2018, 8, 2279.
57. Y. Al-Dhuraibi, F. Paraiso, N. Djarallah and P. Merle, "Elasticity in Cloud Computing: State of the Art and Research Challenges," in IEEE Transactions on Services Computing, vol. 11, no. 2, pp. 430-447, 1 March-April 2018.

2.4. Article nr. #4

This article presents a two-tier method to further improve the ability for an MNO to achieve carbon neutrality.

The main contribution is a two-tier approach that combines EE and CSS methods. Both tiers are applied to the mobile and RAN edges in order to reduce overall CF in the whole networks. EE works as a first tier, in order to reduce overall power consumption in both edges and the remaining CF, which we consider to be the RCF, should be addressed by the second tier aiming to offset it in order to achieve carbon neutrality.

In the end, it is shown that both set of methodologies can be put together in an intelligent way and further ease the carbon neutrality achievement process.

Article details:

- Title: One Step Greener: Reducing 5G and Beyond Networks' Carbon Footprint by 2-Tiering Energy Efficiency with CO_2 Offsetting;
- Date: January 2020;
- Journal: Electronics;
- Scimago/Scopus Journal Ranking: Quartile 1;
- Publisher: MDPI.

electronics

MDPI

Article

One Step Greener: Reducing 5G and Beyond Networks' Carbon Footprint by 2-Tiering Energy Efficiency with CO_2 Offsetting

Luís Carlos Gonçalves [1,2,*], Pedro Sebastião [1,2], Nuno Souto [1,2] and Américo Correia [1,2]

[1] Technology and Information Science Department, ISCTE-Instituto Universitário de Lisboa, Av. Forças Armadas, 1649-026 Lisboa, Portugal; pedro.sebastiao@iscte-iul.pt (P.S.); Nuno.Souto@lx.it.pt (N.S.); americo.correia@iscte-iul.pt (A.C.)

[2] Radio Systems Department, Instituto de Telecomunicações, Av. Forças Armadas, 1649-026 Lisboa, Portugal

* Correspondence: lcbsg@iscte-iul.pt; Tel.: +351-213-130-991

Received: 29 January 2020; Accepted: 28 February 2020; Published: date

Abstract: Fifth generation (5G) and Beyond-5G (B5G) will be characterized by highly dense deployments, both on network plane and user plane. Internet of Things, massive sensor deployments and base stations will drive even more energy consumption. User behavior towards mobile service usage is witnessing a paradigm shift with heavy capacity, demanding services resulting in an increase of both screen time and data transfers, which leads to additional power consumption. Mobile network operators will face additional energetic challenges, mainly related to power consumption and network sustainability, starting right in the planning phase with concepts like energy efficiency and greenness by design coming into play. The main contribution of this work is a two-tier method to address such challenges leading to positively-offset carbon dioxide emissions related to mobile networks using a novel approach. The first tier contributes to overall power reduction and optimization based on energy efficient methods applied to 5G and B5G networks. The second tier aims to offset the remaining operational power usage by completely offsetting its carbon footprint through geosequestration. This way, we show that the objective of minimizing overall networks' carbon footprint is achievable. Conclusions are drawn and it is shown that carbon sequestration initiatives or program adherence represent a negligible cost impact on overall network cost, with the added value of greener and more environmentally friendly network operation. This can also relieve the pressure on mobile network operators in order to maximize compliance with environmentally neutral activity.

Keywords: 5G NR; energy efficiency; carbon footprint; 5G; B5G geosequestration; carbon offset

1. Introduction

Mobile data services will witness unprecedented usage with the advent of the fifth generation (5G) new radio (NR) and Beyond-5G (B5G) networks. A long way has been paved since the first cellular generations, with the fourth generation (4G) already witnessing a very relevant increase of traffic demand. Nevertheless, 5G will be unprecedented: despite being prepared and designed to be optimized by design, with several energy efficiency (EE) techniques combined with several methods to increase overall throughput and user quality of experience (QoE) and overall quality of service (QoS), B5G NR networks will face a multitude of challenges driven by massive data-hungry subscribers and highly performant end user devices capable of generating enormous amounts of traffic, not solely in downlink but also in uplink. 5G NR and beyond will be

fully heterogenous networks (HetNets), with cellular radio access technology (RAT) but also wireless fidelity (Wi-Fi) RAT, where Internet of things (IoT) devices will also play an important role [1]. These networks will have a multitude of different base stations, macrocells, microcells, femtocells, Wi-Fi access points (APs) and relays, making the radio access network (RAN) a very complex architecture. Additionally, a massive number of end user devices are expected to support themselves on 5G NR and beyond, end user or not; e.g., device to device (D2D) or machine to machine (M2M) communications. To support all this, another massification will take place: the deployment of additional mobile base stations, whether they are totally physical or partially virtualized on cloud environments. All of this will make 5G NR and beyond densified networks (DenseNets) [2,3].

All of the aforementioned factors drive the development of new technology, new infrastructures, new processes and network refitting in order to sustain the challenges that are ahead for cellular networks.

From another perspective, mobile services have been witnessing unprecedented growth, bringing new challenges for mobile network operators, from several perspectives:

- Capacity maximization;
- Resource efficiency maximization, particularly spectrally;
- Cost effectiveness and reduction;
- Energy consumption minimization;
- Carbon emissions' minimization.

With all the changes that have taken place, it is expected that energy consumption will rise also unprecedently in 5G and beyond networks. As such, it is of utmost importance to realize that 5G and beyond cellular networks might be on their way to becoming one of the industries that most contributes to environmental impacts through greenhouse gas emissions (GGE).

From sustainability perspectives, global warming has become a top priority on the word's agenda [4,5], and cellular systems do have their contribution, which is expected to rise in the future in the form of carbon dioxide (CO_2) emissions. Most of the observed increase in global average temperatures is due to the observed emissions of CO_2 into the atmosphere. This phenomenon is generally known GGE and occurs naturally, as Earth's radiant heat becomes trapped in the atmosphere due to existing gases and gets radiated back towards the surface, leading to overall climate changes.

Several studies and governmental initiatives exist that address CO_2 emission reduction, primarily those as a direct result of the adoption of the Kyoto protocol resolutions and goals. Although those are out of scope of the current work, CO_2 compensation is not, and is the approach that we address in the current work.

Carbon sequestration (CS) is a term that defines the process of capturing and long-term storage of CO_2. It is a form of geoengineering, which aims to reduce the impacts of greenhouse gas emissions by manipulating environmental processes in order to counteract those effects. Also known as carbon sequestration and storage (CSS), it focuses on the physical and chemical methods of capturing CO_2 from the atmosphere and storing it in another place [6,7]. The most common CSS systems use geological storage, ocean storage or biotic sequestration. In the current article we consider a CSS method to evaluate its capacity of contributing to the reduction of the overall carbon footprint (CF) of normal mobile network operators' (MNOs') networks and activities.

Mobile cellular networks' ever-growing popularity, along with the universe of subscribers which has been increasing exponentially, (and their behavior towards data generation) has become a very relevant contributor to the overall CF [8].

Considering the user plane, the International Telecommunications Union estimated that by 2022, there would be 12.3 billion mobile-connected devices [9], exceeding the world's projected population of 8 billion by one and a half times [8], with expected further increase in traffic demand mainly driven by mobile gaming and video streaming. In this specific plane, just considering mobile traffic alone, the expected annual growth rate is 46 percent from 2017 to 2022, reaching 77.5 exabytes per month by 2022 [8].

As the numbers of base stations (BSs) and remote radio heads (RRHs) increase due to the unprecedent demand for high capacity and throughput, higher power consumption will occur, and therefore, higher CF. Digital Power Group [10], calculated in a 2013 report that 10% of worldwide electricity generation corresponded to such activities. Other estimations forecast that overall electricity consumption and CF due to such networks will represent 51% and 23%, correspondingly, by 2030 [11].As an example, when considering a 4G macrocell BS, approximately 60% of the power consumption is attributed to the power amplifiers [12], electing it as the most crucial component to consider when energy-centric mobile network design, implementation and operation is the focus.

The wireless industry needs significant improvements in the EE of BSs and other network infrastructure to compensate for the increased energy demands due to network growth [13–15]. Therefore, designing energy-efficient communication systems has become a critical issue for 5G and beyond, which promises massive deployments of smart devices served by new infrastructure elements. Especially considering 5G NR, a device will generate an estimated 2.6 times more traffic than a 4G device and this will tend to increase in B5G NR networks [8]. Concepts such as "sustainable green communications" have recently emerged and describe the common trend toward energy-efficient wireless communications systems [16].

Emerging technologies can contribute to reduced power consumption in mobile communication networks, and it is today clear that the telecommunications industry must address CO_2 emissions and become more sustainable. The CF concern in the development and deployment of 5G and beyond is clearly one of the most debated subjects today and deserves a closer look.

2. Contribution

The main research question in this paper focuses on existing EE methods and the fact that most do not suffice for achieving carbon neutrality. As such, we address that aspect with a two-tier model, which we present in this work. This two-tier CO_2 offset model is the main contribution. The 1st tier applies CO_2 reduction through EE techniques, proposed in existing studies aiming to improve overall EE in 5G networks. In that regard, we review the most important contributions in that area and focus on those that maximize CO_2 reduction through the overall reduction of power consumption. As mentioned, there will still be CF after applying such techniques, which we call residual CF (RCF).

The 2nd tier has a sole purpose: to offset RCF. Through the methods that will be proposed, this second layer aims to offset RCF by tightly coupling with CF reduction from the 1st tier. It is shown that it is feasible for mobile network operators to effectively offset total CO_2 emissions with the proposed two-tier model. Furthermore, the impacts of such an initiative, from capital and operation expenditure (CAPEX and OPEX) perspectives, are also quantified.

The 2nd tier will offset the RCF by applying a CO_2 sequestration technique called biotic geosequestration. Trees will be used as carbon-based "sinks" through the process of photosynthesis. The objective is to offset, at least, the RCF produced in the system, considering both the RAN edge (where base stations represent the most energy consuming component) and also the end-user's equipment operation. In this work, we do not focus on offsetting the CF associated to mobile device manufacturing, datacenter operations and daily MNOs routines related with operational tasks. Nevertheless, all are considered and referred along the work. The focus is the RAN and the mobile device edges, especially considering power consumption associated to base stations and mobile end user devices.

3. Organization

We begin our analysis in Section 4, slightly focusing on the need for EE in 5G and B5G networks. Despite being out of the scope of the current work, we believe that there is enough relevancy on the subject that justifies enumerating some of the techniques that have been developed to reduce overall energy consumption. We argue that such techniques are of utmost importance, representing a first tier of CF reduction because they

contribute very relevantly to overall power consumption, and there are methods that transversally focus on all network edges, ranging from the device/subscriber edge/plane to the network edge/plane.

Although EE is very relevant in reducing overall CF, highly dense, heterogeneous and traffic hungry 5G and beyond networks will still result in a relevant RCF after EE deployment. As such, the second tier is focused on in Section 5, introducing the most usual techniques that contribute to overall carbon sequestration and storage. Such techniques will further reduce the RCF from the first tier. Section 6 will focus on estimating the power consumption and CF of both the network and user planes, especially focusing on base stations and end user devices. Based on those estimates, Section 7 presents the method evaluated to demonstrate how CSS can be applied in 5G and B5G cellular deployments, in order to positively offset overall CF and achieve carbon neutrality. Section 8 presents the main results and Section 9 the main conclusions.

4. Energy Efficiency in 5G and Beyond

Two building blocks form the basis for every sustainable system: EE and CF minimization. In this section we focus on some of the most relevant EE techniques which contribute to power consumption reduction, from our point of view.

EE is a term that defines a set of techniques or processes aiming altogether to reduce overall energy requirements. As such, EE can be defined and calculated as the ratio between the quantity of data successfully delivered within a cellular network and the total energy spent in such successful transmission. This means that a system or a component of a system has higher EE if progressively, for the same amount of successfully transmitted data, the energy consumption decreases. It can be expressed as E_{Ef}, given by:

$$E_{Ef}[b/J] = \frac{T_d^S[bps]}{T_e^S[Wh]} = \frac{T_d^S[bps]}{3600 \cdot T_e^S[J]},$$

where T_d^S represents the total successfully transmitted data and T_e^S represents the total energy spent in the successful transmission.

EE plays a crucial role on 5G networks and will be even more crucial on B5G cellular networks in future. As energy consumption of the whole B5G cellular system is expected to increase exponentially, it is of utmost importance to develop techniques that will address that question from a lifecycle perspective.

As such, such a concept must be taken into account right from the design phase on B5G deployments. Thus, it is essential to review some of the most prevalent research works on this field, properly setting the landscape around EE methods for 5G and B5G networks.

Several works have been presented focusing on this subject, reviewing the existing techniques and at the same time stating that energy savings are still far from what would be beneficial and expected [17,18]. Some of the reasons that contribute to such an inability materialize in the shape of several inefficiencies that, when considered altogether, directly impact on overall effort to reduce EE. Some of those inefficiencies are discussed in [19], along with possible and suggested improvements. However, it is also interesting to realize that several smart architectural and topological designs for 5G, which were proposed with resource sharing in mind, offer scalability and flexibility, from a network perspective, but nevertheless, are still insufficient when EE is considered [20–23].

Clearly, it can be seen that, as stated before, EE should be considered right from the start—in the network design phase. Today, it does not make sense to design a cellular network without thinking about flexibility, scalability and especially EE right from the start.

Overall system efficiency must be evaluated at a system level; i.e., both the network and the devices must be properly considered.

4.1. RAN Edge and EE Methods

As previously said, this work is focused primarily in the RAN edge, where base stations are considered the most energy-hungry devices [24]. Several techniques have been proposed, mainly focusing on selective sleep modes and coverage zooming. The former aims to selectively switch some of the BS' radio heads, in

order to reduce energy consumption according to the traffic profile, which is evaluated continuously. The latter focuses on techniques that adapt coverage radius towards cells where there are BSs in sleep mode, in order to compensate and balance traffic load. Sleep modes have been thoroughly explored in several studies surveyed by [25] which show that: (i) switching radio heads on and off can incur additional energy consumption [26]; (ii) clustering BSs can be very beneficial [27,28]; (iii) considering user experience and quality of service, switching on-off optimization can be achieved, trading off between energy consumption, quality of service [29,30] and quality of experience [31]. Other dynamic operational mechanisms have been proposed using relays, site optimization and dynamic switching [32–37].

Regarding coverage zooming, it can be seen as a complement to the above referred-to techniques. Such a technique is very similar to the concept of cell breathing, first introduced with 3G networks, mainly as a result of power control [38,39] adaptation mechanisms. It aims to adjust each cell's coverage according to traffic conditions [40] and it is computed centrally in the core network. Such a mechanism leads to several benefits that impact the overall EE of the cellular network directly, either by balancing the traffic load between congested and non-congested BSs, or by reducing end users' power consumption [41].

Several cell zooming techniques have been present and discussed, with the ultimate aim of increasing EE, focusing, namely, on: (i) centralized and de-centralized algorithms [42]; (ii) non-cooperative game theory [43]; (iii) optimal user association [44]. Most recently, the usage of distributed antennas, namely, coordinated multipoint (CMoP) and energy efficient aware continuous cell zooming strategies, has been proposed [45,46]. When considering 5G and B5G deployments, ultra-dense networks must be taken into consideration and density estimation can be used as a strategy for the overall increase of EE, as firstly presented in [47] and subsequently by [48,49]. User density has also been considered as a parameter to smartly reduce overall power consumption [50]. Data-aware mechanisms have also been focused on, relying once again on BS cooperation in order to optimize the overall cell zooming process, and thus, the EE of the whole network 25. Other mechanisms have also been proposed in [51–58].

4.2. Mobile Device Edge and EE Methods

From and end-device perspective, the aim is to reduce as much as possible the power consumption of mobile devices, not only to extend battery lifetime but also in order to increase EE. Such techniques become even more important if one considers the unprecedented number of devices that will be serviced in a cellular network. One can note that today's behavior towards mobile device usage and consumption patterns has changed radically [7]. Considering, for example, the data exchange between two users which are geographically close to each other, which represents an ever-increasing reality, the possibility of having such data exchange directly between the devices in a D2D fashion can represent several enhancements in power consumption reduction. D2D can be used as a mechanism to decrease power consumption [59,60], by reducing the network hops to only one with the immediate advantage of lower latency; better quality of service and experience; and, from the core network perspective, decrease signaling and overall backbone traffic. Additionally, and probably the most obvious advantage, is that both devices will need less power to transmit the same amount of data, thereby directly increasing EE by reducing energy consumption. However, mobile devices still do not have enough battery life or computational capacity to allow them to perform effectively over long periods of time, which is a requirement for most users. Additionally, for those applications that need high communication or computation power, and low latency, performance will drop severely. As such, an interesting concept has appeared called cyber foraging [61] or computational offloading. Cyber foraging is a technique that enables end user devices to extend their computing power by offloading computation efforts or data processing to more powerful servers located in the cloud, or sometimes, as close as a single hop away. An extensive survey on cyber foraging techniques has been presented in [62], showing clearly the advantages of using such techniques for end user plane's power consumption reduction. Recent studies denote that cyber foraging is still an immature concept from a standardization perspective and have proposed methods and models for its implementation together with protocols for cyber foraging user plane functions [63,64].

4.3. Overall EE Contribution for CF Reduction

For many decades, mobile network operators focused their attention on network optimization and increasing data throughput, special coverage and reducing latency. However, in today's cellular networks, energy consumption reduction (and EE overall) has become one of the prime objectives, due to environmental and economic aspects. We have briefly covered several techniques that show that EE maximization is possible at a network level where there is a set of such techniques and mechanisms already proposed. We have also noted that not only does the network plane need attention, but the end user plane does also, especially the devices themselves. We have presented some of the main techniques that are being discussed in order to minimize power consumption on the end user plane, leading to overall network power consumption reduction.

Current studies indicate that the level of carbon emission due to communication and information technologies has reached up to 10% of overall CF, of which 2%–3.5% corresponds to mobile networks [65–67]. There is no doubt that, considering the actual trends, such values will rise exponentially with the massive requirements of 5G and B5G networks. This means that since, at a given time, EE techniques may cease contributing relevantly to reduce overall carbon emissions, an additional approach must be considered to further reduce overall network CF. It is natural, thus, that CSS techniques and offsetting become the natural second tier in order to achieve a greener cellular network operation. The next section will focus on existing techniques to capture and store CO_2.

5. Carbon Sequestration and Storage

CSS is a process that aims to capture atmospheric CO_2 and store it over the long term, thereby reducing the overall greenhouse gas (GHG) emissions. It consists of actively removing CO_2 from the atmosphere into reservoirs, which can be man-made or natural. In this work we focus on the natural ones, which can be split into several types: biotic (e.g., trees); geological (e.g., underground rock formations and structures); oceanic (e.g., underwater bolsters); and underground sinks, such as saline deposits or gas reserves [68]. Each one of these is important and their main objective is to reduce the overall GHG emissions, taking advantage of natural processes to achieve that goal.

In this work we focus on biotic or biomass CSS, as it is the only one which does not require the development of technology to fulfill its task. Interestingly enough, one of the major contradictions is developing technological methods for reducing GHG emissions without considering that those technological processes themselves contribute to generate additional GHG emissions. As such, the next sections briefly introduce the concepts.

5.1. Geological CSS

Geological CSS is a method for capturing and trapping CO_2 in appropriate rock formations, mainly underground [69]. The capturing process is done in a gaseous form by using physical and chemical methods. Trapping is performed by underground or underwater geological formations, injected by pipelines, but mainly used on underground rock formations [70]. It is a technological process that also creates by itself additional CO_2 emissions and has several drawbacks: rocks have to be carefully chosen; CO_2 can leak into the ocean or surface, causing an impact on marine fauna or generating bolsters of CO_2 near the surface with severe impacts on animals and plants.

5.2. Ocenan CSS

Ocean carbon storage is very similar to geological CSS, but is accomplished totally underwater; namely, in the oceans. The CO_2 is injected in underwater rock formations or bolsters [71]. There are several setbacks to this method, including endangering the environment itself in a harsh way. Ocean water might suffer from acidification when CO_2 is leaked from the ocean beds, reaching the atmosphere anyway. It is a process that

requires advanced technological processes in order to keep pumping and keep the pressure stable, to avoid CO_2 leakage [72,73].

5.3. Biotic CSS

As referred to, the previous two methods present several risks and are not considered very effective, acceptable or sustainable approaches for CSS. Therefore, an alternative is needed, which does not have so many risks, does not involve the use of technology during the sequestration period and, if possible, can increase the CO_2 storage over time.

This is where biotic sequestration, or tree-based sequestration becomes the chosen method. In this case CO_2 is trapped through the natural process of photosynthesis [74–77]. This means that during the capture part there are no setbacks and no side effects, as it is derived from a total natural process.

Trees retrieve CO_2 from the atmosphere and store it in several of their parts, including trunks, limbs, leaves and roots. This behavior constitutes a CO_2 storage technique, as it is removed from the atmosphere and trapped inside the tree. Additionally, as mentioned, there is the possibility of the same tree increasing its storage capacity: as solar exposure increases, more photosynthesis will occur, and more CO_2 will be extracted from the atmosphere and converted into biomass, below and above ground, in the tree itself [78–80]. Photosynthesis converts the energy from sunlight into nutrients that the tree requires and keeps, transforming CO_2 and water into oxygen and glucose [81,82].

This means that the CO_2 storage capacity increases, but only to a certain degree. When the tree becomes older, its storage capacity ceases increasing and remains constant. Nevertheless, as it will be shown in this work, this depends on the type of tree and happens two or three decades after the tree has been planted [83–85]. Table I presents the breakdown in percentages of the total system CF and an estimate each for 2020 and 2025, considering a linear increase of 11 $MtCO_2$-e per year.

Table 1. Carbon footprint (CF) breakdown per system component—2020.

Component	Contribution to Overall CO_2-e
Mobile Device Manufacturing (MDM)	30%
RAN sites' operation (RSO)	29%
Datacenters and data transport (DDT)	19%
Mobile device operation (MDO)	10%
RAN sites' manufacturing (RSM)	4%
MNO activities (MO)	8%
Total (CFTotal) [MtCO2e] (for 2020/2025)	235/290

6. Energy Consumption for RAN and Mobile Edges

This section focuses on CF estimation for both network edges under analysis in this work. After overviewing different proposed EE techniques for both edges, as previously referred to, even after being deployed, there still is RCF that needs to be addressed through a second tier. In order to understand how EE contributes to reducing overall power consumption on both edges, it is crucial to understand what the consumption levels are on both edges first. As mentioned, we will focus on RAN and mobile edges.

6.1. RAN Edge Power Consumption

Regarding base stations, [86] showed that at full system load, the power consumption of a 5G BS can range from 6 W to 1 kW, depending whether it is a femtocell or macrocell, respectively [85,86]. The main reason for a macrocell to have higher consumption than a 5G femtocell is related to the need to cover wider areas, and thus employing higher capacity power amplifiers, which require more power than femtocells. For the

latter, the majority of the power consumption comes from the baseband units [86,87]. Nevertheless, the interesting part is that these units can be disaggregated from the radio elements and virtualized in a C-RAN environment. Just by residing in cloud environments, with the help of NFV, each can be seen as a single virtualized function that can be virtually aggregated, contributing overall to a power reduction on the femtocell side [74]. Such an approach might also be applied to gNBs in 5G NR and beyond networks, considering that NFV combined with the usage of C-RAN principles and cloud edge computing would result in minimizing such power consumption. As such, for beyond 5G networks it is expected that, with the huge massification of small cells, overall power consumption can drop simply by virtualizing part of its functionality. Such functions are not the aim of this work, but it should be noted that NFV and edge computing are two paradigms for beyond 5G NR networks that will certainly contribute to overall power consumption reduction. For the objectives of this work, we will consider the power consumption values previously referred to. According to [24], dense deployments of BSs result in excessive energy consumption levels which can rise up to 60 billion kWh per year.

6.2. Mobile Edge Power Consumption

In both 5G and B5G networks, end devices can be classified under two categories: human-interaction enabled and human-interaction free. The former relates to what are known as smartphones, while the latter relates to all devices able to generate a communication flow, either unidirectional or bi-directional, without any human interaction. Sensor networks and machine-to-machine (M2M) communication fall into this category which is not within the scope of the current document, and thus will not be discussed. On the other hand, smartphones play an important role regarding energy consumption and will be taken into account in this work.

Considering the very relevant and rapid advances in technology, especially as processing capacity and hardware components become more advanced, with sometimes physically larger devices (e.g., touchscreens), energy consumption rises. On the other hand, the number of subscribers has been growing dramatically, demanding quicker access to the Internet, more bandwidth consuming services and applications often requiring near real-time operation. Both will contribute very relevantly to the rise of energy consumption, and consequently, will increase overall CO_2 atmospheric levels and CF. If one considers the estimate that by 2030, there will be more that 50 million connected end user devices, the resulting global CF may become very relevant [88]. Thus, it is of utmost importance to focus on end user devices from this perspective.

Smartphones do not consume energy in a uniform way. Several of their components have different energy consumption needs: usually, during a call, the RF module is the most energy consuming component, whereas the display might become the highest consuming component if video streaming is being used. On the other hand, smartphones often have different operation methods; namely, suspended (when the application processor is idle but the communications processor remains active in background), and the idle state, which is similar to the suspended state, but the display is active and the whole graphical subsystem and processing components. In these two operation modes, the levels of energy consumption are different, with the RF module consuming the most in the suspended state and the display and graphical processor being the most energy demanding pieces in idle mode. Average power consumption figures (excluding the backlight) for three different smartphones have been focused on extensively in [89]. It was shown that when localization services are active and running on background, GPS radio is active and contributes to overall consumption, increasing it. If one considers real time for services such as cloud music streaming, several components will always be active at the same time, raising the level of energy consumption.

Therefore, it is safe to state that smartphone energy consumption profile is heavily dependent on the type of behavior of the subscriber, which implies certain working characteristics, ultimately depending on the types of applications.

Besides application dependence, energy consumption also depends on the brightness level required by the application itself and can increase power consumption 66 times more [89]. Power consumption models

have been proposed for LTE and LTE-A, and it was shown that the downlink data rate (closely related to applications and downlink streaming services) and uplink transmission power are the two most consuming factors. It was also shown that LTE RF module's power consumption varies according to operating bandwidth, up to three times more when working on 15 MHz band compared to 10 MHz [90,91].

Interestingly enough, different smartphone vendors or different brands are also factors to account for, as power requirements and consumption might differ relevantly. Even within the same vendor, different models can have different consumption values. To demonstrate such behavior, we rely on Apple's Environment Reports (AER).

Apple considers a three-year life cycle for their iPhone devices. The results from several AERs, presented on Table 2, show that, in terms of power consumption per year, the minimum value is about 3.6 kWh/year and the maximum is 10.5 kWh/year, for the iPhone SE 32GB and iPhone X 256GB, respectively, according to the 2017 AER. In terms of equivalent carbon dioxide per year, a minimum of 1.8 $kgCO_2$-e/year and a maximum of 5.3 $kgCO_2$-e/year corresponds to the indicated power consumption. These values are close to results from other studies, which estimate average values of 4.5 kWh/year and 2 kWh/year per smartphone [92,93]. Considering the whole smartphone's life cycle, the total equivalent carbon dioxide footprint ranges from 45 $kgCO_2$-e/yr to 100 $kgCO_2$-e/yr, though is typically 40 $kgCO_2$-e/yr according to [94], which is the reference chosen in this work.

On the other hand, another distinguishing factor regarding energy consumption depends heavily on usage patterns. Human usage habits will imply different power consumption patterns. With increasing unbounded access to data and real time data sharing between humans themselves and machines (e.g., M2M communications), energy hungry behavior will manifest itself through the usage of data and power consumption hungry applications, especially multimedia communications and real time collaborative gaming [12].

6.3. Overall Network Consumption

In order to provide a notion of the amount of CO_2-e and the global CF that a mobile communications network is estimated to have by 2020, according to [95] the overall CF will be 235 $MtCO_2$-e, and it will be 290 $MtCO_2$-e by 2025. But, more than the value itself, it is relevant to understand the breakdown of such value. As such, there are mainly six components in a cellular network that contribute to the total amount of CO_2-e produced per year: (i) mobile device manufacturing; (ii) mobile device operation; (iii) RAN sites' manufacturing and construction; (iv) RAN sites' operation; (v) MNO activities; and (vi) datacenters and data transport.

Most recently, the latest estimate of energy consumption of mobile networks circles around about 130 TWh/yr, representing a CF of 110 $MtCO_2$-e/yr for edge RAN alone, added up by 90 $MtCO_2$-e/yr from the mobile edge (including manufacture and use), totaling 200 $MtCO_2$-e/yr [96]. The results from [96] confirm, in a way, the estimations from [95].

This work focuses solely on RAN and mobile computing edges. This means that from Table 1 only three components will be considered, contributing to a total of 43% of whole CO_2-e produced in one year. This value is obtained from summing the components related to RAN sites' operation (RSO) which results in 29%, plus the RAN sites' manufacturing, corresponding to 4%; and finally, mobile device operation representing 10%, as presented in Table 1.

Additionally, we assume that RAN specific RAN sites' operation and RAN sites' manufacturing can be integrated into a single variable which reflects all RAN CF. All CF calculations can be expressed through the following equations:

$$CF_{TOT} = CF_{RAN} + CF_{OTHER} + CF_{MDO} \text{ [MtCo}_2\text{e/yr]}$$

where CF_{RAN} represents the carbon footprint related to the RAN edge; CF_{OTHER} represents the remaining CF related to all other components; CF_{MDO} represents the CF related to mobile device operation (see Table 1) and

$$CF_{RAN} = CF_{RSO} + CF_{RSM} \text{ [MtCo}_2\text{e/yr]}$$

$$CF_{OTHER} = CF_{DDT} + CF_{MO} + CF_{MDM} \text{ [MtCo}_2\text{e/yr]}$$

with CFx representing the carbon footprint related to component X, which can be extracted from Table 1.

In our case, we have considered an edge approach, and as such, we consider the following overall CF:

$$CF_{REdge} = CF_{RAN} = CF_{RSO} + CF_{RSM} \text{ [MtCo}_2\text{e/yr]}$$

$$CF_{MEdge} = CF_{MDO} \text{ [MtCo}_2\text{e/yr]}$$

For the considerations and assumptions of this work, the total amount of CF that must be offset and neutralized will be the sum of both edges we are focusing on

$$CF_{TEdges} = CF_{REdge} + CF_{MEdge} \text{ [MtCo}_2\text{e/yr]}$$

where CF_{TEdges} represents the total carbon footprint for both RAN edge (CF_{REdge}) and mobile edge (CF_{MEdge}).

From what is shown, it can be seen that to meet the objective of turning both edges carbon neutral, successfully reducing and offsetting CF on both ($CF_{Reduced_{REdge}}$ and $CF_{Reduced_{REdge}}$), the overall consumption and CF must be reduced by 43% of the total estimated CF (CF_{Total}) for 2020 and 2025 (Table 1). In this case we have

$$CF_{TotalReduced} = CF_{Reduced\_REdge} + CF_{Reduced\_MEdge}$$

$$\begin{cases} CF_{Reduced_{REdge}} = w_{REdge} \cdot CF_{Total} = 0.33 \cdot 235 = 77.55\ MtoCO2e\ , for\ 2020 \\ CF_{Reduced_{REdge}} = w_{REdge} \cdot CF_{Total} = 0.33 \cdot 290 = 94.70\ MtoCO2e\ , for\ 2025 \\ CF_{Reduced\_REdge} = w_{MEdge} \cdot CF_{Total} = 0.10 \cdot 235 = 23.5\ MtoCO2e\ , for\ 2020 \\ CF_{Reduced\_REdge} = w_{MEdge} \cdot CF_{Total} = 0.10 \cdot 290 = 29.0\ MtoCO2e\ , for\ 2025 \end{cases}$$

where w_{REdge} and w_{MEdge} represent the weight or percentage of contribution for overall CF from RAN and mobile edges, respectively.

Resuming, the total amount of CF that needs to be addressed in 2020 is a total of 101.05 $MtCO_2$-e/yr and 124.7 $MtCO_2$-e/yr by 2025, corresponding roughly to 43% of the overall CF (29% + 10% + 4%) for the corresponding years.

As mentioned, the proposed model is a two tier one, meaning that EE and the carbon offset method in conjunction need to be able to reduce and offset 43% of the overall considered CF. Nevertheless, these are estimated world consumption values, in megatons, meaning that it is important to understand how the whole offsetting process works and how much biotic sequestration it takes to offset the CF of individual elements, such as a base station, a femtocell or an individual smartphone.

7. Carbon Sequestration Estimation

In this section we present the 2nd tier of the proposed method, which consists of CO_2 sequestration and storage in order to offset the RCF after applying the 1st tier EE, previously described.

As mentioned, biotic CSS is considered in this paper. Following our previous work [7], the CSS method is the same in the form of tree planting. For simplicity of the model we do not consider climate changes; tree species, age or size; growth rates; or soil type. As such, we base our proposed methodology on the official UK program called UK Woodland Carbon Code, which is managed by the Scottish Forestry, a Scottish Governmental Agency. This agency is responsible for forestry policy, support and regulation, on behalf of the Forestry Commission in England, the Welsh Government and the Northern Ireland Forest Service [97]. Prior to April 2019 the sequestering data and characteristics were different from the ones today and have been enhanced. This presents us the opportunity to further enhance the model and contribute to the climate effort of reducing GHG emissions or achieving carbon neutrality.

Another different perspective from our previous work is that, considering the objective of the current work and in particular the existence of the first tier of EE, instead of considering collective CSS, we will adopt individual CSS. This means that, instead of assuming two different species as before for carbon offset calculations, we consider only one tree species, the *Fagus sylvatica*, commonly known as beech (BE). Nevertheless, in this section we compare it to conifer species (CON), to comprehensively clarify why BE was chosen. Starting with Figure 1 it is shown that BE has overall more capacity for CSS than a CON over a period of 5 years, as considered in [97]. Spacing and density will still remain the same, which are the two factors that maximize CSS, according to the data from [97]. A second objective of this work is to understand individually, the impact of a single BE tree over the course of a five-year lifecycle period.

Table 2. CF breakdown per apple smartphone.

	Production [%]	Customer Use [%]	Transport [%]	Recycling 100%	Total [kg CO_2-e]	Total Customer Use Life Cycle 3Y [kg CO_2-e]	Total Customer Use Life Cycle per Year [kgCO2-e]	Total Customer Use per Year [kWh]
iPhone X 64GB	80	17	2		79	13.43	4.48	8.9
iPhone X 256GB					93	15.81	5.27	10.5
iPhone SE 32GB	83	12	4		45	5.4	1.80	3.6
Phone SE 128GB					53	6.36	2.12	4.2
iPhone 7 32GB	78	18	3		56	10.08	3.36	6.7
iPhone 7 128GB					63	11.34	3.78	7.6
iPhone 7 256GB					75	13.5	4.50	9.0
Iphone Plus 7 32GB	78	18	3	1%	67	12.06	4.02	8.0
Iphone Plus 7 128GB					74	13.32	4.44	8.9
Iphone Plus 7 256GB					86	15.48	5.16	10.3
iPhone SE 32GB	82	14	3		75	10.5	3.50	7.0
iPhone 6s Plus 32GB	78	18	3		63	11.34	3.78	7.6
iPhone 6s Plus 128GB					70	12.6	4.20	8.4
iPhone 6s 32GB	80	16	3		54	8.64	2.88	5.8
iPhone 6s 128GB					61	9.76	3.25	6.5
iPhone 6	85	11	3		95	10.45	3.48	7.0
iPhone 6 Plus	81	14	4		110	15.4	5.13	10.3

Methodology and Assumptions

From the Woodland Carbon core initiative, one can extract data about estimated capacity of biotic sequestration for certain types of trees from establishment to a total of 200 years. The reporting timespan is divided into 5-year periods, with the aim of minimizing the yearly variation in growing conditions. This means that uniform growth is considered per month, for one year. In the following we will focus solely on the first 5 years. Table 3 shows the assumptions based on the information obtained from [97]. Another assumption is about the type of tree itself. We aim to extract the most CO_2 out of atmosphere as possible, and in order to do that we have considered broadleaf (BL) or hardwood, a type of trees that has a high capacity of photosynthesis and, thus is able to sequester high amounts of CO_2. Additionally, this species is the one that is most fit to European regions, considering the climate [97]. Another option would be conifer, but their capacity of retaining CO_2 is lower when compared with BL. It can be seen from Figure 2 that in the first 5 years BL is able to sequester more than CON; namely, an average of 133%/year over the course of the first five years. Especially one year after establishment, one hectare of BE is able to sequester 167% more CO_2 than the CON. The lookup tables also consider tree thinning. In our case, we do not consider that, because according to the data, it is performed after the first 5 years, while the focus of this work is the first five years [97].

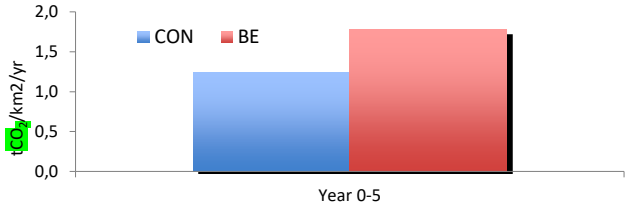


Figure 1. Total CO_2 sequestration capacity in the first five-year period.

Nevertheless, the existing data also shows that the highest CSS capacity occurs while the trees are not yet mature, up to 35–40 years, and still, from a CO_2 sequestration capacity perspective, BL is superior to CON over that period, as shown in Figures 2–4, illustrating the thinning effect on CO_2 sequestration, which basically does not exist until right before a tree becomes mature, at around 25–30 years old [97]. As such, that effect is not considered in this work.

Table 3. Individual sequestration and storage (CSS) model parameters.

Parameter	Assumption
Type of Tree	Broadleaf (BL)
Species	Beech, (BE) Fagus sylvatica (BE)
Considered spacing (m)	1.2
Yeld Class	6
Thinned or non-thinned	both
LifeCycle in Analysis	First 5 years
Growth Rate	Uniform during the lifecycle
Sequestration Capability	Uniform during the lifecycle
Sequestation Quantity [tCO_2-e]	~0.4/year

It can be seen from Figure 2 that in the first 5 years BL is able to sequester more than CON; namely, an average of 133%/year over the course of the first five years. Especially one year after establishment, one hectare of BE is able to sequester 167% more CO_2 than the CON. The lookup tables also consider tree thinning. In our case, we do not consider that, because according to the data, it is performed after the first 5 years, while the focus of this work is the first five years [97]. Nevertheless, the existing data also shows that the highest CSS capacity occurs while the trees are not yet mature, up to 35–40 years, and still, from a CO_2 sequestration capacity perspective, BL is superior to CON over that period, as shown in Figures 2–4, illustrating the thinning effect on CO_2 sequestration, which basically does not exist until right before a tree becomes mature, at around 25–30 years old [97]. As such, that effect is not considered in this work.

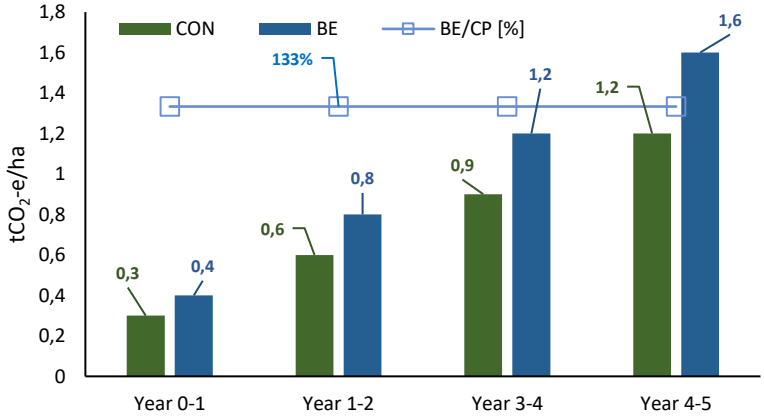


Figure 2. Per year sequestration capacity comparison between broadleaf (BL) and conifer species (CON) in the first 5 years.

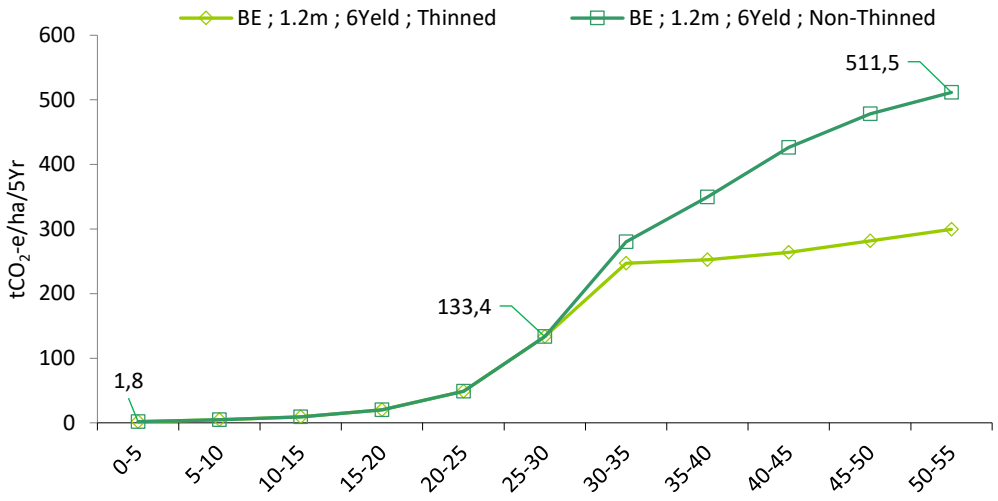


Figure 3. Cumulative carbon standing over time—BL.

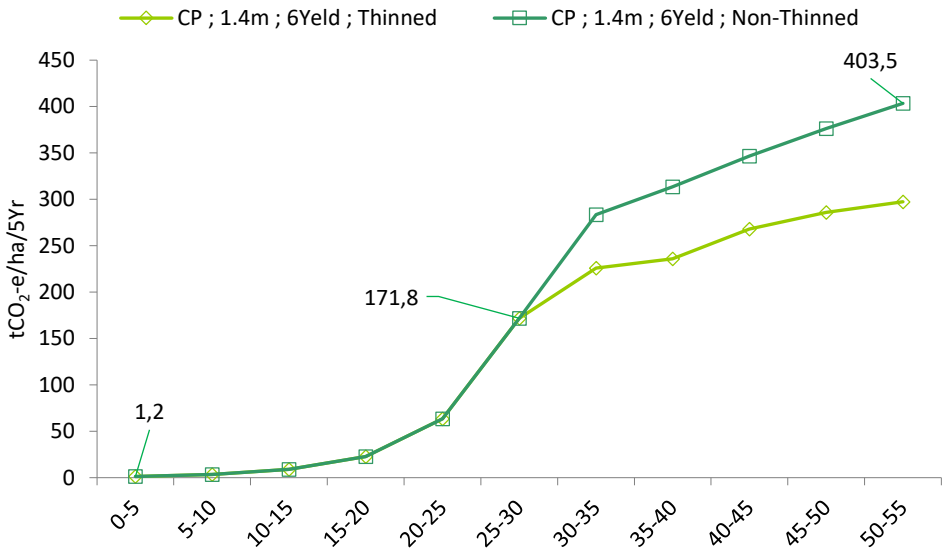


Figure 4. Cumulative carbon standing over time—CON.

With reference to the above, a comparison can now be performed between BE and CON in the first five years, considering uniform growth rate and CO_2 sequestration capacity during the first 5-year period. Table 4 presents the values that will be considered, from that assumption.

Table 4. Individual CSS model parameters.

Parameter	Value
BE sequestration capacity [tCO_2-e] per year	~0.4
CON sequestration capacity [tCO_2-e] per year	~0.2
Area Considered	1 hectare
Tree Spacing	1.2 m
Number of Trees per Hectare	~6700

From Table 4 one derives the information that will become the baseline for the study from now on:

- BE will be considered solely because it is the species that can sequester most CO_2 not only in the first five years but also in long-term, as depicted in Figure 3;
- Each hectare has approximately 6,700 trees considering a tree spacing of 1.2 m;
- All the 6,700 BE trees have the ability to sequester 400,000 $kgCO_2$-e/yr per hectare;
- One single BE tree, considering the uniformity previously referred to, can sequester approximately 60 $kgCO_2$-e/yr.

These are the fundamental parameters and corresponding values used from this point on. The next section will discuss two types of results: understanding how many trees are needed to individually sequester the CFs of a smartphone, a femtocell and a microcell; applying the methodology to a small example 5G NR network in order to understand how both tiers will work.

8. Results and Discussion

As previously referred to, the focus here will be two-fold. First an analysis of the ability to offset individual components will be shown. Secondly a small network simulation will be performed in order to understand how to carbon offset the whole system, achieving carbon neutrality.

Arriving at this point it is important to understand that, as mentioned before, on average EE methods are able to reduce 10% of overall energy consumption. Therefore, we assume a 10% overall reduction of CF due to EE methods. In the third and last subsection an example of a network model is presented, where the two-tier CF reduction methodology is applied, and results and discussion are presented.

8.1. Individual Analysis

We have considered the typical CF of a smartphone life cycle (which is 5 years, aligning with the 5 year period of biotic sequestration that is being considered), which is very close to that we have observed on average in Apple devices that were studied, as presented in Section 6.2 [94].

Table 5 presents a summary of the relevant parameters that were considered for this part of the analysis, considering only RAN and mobile edges, as previously referred to.

Table 5. Average values of individual CFs.

Parameter	Value
Smartphone device [95]	40 kg/0.040 ton
MacroCell [8]	2,531 kg/2.351 ton
FemtoCell [8]	15 kg/0.015 ton

The values for the CFs of both Femto and Macrocells were extracted from [7], assuming an uptime of 8765 h/year and a total power consumption in full load of $P_{FEMTO} = 6$ W and $P_{MACRO} = 1000$ W, and considering the annual CO_2 mass equivalent per kWh in Europe $\gamma_{Europe} = 288.74$ g/kWh. Having characterized the level of CF that each component represents, one can apply the EE and the CSS methods, knowing that

$$CF_{TotalReduced} = CF_{Reduced\_EE} + CF_{Reduced\_CSS},$$

where

$$\begin{cases} CF_{Reduced\_EE} = 0.10 \cdot CF_{TotalReduced} \\ CF_{Reduced\_CSS} = 0.90 \cdot CF_{TotalReduced} \end{cases}$$

which means that EE will be responsible for reducing 10% of overall CF for all components in both edges and CSS will have to try to offset the RCF which is 90% of the whole CF of each element.

Table 6 shows the amount of CF that each tier should perform.

Table 6. CF reduction per tier.

Component	Total CF Reduction [CO_2-e/yr]	
	EE—Tier 1	CSS—Tier 2
Smartphone device (Mobile)	4 kg	36 kg
MacroCell (RAN Edge)	~253 kg	~2,278 kg
FemtoCell (RAN Edge)	1.5 kg	13.5 kg

As such, the values that CSS should reduce are the ones referring to Tier 2. Considering that a single specimen of the considered trees is able to sequester and store approximately 60 kgCO_2-e/year, some initial conclusions can be drawn:

- One single BE tree can offset in one year an amount of CF equivalent to:
  - The yearly CF of four femtocells or;
  - The yearly CF of one smartphone.
- After 5 years that single BE tree is able to offset the equivalent to approximately 300 kgCO_2-e during that year, an amount equivalent to:
  - The yearly CF of 22 femtocells or;
  - The yearly CF of eight smartphones.

Figure 5 depicts how the CF sequestration and storage evolves for a single tree in the first five years. As mentioned, it has a linear and uniform increase per year. Such a CSS increase allows one single tree to offset the yearly CF from one to eight smartphones, or, equivalently, offset the yearly CF of 2–22 Femtocells. It also shows that a single BE tree is not able to neutralize the CF of a single macrocell. At its peak in year 5, the CSS of 299 kgCO_2-e/year of a single BE tree represents only 13% of the whole CF that a macrocell represents. Anyway, if one were to consider partial carbon neutralization, a single tree would partially neutralize 13% of the macrocell's CF. This is related only to the 2nd tier. If one considers the two tiers, EE plus CSS, that value would increase to 23% (10% EE plus 13% CSS). At an average cost of $0.1 (0.09€) /tree, which includes maintenance according to [98], offsetting 23% of a macrocell's CF would have very low and negligible CAPEX (without considering terrain costs, which we do not focus on in this work). Regarding the macrocells, the question is how many trees would it take to totally offset a macrocells' RCF and achieve carbon neutrality?

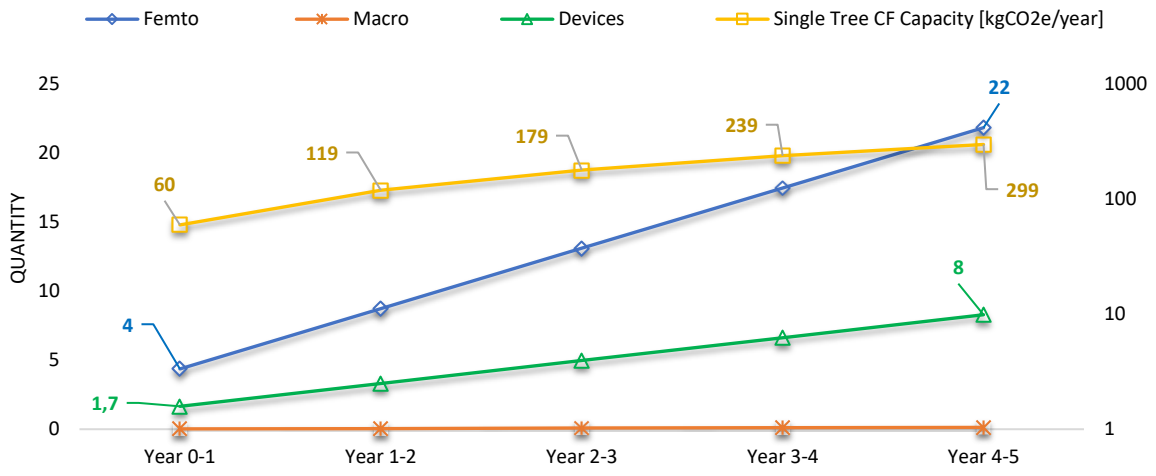


Figure 5. Cumulative carbon standing over time—CON.

If we increase the biotic RCF sequestration capacity by adding additional trees, it easy to see that to fully offset a 5G NR macrocell and achieve carbon neutrality, 40 trees would be needed, as depicted in Figure 6. As it can be seen, in order to completely compensate the CF of a macrocell right after being deployed, 40 trees would have to be planted, and carbon neutrality for that macrocell would be achieved after one year.

Alternatively, those 40 trees could also be considered as the necessary amount to achieve neutrality of 175 femtocells or 66 smartphones. Most importantly, the cost to make a 5G NR macrocell carbon neutral would be around $4.

Another interesting analysis that can be done is whether an MNO should expect to be carbon neutral from the start or plan a program. Figure 7 shows the difference. As it can be seen, if an MNO plans to achieve neutrality, for instance for 10 macrocell at year 0, the number of trees that would be needed is much higher due to their lower CSS. On the other hand, if the MNO plans to achieve neutrality in 5 years when each tree achieves a five times greater capacity for offsetting CO_2, then it would require less trees to be planted on year 0 and the CAPEX would be greatly reduced. For the extreme case of deploying a 1,000 macrocell network and planning to achieve neutrality for all, the difference between doing it immediately (referred as greenfield in Figure 7) or planning to achieve so in a five-year horizon represents one fifth of the cost.

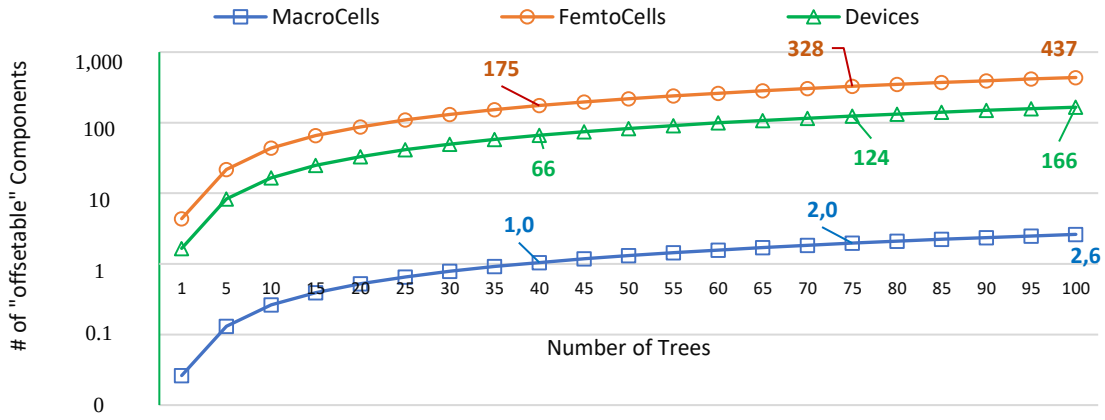


Figure 6. Number of trees required to archive carbon neutrality for each element.

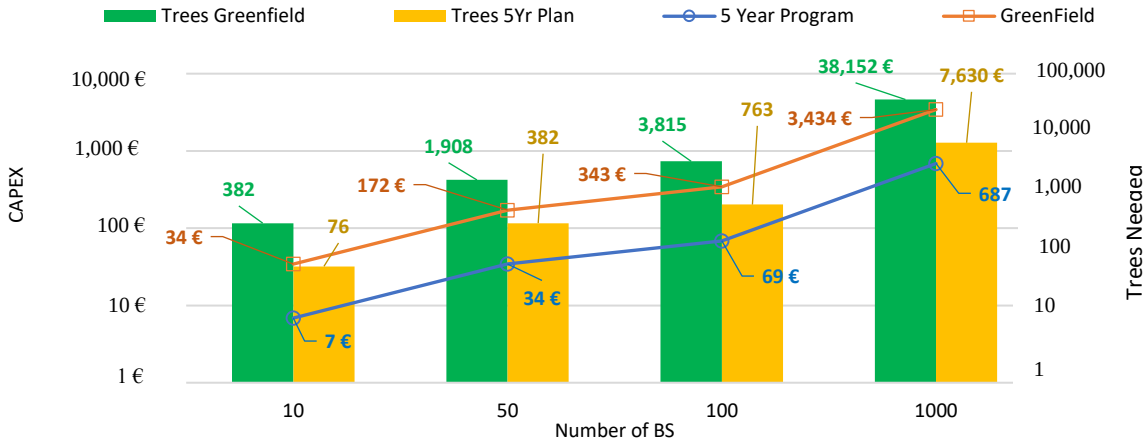


Figure 7. Capital expenditure (CAPEX) difference between 1st year offsetting and a 5 year program.

As such, proper planning for carbon neutrality makes sense, especially if one considers the lifetime of the biotic elements and the fact that up to 35–40 years after initial establishment, this kind of tree continues to increase its CO_2 sequestration capacity.

The MNO can perform its planning from a different perspective, without considering terrain costs, maintenance and other operational aspects. It can develop a partnership with a city hall's government in order for them to plant and maintain the biotic mass within cities; existing terrains; or, e.g., alongside roads and highways. This is an example of a shared business model, which can present very interesting results, and still the MNO will be able to offset very relevantly its CF. Our model shows that, as an example, if one considers a highway of 25 km, BE planting could occur alongside the highway itself and on the central part splitting the lanes in both directions. This would represent, roughly, 75 km of BE planted at a spacing of 1.2 m, resulting in a CSS capacity of 1.12 MtCO_2-e/year or, in other words, the carbon neutrality for 540,000 smartphones. The overall investment in biotic mass would be around 15,900€.

A second approach can be indirect, through the adoption of a carbon offset program. Carbon offset programs are becoming more and more available, allowing companies to invest in biotic sequestration indirectly. This method is not only the greenest but creates revenue for people involved in it by giving them jobs. For the sake of this work, two simple examples are given from [99], where a donation program helps to create an offset program.

One example is the donation of $100, roughly 90€, according to which the program will plant 1000 seedings which results in a capture capability of the equivalent to 40,000 kgCO_2-e/year after 1 year. Such an amount, considering the same type of tree, would be enough to achieve total neutrality for 17 5G NR macrocells, 2,930 femtocells or 1,110 smartphones.

A second program consists of donating $640, roughly 576€, and the program will create a forest with 6,400 trees in order to offset the RCF. This value represents an amount of 256,000 kgCO_2-e/year of carbon sequestration, which can, after five years, represent 1.28 tCO_2-e/year. These 6400 trees are roughly the same number as on the Woodland Carbon Code Forestry program, referred in the previous sections, which can sequester from 400,000 kgCO_2-e/year on the first year to 1.8 tCO_2-e/year in the 5th year after planting. This represents the same to a hectare of 6700 BE separated 1.2 m apart. In this case the amount of sequestered CO_2—for the first year—would allow in bulk to offset and achieve carbon neutrality for 176 macrocells, roughly 29,269 femtocells and 11,100 smartphones, approximately.

If an MNO considers investing a whole hectare, for a 5-year CO_2 offset program starting with a CO_2 sequestration capacity of 400,000 kgCO_2-e/year on the first year and finishing 1.8 tCO_2-e/year, the result would be as depicted in Figure 8.

An MNO that invests in a hectare for instance, to achieve carbon neutrality after five years, in the first year will be able to successfully achieve carbon neutrality for 176 of its macrocells, and in the 5th year, for 878 of its macrocells.

As an example, a small MNO that wishes to introduce femtocells to enhance the capacity of its already EE optimized 5G NR network, can plan its femtocell deployment according to its capacity to achieve carbon neutrality, deploying roughly 29,270 in the first year, and doubling up its capacity year after year until reaching the 5th year, where a total of 146,344 femtocells would be deployed and still achieve carbon neutrality.

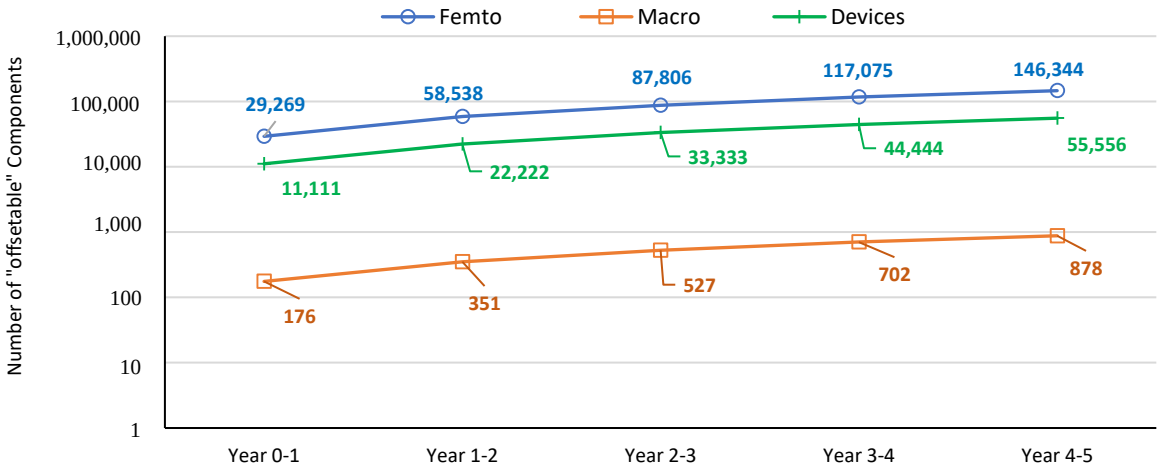


Figure 8. Number of edge components that can be offset per type in a 5-year CO2 offset program.

If we consider the life cycle of a whole cellular generation with the duration of 10 years, the same trees can still be maintained because, as mentioned previously, the trees will still be increasing their CO_2 sequestration capacity up until 30 years old. Figure 9 shows the increase in the number of different element types that could be supported and still be carbon neutral after 10 years. It is important to note that those three categories do not add up; i.e., on the 10th year the amount of CO_2 that can be sequestered is 4.5 tCO_2-e/year, which is enough to achieve carbon neutrality for each one of the represented elements independently—either macrocells, femtocells or devices—but not the sum of any of these components. If one wants to offset all three components, one would have to, in theory, consider a program with a total of 13.5 tCO_2-e/year after the 10th year, which would cost three times more; i.e., roughly 1,730 €. The mixture of a different number of components will be presented in the next subsection, where a simulated scenario is considered.

8.2. 5G NR Deployment Scenario

Accelerated roll-out of 5G NR gives MNOs the opportunity to prevent escalating network energy demand, as it is expected drive to an increase in CO_2 emissions due to unprecedent larger traffic volumes.

By accelerating 5G NR deployments, but especially, preparing for B5G networks, MNOs can set and meet targets from regulators regarding their total CO_2 emissions. That is the main reason why this work focuses on a two-tier mode is that it allows for a first step of energy reduction through network EE and secondly CO_2 emission reduction. With this in mind, in this subsection we present a very simple model of a 5G NR network and focus solely on the number of elements of both edges under analysis. As such, several assumptions are made in order to simplify the system. It is not the aim of this work to fully depict a 5G NR or future network architecture.

The first assumption is related to femtocells: we consider that less than 10 femtocells per cell is not considered a 5G NR DenseNet deployment; at least 10 femtocells should be considered per cell. The system is comprised by a logical hexagonal shaped cell with a gNB in the centre and several spread femtocells, as previously referred to. The total number of cells is seven. Figure 10 is a snapshot of the scenario, where users are considered to be mobile. This sets the baseline for users, considered to be randomly distributed along the different cells. This perspective makes sense from the strict point of view of calculating CF. The whole system is considered to have EE techniques deployed, which will bring down overall power consumption by 10%, thereby reducing equally, the overall CF by 10%.

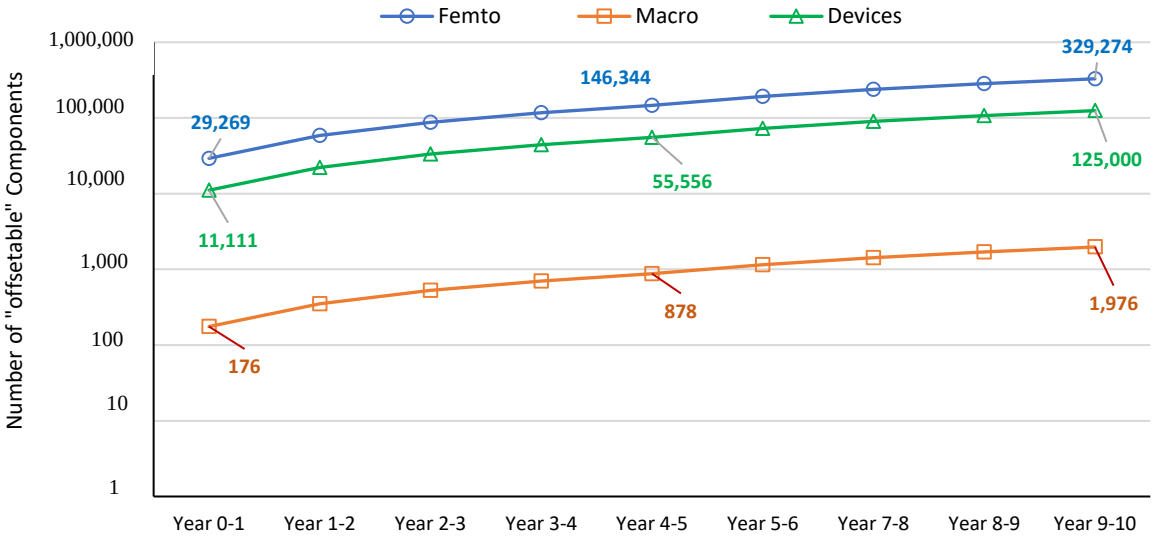


Figure 9. Number of edge components that can be offset per type in a 10-year CO2 offset program.

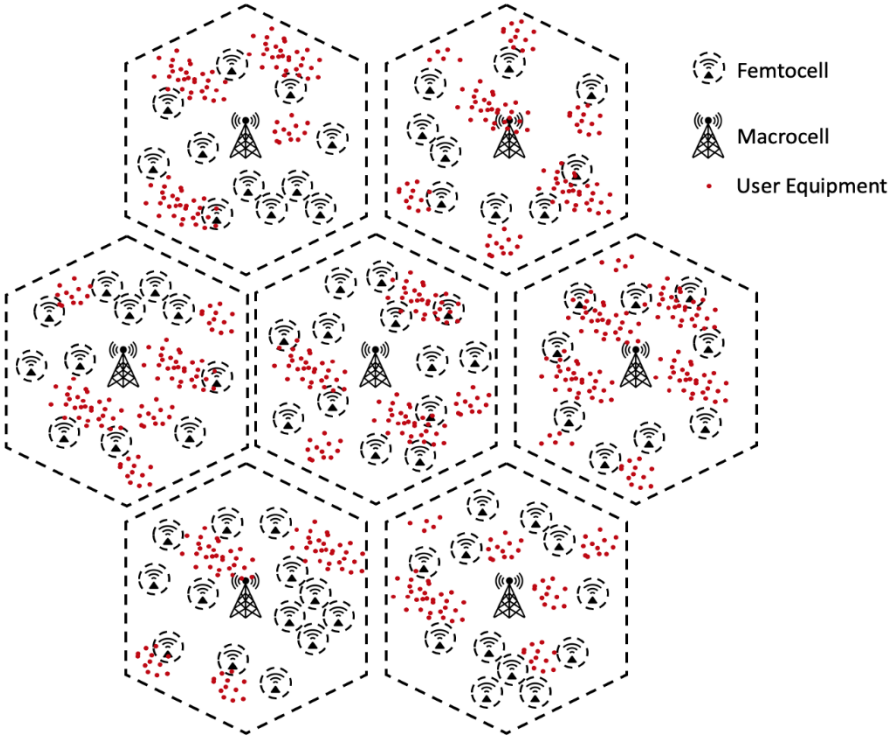


Figure 10. Example system model.

For the sake of simplicity, we do not consider the existence of Wi-Fi access points which might also be considered in a multi-RAT 5G NR deployment.

We define $\mathcal{B}_M$ and $\mathcal{B}_F$ as the set of M macrocells and F femtocells within our proposed scenario, respectively. Particularly, let us define $b \in \mathcal{B} = \mathcal{B}_M \cup \mathcal{B}_F$ the index of a generic base station and also $u \in \mathcal{L}$, as a generic user device belonging to the subscribers set $\mathcal{L}$.

Considering the power consumption of all elements described in previous sections, which are present on the RAT edge, the total CF is given by:

$$CF_{REdge} = \sum_{b \in \mathcal{B}} CF_b$$

From the same perspective, the CF on the mobile edge is given by:

$$CF_{MEdge} = \sum_{u \in \mathcal{L}} CF_u$$

The total system CF is given, therefore, by:

$$CF_{System} = CF_{REdge} + CF_{MEdge} = \sum_{b \in \mathcal{B}} CF_b + \sum_{u \in \mathcal{L}} CF_u$$

where we consider the amount of $F = 20$ femtocells per $M = 7$ macrocells. We also consider $U = 14{,}000$ subscribers in the system, performing high traffic demanding usage of their smartphone, with a mean value of 2000 subscribers per cell.

For the considered system, the CF is evaluated as depicted in Figure 11. From Figure 11, the total CF of the whole system can be obtained as

$$CF_{System} = CF_{REdge} + CF_{MEdge} = 579841 kg/Co2e/Yr,$$

and the RCF after applying EE as the 1st tier is given by

$$RCF_{System} = (1 - \eta_{EE}) \cdot RCF_{System} = 521857 kg/Co2e/Yr,$$

where $\eta_{EE} = 0.1$ represents the reduction factor due to the application of EE technologies to the whole system.

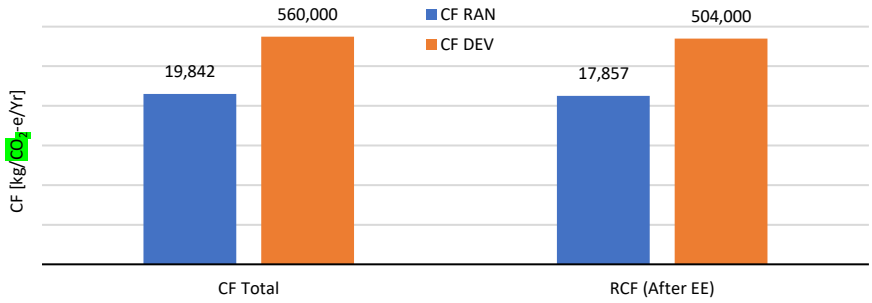


Figure 11. Total CF for the system before and after energy efficiency (EE) 1st tier reduction.

The remaining CF must then be handled by the 2nd tier, which is assumed to be biotic CSS. It was previously explained that a whole hectare of around 6,700 trees is able to offset 1.8 tCO_2-e/year. At this point two different scenarios may be addressed by the MNO: offset immediately all the CF of the system; or progressively reduce it, in a 5-year time span, allowing the trees to increase their capacities and reach the required value after 5 years.

To completely offset the current CF value of 521,857 kg/CO_2-e/year, the whole CF consumption is 30% above the CO_2 offset capacity of 6,700 tree in the first year. In order to completely offset that value, an additional capacity of 122,000 kg/CO_2-e/year is required. This represents the need for an additional 2,333 BE trees, totaling 8,700 trees, with an added price of 183€. As such, the cost for becoming carbon neutral through biotic CSS is 786€.

If the MNO decides to not offset completely in the first year and not invest in additional carbon sequestration capacity (more trees), for example, due to possible high terrain costs which we do not consider in the current work, it can simply wait for the overall capacity to surpass the system target CF, which will occur later-on, assuming this remains constant. In that case, in the first year, 69.5% of the whole CF of both edges is offset and the MNO can wait for trees to grow their storage capacity, as depicted in Figure 11.

As it can be seen from Figure 12, shortly after the first year has passed, the amount of CSS capacity surpasses the requirements for the system, and carbon neutrality is achieved after approximately 1.5 years with an initial CAPEX of 603€, corresponding to one hectare of 6,700 BE trees. Nevertheless, it was considered that the CF of the system remained constant during the 5 years. This might be a little unrealistic, as the number of subscribers rises, and it might be necessary to deploy additional base stations.

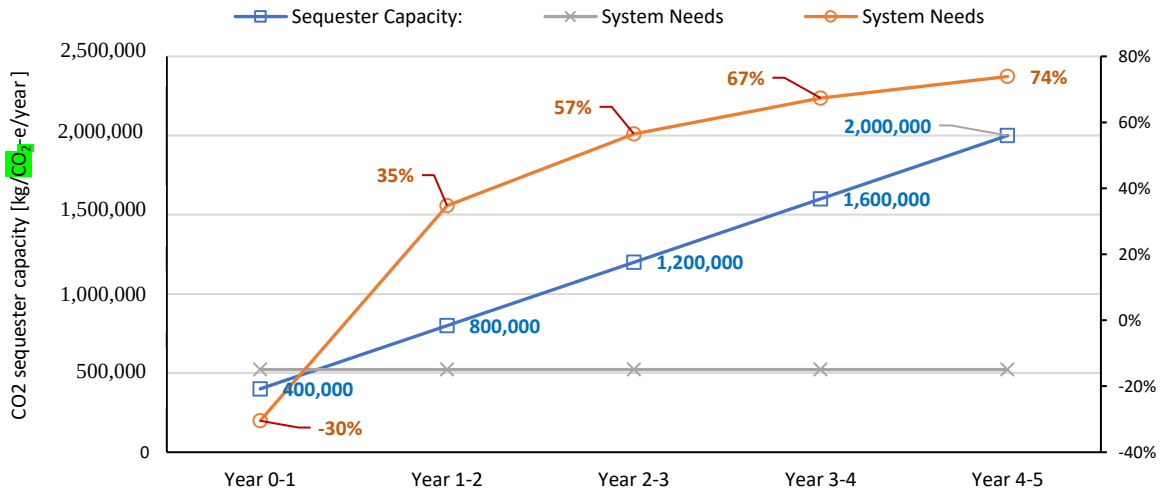


Figure 12. Evolution of CSS capacity for the first 5 years.

As it can be seen from Figure 13, if we consider 100% as being the total carbon offset capacity, any value above that means that the scenario is over the maximum CSS capacity. For example, if one considers 10% of CF rising every year, on the first year it is 44% above maximum CSS capacity (144%), but on the second year, carbon neutrality is achieved because the maximum CSS capacity has increased due to trees increasing their CSS capacity. In this case, the system overall CF represents only 72% of the overall CSS capacity. This means that carbon neutrality has been achieved and there is still 28% carbon credit.

Even if the CF rises 50% or doubles each year, which is something hardly expected in future beyond 5G networks, the increasing sequestration capacity would still be able to compensate and carbon neutrality would be achieved on 2nd and 3rd years, respectively. Note that when we are considering an increase in CF, we assume all components to increase equally in the whole network, as presented in Table 7.

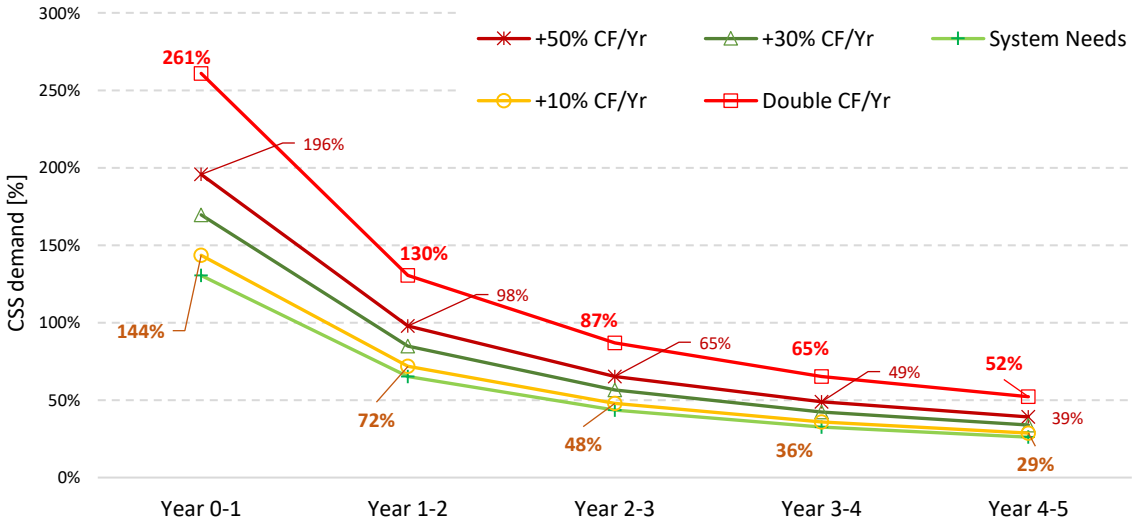


Figure 13. CSS demand considering system's *CF* increase per year.

Table 7. System residual *CF* (RCF), neutrality and system expansion.

System RCF	Carbon Neutral?	System Expansion	Investment in Biotic CSS [€]
Nominal RCF	Yes (2nd year)	Macro: 7 Femto: 140 Subscribers: 12,000	~600
10% Increase in system RCF	Yes (2nd year)	Macro: 8 Femto: 154 Subscribers: 13,200	0
30% Increase in system RCF	Yes (2nd year)	Macro: 9 Femto: 182 Subscribers: 12,000	0
50% Increase in system RCF	Yes (2nd year)	Macro: 10 Femto: 210 Subscribers: 15,600	0
100% Increase in RCF	Yes (3rd year)	Macro: 14 Femto: 280 Subscribers: 36,000	0

What Table 7 shows is that even if the MNO decides to double the number of cells, from 7 to 14, and doubles all components, both tiers will be able to compensate for that growth, and particularly, carbon neutrality is still achievable after the 3rd year of operation, for both edges. In this case, the CAPEX remains the same, approximately 600€, which is the initial investment in one hectare of BE separated 1.2 m apart. This cost is negligible for any MNO, without considering the terrain costs, which can play a relevant role in the whole cost structure. However, for the sake of simplicity and demonstrating the advantages of dual tiering for achieving carbon neutrality, we do not consider terrain cost.

Finally, it is interesting to evaluate on the 5-year time span the maximum expansion that the network system used as a reference would be able to achieve, while still being carbon neutral. It can be seen from Figure 14 that the maximum CF yearly growth that can be supported by the existing biotic CSS is 3.5 times, still with 9% of CO_2 credit. If the chosen reference system is quadrupled, the existing biotic sequestration is not enough anymore, carbon neutrality will not be achieved within the 5 years and a minimum of an additional 4% of CO_2 offset is needed. This can be achieved by acquiring additional biotic capacity, or the MNO has the possibility of choosing to increment the number of years to 6, still taking advantage of the fact that the existing biotic capacity will grow for many years more, as previously explained. It is important to note that, although we have mainly focused ourselves on the first five years, for simplicity and because it is the life cycle period considered for smartphones, CSS capacity would still continue to grow for up to 30/35 years. This means that there would be room to accommodate higher values of RCF as the years advance, still maintaining carbon neutrality.

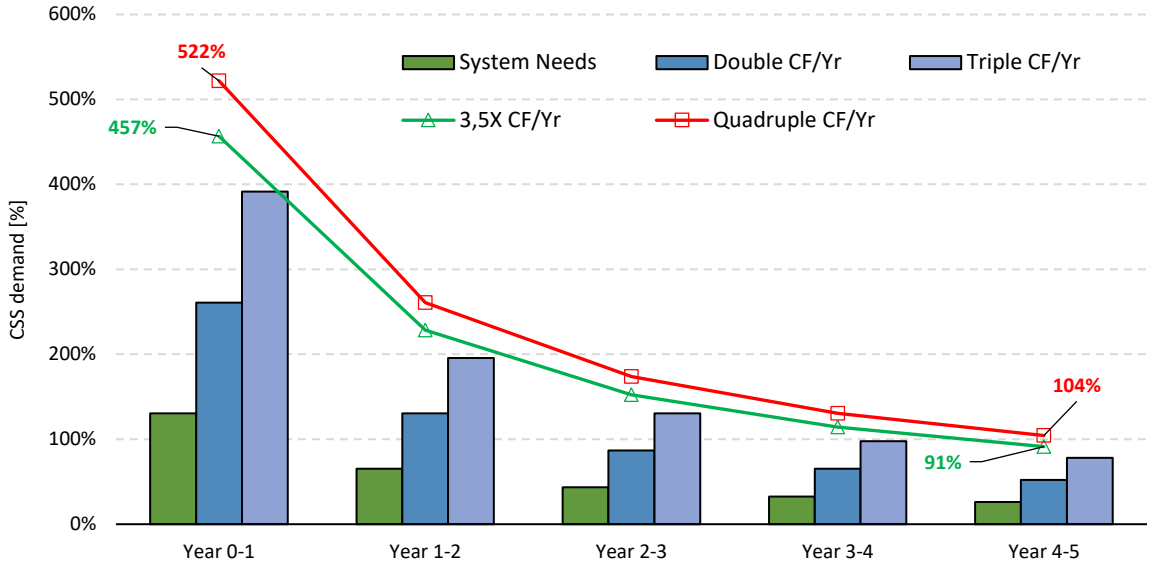


Figure 14. Maximum growth percentage still maintaining carbon neutrality.

9. Conclusions

In this work we proposed a two-tier carbon neutrality method, focused on B5G networks, but that can also be used in 5G NR networks, considering the expected surge in power consumption related to those networks.

A CSS method was proposed which is very flexible; considers the usage of the most recent EE techniques and those to come; and integrates them with biotic CSS methods. It was shown that for both network edges in focus, mobile and RAN, carbon neutrality is achievable for each individual element or set of elements. Both mobile and RAN edges were considered, and the corresponding carbon footprint was calculated in order to evaluate the feasibility of offsetting it. It was shown that both edges are the most relevant in 5G and beyond 5G networks, and that achieving 100% of CO_2 offsetting is possible.

Furthermore, a simple system model was developed and adopted in order to present additional results and analysis about the planning process. Overall, the proposed two-tier methodology contributes further to a greener environment, with the second tier being based on proven natural processes, without any impact or side effects onto nature and ecosystems, being capable to even achieve not only carbon neutrality but also carbon credit, as shown in the considered scenarios, where it achieved values around 9% or above.

As expected, the results show that offsetting CO_2 and becoming carbon neutral can be a process which is not that expensive for an MNO, but it is something that must be carefully prepared and planned upfront. For the upcoming beyond 5G networks, this analysis is primordial and of utmost importance. EE has also come a long way and additional factors may be used in B5G networks, such as the massive use of virtualization and cloud, and edge computing, which are all factors that can help to reduce the overall CF of the systems to come. As shown, it is possible to offload 521,857 kg/CO_2-e/yr with just under 800 euros, which, for the depicted scenarios, represents a quick win for every mobile network operator.

It was also shown that, if an MNO does not desire to develop its own offset activities, there are carbon offsetting programs implemented by third party companies that can be supported. In that case, the cost of implementing a side program aiming to offset as much as possible the CF of MNO's operations is possible at low cost, representing, in the presented scenarios, investments under 1,000 Euro per hectare of planted ground.

Future work will focus on the other edges, and on other EE techniques that can be used for lowering even more power consumption, and more complex B5G scenarios, where different RATs can be added and IoT network layers, in order to evaluate the overall CF and neutral achievability. Additionally, we have not

considered terrain costs, as they would create several degrees of uncertainty and did not contribute relevantly to demonstrate the advantage of having a two-tier system. This is also part of the future work that will further enhance the whole model.

Author Contributions: conceptualization, L.G., P.S., N.S. and A.C.; data curation, P.S.; formal analysis, L.G. and N.S.; project administration, N.S. and A.C.; supervision, P.S., N.S. and A.C.; validation, L.G., P.S., N.S. and A.C.; writing—original draft, L.G.; writing—review and editing, L.G.

Funding: This work was funded by FCT/MEC through national funds and co-funded by FEDER—PT2020 partnership agreement under the project "UIDB/EEA/50008/2020".

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Siddiqi, M.A.; Yu, H.; Joung, J. 5G Ultra-Reliable Low-Latency Communication Implementation Challenges and Operational Issues with IoT Devices. *Electronics* **2019**, *8*, 981.
2. 3GPP TR 21.915, Summary of Rel-15 Work Items V15.0.0, Out. 2019. Available online: https://www.etsi.org/deliver/etsi_tr/121900_121999/121915/15.00.00_60/tr_121915v150000p.pdf (accessed on 05 March 2020).
3. 3GPP TR 21.916, Summary of Rel-16 Work Items V16.0.0, Sep. 2019. Available online: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3493 (accessed on 05 March 2020).
4. The Paris Agreement. Available online: https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement (accessed on 28 December 2019).
5. The European Green Deal. Available online: https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_pt (Accessed on 28 December 2019).
6. Mitrović, M.; Malone, A. Carbon capture and storage (CCS) demonstration projects in Canada. *Energy Procedia* **2011**, *4*, 5685–5691.
7. Le Quéré, C.; Moriarty, R.; Andrew, R.M.; Peters, G.P.; Ciais, P.; Friedlingstein, P.; Zeng, N. 2014 Global Carbon Budget. *Earth Syst. Sci. Data* **2014**, *7*, 47–85.
8. Gonçalves, L.; Sebastião, P.; Souto, N.; Correia, A. 5G Mobile Challenges: A Feasibility Study on Achieving Carbon Neutrality. In Proceedings of the IEEE International Conference on Telecommunications-ICT, Thessaloniki, Greece, 1–8 May 2016; Volume 1.
9. Cisco Annual Internet Report (2018–2023), White Paper, 2020. Available online: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html (accessed on 05 March 2020).
10. Mills, M.P. The Cloud Begins with Coal: Big Data, Big Networks, Big Infrastructure, and Big Power, Digital Power Group, 2013. Available online: https://www.tech-pundit.com/wp-content/uploads/2013/07/Cloud_Begins_With_Coal.pdf (accessed on 05 March 2020).
11. Andrae, A.S.; Edler, T. On global electricity usage of communication technology: Trends to 2030. *Challenges* **2015**, *6*, 117–157.
12. Auer, G.; Giannini, V.; Desset, C.; Godor, I.; Skillermark, P.; Olsson, M.; Fehske, A. How much energy is needed to run a wireless network? *IEEE Wirel. Commun.* **2011**, *18*, 40–49.
13. Zhang, H.; Liu, N.; Chu, X.; Long, K.; Aghvami, A.H.; Leung, V.C. Network Slicing Based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges. *IEEE Commun. Mag.* **2017**, *55*, 138–145.
14. Sohul, M.M.; Yao, M.; Ma, X.; Imana, E.Y.; Marojevic, V.; Reed, J.H. Next Generation Public Safety Networks: A spectrum sharing approach. *IEEE Commun. Mag.* **2016**, *54*, 30–36.
15. Sohul, M.M.; Yao, M.; Yang, T.; Reed, J.H. Spectrum access system for the citizen broadband radio service. *IEEE Commun. Mag.* **2015**, *53*, 18–25.

16. Masoudi, M.; Khafagy, M.G.; Conte, A.; El-Amine, A.; Françoise, B.; Nadjahi, C.; Bodéré, D. Green Mobile Networks for 5G and Beyond. *IEEE Access* **2019**, *7*, 107270–107299.
17. Buzzi, S.; Chih-Lin, I.; Klein, T.E.; Poor, H.V.; Yang, C.; Zappone, A. A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 697–709.
18. Mahapatra, R.; Nijsure, Y.; Kaddoum, G.; Hassan, N.U.; Yuen, C. Energy Efficiency Tradeoff Mechanism Towards Wireless Green Communication: A Survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 686–705.
19. Davaslioglu, K.; Ayanoglu, E. Quantifying Potential Energy Efficiency Gain in Green Cellular Wireless Networks. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 2065–2091.
20. Peng, M.; Yan, S.; Zhang, K.; Wang, C. Fog computing-based radio access networks: Issues and challenges. *IEEE Netw.* **2016**, *30*, 46–53.
21. Peng, M.; Sun, Y.; Li, X.; Mao, Z.; Wang, C. Recent Advances in Cloud Radio Access Networks: System Architectures, Key Techniques, and Open Issues. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 2282–2308.
22. Li, Y.; Chen, M. Software-Defined Network Function Virtualization: A Survey. *IEEE Access* **2015**, *3*, 2542–2553.
23. Fundation, O.N. Software-Defined Networking: The New Norm for Networks. *ONF White Pap.* **2012**, *2*, 2–6.
24. Agiwal, M.; Roy, A.; Saxena, N. Next Generation 5G Wireless Networks: A Comprehensive Survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 1617–1655.
25. Jiang, H.; Yi, S.; Wu, L.; Leung, H.; Wang, Y.; Zhou, X.; Yang, L. Data-Driven Cell Zooming for Large-Scale Mobile Networks. *IEEE Trans. Netw. Serv. Manag.* **2018**, *15*, 156–168.
26. Han, F.; Zhao, S.; Zhang, L.; Wu, J. Survey of Strategies for Switching Off Base Stations in Heterogeneous Networks for Greener 5G Systems. *IEEE Access* **2016**, *4*, 4959–4973.
27. Yu, N.; Miao, Y.; Mu, L.; Du, H.; Huang, H.; Jia, X. Minimizing Energy Cost by Dynamic Switching ON/OFF Base Stations in Cellular Networks. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 7457–7469.
28. Zhang, H.; Cai, J.; Li, X. Energy-efficient base station control with dynamic clustering in cellular network. In Proceedings of the IEEE International Conference on Communications and Networking (CHINACOM), Guilin, China, 14–16 August 2013; p. 384.
29. Samarakoon, S.; Bennis, M.; Saad, W.; Latva-aho, M. Dynamic Clustering and ON/OFF Strategies for Wireless Small Cell Networks. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 2164–2178.
30. Tao, R.; Zhang, J.; Chu, X. An Energy Saving Small Cell Sleeping Mechanism with Cell Expansion in Heterogeneous Networks. In Proceedings of the IEEE Vehicular Technology Conference (VTC Spring), Porto, Portugal, 15–18 May 2016; pp. 1–5, doi:10.1109/VTCSpring.2016.7504126.
31. Ghazzai, H.; Farooq, M.J.; Alsharoa, A.; Yaacoub, E.; Kadri, A.; Alouini, M.-S. Green Networking in Cellular HetNets: A Unified Radio Resource Management Framework With Base Station ON/OFF Switching. *IEEE Trans. Veh. Technol.* **2017**, *66*, 5879–5893.
32. Yuan, Y.; Gong, P. A QoE-orientated base station sleeping strategy for multi-services in cellular networks. In Proceedings of the International Conference on Wireless Communications & Signal Processing (WCSP), Nanjing, China, 15–17 October 2015; pp. 1–5, doi:10.1109/WCSP.2015.7341051.
33. Bhaumik, S.; Narlikar, G.; Chattopadhyay, S.; Kanugovi, S. Breathe to stay cool: Adjusting cell sizes to reduce energy consumption. In Proceedings of the First ACM SIGCOMM Workshop on Green Networking, Hangzhou, China, 18–20 December 2010.
34. K. Manimozhi and V. Vijayalakshmi, "Optimized energy-aware context based switching relay scheme for HetNets," *2017 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, 2017, pp. 1261-1265.
35. Oh, E.; Krishnamachari, B.; Liu, X.; Niu, Z. Toward dynamic energy-efficient operation of cellular network infrastructure. *IEEE Commun. Mag.* **2011**, *49*, 56–61.
36. Son, K.; Kim, H.; Yi, Y.; Krishnamachari, B. Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks. *IEEE J. Sel. Areas Commun.* **2011**, *29*, 1525–1536.
37. Han, T.; Ansari, N. On greening cellular networks via multicell cooperation. *IEEE Wirel. Commun.* **2013**, *20*, 82–89.
38. Yigitel, M.A.; Incel, O.D.; Ersoy, C. Qos vs. energy: A traffic-aware topology management scheme for green heterogeneous networks. *Comput. Netw.* **2015**, *78*, 130–139.
39. Kwak, J.; Son, K.; Yi, Y.; Chong, S. Impact of spatio-temporal power sharing policies on cellular network greening. In Proceedings of the 2011 International Symposium of Modeling and Optimization of Mobile, Ad Hoc, and Wireless Networks, Princeton, NJ, USA, 9–13 May 2011.

40. Luo, S.; Zhang, R.; Lim, T.J. Optimal power and range adaptation for green broadcasting. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 4592–4603.
41. Wang, W.; Huang, Y.; You, L.; Xiong, J.; Li, J.; Gao, X. Energy Efficiency Optimization for Massive MIMO Non-Orthogonal Unicast and Multicast Transmission with Statistical CSI. *Electronics* **2019**, *8*, 857.
42. Ismaiil, K.A.H.; Assaf, B.; Ghantous, M.; Nahas, M. Reducing power Consumption of cellular networks by using various cell types and cell zooming. In Proceedings of the International Conference on e-Technologies and Networks for Development (ICeND), Beirut, Lebanon, 29 April/1 May 2014; pp. 33–38.
43. Niu, Z.; Wu, Y.; Gong, J.; Yang, Z. Cell zooming for cost-efficient green cellular networks. *IEEE Commun. Mag.* **2010**, *48*, 74–79.
44. Le, L.B. QoS-aware BS switching and cell zooming design for OFDMA green cellular networks. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), Anaheim, CA, USA, 3–7 December 2012; pp. 1544–1549.
45. Hu, Z.; Wei, Y.; Wang, X.; Song, M. Green relay station assisted cell zooming scheme for cellular networks. In Proceedings of the International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, 13–15 August 2016; pp. 2030–2035.
46. Zhu, Y.; Kang, T.; Zhang, T.; Zeng, Z. QoS-aware user association based on cell zooming for energy efficiency in cellular networks. In Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC Workshops), London, UK, 8–9 September 2013; pp. 6–10.
47. You, Y.; Sheng, J.; Zhu, Q.; Zhu, C.; Ma, D. A novel cell zooming strategy towards energy efficient based on load balancing in random heterogeneous networks. In Proceedings of the 2017 IEEE 17th International Conference on Communication Technology (ICCT), Chengdu, China, 27–30 October 2017; pp. 522–527.
48. Onur, E.; Durmus, Y.; Niemegeers, I. Cooperative density estimation in random wireless ad hoc networks. *IEEE Commun. Lett.* **2012**, *16*, 331–333.
49. Eroglu, A.; Onur, E.; Oguztüzün, H. Estimating density of wireless networks in practice. In Proceedings of the 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Hong Kong, China, 30 August–2 September 2015; pp. 1476–1480.
50. Yaman, O.; Eroglu, A.; Onur, E. Density-aware cell zooming. In Proceedings of the 2018 21st Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN), Paris, France, 19–22 February 2018; pp. 1–8.
51. Allal, I.; Mongazon-Cazavet, B.; Al Agha, K.; Senouci, S.M.; Gourhant, Y. A green small cells deployment in 5G — Switch ON/OFF via IoT networks & energy efficient mesh backhauling. In Proceedings of the 2017 IFIP Networking Conference (IFIP Networking) and Workshops, Stockholm, Sweden, 12–16 June 2017; pp. 1–2.
52. Nakamura, M.; Takeno, K. Green Base Station Using Robust Solar System and High-Performance Lithium ion battery for Next Generation Wireless Network (5G) and against Mega Disaster. In Proceedings of the 2018 International Power Electronics Conference (IPEC-Niigata 2018-ECCE Asia), Niigata, Japan, 20–24 May 2018; pp. 201–206.
53. Dutta, U.K.; Razzaque, M.A.; Al-Wadud, M.A.; Islam, M.S.; Hossain, M.S.; Gupta, B.B. Self-Adaptive Scheduling of Base Transceiver Stations in Green 5G Networks. *IEEE Access* **2018**, *6*, 7958–7969.
54. Kour, H.; Jha, R.K. Power Optimization using Spectrum Sharing for 5G Wireless Networks. In Proceedings of the 2019 11th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, India, 7–11 January 2019; pp. 395–398.
55. Mi, J.; Wang, K.; Li, P.; Guo, S.; Sun, Y. Software-Defined Green 5G System for Big Data. *IEEE Commun. Mag.* **2018**, *56*, 116–123.
56. Mowla, M.M.; Ahmad, I.; Habibi, D.; Phung, Q.V. A Green Communication Model for 5G Systems. *IEEE Trans. Green Commun. Netw.* **2017**, *1*, 264–280.
57. Mowla, M.M.; Ahmad, I.; Habibi, D.; Phung, Q.V. Energy Efficient Backhauling for 5G Small Cell Networks. *IEEE Trans. Sustain. Comput.* **2019**, *4*, 279–292.
58. Xu, X.; Yuan, C.; Chen, W.; Tao, X.; Sun, Y. Adaptive Cell Zooming and Sleeping for Green Heterogeneous Ultradense Networks. *IEEE Trans. Veh. Technol.* **2018**, *67*, 1612–1621.
59. Memon, M.L.; Maheshwari, M.K.; Saxena, N.; Roy, A.; Shin, D.R. Artificial Intelligence-Based Discontinuous Reception for Energy Saving in 5G Networks. *Electronics* **2019**, *8*, 778.
60. Andrews, J.J.G.; Buzzi, S.; Choi, W.; Hanly, S.V.S.; Lozano, A.; Soong, A.A.C.K.; Zhang, J.J.C. What will 5G be? *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1065–1082.

61. Boccardi, F.; Heath, R.; Lozano, A.; Marzetta, T.L.; Popovski, P. Five disruptive technology directions for 5G. *IEEE Commun. Mag.* **2014**, *52*, 74–80.
62. Goyal, S.; Carter, J. A lightweight secure cyber foraging infrastructure for resource-constrained devices. In Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications, Windermere, Cumbria, UK, 3 December 2004; pp. 186–195.
63. Sharifi, M.; Kafaie, S.; Kashefi, O. A Survey and Taxonomy of Cyber Foraging of Mobile Devices. *IEEE Commun. Surv. Tutor.* **2012**, *14*, 1232–1243.
64. Esposito, F.; Cvetkovski, A.; Dargahi, T.; Pan, J. Complete edge function onloading for effective backend-driven cyber foraging. In Proceedings of the 2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Rome, Italy, 9–11 October 2017; pp. 1–8.
65. Esposito, F.; Paganelli, F.; Fantacci, R. A Decomposition-based Architecture for Distributed Cyber-Foraging of Multiple Edge Functions. In Proceedings of the 2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft), Montreal, QC, Canada, 25–29 June 2018; pp. 247–251.
66. Jia, Z. Energy efficiency analysis of cellular downlink transmission with network coding over Rayleigh fading channels. *KSII Trans. Internet Info. Sys.* **2013**, *7*, 446–458.
67. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, White Paper, 2015–2020, Feb 2016. Available online: https://www.cisco.com/c/dam/m/en_in/innovation/enterprise/assets/mobile-white-paper-c11-520862.pdf (accessed on 05 March 2020).
68. Jahid, A.; Shams, A.B.; Hossain, M.F. Green energy driven cellular networks with JT CoMP technique. *Phys. Commun.* **2018**, *28*, 58–68.
69. Vereecken, W.; Van Heddeghem, W.; Deruyck, M.; Puype, B.; Lannoo, B.; Joseph, W.; Demeester, P. Power consumption in telecommunication networks: Overview and reduction strategies. *IEEE Commun. Mag.* **2011**, *49*, 62–69.
70. Sedjo, R.; Sohngen, B. Carbon Sequestration in Forests and Soils. *Annu. Rev. Resour. Econ.* **2012**, *4*, 127–144.
71. Lal, R. Soil carbon sequestration to mitigate climate change. *Geoderma* **2004**, *123*, 1–22.
72. Friedmann, J. Geological carbon dioxide sequestration. *Elem. Int. Mag. Mineral. Geochem. Petrol.* **2007**, *3*, 179–184.
73. *IPCC–Intergovernmental Panel on Climate Change, Carbon Dioxide Capture and Storage*; Special Report; Cambridge University Press: Cambridge, UK, 2005.
74. Livermont, E.A.; Koh, Y.; Mlambo, T.; Bhawanin, M.; Zhao, B. *Carbon Capture and Storage in Deep Ocean Space for the 21st Century-Guidelines for Implementation in China*; University of Southampton: Southampton, UK, 2011.
75. Shafqat, S.; Kishwer, S.; Qureshi, M.A. Energy-Aware Cloud Architecture for Intense Social Mobile (Device to Device) 5G Communications in Smart City. In Proceedings of the 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 7–9 January 2019; pp. 0739–0745.
76. J. González, N. Tejedor-Flores and R. Pinzón, "A Bibliographic Review of the Importance of Carbon Dioxide Capture in Mangroves," *2019 7th International Engineering, Sciences and Technology Conference (IESTEC)*, Panama, Panama, 2019, pp. 126-131.
77. Raven, J.A.; Karley, A.J. Carbon sequestration: Photosynthesis and subsequent processes. *Curr. Biol. J.* **2006**, *16*, R165–R167.
78. Stewart, C.; Hessami, A. A study of methods of carbon dioxide capture and sequestration–the sustainability of a photosynthetic bioreactor approach. *Energy Convers. Manag. J.* **2005**, *46*, 403–420.
79. Chavan, B.L.; Rasal, G.B. Sequestered Carbon Potential and Status of Eucalyptus Tree. *Int. J. Appl. Eng. Technol.* **2011**, *1*, 41–47.
80. Dubal, K.; Ghorpade, P.; Dongare, M.; Patil, S. Carbon Sequestration in the Standing Trees at Campus of Shivaji University, Kolhapur. *Int. Sci. J. Nat. Environ. Pollut. Technol.* **2013**, *12*, 725–726.
81. Unwin, G.L.; Kriedemann, P.E. *Principles and Processes of Carbon Sequestration by Trees*; Technical Paper; Research and Development Division State Forests of New South Wales, Sydney, Australia, 2000.
82. Lal, R. Carbon sequestration. *Philos. Trans. R. Soc. B Biol. Sci.* **2008**, *363*, 815–830.
83. Mogensen, P.E.; Koivisto, T.; Pedersen, K.I.; Kovacs, I.Z.; Raaf, B.; Pajukoski, K.; Rinne, M.J. LTE-Advanced: The path towards gigabit/s in wireless mobile communications. In Proceedings of the 2009 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, Aalborg, Denmark, 17–20 May 2009.

84. Millard, P.; Sommerkorn, M.; Grelet, G. Environmental Change and Carbon Limitation in Trees: A Biochemical, Ecophysiological and Ecosystem Appraisal. *New Phytol. J.* **2007**, *175*, 11–28.
85. Gorte, R. *Carbon Sequestration in Forests*; Congressional Research Service report for congress; DIANE Publishing: Collingdale, PA, USA, 2009.
86. Auer, G.; Blume, O.; Giannini, V.; Godor, I.; Imran, M.; Jading, Y.; Katranaras, E.; Olsson, M.; Sabella, D.; Skillermark, P.; et al. EARTH Deliverable D2.3: Energy Efficiency Analysis of the Reference Systems, Areas of Improvements and Target Breakdown, Project Deliverable D2.3. Available online: https://cordis.europa.eu/docs/projects/cnect/3/247733/080/deliverables/001-EARTHWP2D23v2.pdf (accessed on 05 March 2020).
87. Sabella, D.; Rapone, D.; Fodrini, M.; Cavdar, C.; Olsson, M.; Frenger, P.; Tombaz, S. *Energy Management in Mobile Networks Towards 5G*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 397–427.
88. Yang, X.; Wang, Z.; Wan, X.; Fan, Z. Secure Energy-Efficient Resource Allocation Algorithm of Massive MIMO System with SWIPT. *Electronics* **2020**, *9*, 26.
89. Global e-Sustainability Initiative (GeSI) Strategy, Accenture, #SMARTer2030: ICT Solutions for 21st Century Challenges. Technical Report, Brussels, Belgium, 2015. Available online: http://smarter2030.gesi.org/downloads/Full_report.pdf (accessed on 05 March 2020).
90. Carroll, A.; Heiser, G. An analysis of power consumption in a smartphone. In Proceedings of the 2010 USENIX Conference on USENIX Annual Technical Conference, Berkeley, CA, USA, 23–25 June 2010; p. 21.
91. Jensen, A.R.; Lauridsen, M.; Mogensen, P.; Sorensen, T.B.; Jensen, P. LTE UE power consumption model: For system level energy and performance optimization. In Proceedings of the IEEE Vehicular Technology Conference (VTC Fall), Yokohama, Japan, 6–9 May 2012; pp. 1–5.
92. Lauridsen, M.; Mogensen, P.; No¨el, L. Empirical LTE smartphone power model with DRX operation for system level simulations. In Proceedings of the IEEE Vehicular Technology Conference (VTC Fall), Las Vegas, USA, 2–5 September 2013; pp. 1–6.
93. Urban, B.; Roth, K.; Singh, M.; Howes, D. *Energy Consumption of Consumer Electronics in US Homes in 2017*; Final Report to the Consumer Technology Association; Fraunhofer Center for Sustainable Energy Systems, Boston, United States of America, 2017.
94. EPRI, Electric Power Research Institute: EPRI Calculates Annual Cost of Charging an iPad at $1.36 2012. Available online: http://www.epri.com/Press-Releases/Pages/EPRI-Calculates-Annual-Cost-of-Charging-an-iPad-at-$1-36.aspx (accessed on 30 December 2019).
95. ERICSSON Energy and Carbon Report, 2014. Available online: https://www.ericsson.com/assets/local/about-ericsson/sustainability-and-corporate-responsibility/documents/ericsson-energy-and-carbon-report.pdf (accessed on 2 December 2017).
96. Fehske, A.; Fettweis, G.; Malmodin, J.; Biczok, G. The global footprint of mobile communications: The ecological and economic perspective. *IEEE Commun. Mag.* **2011**, *49*, 55–62.
97. GSMA Report, 2019 Mobile Industry Impact Report: Sustainable Development Goals Climate Action Deep Dive. Available online: https://www.gsma.com/betterfuture/2019sdgimpactreport/wp-content/uploads/2019/09/SDG_Report_2019_ExecSummary_Web_Singles.pdf (accessed on 05 March 2020).
98. United Kindgom's Forestry Commission, Woodland Carbon Code, Technical Report, July 2018. Available online: https://www.woodlandcarboncode.org.uk/about (accessed on 20 December 2019).
99. Trees for the Future Foundation. Available online: http://treesftf.force.com/checkout/TFTF_Donation_Certificate_Cls (accessed on 20 December 2019).

Page intentionally left blank

Chapter III – Conclusions

This chapter outlies the principal conclusions that can be drawn from the developed work. Section 3.1 summarises the work and focuses on subjects for discussion. It also reinforces, as underlined throughout this thesis document, on the alignment of all the research artefacts in order to produce a coherent aggregated approach to optimization of mobile networks from 4G to 5G and beyond, from a retro-compatibility perspective. In section 3.2 the concluding remarks are presented, mainly focusing on the contributions of the different components of the work, as seen previously, and how all of them are tightly coupled and contribute globally to address the problems at hand. Section 3.3 presents the spectrum of areas to explore and future work to be developed based on the work that was performed.

3.1. Summary and Discussion

The development of the current work followed a standard methodology, that is summarized now. It consisted on 5 different main tasks which are depicted on figure 2. The first task consisted in defining the goals, by clearly stating the research questions, as previously enumerated. The research goals were defined after a thorough contextual and business analysis and contextualization. In this case, as a baseline, 5G NR was the business in question and the context that the work was developed on. To clarify, despite being referred earlier, the research and business goals were, respectively, to: i) identify the best network architecture based on 5G NR that would fit future networks; ii) optimize the overall performance of the existing methods by proposing new ones based on EC, NFV and also cloud environments, namely focusing on ANDSF and cloud assisted optimized traffic steering and offloading; iii) considering the massive network, as well as device edges, to develop a method in order for MNOs to become carbon neutral; iv) improve overall planning, operation and network evolution processes, as well as optimizing subscribers' quality of experience by focusing on their behaviour, evaluating how that would impact the network and if such knowledge could be used to improve future networks. As a general approach we considered as business opportunities all the advantages resulting from the research questions, *e.g.*, churn rate decrease, increased overall performance, higher quality

of service, enabling new services and applications and ARPU increase.

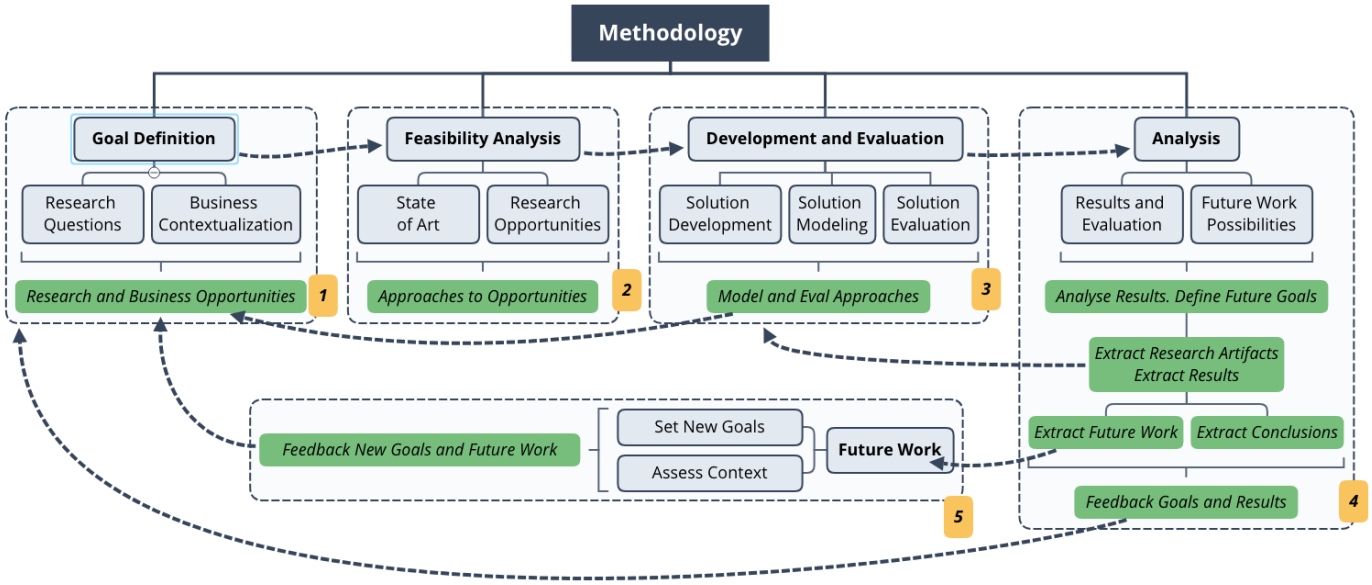


Figure 1 - Overall work methodology

With the research and business opportunities being defined, the second step on our methodology was to perform a feasibility analysis per research opportunity. By taking into consideration the state of the art for each of the research questions, each was identified as a research opportunity and the result from this second phase was a set of approaches on how to implement the research opportunities. From that point on, the third phase which consisted of developing and evaluating was started. This third phase is the most important phase because it is when solutions are developed and modelled to answer the research questions and each of the solution evaluations takes place. Such results in solutions to models and evaluation methods. Also, at this phase it was interesting to observe that some of the models and approaches that were first defined had to be rea-analysed because new opportunities resulted from this process without having properly been thought of before, and had to been re-contextualized in both phase 1 and 2. Clearly, this made the whole methodology more robust as a continuous feedback process was in place.

After having properly evaluated all of the proposed solutions, the fourth phase took place: analysis. In this phase, the proposed approaches to solve the research questions were put to test, evaluated and future work need was identified. The outcome is a set of results for each of the research questions, which we have been presented throughout the document and extensively on the produced articles, but also in Section 1.3, where all contributions were

presented. These contributions come out of the model as research artefacts. Additionally, to extract research artefacts, conclusions are also drawn on each one, and future work is defined, as it is presented in Section 3.3. The future work activities will give birth to a specific set of activities which will consist of specifying new goals, assess new contexts and, naturally, feeding back that information into the first phase of the model again. In our case, this fifth phase is what we consider to be the result of future work analysis and is out of scope in this work, but part of the identified future work. Finally, after going through all the four methodological phases, feedback is drawn on each of the research questions and overall results are drawn.

During this work, several other aspects that the authors were not focusing on appeared in the form of niche results, not being the main research questions themselves. We called these opportunities, and some were addressed through the publication of specific articles, resulting in collateral research artifacts that contributed to enrich the whole research and work even more, *e.g.*, using inter-RAT HO to address poor cell-edge performance in 5G heterogeneous networks. All research artifact, directly resulting from answers to the main research questions or as collateral but constructive opportunities have been properly identified and discussed previously in Section 1.3.

Some of the procedures applied can be greatly enhanced through the identified future work activities. Some methods can be enhanced, others can be widely explored, as new grounds were set in this work.

The resulting methodology framework that was followed might also be improved, especially if one considers the inclusion of advanced algorithmic analysis using data science methodologies. Unfortunately, there was no access to proper datasets with the needed volume of data in order to further explore those aspects in this work. Nevertheless, that was not the main focus of the work.

The authors believe that the achieved results are consistent, the proposed methods are adequate to mobile networks beyond 5G NR and also retro compatible with the 5G. Additionally, results show that the research questions are very relevant and timely, as well as the research artifacts that were produced support the quality of the whole work, not only through the peer reviewing process, but also because some of those artifacts are proposed for the first time and disruptive against the norm. Nevertheless, the aggregation of the work

becomes clear if one thinks in the following order: future networks beyond 5G NR will have increased densification on the RAN level, with highly heterogeneous networks and a number of subscribers using services and applications which requires a large capacity and generates unprecedent amounts of traffic unprecedently. In order to cope with such challenges, optimization should occur on traffic steering and offloading levels, especially at cell edges and for highly mobile users. Machine learning algorithms will help to perform steering decision and data analytics supporting it, over cloud virtualized functions, which are directly accessible through EC. Nevertheless, such networks will represent disruptive paradigms when compared with former generations, meaning that the whole lifecycle must be re-evaluated.

Subscriber behavior analytics will play an important role as subscriber behavior tends to be more selfish, self-centered and user centric. This means that centricity will shift from the CN – concentrated and monolithic – to the user itself, supported by a disaggregated and distributed network function. As so, having the ability to characterize subscribers' segments based on their behavior, the estimated impact that they will have on traffic generation and capacity requirements, will shape new network planning processes, operations, resource allocation and overall network deployment processes. Also, at this point, machine learning (supervised or not) and artificial intelligent will have a primordial role as game changers, towards the concept of self-optimizing intelligent networks. Most important, by knowing the subscribers' behavior, predictive analytics can leverage churn and ARPU minimization and maximization, respectively, as well as evaluate the impact of ceasing or introducing new features over MNOs' subscribers base.

Finally, as this is a very important matter, greenhouse gas emissions will tend to rise, despite the existence of more advanced energy efficient methods. The increase of subscribers, data flows, radio heads, and other components might cast energy efficiency method into the shadows, reducing its overall performance. To address that aspect, carbon sequestration methods will be of utmost importance, and must be performed without additional technology. As so, biotic sequestration as proposed, was shown to be a good and not expensive method to balance the overall carbon footprint, in order to achieve carbon neutrality.

All of these aspects were covered in the current thesis, and are tightly coupled, in order

to enhance and optimize future networks beyond 5G NR, still maintaining compatibility with 5G NR. Most importantly, the developed methods were researched and evaluated in order to be improved in the future, meaning that the majority are future proof and can be rapidly adopted in the forthcoming generations.

3.2. Final Remarks

The developed work showed that by starting and focusing in the network itself, using 5G NR as a basis but aiming to post 5G NR future networks, advanced functions like EC, SDN, NFV and also machine learning and artificial intelligence are the way forward. With such methodologies enabled by the proposed methods in this thesis, it will be possible to enhance even further mechanisms like inter-RAT steering and traffic offloading. It was shown that dense and heterogeneous networks are the best solution, not only to cope with the massive capacity increase that is expected, but also from a cost reduction perspective and also carbon footprint reduction. In this special aspect, which had not been thoroughly addressed yet in what concerns non-technological carbon sequestration and offset, the proposed methodology can be put into practice nowadays already, enabling total carbon offsetting. Further developments over the proposed methods will enable future mobile networks to have greener operation and, with advanced energy efficiency methods in place, extend the carbon offset condition for longer periods of time, thus making MNOs carbon neutral for longer.

Another aspect that was focused in this thesis was EC and network function virtualization, towards the extensive usage of cloud environments. The proposed methods, once again, were developed with 5G NR as a working baseline, but aiming for future networks. In that aspect, the expected massive communications and data transfers will make current centralized cloud architectures insufficient. As so, NFV was studied, with an example of virtualizing the ANDSF function, showing that for future networks only distributed cloud environments will be able to sustain the challenges of networks beyond 5G NR. Advanced steering processes were proposed already with that aspect into consideration, and the results showed the increased performance and reduced cost, but mainly, showed that it is possible to prepare that path and be retro-compatible with 5G NR simultaneously, improving and enhancing 5G networks.

Finally, it was shown that, as never before, it is important for such advanced networks to take into consideration the behaviour of subscribers. Each subscriber is becoming a traffic generator, instead of solely being a traffic consumer. This paradigm change is of utmost importance, as, disrupting with former approaches, uplink capacity begins to be as important as downlink capacity or even more, especially on younger segments of subscribers, as it was shown by the results of the developed work. Youngsters are avid data uploaders, sometimes in real-time and this should be accounted for, from network planning, capacity, operation and evolution perspectives. Such kind of analysis can only be enabled by a concept that has been introduced a couple of years ago which is known as the know your customer (KYC). The work proposed two techniques to leverage such analysis: first an impact model that enables evaluating the impact that a certain subscriber segment has on the traffic capacity of a cellular network. A second work extended the first one and showed that by using advanced clustering, a data science technique, additional information could be extracted from existing one, showing that in fact, there might be user segments that might be being overlooked and to which classical MNOs' processes do not adhere to. It was, therefore, shown that user behaviour is of utmost importance and will constitute one of the main parameters for the success of any future MNO beyond 5G NR, while it might also be applied to existing cellular generations.

The current thesis approach was based on a multi-disciplinary problem approach, resulting in an aggregated whole, with very positive feedback from the whole scientific community, both in journals as well as in conferences.

To notice that two articles were successfully published in two journals and two others are under revision in two other journals by the time this document has been submitted. Seven additional articles were published and presented in international conferences. We believe that aspect, alongside with the different conference articles that were produced are a reliable indicator of the quality of the work produced, evidencing our discoveries. In particular, we would like to point out the carbon neutrality solutions, as well as the subscriber behaviour modelling techniques, further enhanced with advanced clustering. Additionally, as it is discussed in the following section, the work was done with the perspective of further enhancing its different components through future work, which the authors aim to address rapidly. We feel that we should stress that our approaches were

properly supported by results. Despite the problem formulations and solutions of this work being developed with post-5G NR future networks in mind, all of the proposed techniques can be applied to existing real-world 4G or 5G NR deployments and, therefore, have the potential to benefit the MNO's business models and subscribers today.

3.3. Future Work

During the course of developing the current work, several additional optimization ideas appeared. Some of which diverged from the core of the current work. Nevertheless, the focus was on the contributions that were made and the main aspects that this work focused on. The fact that some of the aspects were introduced in this work means that future research and further exploration can be achieved. Although the future work has been partially discussed in the produced research articles, some structuring and improved clarity can be discussed in this section.

Following the different aspects that were focused on in this work, the aim is to further study them, especially in networks beyond 5G NR which was the baseline considered in this work. Focusing on network architecture and 5G NR and beyond networks' enhancements, future work can focus on deepening the concept of EC. According to ETSI, mobile EC provides an IT service environment and cloud computing capabilities at the edge of the mobile network, within the RAN, in close proximity to mobile subscribers. IEEE defines EC as a concept that will place applications, data and processing at logical extremes – edges – of a network, instead of centralizing them. In this work, EC was considered as part of the proposed contributions/novelties of some of the research articles. Future work will focus on the advantages of this kind of technology and further integrate it according to resource management and greener operations.

This was precisely the concept that was followed in this work and needs to be further developed, especially how this concept and the usage of NFV and cloud environments can be further deployed, in order to enable applications that require intensive computation resources and low latency to become even more optimized. As referred, and briefly addressed in this work, in this particular context, adding mechanisms like supervised or non-supervised learning and artificial intelligence the strategy is looking forward: traditional cloud architectures will not suffice and mobile EC with distributed edged cloud

access will be extremely required.

Such will also apply on further work regarding the steering and HOs processes between RATs. This subject has not seen much attention like in the past, and we truly believe that the way forward is through the usage of novel advanced analytics techniques and real time processing of data. The aim in this specific question is to understand, in the context of future networks beyond 5G NR, how can artificial intelligence and machine learning enable more optimized analysis, decisions and traffic steering of offloading processes and strategies. Further work can focus on applying such mechanisms as well as data mining to uncover eventual variables that might relevantly contribute to steering decisions in massively dense and heterogeneous networks, for which less is known today. Such future work will require very specific datasets in order to develop advanced algorithms. We feel that with the technological integration and also heterogeneity of 5G NR edge components, it will be easier than what it is today to find the proper dataset.

In a nutshell, regarding the whole architectural part, further work will focus on proposing enhanced architectures and a framework that can contribute to prioritize resources, planning, investments and costs, while overall uncovering additional variables to optimize networks and further understanding subscribers' behaviour to improve overall quality of experience.

Now focusing exactly on the subscriber's behaviour, we have shown in this work that there is still much to be developed. In this case, just by applying an advanced clustering, results showed that there were significant advantages in applying such techniques. Future work can focus, for sure, on gathering several datasets, with actual data and high volume of observations, such future work could enable evaluating how other mechanisms like artificial intelligence can be used and applied to the knowledge of subscribers' behaviour, translating such knowledge into optimized business models and, tightly coupled with network function which are progressively software-defined and virtualized, allow for auto-tune of the whole network in order to support subscriber-centric perspectives, optimizing the network as a whole, but also the overall quality of experience of subscribers. Also, the aim of this future work is to generalize even more the underlying concepts and conclusions, paving the way to network intelligence driven decisions.

Similarly important but increasingly decisional, carbon footprint offset will play a major

role not only in future beyond 5G networks, but also on existing ones. This is one of the subjects which was focused and where there is large room for additional research, either on the carbon offset process itself but also on additional energy efficiency techniques. The way the current work was structured, not only had retro compatibility as a concern, but also future applicability of the developed work. Such methodology resulted in this work to become a baseline for future developments in some of the focused subjects. Especially, we foresee that in subject areas like energy efficiency combined with CF reduction, behaviour analytics and resource management, by applying techniques like advanced analytics and visualization, machine learning and artificial intelligence, such would for sure enable a more optimized, dynamic and intelligent network, tightly coupled with NFV as well as EC as referred before. As so, future work on this subject will for sure be developed around artificial intelligence and machine learning.

As an example, the authors are finalizing a book chapter focused on those aspects, intitled *"Applications of Artificial Intelligence for 5G Advanced Resource Management"*, with the aim of identifying areas where AI and ML can be used to ever further optimize and enhance post 5G NR networks of the future.

Page intentionally left blank

References

[1] J. Song, T. Yoo and P. J. Song, "Mobility level management for 5G network," 2016 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2016, pp. 940-943.

[2] G. A. Akpakwu, B. J. Silva, G. P. Hancke and A. M. Abu-Mahfouz, "A Survey on 5G Networks for the Internet of Things: Communication Technologies and Challenges in *IEEE Access*, vol. 6, pp. 3619-3647, 2018.

[3] J. Wang *et al*., "Spectral Efficiency Improvement With 5G Technologies: Results From Field Tests," in *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 8, pp. 1867-1875, Aug. 2017.

[4] Anwer Al-Dulaimi; Xianbin Wang; Chih-Lin I, "Machine-Type Communication in the 5G Era: Massive and Ultrareliable Connectivity Forces of Evolution, Revolution, and Complementarity," in *5G Networks: Fundamental Requirements, Enabling Technologies, and Operations Management*, , IEEE, 2018, pp.519-542.

[5] A. Ghosh, A. Maeder, M. Baker and D. Chandramouli, "5G Evolution: A View on 5G Cellular Technology Beyond 3GPP Release 15," in *IEEE Access*, vol. 7, pp. 127639-127651, 2019.

[6] 3GPP TR 21.915, “Summary of Rel-15 Work Items” V15.0.0, Out. 2019.

[7] 3GPP TR 21.916, “Summary of Rel-16 Work Items” V16.0.0, Sep. 2019.

[8] A. Khlass, D. Laselva and R. Jarvela, "On the Flexible and Performance-Enhanced Radio Resource Control for 5G NR Networks," *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, Honolulu, HI, USA, 2019, pp. 1-6.

[9] D. Sattar and A. Matrawy, "Optimal Slice Allocation in 5G Core Networks," in *IEEE Networking Letters*, vol. 1, no. 2, pp. 48-51, June 2019.

[10] W. Lee, T. Na and J. Kim, "How to Create a Network Slice? - A 5G Core Network Perspective," *2019 21st International Conference on Advanced Communication Technology (ICACT)*, PyeongChang Kwangwoon_Do, Korea (South), 2019, pp. 616-619.

[11] C. T. Cicek, H. Gultekin, B. Tavli and H. Yanikomeroglu, "UAV Base Station Location Optimization for Next Generation Wireless Networks: Overview and Future Research Directions," *2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS)*, Muscat, Oman, 2019, pp. 1-6.

[12] B. Li, Z. Fei and Y. Zhang, "UAV Communications for 5G and Beyond: Recent Advances and Future Trends," in *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2241-2263, April 2019.

[13] A. M. Alwakeel, A. K. Alnaim and E. B. Fernandez, "Toward a Reference Architecture for NFV," *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, Riyadh, Saudi Arabia, 2019, pp. 1-6.

[14] C. D. Martino, A. Walid and M. Thottan, "A Cloud-Based Platform Enabling Automation in Resiliency and Performance Testing of SDN," *2018 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, Verona, Italy, 2018, pp. 1-2.

[15] G. Baldoni *et al*., "Edge Computing Enhancements in an NFV-based Ecosystem for 5G Neutral Hosts," *2018 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, Verona, Italy, 2018, pp. 1-5.

[16] M. Gharbaoui *et al.*, "Demonstration of Latency-Aware and Self-Adaptive Service Chaining in 5G/SDN/NFV infrastructures," *2018 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, Verona, Italy, 2018, pp. 1-2.

[17] L. Ma, X. Wen, L. Wang, Z. Lu and R. Knopp, "An SDN/NFV based framework for management and deployment of service based 5G core network," in *China Communications*, vol. 15, no. 10, pp. 86-98, Oct. 2018.

[18] J. Qian, S. P. Gochhayat and L. K. Hansen, "Distributed Active Learning Strategies on Edge Computing," *2019 6th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/ 2019 5th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*, Paris, France, 2019, pp. 221-226.

[19] A. Nanjundappa, S. Singh and G. Jain, "Enhanced multi-RAT support for 5G," 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, 2018, pp. 1-2.

[20] G. Naik, B. Choudhury and J. Park, "IEEE 802.11bd & 5G NR V2X: Evolution of Radio Access Technologies for V2X Communications," in IEEE Access, vol. 7, pp. 70169-70184, 2019.

[21] N. Makris, P. Karamichailidis, C. Zarafetas and T. Korakis, "Spectrum Coordination for Disaggregated Ultra Dense Heterogeneous 5G Networks," 2019 European Conference on Networks and Communications (EuCNC), Valencia, Spain, 2019, pp. 512-517.

[22] J. Zhu, C. Gong, S. Zhang, M. Zhao and W. Zhou, "Foundation study on wireless big data: Concept, mining, learning and practices," in China Communications, vol. 15, no. 12, pp. 1-15, Dec. 2018.

[23] V. P. Kafle, Y. Fukushima, P. Martinez-Julia and T. Miyazawa, "Consideration On Automation of 5G Network Slicing with Machine Learning," *2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K)*, Santa Fe, 2018, pp. 1-8.

[24] M. Jansevskis and K. Osis, "Machine Learning and on 5G Based Technologies Create New Opportunities to Gain Knowledge," *2018 2nd European Conference on Electrical Engineering*

and Computer Science (EECS), Bern, Switzerland, 2018, pp. 376-381.

[25] S. Lagen and L. Giupponi, "Listen before receive for coexistence in unlicensed mmWave bands," 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, 2018, pp. 1-6.

[26] V. Ramaswamy, J. T. Correia and D. Swain-Walsh, "Modeling and Analysis of Multi-RAT Dual Connectivity Operations in 5G Networks," 2019 IEEE 2nd 5G World Forum (5GWF), Dresden, Germany, 2019, pp. 484-489.

[27] Q. Chen, X. Xu and H. Jiang, "Spatial Multiplexing Based NR-U and WiFi Coexistence in Unlicensed Spectrum," *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, Honolulu, HI, USA, 2019, pp. 1-5.

[28] P. B. Oni and S. D. Blostein, "Optimal Node Density for Multi-RAT Coexistence in Unlicensed Spectrum," *2019 16th Canadian Workshop on Information Theory (CWIT)*, Hamilton, ON, Canada, 2019, pp. 1-6.

[29] Z. Machrouh and A. Najid, "High Efficiency WLANs IEEE 802.11ax Performance Evaluation," *2018 International Conference on Control, Automation and Diagnosis (ICCAD)*, Marrakech, Morocco, 2018, pp. 1-5.

[30] D. Lopez-Perez, A. Garcia-Rodriguez, L. Galati-Giordano, M. Kasslin and K. Doppler, "IEEE 802.11be Extremely High Throughput: The Next Generation of Wi-Fi Technology Beyond 802.11ax," in *IEEE Communications Magazine*, vol. 57, no. 9, pp. 113-119, September 2019.

[31] E. Khorov, A. Kiryanov, A. Lyakhov and G. Bianchi, "A Tutorial on IEEE 802.11ax High Efficiency WLANs," in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 197-216, Firstquarter 2019.

[32] A. Malhotra, M. Maity and A. Dutta, "How much can we reuse? An empirical analysis of the performance benefits achieved by spatial-reuse of IEEE 802.11ax," *2019 11th International Conference on Communication Systems & Networks (COMSNETS)*, Bengaluru, India, 2019, pp. 432-435.

[33] D. Kwon and J. Kim, "Opportunistic Medium Access for Hyper-Dense Beamformed IEEE 802.11ax Wireless Networks," *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju, 2018, pp. 198-200.

[34] T. Nakahira, T. Murakami, H. Abeysekera, K. Ishihara, T. Hayashi and H. Nakamura, "Adaptive Multi Radio Interface Control Based on 802.11 AX WLANs," *2018 Asia-Pacific Microwave Conference (APMC)*, Kyoto, 2018, pp. 153-155.

[35] E. Skondras, A. Michalas, N. Tsolis and D. D. Vergados, "A VHO scheme for supporting healthcare services in 5G vehicular cloud computing systems," *2018 Wireless Telecommunications Symposium (WTS)*, Phoenix, AZ, 2018, pp. 1-6.

[36] P. Ren, X. Qiao, J. Chen and S. Dustdar, "Mobile Edge Computing – a Booster for the Practical Provisioning Approach of Web-Based Augmented Reality," *2018 IEEE/ACM Symposium on Edge Computing (SEC)*, Seattle, WA, 2018, pp. 349-350.

[37] N. K. Giang, R. Lea, M. Blackstock and V. C. M. Leung, "Fog at the Edge: Experiences Building an Edge Computing Platform," *2018 IEEE International Conference on Edge Computing (EDGE)*, San Francisco, CA, 2018, pp. 9-16.

[38] X. Wei *et al*., "MVR: An Architecture for Computation Offloading in Mobile Edge Computing," *2017 IEEE International Conference on Edge Computing (EDGE)*, Honolulu, HI, 2017, pp. 232-235.

[39] P. Skarin, W. Tärneberg, K. Årzen and M. Kihl, "Towards Mission-Critical Control at the Edge and Over 5G," *2018 IEEE International Conference on Edge Computing (EDGE)*, San Francisco, CA, 2018, pp. 50-57.

[40] M. Steeg *et al*., "Public Field Trial of a Multi-RAT (60 GHz 5G/ LTE/WiFi) Mobile Network," in *IEEE Wireless Communications*, vol. 25, no. 5, pp. 38-46, October 2018.

[41] F. Liu, G. Tang, Y. Li, Z. Cai, X. Zhang and T. Zhou, "A Survey on Edge Computing Systems and Tools," in *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1537-1562, Aug. 2019.

[42] Y. Li and S. Wang, "An Energy-Aware Edge Server Placement Algorithm in Mobile Edge Computing," *2018 IEEE International Conference on Edge Computing (EDGE)*, San Francisco, CA, 2018, pp. 66-73.

[43] Q. Zhao and M. Gerla, "Energy Efficiency Enhancement in 5G Mobile Wireless Networks," *2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, Washington, DC, USA, 2019, pp. 1-3.

[44] A. Abrol and R. K. Jha, "Power Optimization in 5G Networks: A Step Towards GrEEn Communication," in *IEEE Access*, vol. 4, pp. 1355-1374, 2016.

[45] E. McCune, Q. Diduck, W. Godycki and M. Mohiuddin, "5G New-Radio Transmitter Exceeding 40% Modulated Efficiency," *2018 IEEE 5G World Forum (5GWF)*, Silicon Valley, CA, 2018, pp. 284-288.

[46] F. Han, S. Zhao, L. Zhang and J. Wu, "Survey of Strategies for Switching Off Base Stations in Heterogeneous Networks for Greener 5G Systems," in *IEEE Access*, vol. 4, pp. 4959-4973, 2016.

[47] Y. Cai, Y. Ni, J. Zhang, S. Zhao and H. Zhu, "Energy efficiency and spectrum efficiency in underlay device-to-device communications enabled cellular networks," in *China Communications*, vol. 16, no. 4, pp. 16-34, April 2019.

[48] R. Zhang, Y. Li, C. Wang, Y. Ruan, Y. Fu and H. Zhang, "Energy-Spectral Efficiency Trade-

Off in Underlaying Mobile D2D Communications: An Economic Efficiency Perspective," in *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4288-4301, July 2018.

[49] H. Kour and R. K. Jha, "Power Optimization using Spectrum Sharing for 5G Wireless Networks," *2019 11th International Conference on Communication Systems & Networks (COMSNETS)*, Bengaluru, India, 2019, pp. 395-398.

[50] S. Bhandari, M. Bhandari and S. Joshi, "Spectrum Sharing in Cognitive Radio Networks for 5G Vision," *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, Gangtok, India, 2019, pp. 1-6.

[51] A. N. Al-Quzweeni, A. Q. Lawey, T. E. H. Elgorashi and J. M. H. Elmirghani, "Optimized Energy Aware 5G Network Function Virtualization," in *IEEE Access*, vol. 7, pp. 44939-44958, 2019.

[52] K. Nakura, T. Suehiro, A. Nagasawa, Y. Hirano, S. Kozaki and K. Ishida, "Network resource management in 5G-RAN optical transport," *2019 24th OptoElectronics and Communications Conference (OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC)*, Fukuoka, Japan, 2019, pp. 1-3.

[53] United Kindgom's Forestry Commission, Woodland Carbon Code, Technical Report, Jul. 2018, Retrieved 20th Dec. 2019 from https://www.woodlandcarboncode.org.uk/about.

[54] Trees For The Future Foundation, Online, Available and last accessed Dec. 2019: http://treesftf.force.com/checkout/TFTF_Donation_Certificate_Cls.

[55] M. M. Mowla, I. Ahmad, D. Habibi and Q. V. Phung, "Energy Efficient Backhauling for 5G Small Cell Networks," in *IEEE Transactions on Sustainable Computing*, vol. 4, no. 3, pp. 279-292, 1 July-Sept. 2019.

[56] A. V. Kozinets, G. Y. Alexeevich and V. Y. Alexeevich, "Mobile networks subscribers search time analysis using roaming," *2018 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, 2018, pp. 1-4.

[57] A. Xiong, Y. You and L. Long, "L-RBF: A Customer Churn Prediction Model Based on Lasso + RBF," *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, Atlanta, GA, USA, 2019, pp. 621-626.

[58] A. V. Kozinets, G. Y. Alexeevich and V. Y. Alexeevich, "Mobile networks subscribers search time analysis using roaming," *2018 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, 2018, pp. 1-4.

[59] Hong, J., Ph. D. students must break away from undergraduate mentality. Communications of the ACM, July 2013, Vol. 56 No. 7; pp. 10-11.

[60] Halpern, J. Y., & Parkes, D. C., Journals for certification, conferences for rapid dissemination, Communications of the ACM, Vol.54 No.8, pp. 36-38.