Mikhail Gerasimenko

**Intelligent Resource Allocation in 5G Multi-Radio Heterogeneous Networks**

Tampere 2018
Mikhail Gerasimenko

**Intelligent Resource Allocation in 5G Multi-Radio Heterogeneous Networks**

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB109, at Tampere University of Technology, on the 27th of April 2018, at 12 noon.
Doctoral candidate: Mikhail Gerasimenko
Faculty of Computing and Electrical Engineering
Laboratory of Electronics and Communications Engineering
Tampere University of Technology
Tampere, Finland

Supervisor: Evgeny Kucheryavy, Ph.D., Professor
Faculty of Computing and Electrical Engineering
Laboratory of Electronics and Communications Engineering
Tampere University of Technology
Tampere, Finland

Instructor: Dmitri Moltchanov, Ph.D., Senior Research Fellow
Faculty of Computing and Electrical Engineering
Laboratory of Electronics and Communications Engineering
Tampere University of Technology
Tampere, Finland

Pre-examiners: Tapani Ristaniemi, Ph.D., Professor
Department of Mathematical Information Technology
University of Jyväskylä
Jyväskylä, Finland

Veselin Rakocevic, Ph.D., Senior Lecturer
Department of Electrical and Electronic Engineering
School of Mathematics, Computer Science and Engineering
City University London
London, United Kingdom

Opponent: Periklis Chatzimisios, Ph.D., Associate Professor
Department of Informatics
Alexander Technological Educational Institute of
Thessaloniki
Thessaloniki, Greece
Abstract

The fast-moving evolution of wireless networks, which started less than three decades ago, has resulted in worldwide connectivity and influenced the development of a global market in all related areas. However, in recent years, the growing user traffic demands have led to the saturation of licensed and unlicensed frequency bands regarding capacity and load-over-time. On the physical layer the used spectrum efficiency is already close to Shannon’s limit; however the traffic demand continues to grow, forcing mobile network operators and equipment manufacturers to evaluate more effective strategies of the wireless medium access.

One of these strategies, called cell densification, implies there are a growing number of serving entities, with the appropriate reduction of the per-cell coverage area. However, if implemented blindly, this approach will lead to a significant growth in the average interference level and overhead control signaling, which are both required to allow sufficient user mobility. Furthermore, the interference is also affected by the increasing variety of radio access technologies (RATs) and applications, often deployed without the necessary level of cooperation with technologies that are already in place.

To overcome these problems today’s telecommunication standardization groups are trying to collaborate. That is why the recent agenda of the fifth generation wireless networks (5G) includes not only the development schedules for the particular technologies but also implies there should be an expansion of the appropriate interconnection techniques. In this thesis, we describe and evaluate the concept of heterogeneous networks (HetNets), which involve the cooperation between several RATs.

In the introductory part, we discuss the set of the problems, related to HetNets, and review the HetNet development process. Moreover, we show the evolution of existing and potential segments of the multi-RAT 5G network, together with the most promising applications, which could be used in future HetNets.

Further, in the thesis, we describe the set of key representative scenarios, including three-tier WiFi-LTE multi-RAT deployment, MTC-enabled LTE, and the mmWave-based network. For each of these scenarios, we define a set of unsolved issues and appropriate solutions. For the WiFi-LTE multi-RAT scenario, we develop the framework, enabling intelligent and flexible resource allocation between the involved RATs. For MTC-enabled LTE, we study the effect of massive MTC deployments on the performance of LTE random access procedure and propose some basic methods to improve its efficiency. Finally, for the mmWave scenario, we study the effects of connectivity strategies, human body blockage and antenna array configuration on the overall network performance. Next, we develop a set of validated analytical and simulation-based techniques which allow us to evaluate the performance of proposed solutions. At the end of the introductory part a set of HetNet-related demo activities is demonstrated.
Preface

The research work summarized in this thesis has been carried out at the Department of Electronics and Communications Engineering of Tampere University of Technology (Finland) over the years 2013-2018. This work would not have been possible without the support and contribution of many people, including friends, colleagues, and reviewers who are gratefully acknowledged here.

First of all, I want to express my most profound appreciation to Prof. Evgeny Kucheryavy, who supervised me throughout my research study. Also, I would like to emphasize the role of Dr. Dmitri Moltchanov - as an instructor, he improved my research skills and significantly expanded my knowledge. Also, I am indebted to Dr. Sergey Andreev, whose guidance and experience initialized my interest in research. Special gratitude is dedicated to Dr. Nageen Himayat, from Intel Labs. As an industrial adviser, she played an essential role in practical research supervision and guidance. In addition to this, I would like to separately thank Dr. Alex Pyattaev, Dr. Olga Galinina, Vitaly Petrov, Margarita Gapeyenko, Jani Urama and Anastasiia Voropaeva for their distinct contributions to this research. Finally, I would like to express my sincere gratitude to my friends, family and colleagues for their everlasting encouragement and support.

As financial stability is an important aspect of any solid research, I gratefully acknowledge the generous support received from Tampere University Graduate School. Furthermore, I would like to appreciate Intel Labs and Ericsson Research, together with TAKE-5 and IoT SRA projects for general funding and support of the research projects. The financial contribution by HPY Research Foundation in the form of personal grant is also much appreciated.

Mikhail Gerasimenko. February 14, 2018, Tampere, Finland
Contents

Abstract iii

Preface v

List of Abbreviations ix

List of Publications xi

List of Figures xii

1 Introduction 1
  1.1 Research Motivation .............................................. 1
  1.2 General Background on HetNets .................................. 2
  1.3 Main Contributions and Scope .................................. 3
  1.4 Structure of the Thesis ......................................... 3

2 HetNet Components and Architectural Options 5
  2.1 HetNet RAT Candidates ............................................ 5
     2.1.1 WMANs and WLANs Evolution .................................. 5
     2.1.2 MmWave RATs ..................................................... 6
  2.2 HetNet Architecture Design ...................................... 7
     2.2.1 Network Entities and Their Roles ............................. 7
     2.2.2 Standardization Efforts ...................................... 9
     2.2.3 Other Possible HetNet Enhancements ......................... 9
  2.3 HetNet Applications .............................................. 10
     2.3.1 H2H Traffic Offloading ..................................... 10
     2.3.2 MTC as Part of HetNets ...................................... 11
     2.3.3 Drone Users and Mobile Drone Cells ......................... 12

3 Instrumentation and Methodology 15
  3.1 Scenarios of Interest ............................................ 15
     3.1.1 Multi-RAT Network ............................................ 15
     3.1.2 LTE-based MTC ............................................... 16
     3.1.3 MmWave Network .............................................. 18
  3.2 Simulation-based Environment ................................... 19
     3.2.1 Simulation Environment Comparison ......................... 19
     3.2.2 System Level Simulations ................................... 21
     3.2.3 Protocol level Simulations for LTE-based MTC ............ 22
  3.3 Analytical Framework ........................................... 23
### Contents

3.3.1 Resource Allocation in Three-Tier multi-RAT Network . . . . . . . 23  
3.3.2 MTC-enabled LTE . . . . . . . . . . . . . . . . . . . . . . . . . . . 26  
3.3.3 MmWave Network . . . . . . . . . . . . . . . . . . . . . . . . . . . 27  

4 Performance Evaluation 31  
4.1 Selected Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31  
4.1.1 Multi-RAT Network . . . . . . . . . . . . . . . . . . . . . . . . . . 31  
4.1.2 LTE-based MTC . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33  
4.1.3 MmWave Network . . . . . . . . . . . . . . . . . . . . . . . . . . . 35  
4.2 Trial Activities . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38  
4.2.1 Single UE Tests . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38  
4.2.2 Test Network Infrastructure . . . . . . . . . . . . . . . . . . . . . . 39  
4.2.3 Multi-purpose Automated Vehicular Platform . . . . . . . . . . . . 40  

5 Conclusion and Potential Enhancements 43  
5.1 HetNets Applicability Discussion . . . . . . . . . . . . . . . . . . . . . . 43  
5.2 HetNets as Part of 5G . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44  

6 Summary of Publications 47  
6.1 Publications Description . . . . . . . . . . . . . . . . . . . . . . . . . . . 47  
6.2 Author’s Contribution . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50  

Bibliography 51
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3GPP</td>
<td>Third Generation(3G) Partnership Project</td>
</tr>
<tr>
<td>5G</td>
<td>Fifth Generation</td>
</tr>
<tr>
<td>AP</td>
<td>Access Point</td>
</tr>
<tr>
<td>BBU</td>
<td>Baseband Unit</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>CA</td>
<td>Carrier Aggregation</td>
</tr>
<tr>
<td>CCE</td>
<td>Control Channel Elements</td>
</tr>
<tr>
<td>DC</td>
<td>Dual Connectivity</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communications</td>
</tr>
<tr>
<td>E-UTRAN</td>
<td>Evolved Universal Terrestrial Radio Access Network</td>
</tr>
<tr>
<td>EHF</td>
<td>Extreme High Frequency</td>
</tr>
<tr>
<td>eMBB</td>
<td>enhanced Mobile Broadband</td>
</tr>
<tr>
<td>EPC</td>
<td>Evolved Packet Core</td>
</tr>
<tr>
<td>FDD</td>
<td>Frequency Division Duplex</td>
</tr>
<tr>
<td>H2H</td>
<td>Human to Human</td>
</tr>
<tr>
<td>HARQ</td>
<td>Hybrid Automatic Repeat Request</td>
</tr>
<tr>
<td>HetNet</td>
<td>Heterogeneous Network</td>
</tr>
<tr>
<td>H-CRAN</td>
<td>Heterogeneous Cloud Radio Access Network</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>LAA</td>
<td>Licensed Assisted Access</td>
</tr>
<tr>
<td>LLS</td>
<td>Link Level Simulation</td>
</tr>
<tr>
<td>LoS</td>
<td>Line of Sight</td>
</tr>
<tr>
<td>LPN</td>
<td>Low Power Node</td>
</tr>
<tr>
<td>LPWA</td>
<td>Low-Power, Wide-Area Wireless Technology</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term Evolution</td>
</tr>
<tr>
<td>LTE-A</td>
<td>LTE-Advanced</td>
</tr>
<tr>
<td>LTE-U</td>
<td>LTE-Unlicensed</td>
</tr>
<tr>
<td>LWA</td>
<td>LTE-WiFi Aggregation</td>
</tr>
<tr>
<td>LWIP</td>
<td>LTE WLAN Integration with Internet Protocol Security Tunnel</td>
</tr>
<tr>
<td>mmWave</td>
<td>Millimeter Wave</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
</tr>
<tr>
<td>MCS</td>
<td>Modulation and Coding Scheme</td>
</tr>
<tr>
<td>MTC</td>
<td>Machine Type Communications</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple-Input and Multiple-Output</td>
</tr>
<tr>
<td>NB-IoT</td>
<td>Narrowband IoT</td>
</tr>
<tr>
<td>NR</td>
<td>New Radio</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>RA</td>
<td>Random Access</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>RAN</td>
<td>Radio Access Network</td>
</tr>
<tr>
<td>RAR</td>
<td>Random Access Response</td>
</tr>
<tr>
<td>RAT</td>
<td>Radio Access Technology</td>
</tr>
<tr>
<td>RRH</td>
<td>Remote Radio Head</td>
</tr>
<tr>
<td>RRM</td>
<td>Radio Resource Management</td>
</tr>
<tr>
<td>P2P</td>
<td>Point to Point</td>
</tr>
<tr>
<td>PDCCH</td>
<td>Physical Downlink Control Channel</td>
</tr>
<tr>
<td>PDCP</td>
<td>Packet Data Convergence Protocol</td>
</tr>
<tr>
<td>PLS</td>
<td>Protocol Level Simulation</td>
</tr>
<tr>
<td>PRACH</td>
<td>Physical Random Access Channel</td>
</tr>
<tr>
<td>PUCCH</td>
<td>Physical Uplink Control Channel</td>
</tr>
<tr>
<td>PUSCH</td>
<td>Physical Uplink Shared Channel</td>
</tr>
<tr>
<td>SHF</td>
<td>Super High Frequency</td>
</tr>
<tr>
<td>SLS</td>
<td>System Level Simulation</td>
</tr>
<tr>
<td>TUT</td>
<td>Tampere University of Technology</td>
</tr>
<tr>
<td>TDD</td>
<td>Time Division Duplex</td>
</tr>
<tr>
<td>UAV</td>
<td>Unmanned Aerial Vehicles</td>
</tr>
<tr>
<td>UE</td>
<td>User Equipment</td>
</tr>
<tr>
<td>URLLC</td>
<td>Ultra-Reliable and Low-Latency Communication</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless Local Area Network</td>
</tr>
<tr>
<td>WMAN</td>
<td>Wireless Metropolitan Area Network</td>
</tr>
</tbody>
</table>
List of Publications


# List of Figures

2.1 3GPP 4G network architecture ................................................. 6
2.2 Generalized HetNet structure and entities ...................................... 7
2.3 E-UTRAN-EPC functional split, [1] ........................................... 8
2.4 LWA architecture for collocated (left) and non-collocated (right) cases, [1] 9
2.5 LWIP architecture, [1] .................................................................. 10
2.6 Options for HetNet architecture for an LTE-based network [P2] .......... 11
2.7 IoT applications and requirements [2] ........................................... 12
2.8 LPWA classification, [3] .................................................................. 12

3.1 Scenario 1 layout [4] ..................................................................... 15
3.2 LTE RA procedure details [P5] ...................................................... 17
3.3 mmWave network deployment details [P7] ........................................ 18
3.4 SLS and analytical framework comparison, [4] ................................. 21
3.5 RA procedure and PLS structure [P5] ............................................. 23
3.6 Analytical representation of considered HetNet architecture [P1] ....... 24
3.7 Two-state Markov chain describing the number of non-served user requests [P5] 26
3.8 Connectivity strategies in considered mmWave networks ................... 27

4.1 Analytical model calibration for Multi-RAT scenario [P1] ..................... 31
4.2 Resource allocation strategies performance comparison [P1] ............... 32
4.3 Throughput and fairness comparison for different resource allocation strategies [P1] ............................................................................ 33
4.4 Weighted alpha-fairness control regions [P1] ................................... 34
4.5 PLS calibration [P5] ........................................................................ 34
4.6 Power consumption comparison for different number of users [P5] ........ 35
4.7 RA success probability and power consumption with different BI values [P5] 36
4.8 mmWave system performance with different AP densities [P7] .......... 36
4.9 Interference variations as a function of AP density (left), blocker density (center) and number of AP array elements (right) [P7] ............... 37
4.10 Average throughput when associated with $i^{th}$ closest AP as a function of AP density (left), blocker density (center) and number of AP array elements (right) [P7] ................................................. 37
4.11 Ergodic capacity of different connectivity schemes as a function of AP density (left), blocker density (center) and number of AP array elements (right) [P7] ................................................. 38
4.12 Single UE test prototype implementation architecture [P3] .................. 39
4.13 TUT test HetNet ............................................................................ 40
4.14 Multi-purpose Automated Vehicular Platform, v.1 [5] (left) and v.2 (right) 40

5.1 Options for multi-RAT enabled implementation for LTE-based network [P1] 44
1 Introduction

1.1 Research Motivation

Despite the fact that telecommunications have been in development for more than a century, wireless communication is a relatively modern area of research. Although radio and television broadcast technology started in the first half of the twentieth century, widespread use of bidirectional wireless networks became possible only after the appearance of digital wired telephony and the Internet.

Today, wireless communication is one of the fastest developing technologies in the field of telecommunications. According to a recent report by Cisco [6], the amount of generated data traffic (especially mobile) has grown exponentially. If current forecasts are correct, more than forty percent of all telecommunication traffic will be generated by mobile devices by 2021. Additionally, the deployment of 5G technology is expected to increase the capacity of modern wireless network by 1000x. [7, 8].

The intense growth of wireless system capacity can be explained through three core trends:

- Network densification [9, 10]. The density of deployed base stations (BSs) and access points (APs) correlates with the number of users and their demands. At the same time, average cell size is decreasing, in order to achieve the appropriate capacity boost, which therefore increases the level of interference and control channels load.

- Introduction of advanced spectrum usage techniques and medium access control schemes - these are critical components of flexible load control and interference avoidance.

- Extensive use of millimeter waves (mmWaves)-based radio access. Recent developments make the use of both a super high frequency (SHF) and an extreme high frequency (EHF) spectrum possible across commercial telecommunication networks.

Taking into account growing concurrency of resource use in the electromagnetic spectrum and the limitations of backbone, a capacity boost like the one mentioned above may sound overly optimistic. As already mentioned, one of the options to achieve required throughput, together with low end-to-end latency [11] could be to utilize mmWave technologies, where the bandwidth of up to several GHz could be reserved. However, physical limitations of the channel [12, 13] could mean that additional coordination with conventional cellular networks is required. Moreover, future wireless networks are expected to support multiple application types that will all have variable capacity, latency and reliability requirements. Therefore, cross-communication/inter-working has become
the key enabler of 5G technology development. With this in mind, the primary focus of this thesis are the heterogeneous networks (HetNets) that are one of the core components of smart spectrum usage and aggregation.

1.2 General Background on HetNets

HetNets is a broad term, used to describe a variety of different systems. Commonly, the term is applied to the usage of two or more separate Radio Access Technologies (RATs) that have some level of inter-working between their networks. In previous phases of telecommunications development this inter-working was necessary to connect different generations of cellular networks. To this end, the coexistence of the second generation (2G) and third generation (3G) wireless networks should also be considered to be HetNet, although they were much easier to jointly implement since they were developed by the same standardization group.

Presently the idea of network densification, as discussed above, is the main reason for HetNets development. While the general trend is a move towards a large-scale deployment of smaller low-power and low-cost cells [14, 15] (i.e., micro/pico/femtocells), this brings more challenges in how to support fast-moving users, avoid growing interference and control channel overheads. Therefore, small cells are often considered as capacity boosters with limited signaling capabilities, while the control functions of the network are offloaded to existing macrocell infrastructure [9, 16]. The envisioned 'anchor-booster' architecture theoretically could be used with a number of different RATs, assuming that the control is always deployed in the "anchor" single-RAT network. These types of HetNets are known as multi-RAT networks.

The first attempts to standardize multi-RAT networks, using anchor-booster architecture, were made by a third generation partnership program (3GPP) combining access network discovery and selection function (ANDSF) [17]. ANDSF was initially designed to supply user equipment (UE) with necessary information about non-3GPP networks. Although this information contained policies, ANDSF should be considered as a more advisory and user-centric option, rather than the complex HetNet solution. Later on, "tighter" techniques which assumed multi-RAT integration across the radio access network (RAN)-levels were developed [18]. It is also essential to take into account a variety of applications, e.g., the growing market of machine-type communications (MTC) [19] that require an entirely different network architecture and advanced upper layer solutions. The issues with MTC are caused by the specific shape of MTC traffic and much higher number of connected devices. Despite these difficulties, it is still possible to adapt existing LTE basis to work with MTC applications, although it requires significant changes in medium access control (MAC) layer procedures and modifications in control signaling. On the other hand, support of multiple applications could lead to a situation in which connected users have unbalanced throughput, latency and reliability requirements. In the scale of HetNet architecture, these problems could be addressed by separating different type users on the radio access level and support the traffic aggregation via the set of gateways, connected to the collective central management entity.

Finally, it should be mentioned that the vision of HetNets is continuously evolving, giving new elaboration options for standardization groups and hardware/software developers. For example, the concept of heterogeneous cloud RAN (H-CRAN), detailed in [2] could be considered as an interesting extension of anchor-booster architecture, where part of the anchor functionality is transferred to the cloud network. This idea is especially
interesting, taking into account the growing variety of recently developed RATs, where their interconnections require complex centralized control. Although these concepts might seem futuristic, they are worth evaluating as potential HetNet architecture evolution paths.

1.3 Main Contributions and Scope

This thesis will focus on the analysis of different HetNet aspects, giving both theoretical and practice-oriented evaluation of the overall architecture and individual components. The main contribution is done in the area of RATs inter-working, shown on the example of interconnection of two networks: IEEE 802.11 WiFi and 3GPP LTE, although it is also possible to scale these solutions to other technologies as well. The algorithms and techniques, related to RAT inter-working, are mostly developed for resource allocation in the scope of above-mentioned anchor-booster architecture. In the final chapter of the thesis there is a discussion of the applicability of the proposed solutions.

In addition to this, there is a separate analysis of the mmWave RAT, as the most probable "booster" candidate for future HetNets. Although this part of the research is currently theoretical, the study will concentrate on the system level scale, giving the relative performance of the mmWave and leading to a greater understanding of its place in future wireless networks.

Consideration is also given to the MTC-enabled LTE network as an example of how different applications can impact the individual RAT implementation and overall network performance. It should be mentioned that MTC-related analysis is limited by the performance evaluation of LTE random access (RA) in the massive MTC (mMTC) scenario.

Finally, together with the theoretical models and computer analysis, the prototype HetNet is tested as an example of how considered architecture could be implemented using commercially available equipment.

1.4 Structure of the Thesis

This thesis consists of an Introduction (five chapters) and a compilation of seven publications [P1]−[P7]. There is reference to several co-authored publications, closely related to the topic of this thesis. Furthermore, MTC-related research considered here is partly used in the authors master thesis [20].

In Chapter 2 of the introductory section, a detailed overview of HetNet architectures and individual RAT components is given, including a discussion of the evolution of the most popular wireless standards and their current processes. Moreover, potential HetNet applications are examined to aid analysis of how the system should evolve to support future applications.

Chapter 3 shows the research methodology used in this thesis. It will outline how system-level simulations are used to model modern wireless networks and create a theoretical basis for HetNets analysis. It also outlines the basic HetNet scenarios and appropriate parameters and assumptions.

The performance evaluation of individual RAT components, as well as the complex HetNet environment, are shown in Chapter 4. The chapter starts with the analysis of results,
related to multi-RAT resource allocation, MTC, and mmWave networks. The end of the chapter overviews the HetNet prototype built in the TUT campus, together with a review of the set of tested applications.

Chapter 5 concludes the introductory section, followed by a compilation of the publications. Chapter 5 also contains the discussion of applicability and connects the thesis to other research related to future wireless network development.
2 HetNet Components and Architectural Options

In this chapter, basic HetNet architectural options are discussed. First, there is a review of the evolution of wireless networks and definition of RAT candidates, which could be used as part of HetNets. This is followed by an examination of the standardization efforts concerning the HetNet concept and its RAT components. At the end of the chapter there is a discussion on novel HetNet applications, as well as other possible technology enhancements.

2.1 HetNet RAT Candidates

2.1.1 WMANs and WLANs Evolution

In 1991 the first Global System for Mobile Communications (GSM) network was introduced in Finland. From then on, wireless cellular networks spread exponentially, reaching 318 million subscribers globally by 1998 [21]. Also in 1998, the third generation of cellular networks was released, introducing fully-digital IP-compatible architecture and higher RAN capacity.

In 1997, the first 802.11 standard, or so-called "legacy 802.11" (the term WiFi was coined in 1999) was released, initiating the use of WLANs in both corporate and private user settings. Further developments in the standard were aimed at increasing achievable data rates through the adoption of new techniques (such as MIMO), using different frequencies and bandwidths [22]. The most recent standard, 802.11ay [23] achieved throughput of several Gbps, by working in the 60GHz frequency band and utilizing up to 8.14 GHz of bandwidth.

With each subsequent release this cellular network technology was driven towards the usage of IP, introducing new IP-based services, improving network throughput, radio access latency and energy efficiency [24]. Finally, in the eighth release of 3GPP (2009) fourth generation (4G) cellular standard, long-term evolution (LTE) was introduced together with some extensions to the 3G standard. The modern architecture of 3GPP LTE is shown in Figure 2.1 Today, 3GPP is developing mmWave-enabled radio access, moving towards 28 and 72 GHz frequencies.

The extremely fast rate of technology uptake has meant that different standardization committees have, somewhat inevitably, entered into competition. An example of this is the 802.16 WiMax standard [25, 26], introduced by IEEE in 2004, which works in the licensed band, targeting WMANs as a core segment. Although the standard was overthrown by LTE, WiMax is still used for some applications because of ease of deployment and its
more simple network architecture. Meanwhile, 3GPP has recently developed an LTE-U standard [27], which works in the unlicensed band together with WiFi and increases the average interference level on those frequencies.

Despite this, both 3GPP and IEEE contributors understand that in future it will be impossible to satisfy growing data demand without efficient spectrum usage, and this will involve collaboration between the committees. One of the options for such cooperation requires the creation of inter-standard protocol, where fundamental interactions between the technologies are defined.

2.1.2 MmWave RATs

The recent developments in mmWave RATs were triggered by the possibility of using much higher bandwidths and therefore significantly increasing network capacity [28]. The first step towards commercial usage of higher frequencies was taken by IEEE, who developed the 802.11 ad standard in 2009 [29]. That technology was capable of delivering throughput of up to 7Gbps with a total bandwidth of 2.16 GHz. The 802.11 ad standard released as a follow-up [30, 31] achieved even higher data rates and allows aggregation of up to four 802.11ad channels, as well as usage of a 256-symbol quadrature amplitude modulation scheme (QAM) and MIMO. It is worth mentioning that the WiGig primary area of usage was limited by static applications, e.g., replacing the last mile of Ethernet connections to homes and offices.

3GPP started its standardization process in Release 14, by introducing new radio (NR) technology which was designed to work on frequencies both below and above 6GHz. While the standardization process is ongoing it has become clear that in order to facilitate mobile communication above a 6GHz frequency, significant enhancements in the PHY and MAC of the current 3GPP architecture are required [32].

In mmWave frequencies, located mostly in SHF and EHF bands, diffraction effects are much weaker than in, for example, ultra high frequencies (UHF). This leads to severe signal strength degradation in non-line-of-sight (NLoS) conditions. In practical terms this means that the effects blockage, i.e., the positioning of even a small object between transmitter and receiver, could lead to discontinuities or a complete halt in transmission. In some cases even weather conditions can significantly influence mmWave network performance.
Because of the lack of diffraction [33], it is more beneficial to make antenna directivity patterns as narrow as possible, which will simultaneously increase the antenna gain and theoretically decrease the average interference level across the network. However, a narrow antenna beam-width puts limitations on the UE mobility or at least the necessary modification to allow steerage of the beam in the direction of user trajectory. While mechanical steering is the cheapest option, current technological advancements mean it is also possible to use digital antenna arrays [34]. Although this solution is more expensive, the steering speed of digital antenna arrays is made higher. This distinction could become crucial in environments with frequently changing channel conditions. Moreover, while mechanical steering can be used to maintain P2P links, it is not capable of switching between multiple UEs with frequencies, sufficient to maintain communication sessions.

### 2.2 HetNet Architecture Design

#### 2.2.1 Network Entities and Their Roles

In anchor-booster HetNet architecture [35], the following core entities (shown also in Figure 2.2) require definition:

**Anchor cells.** An anchor cell entity is a conventional cellular base station (eNodeBs in terms of 4G) installed by the operator. In LTE, macrocells are used as "anchors", providing control channels to users in the area of coverage and a backhaul channel to the booster cells. Widely speaking, any entity with backhaul and control capabilities could be considered to be an 'anchor', however, in practice only 3GPP standards currently allow the appropriate signaling.

**Booster cells.** The booster is a last-mile entity; its role is to provide a data connection to the user, while control plane connectivity is partly or entirely transferred to the anchor. In LTE, picocells are usually considered as boosters; although for indoor deployments picocells are used as stand-alone base stations, with full control plane capabilities. The overall idea behind boosters comes from the network densification strategy which first appeared in 2G as microcell technology. Its deployment significantly improved the system capacity in urban environments [36]. Today boosters are not only limited to micro- and picocells. The concept lends itself to the possibility of indoor ‘femtocell’ installation to private customers.
Central management entity. Traffic collected on anchor and booster cells can be forwarded through a central management server that is capable of controlling multiple radio access entities. In theory, the presence of central management is optional, and it is not implemented in some standards (such as WiMax) where most of its functionality is carried out by macrocells. However, in practical 3GPP HetNets, the evolved packet core (EPC) takes the role of the central management entity, taking care not only of conventional functions, e.g., mobility and policy management, but also more advanced HetNet functionality such as access to non-3GPP networks and load balancing. Starting from Release 12 EPC HetNet functionality has partly moved to E-UTRAN (Figure 2.3) cells, allowing RAN-level aggregation of non-3GPP traffic.

![Figure 2.3: E-UTRAN-EPC functional split, [1]](image)

Multi-RAT entities. As discussed above, different RATs can be connected to the EPC over a set of gateways, allowing the network operator to manage several access technologies through one centralized management entity. This concept significantly increases the flexibility of capacity control and also gives the operator the ability to cover new market segments at the same time. For example, the operator is able to build the WiFi network over the city and allow the subscriber to choose between WiFi and 4G, or provide a seamless handover within the combined multi-RAT network. Moreover, assuming appropriate standardization is made, the operator could easily attach new RATs as an extension to the current infrastructure, avoiding significant changes to the core network. Using this logic, the cellular operator could connect sensor-oriented MTC RAT with the unique channel and signaling structure by merely attaching its anchor node to the gateway and making an appropriate software update in the EPC.

User equipment. Finally, HetNet-enabled UE should be developed to support noted potential architecture enhancements. While in some cases it is enough to make a simple software upgrade, some HetNet functions also require new hardware on the UE side. For example, the mmWave technology already discussed has challenges of implementation due to its compact shape, complexity and high price of particular elements (such as antenna array systems) [37]. Alternatively, inexpensive and straightforward MTC devices could be deployed in the thousands, but demand cheap and energy efficient power
2.2. HetNet Architecture Design

supply solutions [38], in addition to the modifications required in RAN signaling and architecture [39].

2.2.2 Standardization Efforts

Besides the ANDSF protocol found in the Release 12, two different strategies of non-3GPP RATs integration have been proposed: LTE-WiFi aggregation (LWA) and LTE WLAN integration with Internet Protocol security tunnel (LWIP).

In LWA, WLAN integration is performed on the Packet Data Convergence Protocol (PDCP) level (Figure 2.4) with two possible WLAN integration strategies - collocated and non-collocated. In the collocated strategy, WLAN AP is built into the LTE eNodeB and works within the coverage area of eNodeB, providing additional capacity and gateway-free traffic transition. In the case of the non-collocated strategy, WLAN AP can be placed in a distant location whilst still offering a significant boost to coverage, while being connected to the eNodeB through xw backhaul.

![Figure 2.4: LWA architecture for collocated (left) and non-collocated (right) cases, [1]](image)

In LWIP architecture, shown in Figure 2.5, the integration is performed on the IP level and requires a gateway and additional tunneling overlay. Conversely, the implementation of this technique does not make specific demands on the AP (such as xw interface support) and gives the operator a greater degree of freedom in their choice of equipment.

Another significant technological innovation has been LTE-U [40], which was developed based on the 12th release LTE [41], with modifications of LTE-A. It works on unlicensed 5GHz frequencies, and although LTE-U cells use the same bands as 802.11 ac, they are much easier to use as booster entities. LAA technology arrived in Release 13 [42] and implies several transmission modes: including DL-only, full TDD, and TDD-FDD CA, where licensed and unlicensed bands can be used simultaneously [43]. The benefit of the LAA scheme over LWA and LWIP is in its more "natural" gateway-free technological integration into the 3GPP stack - this gives the operator more control over LTE-U entities. However, LAA is currently considered to be an alternative solution for WLAN-enabled schemes, due to its competition with 802.11 standards [44].

2.2.3 Other Possible HetNet Enhancements

It could be said that the current trends in Hetnet development are designed to give more functionality to central management entity "anchors", as shown in Figure 2.6, A and B. Further evolution of HetNet architecture could lead to concatenation of a number of "anchors" into a cloud-like structure (Figure 2.6, C). In this setup, called H-CRAN [45, 46, 47], part of the Macro base station functionality is delegated to baseband units (BBU), while radio access part is implemented in the remote radio head (RRH). Furthermore, part of the EPC functionality is transferred to the H-CRAN to speed up delay-sensitive
functions - such as radio resource management (RRM). To implement this functionality some of the central management entities are virtually created inside the H-CRAN. In our example, to control RRM, a 'cloud-RRM' (CRRM) entity has also been created.

2.3 HetNet Applications

2.3.1 H2H Traffic Offloading

While H2H traffic is considered to be the core application for HetNets, it is unclear how exactly the user would benefit from HetNets if it could manually switch between the separated RAT entities. For example, in early ANDSF simultaneous usage of LTE and WiFi was not implied, meaning the user had to perform an inter-RAT handover each time he decided to switch technologies. This problem also occurs within several versions of modern mobile equipment operation systems, making it more difficult for developers to design and implement multi-RAT applications.

In LWA and LWIP solutions, simultaneous transmission over LTE and WiFi is allowed and is referred to as dual connectivity (DC) [48, 49, 50]. In theory, DC over two or more RATs can give a significant capacity boost since it allows for usage of 5G applications such as: ultra-HD video streaming, virtual reality, augmented reality, etc. However, when used blindly this approach is energy-inefficient and when combined with a growing
computation load could significantly limit UE capabilities. Finally, dual connectivity may cause other minor problems (such as unbalanced radio access latency issue) [51], which may occur in all multi-RAT HetNets implementations and should be addressed separately for each case.

### 2.3.2 MTC as Part of HetNets

MTC significantly differs from traditional H2H applications in terms of applications, equipment and traffic shape [52]. In addition to this, there are several development directions for MTC communications. For stand-alone solutions working in the unlicensed spectrum, there are many narrow-band protocols such as LoRa [53, 54] and Sigfox. In general, these protocols allow the implementation of a reliable, low-latency network with a low-end device price and decent connection range and coverage [3]. However, as it is shown in Figure 2.7 the variety of applications within the Internet of Things (IoT) concept leads to different performance requirements [55]. This issue means its not possible to use one overarching technology to cover all IoT applications simultaneously. Therefore it is necessary to give a brief description of licensed solutions, developed by 3GPP in Releases 11-13 and including LTE-MTC and narrowband IoT (NB-IoT).

LTE-MTC (or LTE-M) solutions define a set of modifications to the LTE standard that, when put together with new UE types (Cat 1, Cat 0, Cat M1), decrease the cost and power consumption of the end device [56, 57, 58]. These modifications, including the half-duplex operation and 1.4 MHz bandwidth cover a variety of MTC applications, providing a decent data rate and scalability whilst still connecting to the conventional LTE network either on a device-by-device basis or through the LTE-MTC-enabled gateway [59]. At the same time, the NB-IoT accent tackles the device cost, allowing massive deployments with ultra-low traffic demands per device [2]. Finally, in comparison with conventional sensor networks such as ZigBee, unlicensed and licensed solutions have several common characteristics. These include low power and high coverage, allowing them to be put in a separate category called low-power wide-area (LPWA) (Figure 2.8) [60]. While LPWA standards should theoretically cover all MTC applications [61], in practice there are still a lot of open issues related to energy efficiency, coverage, cost per device and security.

![Figure 2.6: Options for HetNet architecture for an LTE-based network][P2]
2.3.3 Drone Users and Mobile Drone Cells

Drones are often referred to as Unmanned Aerial Vehicles (UAVs) and are further prospective users of 5G networks that cannot be directly classified as MTC devices. Although commercial UAVs tend to work on R/C and ISM bands, cellular networks can potentially be used as the drone communication provider [62]. In this case, the potential communication range is bound to the coverage of the cellular network, which usually means much broader horizontal distances. However, at the same time, the drone elevation height becomes limited because cellular RATs are primarily designed to provide coverage.
to ground users. Additionally, drones have higher requirements for latency and reliability and are ITU-classified as ultra-reliable and low-latency communication class (URLLC), which means that they should be served separately from the conventional cellular users or at least have a higher priority in RRM-related procedures. There have been notable discussions within research communities into using UAVs as a mobile booster cell [63, 64]. Drones are already used to support essential communication in emergency situations\(^1\). Other scenarios demonstrate multiple drone cell deployment to improve capacity and coverage of conventional cellular networks. This will require solutions to the issues related to drone positioning, energy efficiency and backhaul between UAV-based AP and core network [65].

\(^1\)https://arstechnica.com/information-technology/2017/11/att-drone-brings-lte-access-to-hurricane-damaged-puerto-rico/
3 Instrumentation and Methodology

This chapter will focus on the methodology developed during the research for this thesis. The chapter starts with the consideration of different scenarios that each represent a set of HetNet architectural aspects. The instrumentation will be reviewed in order to evaluate HetNet performance, which will then include a set of computer simulation tools and analytical models.

3.1 Scenarios of Interest

The previous chapter discussed the available options of HetNets architecture as well as possible development directions. This section will describe scenarios studied in the scope of this thesis. In Subsection 3.1.1 a simplified HetNet deployment is considered with two active RATs: LTE and WiFi. In Subsection 3.1.2 there is a description of an LTE-based MTC scenario, which the author investigated in his master thesis [20]. Finally, in Subsection 3.1.3, an example of mmWave-based network deployment will be shown (detailed in [P7]).

It should be noted that the multi-RAT scenario, considered in Subsection 3.1.1 is related to the inter-working of existing technologies in the frame of HetNets, while subsections 3.1.2 and 3.1.3 describe issues related to the individual RAT components.

3.1.1 Multi-RAT Network

The above is a generalized HetNet with two interconnected RATs, further referred to as Scenario one and shown below in Figure 3.1.

![Scenario 1 layout](image)

**Figure 3.1:** Scenario 1 layout [4]
Chapter 3. Instrumentation and Methodology

In contrast with traditional 3GPP nineteen cell scenario [66], Scenario One considers the deployments limited by the coverage of one macrocell. Within this area, we deploy Pico LTE eNodeBs and WiFi APs (both referred as LPNs). By changing the number of deployed LPNs, the cell density of the considered scenario is controlled. Other primary parameters of the system are summarized in Table 3.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTE/WiFi configuration</td>
<td>10 MHz FDD / 20 MHz</td>
</tr>
<tr>
<td>Layout</td>
<td>1 macro cell, several LPNs</td>
</tr>
<tr>
<td>Macro/LPN-UE pathloss model</td>
<td>ITU UMa/UMi [67]</td>
</tr>
<tr>
<td>Macro/LPN antenna gain</td>
<td>17/6 dB</td>
</tr>
<tr>
<td>Macro/pico/WiFi max. power</td>
<td>43/23/20 dBm</td>
</tr>
<tr>
<td>UE max. power</td>
<td>23/20 (LTE/WiFi) dBm</td>
</tr>
<tr>
<td>LTE/WiFi power control</td>
<td>Max power</td>
</tr>
<tr>
<td>UE/macro/LPN antenna height</td>
<td>1.5/25/10 m</td>
</tr>
<tr>
<td>UE noise figure/feeder loss</td>
<td>5 dB / 0 dB</td>
</tr>
<tr>
<td>Traffic model</td>
<td>Full-buffer</td>
</tr>
<tr>
<td>LPN/UE deployment type</td>
<td>Uniform [68]</td>
</tr>
<tr>
<td>LPN/UE-macro distance</td>
<td>&gt; 75/35 m [69]</td>
</tr>
<tr>
<td>LPN/UE-UE distance</td>
<td>&gt; 40/10 m [69]</td>
</tr>
<tr>
<td>Trials per experiment</td>
<td>1000</td>
</tr>
</tbody>
</table>

The core task, discussed in the publications [P1]-[P4] is to create an adjustable RRM strategy assuming either dual connectivity or inter-RAT handover. For this scenario, only H2H greedy elastic traffic has been considered, but the proposed solutions may also work with other traffic types. In addition to this, the following assumptions were applied for Scenario 1:

- A user may access only one serving entity per RAT.
- User locations are fixed and assumed to be known by the central management entity.
- Communication is not unidirectional. In the simulator tool (expanded on later) only a UL side was implemented when this research was being conducted, although current techniques do allow bidirectional communication analysis.

### 3.1.2 LTE-based MTC

In the scope of this thesis, LTE-based MTC modeling was limited by performance evaluation of that particular protocol. For the purposes of the thesis the scenario is referred as **Scenario two**. The essential research goal of Scenario Two was to test how well the conventional LTE control channels are adapted for mMTC deployments. The parameters set out for Scenario Two are summarized in Table 3.2. To better understand particular parameters in the Table 3.2 it will be necessary to go through the random access (RA) procedure description, detailed in [70]. Below is an explanation of the basic principles of LTE RA, necessary to understand the problem statement.

The core part of the LTE-RA procedure is depicted in Figure3.2 and UE-eNodeB message exchange is shown in Figure3.5. The procedure starts when UE transmits the RA preamble - chosen from 64 pseudo-random sequences (Msg 1). Within the preamble, UE then sends a specific id, which eNodeB will also record in Msg 2 to confirm the reception of the Msg
3.1. Scenarios of Interest

Table 3.2: Scenario two deployment parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell bandwidth</td>
<td>5 MHz</td>
</tr>
<tr>
<td>PRACH Configuration Index</td>
<td>6</td>
</tr>
<tr>
<td>Total number of preambles</td>
<td>54</td>
</tr>
<tr>
<td>Max. number of preamble transmissions</td>
<td>10</td>
</tr>
<tr>
<td>Number of UL grants per RAR</td>
<td>3</td>
</tr>
<tr>
<td>Number of CCEs allocated for PDCCH</td>
<td>16</td>
</tr>
<tr>
<td>Number of CCEs per PDCCH</td>
<td>4</td>
</tr>
<tr>
<td>Ra-ResponseWindowSize</td>
<td>5 ms</td>
</tr>
<tr>
<td>Mac-ContentionResolutionTimer</td>
<td>48 ms</td>
</tr>
<tr>
<td>Backoff Indicator</td>
<td>20 ms</td>
</tr>
<tr>
<td>Probability of successful delivery for Msg 3/Msg 4</td>
<td>0.9/0.9</td>
</tr>
<tr>
<td>Max. number of HARQ Tx for Msg 3 and Msg 4 (non-adaptive HARQ)</td>
<td>5</td>
</tr>
<tr>
<td>Number of MTC devices</td>
<td>5K, 10K, 30K</td>
</tr>
<tr>
<td>Number of available subframes for device activation</td>
<td>10K, 60K</td>
</tr>
<tr>
<td>Periodicity of PRACH opportunities</td>
<td>5 ms</td>
</tr>
<tr>
<td>RAR response window</td>
<td>5 ms</td>
</tr>
<tr>
<td>Preamble transmission time</td>
<td>1 ms</td>
</tr>
<tr>
<td>Preamble processing time at eNodeB</td>
<td>2 ms</td>
</tr>
<tr>
<td>Processing time before Msg 3 transmission</td>
<td>5 ms</td>
</tr>
<tr>
<td>Time of transmission of Msg 3, waiting, and reception of Msg 4</td>
<td>6 ms</td>
</tr>
<tr>
<td>Power consumption in inactive state</td>
<td>0.0 mW</td>
</tr>
<tr>
<td>Power consumption in idle state</td>
<td>0.025 mW [71]</td>
</tr>
<tr>
<td>Power consumption of processing and Rx</td>
<td>50 mW [71]</td>
</tr>
<tr>
<td>Power consumption during Tx</td>
<td>50 mW [71]</td>
</tr>
</tbody>
</table>

1 from each particular user. After Msg 2, UE sends a "RRC connection request" message (Msg 3) to the eNodeB, which is then acknowledged with "RRC connection set-up" (Msg 4).

![LTE RA procedure details](P5)

The most critical part of this research is the transmission of Msg 1. If two or more UEs choose the same preamble within the same slot, a collision will occur. This will mean that UE will need to re-transmit the message after waiting for a period of time within the backoff window. While in conventional LTE applications the physical random
access channel (PRACH) capacity is sufficient to support even heavily-loaded scenarios, in LTE-MTC there could be thousands of UEs trying to gain access to the channel at the same time, potentially causing a PRACH overload.

As well as the LTE-MTC Msg 1 overload scenario the research also considered the limitations of Msg 2 capacity (which is bound to the number of Control Channel Elements (CCEs) allocated in Physical Downlink Control Channel (PDCCH)) and power consumption of the overall procedure. Finally, there is a brief discussion of the overload-avoidance methods, which would be applicable in the mMTC scenario without any necessary hardware alterations.

The considered model is studied under following assumptions:

- Only MTC-related traffic models from [72] are considered.
- RA power ramping effect is approximated analytically (see 3.3.2)
- Msg 3 and Msg 4 are considered to be successfully received based on the probabilities \( \pi_3 \) and \( \pi_4 \) respectively.

### 3.1.3 MmWave Network

Details of the analyzed mmWave network are shown in 3.3. As described in Section 2.1.2, this mmWave channel propagation model differs from one in conventional cellular networks, leading to necessary modifications in the AP and UE antenna configurations. In this scenario, mmWave APs and potential signal blockers are dropped according to Poisson distribution with intensities \( \lambda_A \) and \( \lambda_B \) respectively (Figure 3.3, left). Blockers represent users not connected to a mmWave network who obstruct the LoS connection between mmWave AP and active UE. Additionally, mmWave APs are equipped with digital array systems and can switch between different array configurations, in order to maintain full 360-degree coverage within the cell range (3.3, center). Horizontal and vertical beamwidths of a mmWave antenna are determined with respect to the number of appropriate array elements (3.3, right), and the beam direction depends on the element spacing and array codebook configuration.

![Figure 3.3: mmWave network deployment details](P7)

From now on, mmWave network deployment, discussed in this thesis, will be referred as **scenario three**. Default system configuration settings, related to Scenario three, are
shown in Figure3.3. Besides the previously discussed number of array elements there are several critical parameters that require explanation. This study, for example, was to show user statistics deployed at the distances higher than $R_b$. If the UE is located outside the zone defined by $R_b$ it is assumed to be in outage in case of blockage. The blocker and UE dimensions, together with the mmWave AP height, define the probability of a blockage in detail as studied in [73, 74, 75]. Finally, the received SINR is affected by transmission power, the probability of blockages occurring and characteristics of the channel model [76]. These in turn depend on the central frequency, the distance between UE and AP and the path-loss exponent.

<table>
<thead>
<tr>
<th>Table 3.3: Scenario 3 deployment parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>Operational frequency</td>
</tr>
<tr>
<td>Height of AP</td>
</tr>
<tr>
<td>Height of blockers</td>
</tr>
<tr>
<td>Height of UE</td>
</tr>
<tr>
<td>Blocker radius</td>
</tr>
<tr>
<td>SNR blockage threshold</td>
</tr>
<tr>
<td>Outage radius</td>
</tr>
<tr>
<td>Interference threshold</td>
</tr>
<tr>
<td>Transmit power</td>
</tr>
<tr>
<td>Path loss exponent</td>
</tr>
<tr>
<td>AP array</td>
</tr>
<tr>
<td>UE array</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>AP/UE attenuation coefficients</td>
</tr>
<tr>
<td>Intensity of APs</td>
</tr>
<tr>
<td>Intensity of blockers</td>
</tr>
<tr>
<td>Speed of blockers</td>
</tr>
<tr>
<td>Array switching time</td>
</tr>
</tbody>
</table>

The main goal of the conducted mmWave RAT research is to evaluate stand-alone network performance, taking into account the impact of four core aspects: cell density, human blockage, multi-connectivity and antenna array configuration. To complete the description of Scenario three there is a list of assumptions made as part the study:

- Only the LoS component of the channel propagation model is taken into account. Multi-path components, as well as NLoS conditions, are not considered.
- Only the horizontal part of the array directivity model is considered. The considered array is linear, although the model could also be extended to take into account 3D array directivity.
- Frame structure and control signaling is not modeled directly but replaced with appropriate timings.

3.2 Simulation-based Environment

3.2.1 Simulation Environment Comparison

In modern telecommunications, network equipment developers and operators are trying to minimize possible losses that can arise from abnormal equipment behavior and flaws
in network performance. Even during the UE development process simple prototyping and pre-sales debugging are often not sufficient and can cause unexpected failures\(^1\), which in combination with poor risk-management have the potential to lead to company bankruptcy. In networking, prototyping is often either impossible or limited to sizes well below the scale of deployment of the final version, while potential losses are sometimes unpredictable and could cause government-level consequences. It is crucial, therefore, to consider all possible methods of telecommunication systems evaluation before the final commercial product is deployed.

Besides prototype implementation, there are two main options to test any new technology; analytical models and computer simulations. While analysis is crucial in fundamental research, in practical networks modeling the mathematical abstraction becomes too complicated or has too many assumptions to reflect realistic behavior. In telecommunications, analytical models are sufficient to evaluate upper and lower bounds of the considered network capabilities, but not detailed enough to take into account all essential protocol features. Network simulators, on the other hand, are positioned as a trade-off solution given the time and complexity of the resources and the subsequent accuracy of evaluation.

Simulation tools considerably differ in their purposes, complexity and application area. On a basic level, we could split the computer modeling-based simulator tools into two major classes: general-purpose and technology-specific. Technology-specific modeling tools are designed to evaluate the performance of either particular protocol stack or the combination of multiple technologies in the particular environment. Technology-specific simulators are usually developed by the company R&D units and intended for internal use only. In contrast, general-purpose simulators are often created on the open-source basis, or sold as an end product, available for general public. For example, Riverbed Opnet \(^2\) could be considered as an instance of a commercial general-purpose simulator with user-friendly interface and supplementary technical support, while NS3 \(^3\) is an open-source platform with the worldwide community, sandbox-like basic functionality and a number of community-driven modules, integrated within the common framework.

In addition to basic comparisons, there are several sub-classes of simulators that are designed to solve a narrower set of tasks:

- **System level simulators (SLS).** Designed to evaluate the technology capabilities on a system-wide scale, SLS are usually developed at the final stage of the risk-management chain in order to approximate network performance in the busy hour/overload conditions and to observe possible weak spots that are not visible at the prototype stage. SLS development usually requires a high level of cooperation between diversified developer teams, driven by demand arising through the coexistence of several protocol stacks and advanced channel models within one simulated environment.

- **Link layer simulators (LLS).** LLS are usually created in the early stages of development in order to obtain a realistic channel picture of P2P links. LLS design is usually based on the measurements made in certain environments, or is reliant on the appropriate methodology to describe essential system parameters and assumptions. LLS results can be used as an input for SLS to reduce simulator complexity and computation time.

\(^1\)https://www.wired.com/2017/01/why-the-samsung-galaxy-note-7-kept-exploding/
\(^3\)NS3 simulator project website, https://www.nsnam.org/
Protocol level simulators (PLS). Developed to model specific protocol details, PLSes could use the same general-purpose platform as the appropriate SLS or may be designed as a separate tool, with the results used in SLS to help reduce computation time and omit unnecessary code complexity.

### 3.2.2 System Level Simulations

As an example of the necessity of system level simulations, in Figure 3.4 comparison between the capabilities of LTE-WLAN SLS and a simplified analytical environment and architecture is shown.

![Diagram showing comparison between SLS and analytical framework](image)

**Figure 3.4:** SLS and analytical framework comparison, [4]

Although it is an isolated example, the comparison is representative enough to show basic differences between the approaches:

- While in SLS the considered protocol details and signaling are modeled with the minimal level of abstraction. In the analysis protocol details are often omitted or replaced with appropriate statistical models.

- SLS traffic could either be simulated based on appropriate models or generated from real traffic samples. In the analysis the traffic generator has been replaced with well-known distributions where individual packets are not usually modeled.
Chapter 3. Instrumentation and Methodology

- The SLS channel model is based on real measurements or LLS channel predictions and is then simulated for each user individually, taking into account per-UE interference and realistic mobility. In an analytical environment, the channel model is usually replaced with statistical abstraction.

- SLS traffic generation requires per-packet (or at least per user) statistics collection, while analytical results follow the output of the final equations and do not require per-packet parsers.

In the research work summarized in this thesis, we used our own SLS called Wintersim\(^4\). Wintersim has been developed at Tampere University of Technology (TUT) and is employed for analysis of advanced wireless networks. It has a modular structure which allows the integration of new protocols into the framework without the necessity to change core architecture. Wintersim is event-driven, meaning that it skips empty slots and models only the set of time instances that contain actual events. The event is indicated by any change in the simulated environment (such as data transmission, channel or mobility variation or traffic arrival). The event-driven modeler design structure is usually more complicated than time-driven, but it also allows for a decrease in computation intensity, which is a crucial parameter for numerous system-wide scenarios.

Wintersim was initially developed using three programming languages: C++, Python, and Matlab (for results parsing), with further migration to pure Python in order to decrease code complexity. For heavily-loaded scenarios, there is an option to utilize MobgoDB\(^5\) as an output results storage, which simplifies parsing procedures and allows seamless multi-user access to the collected statistics. Moreover, the Wintersim SLS tool supports multi-threading and remote operations via SSH. However the SLS currently only works on Linux machines and requires some degree of knowledge of the Linux environment.

There are a number of protocol stacks entirely or partly implemented in Wintersim including IEEE 802.11 g/n/ac/ad, 3GPP LTE (release 12 and partly 13) and IEEE 802.16 WiMax. It is also used to test upcoming releases of the standards under development, such as 3GPP mmWave NR \([77]\) and analyze recently-standardized architectures such as LWA and LWIP (both described in the previous chapter). Finally, it should be mentioned, that although higher-level protocols, such as TCP and UDP are also implemented in the SLS, the core functionality of Wintersim is aimed to perform a MAC and PHY-level evaluation.

### 3.2.3 Protocol level Simulations for LTE-based MTC

In Figure 3.5, the Protocol Level Simulations (PLS) structure, developed to model LTE RA procedure are shown in combination with the appropriate message exchange \([78]\). Although PLS was designed to simulate a particular procedural behavior in a similar vein to Wintersim SLS, it has an event-based modular structure, a per-packet traffic generator and statistical record capabilities. The reasoning behind these design features is that besides a simplistic environment, the PLS is intended to simulate densely populated MTC scenarios with several thousand active connections, which puts strict limitations on the individual event computation time. Moreover, the simulator is designed in such a way that all other modeled components, such as the channel, frame structure and higher-level protocols, are either omitted or replaced with the appropriate abstraction.

---

\(^4\)First version of Wintersim could be found at [http://winter-group.net/download/](http://winter-group.net/download/)

\(^5\)MongoDB project website, [https://www.mongodb.com/](https://www.mongodb.com/)
3.3 Analytical Framework

3.3.1 Resource Allocation in Three-Tier multi-RAT Network

Although this example might seem too specific, it shows the general trend in the protocol-oriented simulators development process. The architecture of PLS usually depends only on the simulation conditions and considered protocol architecture, while structural modifications and functionality beyond initial requirements are usually not supported. Nevertheless, if the PLS was developed in the framework of a general-purpose simulator, it would be possible to use it in the future as a separate module of large-scale SLS.
The first approach, called **max-min fairness**, was initially developed by Bertsekas and Gallager in [81]. The objective function of this method was formulated as a maximization of minimum per-UE resource allocation, $\sum_{p=1}^{P_d} x_{dp}$, $d = 1, 2, \ldots, N$, where $N$ stands for the number of users, $P_d$ represents the number of paths (RAT links) through which the traffic of each user $d$ could be delivered to the aggregator. The output solution $\hat{x}$, will be lexicographically at maximum between all the possible allocation vectors. While being relatively simple to implement, the solution is also the LP type, which means a relatively low computation complexity that allows efficient scaling for network-wide deployments. At the same time, max-min optimization could prove to be sub-optimal to overall system capacity, because the system will always try to maximize the fraction of resources given to the UE with poor signal quality.

The second approach, called **proportional fairness**, is usually positioned as a trade-off between fairness and capacity maximization-oriented resource allocation techniques. A simple, proportionally fair objective criterion could be defined as:

$$\sum_{d=1}^{N} \log x_d \rightarrow \max,$$

(3.1)

The logarithm function in (3.1) plays a crucial role in balancing allocation fairly: the output of the logarithm drops dramatically if the input value is too small and does not grow significantly if the input is too big. At the same time, the usage of PF in that
formulation is appropriate to challenge. Firstly the logarithm function does not give the necessary flexibility if the operator wants to tune the network resource allocation to increase fairness or capacity. Moreover, in the case of the wireless networks, the fairness emphasis in RRM could lead to a substantial degradation of the overall network performance [82]. In addition to this, the objective function in (3.1) belongs to the convex optimization set of problems which is more complicated, in terms of solution algorithms, than the linear programming discussed previously.

Therefore, in [P1] more optimization criteria are demonstrated, called weighted alpha-fairness. The technique has been developed on a max-min approach with modifications, enabling flexible control of fairness/capacity ratio. To understand the proposed criteria requires analysis of the solution Algorithm 1, detailed in [publication access].

**Algorithm 1** The weighted max-min algorithm for resource allocation problem for capacitated fair networks.

1. Modify the initial vector of SEs \( (s_{dp0}) \) to receive control over the system fairness/throughput. \( s_{dp} = f(s_{dp0}) \), where \( f(s_{dp0}) \) is the selected control function (e.g., \( f(x) = \alpha^x \))
2. Estimate \( \Delta \) as a solution to the LP problem and
   - set \( n = 0, \Delta^{(0)} = \Delta \);
   - define \( Z_0 = \{1, 2, \ldots, N\} \).
3. Set \( n = n + 1 \) and for each \( d \)
4. Check the throughput allocation. The users whose allocation can be increased are defined as a subset \( Z_n \subseteq Z_0 \).
5. if \( Z_n = \emptyset \) then
6. go to 13
7. else
8. \( Z_{n-1} = Z_{n-1} \setminus Z_n \).
9. end if
10. Solve the following LP problem
    
    \[
    \text{max } \Delta, \\
    \text{subject to:}
    \]
    
    \[
    \sum_{p=1}^{P_d} s_{dp} x_{dp} = h_d, \quad d = 1, 2, \ldots, N, \\
    \Delta - h_d \leq 0, \quad d \in Z_n, \\
    \Delta^{(k)} - h_d \leq 0, \quad k = 0, 1, \ldots, n - 1, \\
    \sum_{d=1}^{N} \sum_{p=1}^{P_d} \delta_{xdp} x_{dp} = B_e, \quad e = 1, 2, \ldots, E,
    \]
11. Set \( \Delta^{(n)} = \Delta \)
12. go to 3
13. Apply function \( f_1(x) \) that will take into account the SE vector modifications considered in step 1. With the said function applied, the final user throughput vector \( h_{d1} \) will have a form of \( h_{d1} = h_d s_{dp0} / s_{dp} \).

The max-min optimization problem is solved in step ten of the Algorithm 1. It should be noted that the capacity constraints of the optimization function in step ten are defined according to eNodeB/AP total allocated bandwidth \( B_e \) and output throughput.
on each RAT $h_d$ delivered to the appropriate user $d$. On top of this, the modified vector of per channel spectral efficiencies $S_{dp}$ is taken into account in the total per-user throughput calculations. The spectral efficiency vector is modified according to the function $f(x) = \alpha^x$, where coefficient $\alpha$ determines the fairness/throughput balancing ratio. Thereby, by varying coefficient $\alpha$ the operator could switch the network towards more efficient bandwidth utilization or a more even distribution of available resources.

Finally, in [P1] we also consider two simple heuristic approaches, which are used as the baseline comparison options for the weighted alpha-fairness scheme under proposal. The first approach, called ’max-usage’, represents the network performance if the centralized control is absent, and UE is connected to all available HetNet tiers. In that case, assuming greedy elastic traffic, UE will utilize as much resource as the appropriate RAT schedulers allow. The second approach, called ’WiFi-preferred’, is based on the algorithm initially proposed in [P4]. This technique could be viewed as network-assisted because it implies that the UE is choosing to which tier he should connect based on information provided by the central management entity. However, to make the system more flexible, the signal threshold is set to be received on each tier so that it regulates the minimum connection quality and enables alterations of each tier’s coverage area. In these conditions, the primary strategy UE is to connect to the WiFi network first, assuming higher data rates and a lower load - although it will depend on several factors, including regional user density and threshold levels mentioned above.

3.3.2 MTC-enabled LTE

Analysis of the RA procedure in massive-MTC is divided into two parts: studies on overload control performance and research of RA energy efficiency.

The analysis of overload performance starts with the calculation of the average time required to complete the RA procedure,

$$E[\tau] = E[\tau^{(1)}] + E[\tau^{(2)}],$$

where $E[\tau^{(1)}]$ and $E[\tau^{(2)}]$ are the average transmission and processing delays of Msg 1-2 and Msg 3-4 respectively. While Msg 3-4 processing time is based solely on the probability values, defined in Table 3.2, the computation of Msg 1-2 is more complicated and implies two stages for cases both with and without collisions. The analysis with no collisions takes into account only the power ramping effect and gives the average time required to process Msg 1-2 for a single UE system. The modeling of collisions is done with abstraction of memory effects and is based on the two-state Markov chain, shown in Figure 3.7, where states represent the number of pending user requests, while the transitions show the probability of new service request ($\Pi$) arrival and successful Msg 1-2 reception ($\tilde{\mu}$).

![Figure 3.7: Two-state Markov chain describing the number of non-served user requests [P5]](image-url)
In addition to the average service time in studies presented in [P5] there is also a description of the probability of RA procedure failure, which happens if the maximum number of preamble transmission attempts is reached (see Table 3.2). These calculations are especially crucial for the beta distribution-based traffic, which represents an emergency scenario where devices are trying to get the network access at the same time and therefore affecting the RACH instantaneous load. However, beta distribution-based traffic is not analyzed due to the time-dependent $\Pi$. Thereby the emergency scenario is modeled only in PLS.

The energy consumption performance is obtained by setting certain consumption values for inactive ($P_0$), idle ($P_1$), RX ($P_2$) and TX ($P_3$) states depicted in Figure 3.2. While it is self-evident that the total energy consumption will be lower if the transmission time is minimized, it is also important to analyze the system performance during the waiting period. The idle state power consumption is particularly interesting in connection with the overload avoidance mechanisms under scrutiny, which imply much longer backoff windows or usage of initial backoff [83] inserted before the transmission of Msg 1. In these conditions, the idle period could last up to several seconds and therefore will influence the total power consumption of the device.

### 3.3.3 MmWave Network

Research related to Scenario Three could be divided into two parts: LoS blockage analysis and connectivity-related performance evaluation. While the first part is detailed in [73], in this thesis, we will concentrate on the second part, studied in [P7].

The analysis starts with the calculation of the average channel capacity of the user associated with $i^{th}$ nearest mmWave AP. Using Shannon’s formula, it could be defined as:

$$ E[R_i] = cB \log \left(1 + \frac{P_{R,i}}{N_0 + I}\right), \quad (3.3) $$

, where $P_{R,i}$ is received signal power, calculated based on the channel propagation model, $N_0$ is the noise level, $B$ is the channel bandwidth, $c$ - MCS leveling coefficient and $I$ is the average received interference level. It is particularly interesting that in mmWave networks the calculation of the interference depends not only on the AP and UE density and power but also from the parameters of the array system, used on both sides. While the complete analysis of interference is detailed in [P7], here is a brief summary of the three different connectivity approaches, used to provide data plane access.

In Figure 3.8 A, basic static connectivity option, referred in [P7] as "static, nearest AP" is shown. This is the most simple strategy, according to which UE chooses nearest AP
based on the appropriate distance and does not take into account the received signal strength, as it is usually done in conventional cellular networks. Assuming that handovers are not allowed, the ergodic UE session capacity could be calculated as

\[ C = w_1 p_{A,1} E[R_1], \]  

(3.4)

where \( E[R_1] \) is the average channel throughput. This assumes that UE is always connected to the nearest AP, \( p_{A,1} \) (the proportion of time) when the LoS is not blocked, and \( w_1 \) is the probability that blockage does not last longer than transmission session. The reader should keep in mind that one of the assumptions made in this analysis implies that in the case of a blockage, the transmission is stopped and the appropriate channel capacity becomes equal to zero. Therefore, if the connection choice is based only on the distance between UE and AP, the UE could be blocked entirely during the whole session with probability \( 1 - w_1 \), the value of which is connected to the blocker geometry, population density and relative UE-AP height [73].

The second connectivity strategy (Figure 3.8, B), referred in [P7] as "static, LoS AP" is based on the signal strength, which allows UE to avoid blocked connections, although handovers between different APs are still not allowed. The ergodic channel capacity, in this case, could be calculated as follows

\[ C = \sum_{j=0}^{N} q_j w_j p_{A,j} E[R_j]. \]  

(3.5)

In (3.5), the coefficient \( q_j \) is introduced, representing the probability to be connected to AP \( j \). The value of \( q_j \) depends solely on the probability of chosen AP blockage. Here \( N \) represents the degree of multi-connectivity and in theory could be equal to infinity, while in practice its value is bounded the receiver sensitivity.

Finally, in the third approach, shown in Figure 3.8 C, handovers are allowed. Referred to in [P7] as "dynamic", this strategy enables UE re-connection in case of an LoS blockage. The ergodic capacity, in this case, depends on the degree of multi-connectivity. For \( N = \infty \) it could be calculated as follows

\[ C = \sum_{j=0}^{N} \pi_j w_j \frac{E[R_j] E[A_j]}{E[L_j]}. \]  

(3.6)

Here, instead of relative time spend in a non-blocked state (\( p_{A,j} \)), we have used the proportion of active transmission time \( E[A_j] \) and a total session duration time \( E[L_j] \). In the case of ideal instantaneous handovers \( E[A_j] = E[L_j] \) the capacity depends only on AP density and channel conditions. However, in practical mmWave systems handover duration depends on the beam search procedure speed, thereby the overall performance is also connected to the appropriate protocol implementation.

In conclusion it should be mentioned that that in Scenario Three mmWave networks were evaluated outside of the context of HetNets. In the next step Scenario Three will be
modified to take into account the possibility of assistance from the conventional cellular networks. In anchor-booster HetNet architecture, mmWave APs are usually envisioned as boosters, providing user-plane connectivity, while 4G macrocells can be assigned as anchors, giving reliable control-plane signaling and coverage.
4 Performance Evaluation

This chapter will center on a comprehensive evaluation of developed methodology, starting with the presentation and discussion of the selected results, related to three chosen scenarios. At the end of the chapter, there will be a short description of the developed prototype network that has been built at the TUT campus.

4.1 Selected Results

4.1.1 Multi-RAT Network

The performance assessment of three-tier multi-RAT network starts with the calibration of the analytical environment with a large-scale SLS tool. Comparison of these two instruments is based on two performance indicators shown in Figure 4.1: percentage of UEs associated with each tier and per-UE throughput. The default RRM scheme used on the calibration stage is "max-usage" described in Subsection 3.3.1. There are 60 UEs and five (of each) LTE pico eNodeBs and WiFi APs uniformly distributed within the macrocell coverage area. Other parameters, related to Scenario One are given in 3.1.

The difference in the percentage of association was marginal and explained by a number of details, taken into account in SLS and abstracted during analysis. Although throughput comparison does not exactly match due to the presence of interference in SLS, it is still valid to claim the correctness of the analytical model, at least for the area below 25% mark. The small step in the analysis CDF observed at around twenty-five percent shows the difference between the performance of UEs with macrocell coverage only and all other users. In SLS, this step is absent because of interference between neighbor

Figure 4.1: Analytical model calibration for Multi-RAT scenario [P1]
Chapter 4. Performance Evaluation

LTE picocells which caused significant performance degradation [84] and shrunk the difference in performance for the UEs with different connectivity capabilities. Based on these observations it can be concluded that the test environment was sufficient to evaluate the proposed resource allocation strategies in this scenario. For the simplicity of implementation, in further research related to the multi-RAT RRM, we are using only the analytical model discussed above.

After the calibration there will be an evaluation of the proposed resource allocation strategies. In Figure 4.2 throughput comparison of two chosen heuristic algorithms and max-min fairness technique is shown. At a glance, the fifty percentile value of all three schemes is nearly equal. However, one may observe a dramatic difference in the five percentile, as well as an imbalance between the throughput level of five and ninety-five percentile UEs for heuristic schemes. It is especially noticeable for the "WiFi-preferred" scheme, where the inequality between resource allocation for UEs with WiFi connectivity and all others is vast. As predicted, the max-min scheme gives the best performance for cell edge users and at the same time limits the resource consumption for users with the best connectivity and channel conditions.

On top of basic CDF comparison, the numeric representation of fairness could be defined in a form of Jain’s index:

$$J = \left( \sum_{i=1}^{N} x_i \right)^2 \frac{1}{N \sum_{i=1}^{N} x_i^2}$$  \hspace{1cm} (4.1)

In Figure 4.3 a comparison of Jain’s fairness index for different RRM strategies is shown together with the calculations of so-called weighted throughput (multiplication of average throughput by the Jain’s index). The observation showed that when the number of deployed LPNs increased, the average performance of all three schemes grew unequally. Initially, the WiFi-preferred strategy seemed to be the most beneficial for average per user throughput, although its fairness remained limited. However, as the LPN density increased, the connectivity capabilities of the other two schemes grew, and the average
4.1. Selected Results

Throughput performance aligned. At the same time, the fairness improvement of all three systems stopped - a further LPN number increment did not give additional coverage and only affected network capacity. Nerveless, the fairness of max-min technique at ultra-high LPN densities was significantly better, which in combination with decent average throughput made it the preferable resource allocation strategy to be used in that particular scenario.

Finally, the performance of the weighted alpha-fairness algorithm must be evaluated. In Figure 4.4, the performance comparison for two alpha-control functions is presented. As described in Subsection 3.3.1, the tuning of the fairness-to-capacity ratio was achieved by modifying the set of input parameters with the specific weighting function, controlled by the coefficient $\alpha$. In both $x^\alpha$ and $\alpha^x$ functions, higher values of $\alpha$ changed resource allocation behavior towards more fairness while function $\alpha^x$ gave significantly wider control limits. In $x^\alpha$, if $\alpha = 1$ the scheme worked as conventional max-min, while $\alpha = 0$ switched the network into ‘max-usage’ mode, where UE channel conditions were not taken into account. The $\alpha^x$ function went even further, allowing the operator to limit the performance of UEs with poor channel conditions by setting $\alpha < 1$. In both cases, the control region was confined by coverage and capacity conditions of the network. In case of fairness maximization, there is a limit after which its performance could not be improved any further due to poor connectivity capabilities of particular UEs. At the same time, the growth of maximum network capacity was proportional to the sum of bandwidths of the involved RATs.

4.1.2 LTE-based MTC

The calibration for LTE-based MTC scenario was done according to the 3GPP methodology document [85] (Figure 4.5). Although the methodology only gave the ten and ninety quantiles, it was enough to confirm the validity of the developed simulation tool. For traffic type one (uniform time of arrival distribution) there was a perfect match between PLS and methodology performance, while for traffic type two (beta distribution-based), there was about a ten percent mismatch observed in the upper quantile.

The comparison of PLS results and related analysis has been made for Traffic Type One only due to limitations within the test environment. The results of power consumption measurements for the tools used here are shown in Figure 4.6. In both simulation and
Figure 4.4: Weighted alpha-fairness control regions [P1]

Figure 4.5: PLS calibration [P5]

analysis, there were no marginal power consumption variations between 10K and 30K cases. The same states for the other metrics are summarized in Table 4.1.

In the presence of Traffic Type Two, the performance of the RA procedure degraded as shown in Figure 4.7. In case of simultaneous traffic arrivals from 30K users, the Msg 1 collision probability grew dramatically, reaching fifty percent when the default backoff indicator (BI) was equal to twenty milliseconds. Moreover, because of the small size of the backoff window, there was a high probability that users would re-transmit the preamble within the next subframe, leading to consequent preamble collisions and RA failures after the number of maximum re-transmissions was exceeded. Notably, to overcome this effect, the value of the backoff window had to be increased to at least three hundred milliseconds. At the same time, to completely negate access failure a probability BI value of at least 3000ms was required. Finally, it is worthy of note that even when non-zero
4.1. Selected Results

Table 4.1: Random access procedure performance for traffic type one

<table>
<thead>
<tr>
<th>Number of devices</th>
<th>5000</th>
<th>10000</th>
<th>30000</th>
<th>Results origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collision Probability (%)</td>
<td>0.01</td>
<td>0.03</td>
<td>0.22</td>
<td>Methodology [85]</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.03</td>
<td>0.23</td>
<td>Simulation</td>
</tr>
<tr>
<td>Number of preamble Tx attempts</td>
<td>1.43</td>
<td>1.45</td>
<td>1.50</td>
<td>Methodology [85]</td>
</tr>
<tr>
<td></td>
<td>1.43</td>
<td>1.44</td>
<td>1.50</td>
<td>Simulation</td>
</tr>
<tr>
<td></td>
<td>1.44</td>
<td>1.47</td>
<td>1.57</td>
<td>Analysis</td>
</tr>
<tr>
<td>Access delay (ms)</td>
<td>25.60</td>
<td>26.05</td>
<td>27.35</td>
<td>Methodology [85]</td>
</tr>
<tr>
<td></td>
<td>25.70</td>
<td>26.00</td>
<td>27.10</td>
<td>Simulation</td>
</tr>
<tr>
<td></td>
<td>25.90</td>
<td>26.40</td>
<td>28.45</td>
<td>Analysis</td>
</tr>
</tbody>
</table>

Figure 4.6: Power consumption comparison for different number of users [P5]

power consumption (see Table 3.2) during the backoff waiting period was assumed, the overall energy efficiency grew significantly along with the increment of the size of the backoff window.

As a summary for the MTC-related part, it can be concluded that usage of RACH-based operations in mMTC conditions required overload avoidance techniques. However, while the backoff increment was suitable for delay-insensitive scenarios it would be more appropriate to apply other methods for delay-critical MTC applications. [P6].

4.1.3 MmWave Network

Although large-scale SLS toot was not directly used in the mmWave scenario evaluation, in the baseline scenario the analytical model was benchmarked with the simplified computer simulation as shown in Figure 4.8 (simulation results are marked with dots). In the figure several channel performance metrics calculated for different AP densities are also presented. It could be inferred from the plot that although the interference level grew significantly the overall SINR was mostly affected by the received signal strength value. This happened because of the marginal base value of interference which initially did not exceed the noise level. Low interference level, in turn, was explained by the mmWave sessions spatial separation, which was itself enabled by the narrow directivity patterns of the antenna arrays.
In mmWave systems, the interference level depends not only on the density of AP, but also on the geometry of the scenario. Static obstacles, such as buildings or trees were not considered in this research and should be studied by using advanced simulation tools. Here we investigate the effect of dynamic human blockers, that can not only interrupt the active sessions but efficiently shield other users from potential interference from other APs. As it is shown in Figure 4.9 (left, central), the high density of blockers significantly decreases the amount of received interference, especially for the low AP density deployments. At the same time, the number of antenna elements (Figure 4.9, right) does not affect the interference level. Despite higher gains, the higher number of elements shrinks the effective antenna beamwidth which therefore decreases the probability of that particular transmission causing interference to other users.

It has also been noticed that average throughput directly depends on the number of antenna array elements as it is shown in Figure 4.10 (throughput values are calculated per one Hz of bandwidth). This behavior is caused by the higher beam-forming gain of
4.1. Selected Results

the antenna array with more elements, assuming ideal beam-searching and beam-tracking are implied. The other interesting effect could be observed in the central subplot of Figure 4.10. Assuming the \( i^{th} \) AP closest to UE is not blocked, the interference level on the link decreases, because the effects of blocking reduce the probability of interference from the other active session. Finally, the left plot of Figure 4.10 shows the relation of throughput to higher AP densities. The average capacity, predictably, increases to the point of \( \Lambda_A = 10e^{-4} \), while further densification of the deployment causes performance degradation because the interference finally starts to influence the connection quality.

In Figure 4.11, the ergodic capacity of three different connectivity strategies (see Subsection 3.3.3) is shown. Observations found the performance of static schemes to be nearly equal. Two phenomena could explain this: the non-linear growth of blockage probability with the AP-UE distance and marginal capacity variability for several of the nearest APs. In practice, this means that when the UE connects to the more distant non-blocked AP at the beginning of the session, the connection could be still blocked during transmission, whilst the average session throughput would remain the same.

At the same time, the dynamic scheme predictably outperforms both static options. Furthermore, in the left plot of Figure 4.11 the optimal value of AP density was observed for all three schemes. The appearance of the optimal point was explained by the existing trade-off between the received signal strength and growing level of interference.

The effect of blockage, shown in 4.11(central) has a limited impact on the total ergodic capacity. Although the capacity of all three schemes initially starts to grow with the
higher blockers density due to interference suppression, eventually the blockers start to affect the active session as well. This leads to a slow degradation in performance of both static schemes. At the same time, the dynamic scheme also suffers from blockers due to non-zero handover time. Finally, the number of antenna elements also marginally affects the total channel capacity (4.11, right).

4.2 Trial Activities

4.2.1 Single UE Tests

The implementation of Multi-RAT demo network was initiated by providing single UE tests, focusing on flow switching and DC scenarios described in Subsection 3.1.1. Although single-UE tests do not allow network performance estimation, it is still feasible to conduct research in this direction to evaluate practically-oriented solutions that could then be used in deployment on a broader scale.

One of the first attempts to build a WiFi-preferred switching mechanism was made using Rohde&Schwarz cmw500 radio communication tester, serving as LTE eNobdeB\(^1\) and OpenWRT-enabled\(^2\) WiFi AP. During the setup phase, both WiFi AP and cmw500 were connected to a central management entity (gateway server), which bridged two networks together. Next, UE starts the transmission session on WiFi and performs periodic channel measurements. When the WiFi signal quality decreases, UE switches to the LTE network, although the channel measurements of a WiFi session continues. In these conditions, the re-connection speed depends on the multiple factors: measurement reports periodicity, handover triggers setup and RAT entry procedures speed. However, in our test, the UE already had the active LTE session before handover was triggered, which accelerated the switching procedure. At the same time, the Rohde&Schwarz shielding box was used to decrease WiFi signal strength artificially, and the link quality degradation was almost instant, which also rarely happens in a real environment.

In the next part of the single UE test, the aim was to utilize both WiFi and LTE connections at the same time. However, modern mobile platforms often have limited access to the interface drivers, as well as the capabilities to maintain multiple RAT connections at the same time. To enable these features, in our tests we used Jolla phones\(^3\)

---

1. winter-group.net/rohde-schwarz-tutorial/#more-805
2. OpenWRT project website, https://openwrt.org/start
4.2. Trial Activities

running Linux-based SailfishOS\(^4\) that allow the construction of generic Linux software and custom kernel modules. To aggregate LTE and WiFi links on the UE side, OpenvSwitch\(^5\) software was used. Although OpenvSwitch requires that both links are able to handle Ethernet headers that are not supported by an LTE connection, it is possible to overcome this problem by utilizing GRE tunnels. In the initial solution, an additional VPN tunnel was deployed to create a joint network between the radio links. On top of that an OpenFlow-based controller\(^6\) was designed to maintain OpenvSwitch forwarding table that then allowed the traffic to split flexibly between connections based on the designed criteria, e.g., signal strength. The overall topology of the proposed solution is shown in Figure 4.12.

4.2.2 Test Network Infrastructure

In the next step the test network, utilizing a number multi-RAT LPNs, was deployed at the TUT campus (Figure 4.13). On the radio access level, Cisco AIR-LAP1142N and Ericsson RB6402 Pico eNodeBs were used to provide WiFi and LTE connectivity respectively. The LTE access network was connected to the EPC, (located in Aalto University) through a VPN connection, while WiFi traffic was forwarded directly to the application servers, located in TUT. Although such topology could create unbalanced latency performance, the distribution of resources was still fair concerning throughput and coverage, assuming all LPNs were deployed on the same floor.

By having full control over the application server and WiFi network, together with partial access to LTE radio access and EPC functions, it is possible to implement different multi-RAT connectivity techniques, e.g. the LWIP described in Chapter 2. Furthermore, the site has the capability to test several 5G-enabled scenarios, including MTC, network-assisted D2D and conventional H2H applications.

The network was built with the support of Ericsson Research Finland within the scope of 5th Evolution Take of Wireless Communication Networks (TAKE-5) project\(^7\).

---

\(^4\)https://sailfishos.org/
\(^5\)http://openvswitch.org/
\(^6\)https://www.opennetworking.org/
\(^7\)http://take-5g.org
4.2.3 Multi-purpose Automated Vehicular Platform

To test the deployed network capabilities for MTC-related traffic, two versions of automated vehicular platform was developed, as shown in Figure 4.14. The first version of the platform is equipped with two low-power electric motors, used for driving and steering, two cell lithium polymer batteries, custom motor controller and raspberry pi model 3 with integrated WiFi interface and LTE dongle. In addition to this, Raspberry Pi Camera Module v2 and infrared distance meter were installed, both connected to Raspberry Pi. The platform is connected to the application server through both LTE and WiFi RATs, utilizing connectivity scheme shown in Figure 4.12. On the application side GUI was developed, allowing the operator to obtain real-time video from the camera and control the platform using a keyboard. Furthermore, GUI can print RAT-related statistics such as latency and connection throughput. It also allows manual switching between the interfaces. The second version of the platform (Figure 4.14, right) is equipped with more powerful chassis, camera and sensor subsystem, together with advanced single-board processing unit which allows fully unmanned vehicle operation and possibility to connect 5G mmWave transceivers.

The general idea behind this prototype is to enable automated mobility of the vehicular platform, with periodic measurement reports to the operator and provides simulation of
low-throughput mobile MTC traffic. In case of an emergency, e.g., when the proximity sensor detects an obstacle, the platform stops and switches into manual mode, in which the operator has full control over steering and driving. From that point, requirements for throughput and latency increase significantly in comparison with the automated mode and multi-RAT capabilities of the network could be utilized. During this research, two basic approaches were tested for traffic management: WiFi-preferred RAT switching and DC over OpenvSwitch and multi-path TCP protocol. Further work and testing will continue and this research will require the more advanced platform. This would enable unmanned driving and precise indoor positioning.

For more details, please read the MSc Thesis by the authorship of Jani Urama [5], who is the core software developer for this scenario.
5 Conclusion and Potential Enhancements

The conclusion will be a discussion of the investigative work presented in this thesis, and outlines the following research goals that have been achieved:

- The evolution of the HetNet concept was analyzed, together with existing standardized and conceptual HetNet architectures.
- Multi-RAT radio resource management mechanisms were developed and compared.
- MTC was considered in the context of HetNets. MTC-enabled LTE RA procedure was evaluated.
- An analytical model for mmWave network performance evaluation was created. Created methodology provided insight into the network capabilities and limitations.
- A prototype HetNet was deployed in the TUT campus. Several different multi-RAT traffic switching and splitting strategies were implemented and tested on H2H and MTC applications.

During this work, our research group also developed several modules for a large-scale SLS, PLS and application prototypes tested in the deployed HetNet. All these instruments, together with the analytical methodology, will be used in future studies.

5.1 HetNets Applicability Discussion

In theory, the deployment of HetNets gives significant benefits in terms of resource management flexibility, QoS, application variability and network capacity. However, in practical application many unsolved issues will need to be addressed. Traffic variability, as was shown in the MTC example, can be sometimes difficult to adapt to existing standards that work in the new applications. For the new developed RATs, such as mmWave-based NR, it is essential to evaluate the role and capabilities of upcoming technology and its relation to the existing infrastructure. However, the most challenging part is the integration of RATs, developed by different standardization groups. It is often unclear how such RATs should be connected, starting with the gateway design and continuing with the level of integration and control. Multi-RAT resource allocation schemes that were considered in this thesis were not bonded with particular implementation architecture, although every architecture had a unique set of practical limitations that should be taken into account. For example, in Figure 5.1, the architecture of a practical packet-based
centralized HetNet is shown. In this case, RAT interconnection is implemented on LTE PDCP layer, and resource allocation is performed on a per packet basis. Here, the appropriate analytical evaluation should take into account queuing and individual packets, which significantly complicates problem formulation and solution. Further analysis of given system is presented in [P1].

Figure 5.1: Options for multi-RAT enabled implementation for LTE-based network [P1]

5.2 HetNets as Part of 5G

While the definition and concept of 4G networks is only related to cellular technologies, 5G is a much broader term that includes and unites multiple technologies under one framework. In fact, the definition of 5G often varies and depends on the point of view of the particular company. Despite these variations, the timeline for the development 5G-enabled networks is defined by standardization committees.

For example, according to the current ITU plan, deployment of 5G-enabled networks will start by 2020 [11]. Recently ITU also confirmed a list of minimum requirements for IMT-2020 radio interfaces [86]. These requirements touch three primary system subsets: enhanced mobile broadband (eMBB), URLLC and massive machine type communications (mMTC). It is worthy of note that peak throughput requirement for eMBB technologies is set to 20Gbps and 10Gbps for DL and UL respectively. These values could be achieved either by utilizing large transmission bandwidth, which could be used solely on mmWave frequencies or by combining the channels of multiple RATs for one transmission, i.e., building a multi-RAT HetNet. On the other hand, pure mmWave networks would not satisfy another condition set for eMMB: support of high-speed users (up to 500 km/h) in rural areas, which means that a mmWave network should at least allow for handovers to a conventional cellular network. Furthermore, the minimum user plane latency requirements for eMMB network is set to 4ms (and 1ms for URLLC) that could not be supported with
10ms LTE frame and would require the use of more advanced scheduling techniques. In
addition to this, it is assumed, that an appropriate mMTC network will support up to
one million devices per square kilometer. Nevertheless, both density and latency demands
could be satisfied by utilizing the ultra-dense HetNets concept; where multi-RAT boosters
that have been widely deployed within the coverage area provide sufficient user plane
delay and capacity and play the role of gateway units for mMTC devices. Macro eNodBs
are then used as anchors and give control plane connectivity, which has relaxed latency
and throughput requirements.

Although the ITU documents examined do not directly refer to the term of 5G, they
depict the basis for the next generation of wireless networks. Even basic elemental analysis
shows that the concept of HetNets plays the unifying role, allowing the interconnectivity
of all 5G components together.
6 Summary of Publications

6.1 Publications Description

The main publications used in this thesis are referred to as [P1]-[P7]. The publications include four works ([P1], [P2], [P5]) published or submitted ([P7]) in scientific journals whilst the remainder are conference papers. This section will clarify the contribution of each of the publications.


Description
In [P1] we investigate the impact of different radio resource management schemes on the three-tier multi-RAT network performance. The conducted research output is based on both simulation and mathematical modeling. The results include the comparison of three different resource allocation strategies, assuming permitted simultaneous connectivity to all three radio access technologies. Also, we propose our optimization algorithm, which enables flexible control of fairness-to-capacity ratio over the considered network. Further, we test the performance of different control functions, used in the developed algorithm, and define the appropriate limitation regions. At the end of the paper, we provide applicability discussion of the HetNets resource allocation strategy and propose the alternative analytical approach, which takes into account practice-related issues assuming 3GPP LTE-WiFi aggregation architecture.

This paper is a collaborative work of the author and his supervisor with Dr. Dmitri Moltchanov and Dr. Sergey Andreev from the same research group in Tampere University of Technology (Finland), and his industrial advisers Dr. Nageen Himayat, Dr. Shu-ping Yeh and Dr. Shilpa Talwar from Intel, Labs (USA).


Description
In [P2] we scale the solutions, developed in [P1], to the Heterogeneous Cloud Radio Access Network (H-CRAN) concept. At the beginning of the publication, we discuss the evolution of HetNets taking as the example the standardization path of 3GPP
LTE and IEEE WiFi interconnection architectures. In this context, H-CRAN is considered as one of the possible HetNet development directions in which the central management entity functions are partly or completely transferred into a distributed network of RAN entities. The paper then extends the evaluation of the RRM schemes, initially considered in [P1]. Finally, at the end of the publication, we also propose the demo network prototype architecture based on the solution that is then developed and tested by our research group.

This paper is a collaborative work of the author and his supervisor with Dr. Dmitri Moltchanov, Roman Florea and Dr. Sergey Andreev from the same research group in Tampere University of Technology (Finland), and Dr. Nageen Himayat, Dr. Shu-ping Yeh and Dr. Shilpa Talwar from Intel, Labs (USA).


**Description**

In [P3] we continue the discussion related to the multi-tier HetNet resource allocation problem. Here we introduce the modified max-min optimization algorithm, capable of performing priority-aware multi-RAT RRM. In practice the priorities represent subscription classes, which a mobile network operator could set according to its pricing policy. Subsequently, we evaluate the performance of different prioritization approaches and compare the algorithm with unbiased max-min-based resource allocation. At the end of the publication we also extend the prototype architecture that was initially proposed in [P2]. It should be mentioned that the prototype-related solutions, described in [P3], are now used in the TUT multi-RAT demo network.

This paper is a collaborative work of the author and his supervisor with Dr. Dmitri Moltchanov, Roman Florea and Dr. Sergey Andreev from the same research group in Tampere University of Technology (Finland), and Dr. Nageen Himayat from Intel, Labs (USA).


**Description**

In [P4] we evaluate the performance of heuristic RAT-switching (inter-RAT handover) strategies in a two-tier multi-RAT environment. In particular, we develop and compare three RAT-switching approaches, which enable the flexible control of the user population in a specific RAT, by taking into account the necessary performance indicators including serving entity load, UE channel conditions and particular tier coverage area. To compare the proposed schemes, we used three core metrics: number inter-RAT re-connections, average, and 5% per-UE throughput. In contrast with other publications, in [P4], the research is based only on the simulation tool results. At the end of the [P4], we also study the influence of hysteresis on the proposed inter-RAT handover schemes performance.

This paper is a collaborative work of the author and his supervisor with Dr. Sergey Andreev from the same research group in Tampere University of Technology.
(Finland), and Dr. Nageen Himayat, Dr. Shu- ping Yeh and Dr. Shilpa Talwar from Intel, Labs (USA).


**Description**

In [P5], we study the influence of mMTC traffic on LTE random access procedure performance. The study starts with the description of analytical environment and protocol level simulation (PLS) tool, developed specifically for this scenario. Further, we benchmark our tools with the appropriate methodology documents. Next, we consider RACH overload scenario, where mMTC devices are trying to get access trough PRACH, assuming on the beta-distribution-based traffic arrivals. Performance evaluation of the RACH overload scenario is based on several criteria, including a preamble collision and access-success probability, total random access procedure delay and power consumption. In addition to this, we also test basic overload-avoidance algorithms and evaluate the network performance in non-overloaded ("regular") conditions.

This paper is a collaborative work of the author and his supervisor with Vitaly Petrov, Olga Galinina and Dr. Sergey Andreev from the same research group in Tampere University of Technology (Finland).


**Description**

In comparison with [P5], where PRACH was used to establish the initial connection, in [P6] we evaluate the performance of PRACH and PUCCH as data access (scheduling) providers for MTC devices. We also consider the novel data access scheme, working over PUSCH and developed specifically for mMTC-enabled LTE deployments. The performance assessment is made by utilizing both PLS tool and analysis and takes into account conventional LTE users. Comparison of considered techniques is based on the calculations of resource and power consumption, as well as the overall data access delay.

This paper is a collaborative work of the author and his supervisor with Dr. Sergey Andreev, Vitaly Petrov and Olga Galinina from the same research group in Tampere University of Technology (Finland), and Anna Larmo, Dr. Tuomas Tirronen and Johan Torsner, from Ericsson Research (Finland).


**Description**

In [P7] we consider mmWave network-only scenario, assuming a randomized distribution of mmWave access points and human blockers. First, we evaluate the
mmWave-specific system parameters, related to channel and antenna modeling, which are necessary to capture the essential differences of mmWave-based wireless communications when in comparison with conventional cellular deployments (working on the frequencies below 6GHz). Here, we also reuse the mathematical model of human body blockage, considered in our previous scientific papers. This publication also further develops the analytical framework and allows us to estimate the interference level and capacity of the system. Finally, we evaluate the performance of three user connectivity options, assuming varying AP, blocker and antenna configurations.

This paper is a collaborative work of the author and his supervisor with Dr. Dmitri Moltchanov, Margarita Gapeyenko and Dr. Sergey Andreev from the same research group in Tampere University of Technology (Finland).

### 6.2 Author’s Contribution

The core publications, included in this thesis, were done in the Department of Electronics and Communications Engineering (ELT) in Tampere University of Technology (TUT), Finland. In addition to this, the author participated in research visits to Brno University of Technology (Czech Republic) and Intel Labs (USA), where he completed part of the simulation and prototype development-related assignments. The author of this thesis is a primary contributor to [P1]-[P5] and [P7]. The thesis-related research done by the author was guided by his supervisor, Professor Evgeny Kucheryavy, and Instructor, Dr. Dmitri Moltchanov. Dr. Nageen Himayat (Intel Labs, USA) supported the author with practice-related contributions, participating in research activities as an industrial adviser. Finally, the results summarized in this thesis were developed in tight collaboration between the author, his TUT research team, and international colleagues. The section below will summarize the author’s contribution in all included publications, [P1]-[P7].

In [P1], the author was responsible for the development of the analytical part, related to the weighted alpha-fairness algorithms, as well as the evaluation of all results, presented in the publication. Furthermore, the author also developed heuristic algorithms, and contributed to the problem statement and envisioned HetNet architecture description. In [P2] and [P3], the author described the system model and made a numerical performance evaluation. In [P4] the author was responsible for the algorithms development, system model description, and system model simulations. In [P5], the author was the principal developer of the simulation tool used for performance evaluation. Additionally he also described the practical part of the problem statement related to the procedure under consideration, as well as the solutions presented and discussed at the end of the publication. In [P6], the author’s contribution was limited by the simulation-based evaluation of the proposed solution, although the author also contributed to the protocol-related description of the problem statement. Finally, in [P7], the author participated in the development of the system model, problem statement and results evaluation.
Bibliography


[13] “Study on channel model for frequencies from 0.5 to 100 GHz,” 3GPP Technical recommendation (TR) 38.901, 2017.


