BRNO UNIVERSITY OF TECHNOLOGY

Faculty of Electrical Engineering and Communication

TAMPERE UNIVERSITY

Faculty of Information Technology and Communication Sciecnces

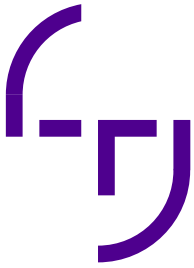
DOCTORAL THESIS

Brno, 2025 Ing. Radek Možný

BRNO UNIVERSITY OF TECHNOLOGY

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

DEPARTMENT OF TELECOMMUNICATIONS



TAMPERE UNIVERSITY

FACULTY OF INFORMATION TECHNOLOGY AND COMMUNICATION SCIENCES

UNIT OF ELECTRICAL ENGINEERING

CELLULAR NETWORK CAPACITY PLANNING FOR FUTURE SMART-GRID SERVICES

DOUBLE DEGREE PROGRAMME BETWEEN BRNO UNIVERSITY OF TECHNOLOGY AND TAMPERE UNIVERSITY

DOCTORAL DISSERTATION

AUTHOR	**Ing. Radek Možný**
ADVISORS	**Ing. Pavel Mašek, Ph.D.** **Dr. Dmitri Moltchanov**

BRNO 2025

ABSTRACT

The increasing demands of smart-grid applications in the landscape of Industrial Internet of Things (IIoT) require optimized communication networks, particularly through Narrowband IoT (NB-IoT) and Long Term Evolution for Machine-Type Communication (LTE-M or eMTC). This dissertation provides novel strategies to enhance these technologies to meet the critical requirements of reliability, low latency, and efficient traffic management, including permanent connectivity in smart-grid deployments driven by regulatory entities, such as the European Parliament and its directives.

A core component of this research is an comprehensive measurement campaign conducted to assess the performance of NB-IoT and eMTC in both controlled environments and realistic smart-grid conditions with heavy traffic loads. The results reveal that while NB-IoT offers superior coverage, it struggles to meet the high-throughput and low-latency demands of smart-grid networks. These findings highlight the need for optimized message distribution on single or a Multi-Radio Access Technology (Multi-RAT) approach to optimize performance by leveraging both NB-IoT and eMTC.

To address the specific challenges posed by smart-grid deployment, several models have been introduced to assess the stability and delay performance of NB-IoT networks based on the gathered results. The generalized Markov chain model highlights the limitations of the NB-IoT in supporting End Devices (EDs) with permanent connectivity, leading to significant delays under high traffic loads, even for 100 EDs. By increasing radio resources in the first model and optimizing message transmission intervals in the simplified model, this research improves the delay performance and balances the network load more effectively, enabling the coexistence of densely deployed traditional sensors and several hundred permanently connected smart meters. Additionally, techniques such as Early Data Transmission (EDT) have been explored, demonstrating substantial latency reductions of more than 50% and enhanced efficiency for smaller message sizes.These optimizations demonstrated efficacy in meeting the smart-grid latency requirements.

Furthermore, an optimal RAT association algorithm is proposed to manage the communication load between the NB-IoT and eMTC EDs to increase communication reliability. This approach ensures more efficient operations by dynamically balancing traffic across technologies, leading to a reduction in latency and an increase in system reliability compared to single-RAT deployments. The algorithm successfully meets the strict latency requirements of smart-grid systems, enabling reliable communication even under diverse geographical conditions and varying deployment densities reaching up to 3,000 EDs.

KEYWORDS

5G IoT, Latency, eMTC, Mobile IoT, mMTC, Multi-RAT, Reliability, NB-IoT, Optimal Associations, Smart-Grids

MOZNY, Radek. *Cellular Network Capacity Planning for Future Smart-Grid services*. Brno, 2025. Doctoral thesis. Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Telecommunications. Advised Dr. Pavel Masek, and Dr. Dmitri Moltchanov.

Author's Declaration

Author:	Ing. Radek Mozny
Author's ID:	173711
Paper type:	Doctoral thesis
Academic year:	2024/25
Topic:	Cellular Network Capacity Planning for Future Smart-Grid Services

I declare that I have written this paper independently, under the guidance of the advisor and using exclusively the technical references and other sources of information cited in the paper and listed in the comprehensive bibliography at the end of the paper.

As the author, I furthermore declare that, with respect to the creation of this paper, I have not infringed any copyright or violated anyone's personal and/or ownership rights. In this context, I am fully aware of the consequences of breaking Regulation § 11 of the Copyright Act No. 121/2000 Coll. of the Czech Republic, as amended, and of any breach of rights related to intellectual property or introduced within amendments to relevant Acts such as the Intellectual Property Act or the Criminal Code, Act No. 40/2009 Coll. of the Czech Republic, Section 2, Head VI, Part 4.

Brno ....

author's signature*

*The author signs only in the printed version.

ACKNOWLEDGEMENT

This research was conducted as part of a double-degree study program at the Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, Czech Republic, and the Unit of Electrical Engineering, Faculty of Information Technology and Communication Sciences, Tampere University, Finland, between 2019 and 2024.

This doctoral thesis was developed with the support of various AI tools, including Paperpal, Grammarly, Writefull, and ChatGPT, which assisted in refining the language and ensuring clarity, coherence, and consistency throughout the document. It is important to note that the content, research findings, and all substantive text within this thesis were entirely generated by the author. The AI tools were not used to create or generate any original content or ideas presented in this work.

The research described in this work was supported by the Technology Agency of the Czech Republic under project no., FW03010424, grant TN02000067 under program National Competence Centre, and the TREND Programme, project no. FW07010004.

I would like to extend my sincere appreciation to my supervisors, Ing. Pavel Mašek, Ph.D. at Brno University of Technology, and Assoc. Prof. Dmitri Moltchanov at Tampere University, for their invaluable guidance, support, and assistance throughout my doctoral studies. Their contributions were indispensable.

I extend my gratitude to the esteemed faculty members of Brno University of Technology, particularly those affiliated with the WISLAB, CyberGrid, and SmartGrid research groups, for their unwavering support and contribution to the successful execution of experiments and the completion of academic writing tasks. Their camaraderie and encouragement proved invaluable during the challenging times of my research. Additionally, I express my appreciation to my peers at Tampere University for their warm welcome and hospitality during my research stay to Finland. Their support and friendship made the experience both successful and enjoyable.

Ultimately, the completion of my work would not have been possible without the unwavering support of my family, loved ones, closest friends, and, above all, my dear wife, Michaela. I extend my sincerest gratitude to all of you for being there for me during my most challenging moments.

Contents

List of Figures

List of Tables

1 Introduction

This chapter presents a comprehensive analysis of the current state of the art in the field of Massive Machine-Type Communication (mMTC) and its implementation within the context of smart-grid systems. It aims to identify and address research gaps in the literature and propose a clear methodology for conducting further research in this area. Remainder of this chapter is organized as follows. The first section discusses the research gaps and motivation for this study, based on the available literature. The second section outlines the research objectives and questions, and defines the exact methodology that will be used to answer them. The chapter concludes with a summary of the key contributions and novelty of this research.

This dissertation, covering the most significant developments achieved over the five-year Ph.D. period from 2019 to 2024, consists of seven chapters. In addition to Chapter 1, containing Introduction, and 7 presenting the Conclusion, the five core chapters summarize key contributions.

The author of this thesis documented and disseminated the essentials of the contributions in the form of scientific publications in recognized journals and proceedings of international conferences in order to undergo peer review and result verification. Believing that the perspectives of industry and standardization communities strengthen the novelty and impact of academic research, some of the important outcomes of this thesis were obtained in collaboration with industrial partners.

Throughout this thesis and subsequent chapters, references to the published works of the thesis author have been extensively utilized. The list of papers included in this dissertation is presented in the first part of the bibliography [1–7]. The first listed work [1], which has been published, serves as the main source for this thesis, as it establishes the critical goals, majority of the measurement campaigns, and optimization proposals presented in the following chapters. The second mentioned work [2], which has been recently accepted for publication, constitutes the second main source of this thesis, defining the second primary optimization proposal. Although the thesis author is not the primary author of this paper, his contributions were pivotal in defining the research topic and ensuring the practical execution of the study. His involvement was critical to shaping the paper's focus and translating conceptual ideas into actionable, practical outcomes that were foundational to the study's success. The remaining key authors' publications [3–7], published at conferences, served to disseminate preliminary outcomes. The author of this thesis played a crucial role in the development of the methodology, planning, execution, and evaluation of the experiments, as well as in the support of optimization mechanism proposals and their implementation.

Additional publications generated during doctoral studies can be found in the

other references [8–18]. Although not the primary sources for this research, they were essential for providing support and material during the completion of this research.

1.1 Motivation and Rationale

Although wireless technologies have become widespread and advanced significantly over the past decade, the demand for emerging applications continues to increase at an even faster pace. In recent years, Machine-to-Machine (M2M) applications have seen the greatest growth in communication systems, and the trend of various networks forming the Internet of Things (IoT) is gaining momentum. Conventional Human-to-Human (H2H) technologies create a network of mutually connected services that form heterogeneous networks. However, the wide variety of technologies that are suitable for different services and use different communication paradigms and frequency spectra makes the development and optimization of heterogeneous applications challenging. Consequently, there are many challenges to overcome, such as providing a scalable network with the capacity to serve billions of expected communication devices in a power-efficient, secure, and reliable manner. Heterogeneous systems are becoming increasingly popular in areas such as smart cities, smart homes, Industry 4.0, smart-grids, and wearables [19–23].

1.1.1 Cellular Technologies in the Context of IoT

With the widespread adoption of wireless connectivity, it is crucial to offer various services for novel applications with unique network requirements and Key Performance Indicators (KPIs) [19, 24]. Although wireless communication was primarily designed for H2H interactions, significant progress has been made in autonomous M2M communication scenarios that have distinct network requirements [19]. Fifth-Generation (5G) networks can facilitate these applications owing to their heterogeneity. 5G categorizes all applications and services based on their use cases, frequency utilization, and drivers, and divides them into three categories, each with different stringent requirements to suit its Network as a Service (NaaS) paradigm.

- **Enhanced Mobile Broadband (eMBB)** provides high data throughput by utilizing extended frequency carriers not only below but also above 6 GHz, offering a limited communication range. This technology is well-suited for applications such as Virtual Reality (VR) and Augmented Reality (AR) [19].
- **Ultra-Reliable Low-Latency Communication (URLLC)** is a critical component for delicate communication scenarios that require extremely low

latency, such as remote surgery or tactile internet, with stringent reliability standards of up to 99.9999 % [19, 21, 24].

- **massive Machine-Type Communication (mMTC)** refers to a type of communication that supports a scalable network with a dense deployment of End Devices (EDs), such as sensors and actuators, that transmit limited amounts of data. This communication technology has various use cases, including eHealth, smart factories, smart-grids, and Industrial IoT (IIoT). An example of this type of communication is shown in Fig 1.1 [19].

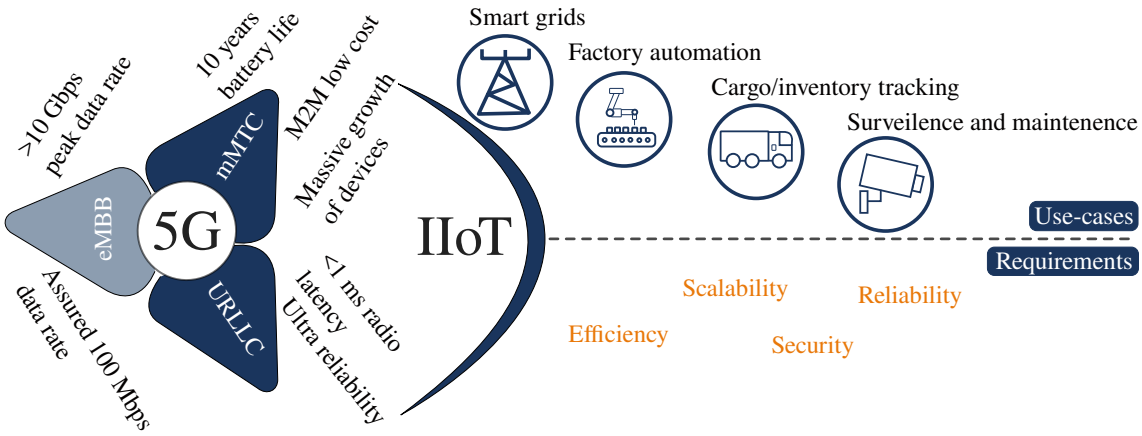


Fig. 1.1: Main directions of heterogeneous 5G with respect to IIoT diverse use-cases and requirements [19].

According to the mMTC paradigm, Low-Power Wide-Area (LPWA) technologies are a subset of end-to-end connectivity solutions that are specifically designed for industrial sector scenarios that require low-speed, power-efficient, and robust communication schemes. These technologies gained popularity in the market due to the increasing number of connected devices. Thus, this sector is a market mover [21].

LPWA Technologies

The IoT is a concept that envisions seamless global connectivity. The widespread deployment of the IoT, comprising a diverse array of connected devices, is expected to reach a massive scale in the near future. Furthermore, wireless connectivity through LPWANs is a critical component of future communication scenarios (refer to the typical LPWA infrastructure in Fig. 1.2). Predictions of Ericsson's vision estimated over 9 billion mobile subscriptions by 2030 [25], whereas authors in [23] extends even further, with up to 17 billion mobile subscriptions and over 90 billion M2M connected devices by 2030 forming IoT ecosystem. These projections raise the question of which devices will become connected and which are not yet online. The initial phase of deployment involves early adopters utilizing smart meters for gas,

water, and electricity, as well as monitoring devices such as smoke sensors and fire alarms. As a result, the number of conventional applications, including connecting utility meters to aid in distribution and billing, is expected to grow. This expansion in the industrial IoT sector, particularly in smart-grids and smart metering, creates opportunities for delay-tolerant data collection from remote devices [26–28].

The growth of connected devices that transmit M2M data is supported or managed by recent work done by the 3rd Generation Partnership Project (3GPP) development organization in Release 13, as demonstrated by Masek et al. [29]. The design of new Cellular IoT (CIoT) or Mobile IoT technologies has been re-evaluated based on the requirements of IoT vision, and new radio access technologies operating in the licensed frequency spectrum have been developed and released within a short timeframe, as reported by Martinez et al. [30]. These new technologies, known as Long Term Evolution for Machine-Type Communication (LTE-M) also know as Enhanced Machine-Type Communication (eMTC), and Narrowband IoT (NB-IoT), are designed to provide communication for IoT in the cellular infrastructure. However, even already deployed Non-3GPP solutions like Long Range Wide Area Network (LoRaWAN) contribute to this diverse field, each bringing its own set of benefits and drawbacks.

Fig. 1.2: LPWA infrastructure.

1.1.2 Emerging Challenges and Solutions for 5G IoT Wireless Systems

Owing to the significant increase in the number of MTC devices, it is crucial to develop new mechanisms that can efficiently serve a large number of devices with high power efficiency, whether they are battery-operated or zero-energy devices. Additionally, in ultra-dense deployments with up to 100 devices per cubic meter [19, 31], these devices are expected to transmit low volumes of data in a sporadic, bursty, or periodic manner. However, in traditional scenarios with simple sensors, only a small portion of the devices are currently active, with the majority in power-saving modes [19, 31, 32].

Optimizing the network capacity of such a system presents a significant challenge, and it is difficult to propose a single killer application or Radio Access Technology (RAT) as a universal solution. A more practical approach is to develop tailored solutions for specific cases based on the standardized framework of 5G technologies [19]. For instance, the authors of [33] proposed a potential solution to enhance network capacity through cell densification, edge computing, advanced modulation and coding schemes, and Non-Orthogonal Multiple Access (NOMA). These ideas have been adopted in recent years, as evidenced by [34] and 6G Vision White Papers [19].

This work delves into potential solutions for Beyond 5G (B5G) systems, particularly focusing on the current deployment challenges. Specifically, it examines network evaluation and optimization for the rollout of smart meters, mainly in the European Union (EU). The use case employs the LPWA communication technology NB-IoT or eMTC, which is considered one of the best-suited options for transmitting electricity meter data [32]. This is owing to its beneficial parameters, such as robust communication and long range. However, the main problem lies in the communication state of the device. In a typical mMTC scenario, most devices are in sleep mode, and the scenario depends on always open communication channels for possible immediate electricity meter action. As a result, it is considered a Mission-critical MTC (cMTC) [19] but with less strict requirements [21, 24]. The question remains: Will the currently deployed NB-IoT or eMTC technology be able to withstand such a use case where some devices might be in dead-zone coverage areas (under the surface, valleys, forests, deep-indoor)? If the network capacity and parameters cannot be optimized to enable sufficient service provisioning, the communication chain will fail to provide sufficient service.

An alternative perspective on this matter is to concentrate on optimizing the message transmission of the mMTC device from a slightly different angle. As previously stated, there is no single application that is considered a "killer app" for mMTC use cases, and the same applies to RATs. However, certain applications may require

a single RAT because of the device's lower complexity and reduced cost. A potential solution could be to implement smart load balancing between RATs, which could enhance overall coverage, performance, and power efficiency, and reduce operational costs [19, 35]. Another challenge in device optimization is for delay-tolerant applications. In mMTC, a low delay is not the KPI; however, some applications, such as smart meters, require a response from the device to a server within 10 s [21]. Therefore, the second question to be addressed is to explore possible ED optimizations to satisfy the aforementioned requirements. The need to maximize mMTC for smart electricity metering is heightened by the European Parliament Directive 2019/944, which unequivocally stipulates that smart metering systems will be implemented across EU member states starting from 2024 [36, 37]. This regulation imposes significant pressure on electricity distributors and telecommunication operators, who must ensure the transmission of data via cutting-edge wireless communication technologies. The smart metering use case, which involves remote reading, monitoring, and management of electricity meters, goes beyond the original goals of the IoT and Industry 4.0. As the smart metering scenario falls under the category of critical infrastructure, it requires continuous connectivity to electricity meters rather than just on-demand access. Furthermore, there are strict guidelines regarding the maximum amount of time the smart meters must be accessible, such as the 15-minute blackout recovery time for all electricity meters [38].

In conclusion, the issue of enhancing network capacity and improving the performance of EDs to meet smart metering requirements and function effectively in challenging signal environments was emphasized. The subsequent sections explore potential solutions for these problems in the context of B5G and current smart meter deployment scenarios.

1.1.3 5G IoT as an Enabler for Smart Metering Use Applications

The industry segment pertains to a newly introduced group of IoT applications that necessitates stable, secure, and permanent connectivity with UE. A prime example of such a requirement is the emerging smart metering technology [39]. This includes electricity meters that have already been deployed, whether they lack communication modules or have legacy 2G modules that allow data communication via the General Packet Radio Service (GPRS). As enhanced smart metering technology demands permanent connectivity through sustainable wireless communication, it has become a critical issue that requires dialog between mobile operators and electricity distributors. However, the current situation is the opposite, with network operators seeking effective network utilization and distributors accustomed to permanently utilizing network resources and broadcast communications. The smart metering de-

ployment strategy in the EU is even more demanding, with every deployed electricity meter for locations with an annual consumption above 6 MWh requiring a smart meter capability, or a communication module for remote metering and management, starting from 2024 [40].

In recent years, the introduction of new industrial communication scenarios, such as smart metering, has diverged significantly from the scenarios envisioned for NB-IoT technology in the past decade. The 3GPP standardization body primarily focused on covering IIoT communication scenarios for UEs performing asynchronous data transmissions, resulting in purely stochastic traffic patterns. Consequently, discussions between telecommunication operators and industry companies have commenced [24, 41].

The energy distribution sector, specifically smart-grids, is pushing the boundaries of conventional communication systems by incorporating cutting-edge Mobile IoT technologies. Photovoltaic installations have increased rapidly over the past decade, leading to the emergence of new segments. Electromobility and smart electricity meters require reliable and constant connectivity, which is achieved by the development of intelligent microenergy grids that are part of the 5G platform. It is not just the communication between Electric Vehicles (EVs) and local charging stations that is being enhanced, but also the communication between the grid and the electricity distributor. This includes verifying the user, approving the charging process, and finalizing the transaction with a payment [42]. Moreover, charging stations at home or work, such as in underground garages, require permanent connectivity with the distributor's network [43].

The use of smart meters provides novel communication approaches to light. Notably, the intervals for data collection set by electricity providers range from 5 to 15 min and can be even more frequent in specific circumstances (e.g., when managing the meter in response to a critical failure such as a power grid overload). In less-demanding cases, data can be collected every hour. It is worth noting that these nonstandard situations, which require rapid response times, are the main drivers of the need for constant connectivity in two-way communication [44–46].

Given the brand-new set of requirements for smart metering applications using Mobile IoT technologies NB-IoT and eMTC, telecommunication operators have started assessing the performance of these Mobile IoT technologies [29]. The completion of proof-of-concept trials has been impeded by two main factors: the scarcity of fully operational smart meters with integrated Mobile IoT modules, and the current level of Mobile IoT technology integration from the operators' standpoint. To address this deficiency, it has become necessary to generate initial data through simulations and develop digital twins of anticipated scenarios.

Network Capacity Planning and Optimization

Several methods have been proposed to address the ED density and capacity requirements of mMTC in B5G systems. In [19, 31–34], the authors presented similar ideas and approaches. One possible solution for extending coverage is to use newly designed coherent modulation and coding schemes that can meet the requirements for low-power transmission and robust communication. However, this method has limitations, as it can only achieve a receiver sensitivity of up to $-80\,\mathrm{dBm}$, which may not be sufficient for remote areas. In addition, 4G random-access procedures are insufficient for mMTC because of their demanding synchronization requirements and high power consumption. Therefore, new approaches with newly designed security mechanisms, such as the proposed Physical Unclonable Function (PUF) [19], should be implemented to address the security challenges of constrained devices.

The second strategy focused on the access procedure. Researchers from [19, 33] have recommended a straightforward cell-free or Grant-Free (GF) NOMA approach with no explicit connection to the base station and with the potential for joint (aggregated) reception, resulting in reduced communication overhead and radio resources. However, the authors of [34] highlighted that legacy orthogonal resource allocation is more demanding in terms of synchronization, but also decreases interference. In contrast, non-orthogonal resource allocation enhances spectrum efficiency, although it also increases interference and requires more complex decoding schemes. Consequently, researchers have proposed solutions that combine both approaches based on the specific needs of the application.

The final suggested approach is network densification. According to a study by [31], increasing the number of base stations in specific areas enhances the Uplink (UL) capacity and reduces interference, consequently lowering the necessary transmission power for communication. This development, coupled with low-power non-orthogonal receivers and the potential for zero-energy devices, could be crucial for the deployment of billions of low-power sensors in urban areas, as described in [19]. However, it should be noted that in some remote locations, the need for long-range communication may necessitate higher power consumption, owing to its associated drawbacks.

Smart-grids pose significant challenges to network capacity optimization. They require an extended communication range and latency, which necessitates that EDs maintain constant connectivity with the network in a Radio Resource Control (RRC)-connected state and poll EDs at regular intervals [47]. This leads to batch arrivals at the air interface, resulting in increased latency and violation of 5G mMTC requirements [48, 49]. The communication infrastructure's primary requirement is dictated by the reading pattern established by the Head-End System

(HES), managed by electricity distributors. Typical intervals for reading requests fall between 15 and 60 min.

Communication technology for smart meter deployments must be reliable and able to function in areas with weak signal, which is why the NB-IoT was initially selected. The communication infrastructure of electricity distributors follows traditional principles and relies on data communication protocols, such as Device Language Message Specification (DLMS) and Companion Specification for Energy Metering (COSEM), which require a secure communication path and a permanent connection that can only be provided by licensed communication technologies.

To provide a real-world example of the aforementioned requirements, the amount of data generated by an electricity meter typically ranges from 15 to 20 MB per month. This amount of data is significantly higher than the initial assumptions made in IoT scenarios, where only tens or hundreds of bytes were considered for transmission. In addition, radio conditions play a crucial role in determining the complexity of data transmission. Consequently, it was quickly realized that NB-IoT may not be suitable for handling large amounts of data traffic. As a result, several years after the introduction of NB-IoT, eMTC has emerged as a promising solution for such use cases [29, 50].

The suggested strategy deviates from the conventional mMTC deployment, necessitating the development of a novel network capacity model to offer an optimal dimensioning solution. To address this, the proposal entails creating a mathematical model that encompasses multiple scenarios to address this issue [19, 51].

Multi-RAT Approach for Enhanced Service Provisioning

An earlier section emphasized the importance of optimizing EDs to be energy-efficient and cost-effective, while still being able to communicate in signal-dead zones. This section explores potential solutions to these challenges in both 5G and emerging smart meter deployment scenarios. Machine learning (ML) methods incorporated into the device offer promising mechanisms to address this issue, enabling ED to optimize its operations on the fly [19, 52]. However, the complexity of this device poses a problem. Most constrained devices are considered; therefore, the mechanism cannot be as intricate as many deep learning algorithms [52]. One potential solution to the computational complexity is the digital twin of the device, as suggested in [19]. Another approach is to use less complex algorithms for the devices [52]. On the other hand, one proposed solution to improve communication delay, cost, and power efficiency is to implement a Multi-RAT with predictive switching between RATs, providing smart load balancing, which could offer even more seamless communication.

Another question is whether 5G systems can act as enablers for new communication use cases in mMTC scenarios in which traditional communication technologies such as 2G, 3G, and 4G cannot be achieved. A potential approach could involve a Multi-RAT solution that incorporates both 3GPP technologies utilizing licensed spectrum and non-3GPP operating in license-exempt spectrum. Most LPWAN technologies, such as LoRaWAN, IEEE 802.11ah, and IEEE802.11p, are classified as non-3GPP RATs. To facilitate compatibility between 3GPP systems (NB-IoT and eMTC) and non-3GPP systems, the Non-3GPP Inter Working Function (N3IWF) was developed, enabling access to the 5G Core Network (CN) [53].

The market offers various LPWAN technologies with unique features, leading to the potential for multifunctional devices such as sensors, actuators, and aggregation gateways to satisfy the demands of next-generation IoT devices operating in emerging 5G and B5G networks [54]. Although soldering multiple RAT transceivers on a single board has been explored previously [55], current chipset manufacturers offer short- and long-range communication capabilities [56, 57]. However, these chipsets do not enable integration of recent LPWAN technologies (3GPP and non-3GPP) on a single chip, which is necessary for future IoT devices.

LPWAN technologies, such as LoRaWAN, Sigfox, and Weightless, provide an excellent communication range. However, these methods have other limitations. For instance, license-exempt technologies face strict duty-cycle regulations, which can lead to higher probabilities of outages owing to interference. In comparison, licensed technologies like NB-IoT and eMTC do not face such limitations, but they come with their own set of challenges. The most crucial concern is striking the right balance between required data transmission rates and energy usage. According to Mikhaylov et al. [19, 58], achieving an optimal balance between these two factors is crucial for successful deployment of LPWAN technologies.

To attain this balance, using multiple radio interfaces on a single ED to enhance capacity through load balancing and to improve network availability is an option [58]. Nevertheless, the question arises as to which RAT should be used for data transmission at a specific moment.

Unexpected on-demand readings and remote management tasks can occur due to severe system failures, such as power grid overloads. In these exceptional circumstances, electricity meters (EDs) and HES require both lower latency Downlink (DL) communication and permanent two-way connectivity to function effectively. [44–46]. To enhance reliability and streamline ED design, solutions with multiple LPWAN interfaces have been proposed [35, 59]. However, the challenge of selecting the optimal RAT arises because operating over multiple interfaces necessitates minimizing the latency at the air interface. However, reliance on well-established 3GPP infrastructure seems to be the best approach.

1.1.4 Thesis Motivation and Goals

Research aim of this work concerns optimizing the capacity of the mMTC Mobile IoT network and enhancing end devices, which aligns well with current industry trends. The global Mobile IoT connections are projected to significantly increase, with a substantial growth driven by the adoption of cellular connectivity across various sectors such as utilities, automotive, industrial, retail, and healthcare. This growth emphasizes the need for research into improving the performance and efficiency of IoT technologies, including LPWA technologies like NB-IoT and eMTC, which are pivotal for applications like smart metering.

This research focuses on performance characteristics, implementation challenges, and specific enhancements required for modern IoT applications, all of which are critical in the context of the ongoing development and deployment of 5G and future 6G networks. This alignment suggests that this research could contribute valuable insights and solutions to the ongoing development of these technologies.

1.2 Research Objectives and Methodology

To achieve the goals of this doctoral thesis, *Research Objective* (RO) has been given with four corresponding *Research Questions* (RQx).

RO: Research of ambitious field targeting optimizations of the mMTC Mobile IoT network capacity and enhancements of the end device behavior in case of emerging smart-grid services.

This study aims to explore and advance the optimization of mMTC Mobile IoT network capacity and the enhancement of end devices in emerging IoT use case of smart-grids. Central to this objective is the investigation into the deployment feasibility of LPWA technologies, particularly NB-IoT and eMTC, in stringent smart-grid applications, which presents more demanding requirements for transmission delay and type of traffic than initially anticipated.

- **RQ 1** : *What are the specific performance characteristics and limitations of current cellular technologies in emerging use cases, such as smart-grids, and where do the most significant gaps exist?*
  Obtaining a comprehensive understanding of the practical performance of currently deployed technologies, such as LPWA technologies NB-IoT, and eMTC, is crucial in determining their suitability for various applications. This question aims to investigate which use cases are feasible and whether there is a necessity for optimization to ensure their successful implementation. Smart-grids, which require permanent connectivity and stochastic data traffic, are mentioned as a key use-case where these technologies are applicable.

- **RQ 2**: *How do future smart-grid applications evolve in terms of their requirements and deployments, and how do NB-IoT and eMTC perform in fulfilling these demands? What are the limitations of these technologies in this regard?*
  This question delves into the suitability of NB-IoT and eMTC technologies for smart-grids applications, which are becoming increasingly sophisticated. It seeks to assess how well both technologies adapts to the specific needs of smart-grid, such as long-range communication, reliable data transmission and delay tolerances. The exploration includes an analysis of the inherent limitations of NB-IoT and eMTC in this context, such as bandwidth constraints, latency issues, or any potential challenges in scalability and interoperability with existing infrastructures. Additionally, the study examines in-depth the current status of smart meter installations worldwide and the European Union's regulations and specifications for smart metering systems. This understanding will help in identifying areas where both technologies may need enhancement to effectively serve the demands of smart-grids in the IoT landscape.
- **RQ 3**: *What are the effective strategies to improve the performance of cellular technologies NB-IoT to meet the continuous connectivity, transmission reliability, and communication delay demands of smart-grids applications?*
  This inquiry focuses on identifying and developing strategies to enhance the performance of NB-IoT, predominant LPWA technologies, specifically for smart-grids applications. It involves exploring various technical and network optimization strategies, such as advanced optimized traffic management algorithm considering coexistence of regular sensoric traffic with stochastic traffic of smart-grid applications. The goal is to refine this technology to ensure effective provision of continuous connectivity, and communication delay. The research will also consider how these enhancements can be implemented within the existing cellular infrastructure, evaluating their feasibility, and potential impact on the overall network performance.
- **RQ 4** : *What specific enhancements in Mobile IoT technologies are necessary to meet the unique requirements of contemporary smart-grid systems if synergistically combined in Multi-RAT for optimal performance?*
  The research inquiry aims to explore the feasibility of implementing Multi-RAT through the use of NB-IoT and eMTC technologies to address emerging smart-grid applications. Specifically, the investigation seeks to explore how to optimally distribute the workload between two technologies, designating one as the primary system and the other as a backup. The goal is to determine the optimal number of devices to employ without compromising the smart-grid requirements, thereby enhancing communication reliability. This investigation focuses on striking the right balance between these technologies to achieve

optimal performance and redundancy.

To achieve the objectives outlined in this dissertation, a general strategy was implemented to address the aforementioned questions. This research will utilize a combination of literature review utilizing both academic research to observe future trends and industry white papers, and outputs of research projects with industry partners to observe current deployments and near future trends. This research also implies empirical measurements to identify or confirm research gaps pinpointed by the literature review and simulation-based analysis utilizing inputs from measurements to address the research questions. The approach includes:

1. **Literature Review:** A thorough examination of prior research centered on mMTC, with a particular emphasis on NB-IoT and eMTC, has been conducted. This review evaluates broader patterns in the realm of IoT, while also concentrating on the applications of these selected technologies in the context of smart-grids. The review encompasses both potential enhancements to these technologies as well as their current deployment and deployment guidance. Furthermore, the evaluation of smart-grid deployments and requirements from European directive, energy distributors and telecommunication operators is conducted to establish testing and optimization thresholds for the chosen technologies.
2. **Empirical Measurements:** Undertaking performance evaluation campaigns to gauge the capabilities and limitations of NB-IoT and eMTC in real-world situations. This involves assessments of both technologies under ideal conditions, as well as a stress test for NB-IoT, simulating heavily congested radio networks. Conducting such extensive measurements necessitates elaborate infrastructure that encompasses the hardware and software design of remote servers and prototype EDs.
3. **Simulation-based Analysis:** Developing mathematical models and simulation tools to analyze and optimize network performance and end-device behavior in smart-grid applications is a critical aspect of the research. This includes the development of optimization strategies that consider Multi-RAT approaches that can effectively balance the load between primary and secondary wireless technologies. The second approach involves the evaluation of a Single-RAT smart-grid application that must take into account the coexistence of two different types of devices. The first type of device is represented by mid-range smart meters that have strict connectivity requirements, while the second type is represented by traditional battery-operated sensors that have sporadic and periodic communication patterns. It is essential to optimize the communication patterns of these devices to achieve good performance and meet the communication requirements of the smart-grid.

1.3 Key Contributions and Novelty

The main *Contributions* (Cx) proving the novelty of the thesis are organized in three groups mapped onto the four **RQs** in **RO** formulated in Section 1.2.

C1. 5G IoT Applicability Assessment for Smart-Grids

A detailed overview of Mobile IoT technologies, NB-IoT and eMTC, focusing on their performance and evolution across 3GPP Releases. This analysis identifies potential bottlenecks for smart-grid use cases, aiding integrators in assessing present and future applicability. The smart-grid use case is further examined in the context of current and anticipated global deployments, with an emphasis on EU regulations and future implementation trends. The key contribution of this description (referring to in **RQ1**) is a comprehensive elaboration of both the use case and technologies for service provisioning, which highlights their applicability and possible constraints. The innovative contribution is supported by exact requirements, e.g. guaranteed 11 kbps of data throughput, 20 MB of data transmitted per day or a communication delay threshold of 2 s, which place a significant strain on both technologies in the current deployments. Thus, the novel optimizations and enhancements in future 3GPP releases should significantly improve the service provisioning of both technologies on either the operator and the customer sides.

C2. Performance Evaluation of Current mMTC Technologies

To address the performance of NB-IoT and eMTC technologies with regard to smart-grid requirements, as stated in **RQ2**, this thesis adds to **C2** by carrying out a measurement campaign and contrasting NB-IoT (LTE Cat NB2) and eMTC (LTE Cat M1). Measurement scenarios were designed to identify limitations and provide unique statistical data as a basis for novel optimizations of the NB-IoT for data-intensive applications and to provide the first insights into the performance of eMTC, an alternative technology. The key contributions of **C2** are:

- Unique report and analysis of communication throughput, delay, and jitter for the deployed network configuration across the whole communication budget of NB-IoT and eMTC. This report provides novel recommendations and statistics for service fine-tuning for time-sensitive data services.
- The results of the experiment reveal novel insight that both NB-IoT and eMTC can provide adequate services to meet the requirements of the smart-grid. However only in perfect radio conditions and not congested the network. By this perspective eMTC is a better choice for smart-grid applications owing to its higher data transfer rate, lower latency, and communication delay. NB-IoT is more suitable as a backup connection when there is poor coverage, owing to its longer communication budget. However, the second scenario emulating heavy load on radio link utilizing 200 EDs under single cell confirms that lower

data transfer rates and a narrower frequency bandwidth of NB-IoT results in Random Access (RA) congestion, leading to significant communication delays. The use of NB-IoT in such case without optimization is proven impractical.

C3. 5G-IoT Radio Interface Communication Delay Enhancements

To enhance the communication efficiency of congested radio interfaces in mMTC technologies by decreasing communication delay, several proposals have been put forth to address **RQ3**, based on results from measurement campaign. Initially, we examined the influence of the new smart meter traffic coexisting with conventional sensors, thereby merging both stochastic and regular traffic types. To achieve this, a comprehensive and novel mathematical model was developed based on data from a measurement campaign. From this model, a simplified version was devised to address specific requirements for optimal message dissemination within 15-minute interval. Finally, this study explores the future optimization of communication procedures based on the Early Data Transmission (EDT) mechanism outlined in 3GPP Release 15. While the first two concepts are immediately applicable as they are not dependent on future 3GPP releases, the latter is subject to such constraints.

- **C3.1 Comprehensive Model for NB-IoT Smart Metering Traffic**
  The novel models developed in this work can be utilized to assess the resources required to be allocated to the NB-IoT system to fulfill specified delay guarantees of merged stochastic and regular traffic types. The key contributions of **C3.2** are summarized as follows:
  - A innovative mathematical model that captures the behavior of the system when conventional and new types of EDs requiring consistent connectivity coexist and share the NB-IoT air interface. By employing state aggregation method and its corresponding numerical algorithm, a simplification of the model enables evaluating the average communication delay for the considered stochastic and regular traffic types.
  - The unique numerical results reveal the following: (i) The NB-IoT system remains stable for 72×10^4 conventional and 9×10^3 permanently connected EDs in the system; (ii) the required 2 s delay performance for permanently connected EDs is not met when there are more than 100 such UEs in the system, and (iii) by increasing resources, for example, PRBs to three, the latency requirements are met for 240 EDs.
- **C3.2 Optimal Spreading of NB-IoT Smart Metering Traffic**
  To provide a simplified and more practical model, a straightforward and innovative method is proposed to distribute message transmission time slots for regular traffic in the presence of conventional random transmissions. We calculated the average message transmission delay for both types of traffic as a function of the system parameters. By employing both analytical and simu-

lation techniques, we determine the optimal spreading interval and assess the overall performance of the system. The key contributions of **C3.1** are:

  - The widening of communication intervals aims to enhance the delay performance of both considered traffic types in the NB-IoT deployment.
  - The unique optimal spreading interval and associated mean message transmission delay for NB-IoT network can be estimated using a mathematical model parameterized by the results of a measurement campaign.
  - Novel network capacity planning for the NB-IoT is based on the observation that the optimal spreading interval and mean message delay of regular traffic are directly proportional to the number of EDs.

- **C3.3 Future Optimization Strategies**
  This study examines the NB-IoT features introduced by 3GPP in Release 13 as well as Release 15's EDT for smart-grid scenarios. The innovative simulation results demonstrate that EDT is a powerful feature for reducing transmission delays by more than 50% and improving spectrum efficiency with reduced overhead. The significant decrease in latency was especially for smaller messages (less than 97 B in UL and 57 B in DL), and our novel results confirm that a NB-IoT (NB1) network can manage 1,000 smart meters per single base station communicating in a 15 min window.

C4. 5G-IoT Radio interface communication reliability enhancements

To improve the dependability of mMTC communication technologies and address **RQ4**, a novel theoretical model and corresponding algorithm based on data from measurements were developed to balance the traffic load between NB-IoT and eMTC. The key contributions of **C4** are summarized as follows:

- The innovative outcomes demonstrate that the optimized utilization of Multi-RAT EDs significantly increases the number of EDs operating in the polling-based mode. Already for 500 EDs using single eMTC technology, the mean access delay exceeds two seconds, which conflicts with the minimum requirements of smart-grid applications. Conversely, Multi-RAT EDs running the developed algorithm enhance the service capacity by up to six times (up to 3,000 EDs) while still adhering to the two-second latency budget.
- A unique observation that (i) in the RRC connected state, within standard NB-IoT and eMTC parameter values, an optimal load balance point always exists between the two technologies. (ii) For EDs with purely random arrivals, which are typical of those that go through idle and connected states, eMTC technology is always the preferred choice and NB-IoT should only be used when eMTC is fully occupied. (iii) By utilizing the optimized usage of Multi-RAT EDs, it is possible to support a higher number of EDs in polling-based mode, up to six times more than with a single NB-IoT.

2 Cellular Technologies for Smart-Grid Applications

This chapter commences with a comprehensive review of mobile system history, tracing the developmental trajectory from outdated mobile systems to the latest technological advancements in 5G, while also touching upon future innovations. The initial portion of the chapter primarily emphasizes key technological improvements with highlights of the M2M communication direction and is subsequently expanded upon with a technical overview and roadmap of the Mobile IoT technologies of NB-IoT and eMTC. The final section of the chapter is dedicated to the deployment guidelines of both technologies and a comparison of their current forms up to 3GPP Release 17. For the purpose of this thesis, in terms of performance evaluation, the focus of the last part is on deployment in 3GPP Release 13 and 14, with a brief mention of the anticipated improvements in Release 15, which not only adhere to the 3GPP's definition in its entirety but also exclude features that lack practical applicability.

2.1 From Legacy Towards Fullscale 5G

The development of next-generation heterogeneous systems for 5G and B5G necessitates the integration of various technologies that cater to specific service requirements. One way to classify these technologies is based on their frequency spectrum. The standardized division includes the Frequency Range (FR) 1, which encompasses low frequencies (sub-GHz) and mid frequencies (1 –7 GHz), and the FR2, which features high frequencies (above 7 GHz). Frequencies above 7 GHz are well-suited for precise positioning, short-to-mid-range data transmissions due to high propagation losses, high data rates due to high-order modulations, and up to 400 MHz transmission Bandwidth (BW). These properties make them ideal for eMBB applications. However, the same frequencies do not provide sufficient capabilities for Mobile IoT use cases outlined in this thesis, as they necessitate high data throughput of less reliable and robust communication schemes suitable for data streaming. The mid-band frequencies, which were employed in legacy mobile systems of 2G to 4G, serve as a middle ground between coverage and data throughput. Nevertheless, industry requirements are moving towards end-to-end connectivity, and sub-GHz frequencies are the most suitable for these purposes. Technologies like the LPWA sub-group from the mMTC group (Mobile IoT), which utilize sub-GHz frequencies, gained momentum in the market in the last decade. Although this work focuses on long-range technologies with end-to-end connectivity provisioning and predominantly star

topology deployment, it is essential to note that for smart home scenarios or industry scenarios with aggregation points, short/mid-range technologies such as WiFi, BLE, Zigbee, Matter, or WiSun can be considered [19, 22].

As previously mentioned, the utilization of sub-7 GHz or even sub-GHz frequency spectrum presents undeniable advantages for mMTC. To provide a comprehensive overview of the evolution of wireless mobile networks, following text delves into the progression from legacy human-centric 2G systems to heterogeneous 5G networks. Although the description includes all major enhancements in 5G systems including those utilizing above 6 GHz frequency spectrum, further elaboration on these enhancements is beyond the scope of this study. Instead, the focus is on a specific group of sub-GHz Mobile IoT or LPWA technologies and their associated features.

Reflections on the Historical Development of Mobile Systems

Today, the market for the IoT, particularly in mobile smart metering, is still largely occupied by legacy 2G technologies, such as Global System for Mobile Communication (GSM) extended through the GPRS with data rates of up to tens of kilobits per second (kbps) and later introduced Enhanced Data rates for GSM Evolution (EDGE) technology, which utilizes eight-state Phase Shift Keying (PSK), stepping up data rates of up to 200 kbps. This technology is better suited to the requirements of smart metering compared to 4G in terms of complexity, cost, communication robustness, and data throughput but lacks the crucial security and other improvements. As a result, 2G IoT is on the decline, with the forecast of being completely phased out within the next decade [60–62].

The fourth generation of mobile networks, commonly referred to as 4G, remains the most widely utilized technology today. This generation of networks provides a unified telecommunications service provisioning, which is primarily based on the Internet Protocol (IP) and is packet-oriented. In this system, the Base Station (BS) controllers have been removed from the infrastructure, and a new access solution, Orthogonal Frequency-Division Multiple Access (OFDMA), has been implemented. The improved infrastructure, which consists of only one element on the access network side (Evolved NodeB or eNodeB), has resulted in low latency of tens of milliseconds, which is suitable for real-time services [62–64].

The foundation of LTE was established in 3GPP Release 8; however, the true 4G network was first introduced in Release 10 (LTE-Advanced) with a data rate of over 1 Gbps. This was achieved through the use of Carrier Aggregation (CA) with subsequent optimization in Release 11. Additional enhancements were made in the form of Multiple-Input Multiple-Output (MIMO) operations in the UL/DL, new frequency bands, or self-optimizing networks to improve human-centric service

provisioning for mobile broadband [62–64].

When it comes to smart metering, the current trend with legacy 4G technologies offers LTE categories 1 or 1-bis, which are capable of meeting smart meter requirements. These technologies are built on well-established infrastructure and utilize the low cost of category 1 devices. However, upcoming Mobile IoT or 5G IoT standards significantly surpass these legacy 4G technologies in terms of long-term support, the number of connected devices, energy efficiency, communication budget, and more [62–64].

Early Implementations of 5G Body

Prior to the advent of 5G technology, mobile networks were primarily developed for human-centered communication. However, the shift toward M2M communication was initiated with the release of 3GPP Release 10. Despite the introduction of this technology, the mobile network standards defined by 3GPP are not currently available. The necessity for power-efficient, and dependable low-data-rate transmission at low frequencies has resulted in the emergence of a multitude of long-range wireless technologies. As a consequence, with the emergence of 5G and LPWANs, these legacy technologies have lost momentum and are now only utilized in specific situations [65].

In the 3GPP Release 13, innovative techniques and enhancements were introduced to facilitate mMTC services made possible by cutting-edge technologies such as NB-IoT and eMTC. These enhancements included power-saving operational modes, robust communication measures using repetitions and robust modulation and coding schemes, and a streamlined protocol stack that is still interoperable with and deployable to legacy 4G systems. Moreover, Release 14 continues to concentrate on optimizing mMTC for future 5G integration [22].

The major change in priority came in Release 15, where the focus was given to another part of 5G verticals, namely eMBB. This was issued by introducing Non-Standalone (NSA) 5G with current deployments utilizing frequency carriers below 6 GHz (1.8 to 3.6 GHz) and bandwidth up to 100 MHz enabling increased data throughput. The NSA deployment indicates the novelty is in the radio access network part, but the core network and signaling are delivered by 4G network. Going towards releases from Release 15, the trend goes toward eMBB evolution by introducing pure standalone 5G utilizing frequency carriers in frequency spectrum up to several tens of GHz (up to 52.6 GHz) known as mmWaves. Release 15 also introduces basic New Radio (NR) URLLC support further enhanced in following releases. This could be an enabler for use-cases such as VR or AR enabled by small-time units, edge-computing, and network slicing mechanisms enabling desired

features of extremely low latency and superior reliability [19, 23, 26, 54, 56, 65].

Release 16 shifts focus from primarily eMBB toward the industrial use-cases. Here is introduced integration of mMTC technologies NB-IoT and eMTC in 5G core, enabling the 5G IoT with further IIoT enhancements. Release 16 also brings desired Non-Public Networks (NPNs) that aims for private company infrastructures replacing or competing with WiFi 7 or UWB. However full 5G NPN package already in Release 16 combines strong side of both mentioned competitors and outperforms them with additional enhancements such as mobility, security, network robustness, data throughput, low latency etc. As the hight frequencies fo 5G, FR2 especially, which brings environmental challenges in terms of radio propagation, Release 16 introduced new type of wireless distributed infrastructure Integrated Access Backhaul (IAB) to help overcome this issue [19, 23, 54, 56, 65].

Advancements in 5G Deployments and the Transition Toward 6G

Later Release 17 introduces three crucial innovations. Beside eMBB enhancements in form of use of FR2 frequency spectrum moving up to 71 GHz fo mmWaves introduces new major communication technology NR-Light known as Reduced Capability or RedCap, which serves as middle ground and universal solution for all 5G use-cases since it draws specifics from all three 5G pillars eMBB, URLLC and mMTC. This technology is very capable for all mid-level use-cases such as advanced smart meters and wearables. Third major enhancement of Release 17 is enabling affordable and less complex mobile to satellite communication in form of Non-Terrestrial Networks (NTNs) for both high throughput NR and robust IoT (eMTC/NB-IoT). During the completition of this thesis, Release 18 was the last with complete definition introducing beside enhancements of previous releases the first AI/ML features to physical layer of 5G RAN together with also support of NR UAVs and enhancements for Security and Public Safety (NSPS) vertical. With Release 18, 3GPP body enters the 5G Advanced state. Upcoming Release 19 is in its initial state taking actions to prepare ground for 6G coming in 3GPP Release 21 [19, 23, 54, 56, 65].

Overall continuous 5G evolution based on standardization of 3GPP releases is as illustrated in Fig. 2.1. The next steps in the evolution in 3GPP body from 3GPP Release 19 are expected to be further enhancements of NTNs and next phase of AI/ML models adoption in the 5G advance networks. Early studies of this release also indicates support of metaverse in enhanced Mixed Reality (XR) use-cases or NR-enabled wake-up-signal, wake-up-receivers and ambiently powered IoT devices further going toward green IoT and sustainable mobile networks [19, 23, 54, 56, 65].

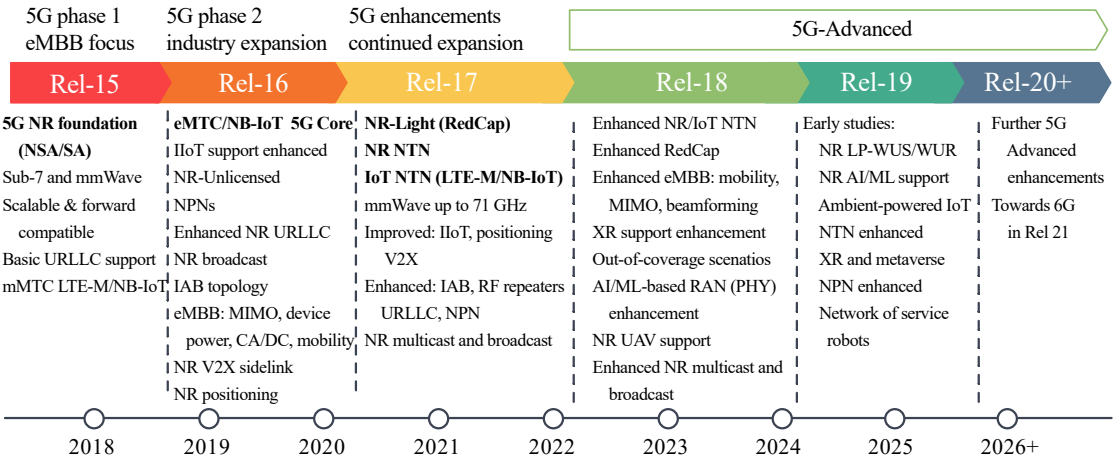


Fig. 2.1: 5G and beyond evolution from the perspective of 3GPP releases [66].

2.2 Enhancements and Evolution of 5G IoT

As the notion of IoT shapes the idea of MTC for low-power end devices, such as sensors, servers, and data processing systems, it has become evident that traditional wired and wireless communication technologies, such as 2G, 3G, and 4G, which focus primarily on mobile broadband, cannot handle data transmissions with the necessary Quality of Service (QoS) [67–69]. The development of these systems towards eMBB aims to meet the principal performance requirements as defined by the ITU-R M.2410, which specifies stable data rates in hundreds of Mbps with peak rates in tens of Gbps or communication delays in tens of ms [24].

In contrast, the primary requirements for IoT and Industry 4.0 involve reliable communication over an extended range rather than high-speed transmission. The goal is to achieve a dense coverage of 1 million UEs per km^2, with low communication loads from devices that are mostly in low-power consumption states [24]. Consequently, a narrow group of technologies collectively known as LPWA technologies has emerged in the past decade and gained popularity owing to their ability to transmit small data blocks over long distances with minimal power consumption, particularly in the context of IoT and Industry 4.0 [67–69].

LPWA technologies aim to deliver cost-effective, energy-efficient, and scalable solutions for MTC or mMTC. Current LPWA technologies fall into two main categories based on their frequency spectrum usage: those requiring licenses and those that are license-free. The former includes 3GPP NB-IoT and eMTC technologies, which make up the Mobile IoT group. The latter encompasses various technologies operating in unlicensed frequency bands, with LoRaWAN being the most notable example. Because license-exempt technologies share frequency spectrum with other unsynchronized systems, there is a potential for interference occurrence [70]. How-

ever, this work primarily focuses on the Mobile IoT technologies due to their industry momentum.

NB-IoT and eMTC, two prominent LPWA technologies, have become essential components of the 5G cellular systems. Initially defined by the 3GPP in Release 13 and integrated into the LTE infrastructure, these technologies were introduced as alternatives to the existing long-range communication technologies, such as LoRaWAN. Both technologies are defined based on the 3GPP LTE standard, with additional features specifically tailored to meet the requirements of the industrial sector and demanding data transmissions. Unlike technologies such as LoRaWAN, and Wireless M-BUS, which use license-exempt frequency spectrums, the 3GPP cellular technologies for IoT and Industry 4.0 employ dedicated (licensed) frequency spectrum, similar to those used in 4G cellular systems [71, 72].

The official integration of both technologies into the 5G ecosystem was defined in 3GPP Release 15, with real integration into 5G infrastructure occurring in Release 16. With each new 3GPP Release, the array of features continues to expand, driven by anticipated communication scenarios that demand simple, performance-constrained, battery-operated devices. Mobile IoT technologies enable various applications, including remote monitoring and measurement, asset location tracking, real-time surveillance, and smart metering. These use cases represent just a few examples of the possibilities opened up by these advancements [73, 74].

2.2.1 Evolution of Narrowband Internet of Things

Compared to conventional 4G (LTE) communication systems designed for human use, NB-IoT, defined in Release 13, offers several game-changing features. These include extended coverage, enhanced power-saving modes, and a more limited set of functionalities. Thanks to these features, NB-IoT can connect devices in difficult positions and enable long battery life while reducing device complexity. By 2018, NB-IoT had become the preferred cellular communication technology for remote metering across Europe as the IoT market continued to evolve [62, 75, 76].

NB-IoT is specifically engineered to meet LPWAN requirements, including enhanced coverage of up to 164 dB (20 dB more than traditional LTE systems), extended battery lifespan exceeding ten years, and communication latency under 10 seconds. Leveraging LTE numerology, NB-IoT can be implemented within the same frequency band as existing LTE, utilizing a 180 kHz spectrum segment (single Physical Resource Block (PRB)) in in-band mode. This configuration divides the spectrum segment into channels with either 3.75 kHz (UL only) or 15 kHz (both UL and DL) spacing, balancing extensive coverage and high UL capacity within the constraints of narrow spectrum deployment. Additionally, NB-IoT can be deployed

in the LTE guard band or as a standalone channel, as illustrated in Fig. 2.2 [77, 78].

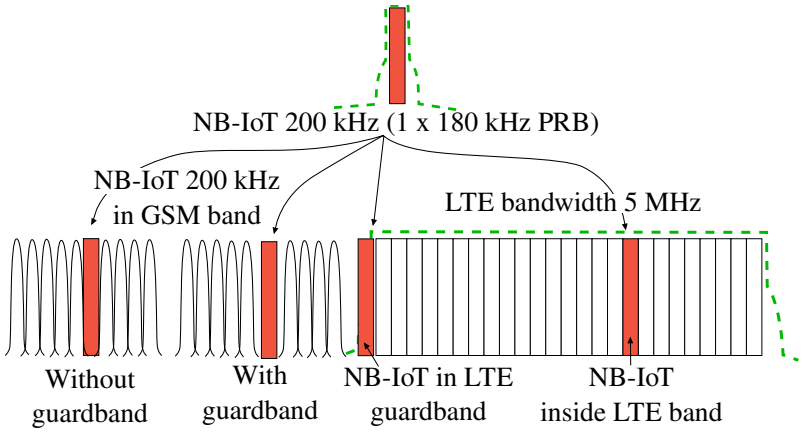


Fig. 2.2: The possible NB-IoT deployments.

Coverage Enhancements

The main objective of LPWAN is to extend coverage, which is mainly achieved in the case of NB-IoT by utilizing low-rate Modulation Coding Schemes (MCS) and repetitions. Notably, the Narrowband Physical Random Access Channel (NPRACH) that employs single-tone transmission with a 3.75 kHz subcarrier spacing and up to 128 repetitions. Additionally, formats 0 and 1, which utilize 66.67 μs and 266.7 μs cyclic prefixes, respectively, allow for cell radii of up to 10 km and 40 km. Furthermore, UL transmissions supported by a Narrowband Physical UL Shared Channel (NPUSCH) may also use up to 128 repetitions. NPUSCH can use 3.75 or 15 kHz spacing with $\pi/2$-Binary-Phase Shift Keying (BPSK) or $\pi/4$-Quadrature Phase-Shift Keying (QPSK) modulated signals to reduce the Peak-to-Average Power Ratio (PAPR) in single-tone mode. Another option is to employ a 15 kHz spacing with QPSK modulation in multi-tone mode. Moreover, NPUSCH supports Transport Blocks (TB) of up to 1,000 bits in Release 13. It also uses a Demodulation Reference Signal (DMRS) specifically for channel estimation in the frequency domain. In contrast to the UL, which comprises NPRACH and NPUSCH signals, the DL frame structure incorporates as many as eight distinct signals/channels. These define 15 kHz spacing and utilize QPSK modulation, with a maximum TB size of 680 bits [68], [69], [79].

As previously stated, the fundamental concept of extended coverage revolves around repetitions. However, the broader range comes with an increased latency, which can quickly surpass the 10 s limit when the message exceeds the actual TB size. To address this issue, NB-IoT introduces three Enhanced Coverage Level (ECL)

classes that define the TB sizes and maximum number of repetitions for each channel. These classes are illustrated in Fig. 2.3, where the system moves to the appropriate ECL based on threshold values. The RSRP and SINR values determine these threshold levels, and it is important to note that the ECL values can be adjusted according to network operator requirements, as shown in the example provided [68].

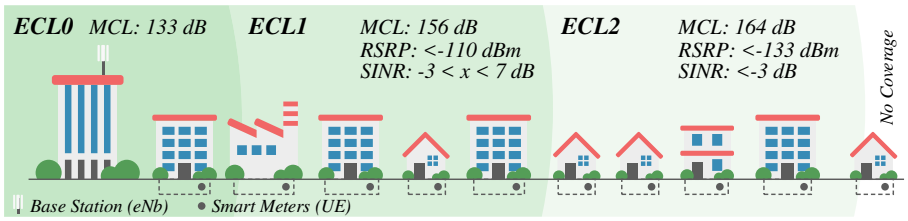


Fig. 2.3: Coverage enhancement levels [3].

Power Consumption Enhancements

The design of NB-IoT protocol is based on LTE and its procedures, which were substantially simplified, but the signaling in early 3GPP Releases 13 and 14 with full random access procedure and not relaxed cell measurements still remain rather power-hungry. However, improved energy efficiency is ensured by power saving modes defined in 3GPP Releases 12 and 13 that allow extending the paging period or even completely turn off the radio interface [22].

The latter mentioned mechanism is Power Saving Mode (PSM) in which the device turns off the radio part and leaves only RTC running while entering deep-sleep mode with a duration defined by timer T3412. The device is still registered to the mobile network, but it must periodically (when the timer expires) report to BS via Tracking Area Update (TAU) message. The device is subsequently reachable only in a short periodical interval of activity defined by timer T3324 [80, 81].

The second mechanism is extended Discontinuous Reception (eDRX) which allows the device to listen for incoming messages in pre-specified paging intervals ranging from 2.56 s up to almost 3 hours. Nevertheless, connecting to the network or sending data leads to increased NB-IoT radio power consumption due to extensive signaling and high transmit power (up to 23 dBm for UE power class 3). This can result in significant current draw, potentially reaching 300 mA at peak for a 3.3 V system. To address this issue, numerous optimizations and new features have been introduced to NB-IoT in recent years, including the EDT mechanism implemented in Release 15 [19, 80, 81].

NB-IoT procedure with illustrated power consumption according to the 3GPP Release 13 is depicted in Fig 2.4, which completes the remaining NB-IoT channels

utilized before the data transmission. The Narrowband Reference Signal (NRS) is utilized for assessing download channel conditions, which is essential for measuring signal strength and quality. Additionally, Narrowband Primary Synchronization Signal (NPSS) and Narrowband Secondary Synchronization Signal (NSSS) are employed to align UE with the NB-IoT cell. These signals enable UE to identify the Physical Cell Identity (PCI) through a synchronization algorithm during initial acquisition, without prior knowledge of deployment mode, based on an 80 ms repetition interval in specific subframes. The Master Information Block (MIB) is broadcast every 10 ms via the Narrowband Physical Broadcast Channel (NPBCH). Meanwhile, the System Information Block (SIB), along with UL grant and DL scheduling information, is transmitted through a Narrowband Physical Download Control Channel (NPDCCH) with a fundamental Transmission Time Interval (TTI) of 1 ms. The actual unicast data from eNB to UE is conveyed via a Narrowband Physical DL Shared Channel (NPDSCH). Upper layer packets are divided into one or more Transport Blocks (TBs) and transmitted sequentially [68, 79, 82, 83].

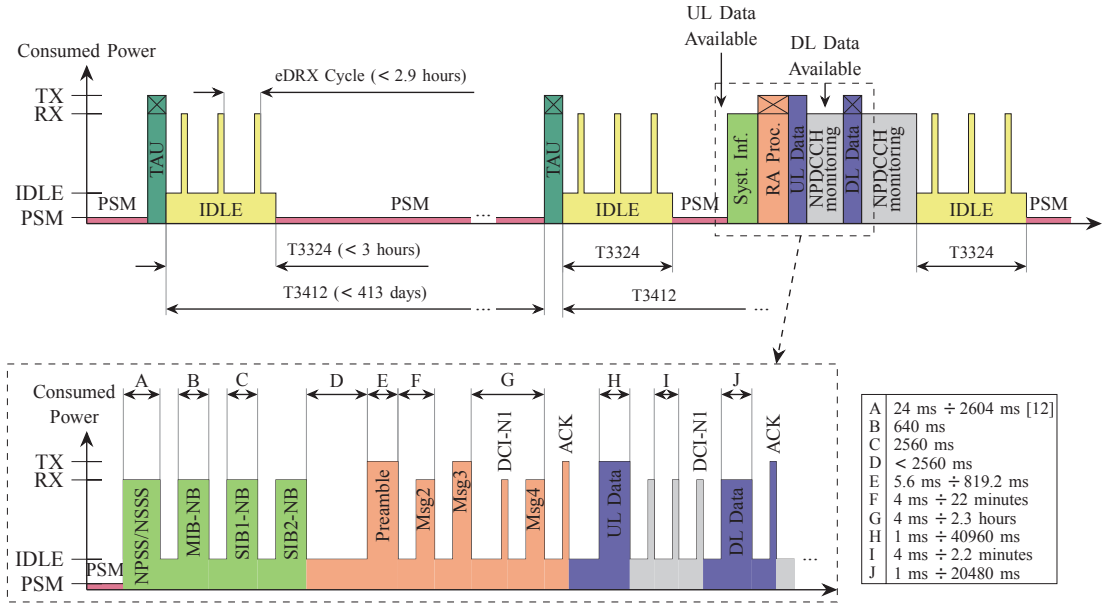


Fig. 2.4: Operational modes for NB-IoT (3GPP Release 13): (i) TAU, (ii) IDLE with (e)DRX, (iii) PSM and (iv) data transmission [29].

Data Transmission Enhancements

To achieve the objective of serving numerous UEs through a single eNB, Mobile IoT necessitates reducing signal overheads, particularly in the RAN. In order to achieve this, Release 13 introduced Mobile IoT Evolved Packet System (EPS) optimization in both the Control Plane (CP) and User Plane (UP). Both optimizations included

additional improvements in the form of an Attach procedure without a PDN context, allowing UEs to remain in an inactive connection state for an extended period with minimal data communication in congested RAN areas [62, 83, 84].

User Plane Optimization: The *RRC Resume Procedure*, a feature of NB-IoT technology, is implemented following an initial RRC connection that sets up radio bearers and the Access Stratum (AS) security context. When the User Equipment (UE) is in the *RRC Idle* state, it stores the context at the UE, eNB, and Mobility Management Entity (MME). This stored context is then utilized to reestablish the connection when new traffic arises, as illustrated in Fig. 2.5. For context restoration, the UE must furnish the Resume ID employed by the eNB to access the stored UE data. This mechanism enables the UE to circumvent AS security setup and RRC reconfiguration during each data transmission. Data transfer optimization is facilitated through the utilization of a user plane, which permits subsequent communication via established data pathways. Thus, the *RRC Resume Procedure* can be used for both short- and long-duration data transactions [22, 83, 84].

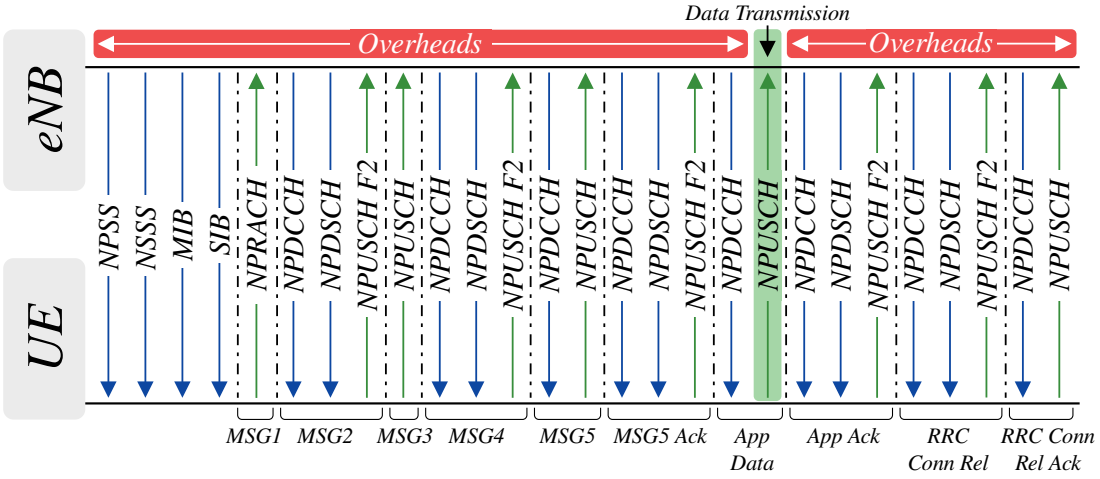


Fig. 2.5: NB-IoT: User Plane Optimization [3].

Control Plane Optimization: This optimization enables the UE to employ CP to transmit data packets packaged in Non-Access Stratum (NAS) signaling messages to the MME. Consequently, the UE bypasses AS security configuration and UP bearer establishment (Fig. 2.6). This approach is particularly advantageous for brief data transfers. In addition, to minimize communication overheads and consequently reduce power consumption, the introduction of Non-IP Data Delivery (NIDD) over CP has been implemented [76, 83, 84].

Last to mention is that NB-IoT as specified by 3GPP Release 13 provides limited mobility through cell reselections in RRC Idle mode only, without handover

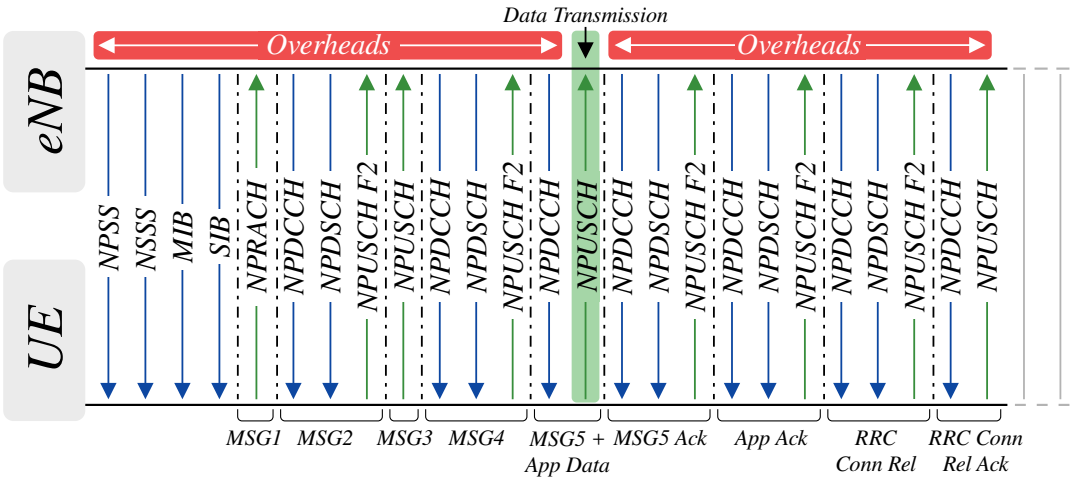


Fig. 2.6: NB-IoT: Control Plane Optimization [3].

functionality. Moreover localization options are also limited, enabling only Timing advance (TA)-based localization techniques, which are readily available without additional hardware or investment, relying on a database of BS coordinates and frequent radio communication for TA value adjustment. This makes them ideal for IoT applications that do not require high accuracy [62, 76, 83, 84].

3GPP Release 14 Advancements

Building on the foundation established in Release 13, 3GPP Release 14 introduced several key enhancements to NB-IoT and the new *Category NB2* [62].

UE Power Class 6, paging and random access on non-anchor carriers: Release 14 brought notable improvements in terms of coverage and power consumption, particularly through the introduction of a new UE power class with a maximum output power of 14 dBm. This addition caters to devices with specific power constraints and facilitates more efficient usage in urban areas with dense base station coverage. Additionally, Release 14 introduced support for paging and random access procedures on non-anchor carriers, which enhances flexibility and improves network performance [62, 75, 76].

Release Assistance Information (RAI): In terms of data communication and radio resource utilization, essential features were defined for CP optimization in the form of RAI in NAS signaling. This allows the UE to notify the MME that no further data transmission is expected, and a single DL data transmission following this message is anticipated. Consequently, the MME can immediately initiate the release procedure, thereby reducing the duration of the UE waiting period for potential data transmission [62, 75, 76].

Increased data throughput, Single-cell Point-to-Multipoint (SC-PTM) multicast: Additionally, Release 14 brings two other optimizations to data transmission. Firstly increasing the peak data rates for NB-IoT UE improves the overall efficiency and usability of the technology for a broader range of applications. This is achieved by introducing a TBS increase in both the UL and DL shared channels to 2536 bits, together with the introduction of additional Hybrid Automatic Repeat Request (HARQ) (maximum two HARQ possible). This leads to a peak data rate increase from 62 kbps to 159 kbps in the UL and from 27 kbps to 127 kbps in the DL. The second significant enhancement is the support for multicast DL transmission using SC-PTM feature. This allows the efficient distribution of the same data to multiple devices, which is particularly useful for applications such as firmware updates and broadcast services [62, 75, 76].

Improved positioning, connected-mode mobility: Release 14 presents a number of improvements to positioning services. These services employ Enhanced Cell-ID (E-CID) and Observed Time Difference of Arrival (OTDOA) techniques. E-CID positioning increases the accuracy of the TA-based method by incorporating additional measurement information such as Narrowband Reference Signal Received Power (NRSRP) and Narrowband Reference Signal Received Quality (NRSRQ). This is particularly important for applications that require precise location data, such as asset tracking and navigation. Although these enhancements do improve accuracy compared to the previous Release 13 method, they are still not as precise as 5G NR techniques owing to NB-IoT communication delay constraints and narrow bandwidth. Moreover, the connected-mode mobility functionality of Release 14 increases UE mobility, facilitates handovers, and ensures that devices continue to operate at peak performance levels even in highly mobile conditions [62, 75, 76].

Relaxed reselection monitoring: Release 14 offers a novel mechanism for cell reselection in the NB-IoT that is designed to decrease the power consumption of IoT devices. Conventional cell reselection processes typically involve frequent assessments of neighboring cells, which can be energy-intensive. With the relaxed monitoring approach, the frequency and magnitude of these assessments are reduced, enabling devices to preserve battery life while maintaining a reliable connection. This feature is especially beneficial for IoT devices that remain stationary or semi-stationary, such as smart meters and environmental sensors, because they do not frequently change their position [62, 75, 76].

3GPP Release 15 Advancements

3GPP Release 15 places a strong emphasis on diminishing latency and energy consumption. This release also introduced enhancements to measurement accuracy,

with the aim of improving network performance and dependability. Additionally, 3GPP Release 15 made advancements in random access reliability and range, which are critical for device connectivity and network access [62, 75, 76].

Support for small cells, access barring, and TDD: Release 15 featured significant enhancements to support small cells, enabling greater network densification and increased capacity. The inclusion of Time-Division Duplex (TDD) support provided greater flexibility in spectrum usage, allowing networks to adjust to fluctuating traffic patterns. Additional improvements to ECL access barring and eDRX mechanisms were also introduced to effectively manage network access during congestion and optimize power-saving features [62, 75, 76].

Battery Efficiency Security for low Throughput (BEST): BEST is a highly efficient encryption mechanism that enables data to be encrypted with minimal overhead, making it well-suited for low-power IoT devices. It is particularly effective for control plane transmissions using 3GPP Authentication and Key Agreement (AKA) symmetric cryptography. This enhancement ensures that data remains secure without significantly impacting battery life, which is crucial for IoT applications where devices often have to operate for extended periods on limited power sources [62, 75, 76].

Scheduling Request (SR): NB-IoT was introduced with an enhanced SR mechanism to improve the efficiency of UL data transmission. This enhancement reduces latency and control signaling (transmissions necessary for requesting UL resources), supports sporadic data transmission, and optimizes power consumption, making it particularly suitable for IoT applications that require extended battery life and reliable data transmission [62, 75, 76].

Reduced System Acquisition Time: Release 15 implemented mechanisms to reduce latency by improving the BS search procedures and system information acquisition across all operational modes. This was achieved by (i) increasing the repetitions of NPSS/NSSS/NPBCH, minimizing false detections, and enhancing correlation properties. (ii) SIB1-NB is repeated in other subframes or carriers to ensure reliable system information delivery. (iii) Introducing mechanisms such as WUS, enabling devices to skip reading MIB-NB, SIB1-NB, and SI messages, thus reducing the time and power required for initial access and ongoing operations [62, 75, 76].

Wakeup Signal (WUS): In 3GPP Release 15, the WUS enhances NB-IoT power efficiency by allowing devices to stay in deep sleep mode until a specific signal from the eNodeB instructs them to monitor the NPDCCH for paging. If the WUS is not received, the device skips the paging procedures, keeping hardware off longer and saving the energy needed for decoding NPDCCH and NPDSCH. This low-complexity signal is detected with minimal energy, and devices wake up during synchronized windows, optimizing the transitions between sleep and active

states. Configurable wake-up windows balance power use and latency, extending battery life for IoT applications with infrequent data transmission, such as low-end smart meters and environmental sensors, making the NB-IoT suitable for long-term deployments [62, 75, 76].

EDT: The introduction of the EDT mechanism has greatly enhanced the utilization of frequency spectrum. In RRC Idle mode, the UE is able to send messages during MSG3, also known as the RRC Early Data Request in EDT. The RAP can be terminated by the RRC Early Data Complete message once the UL transmission is successful. It's important to note that this message may include DL information, such as application confirmations, which prevents the UE from moving to a connected state unless the MME or eNB determines otherwise. EDT makes use of a predetermined set of NPRACH resources. The EDT's backing and the new NPRACH scheme have made it possible to extend coverage from the original 40 km defined in Release 13 up to 100 km [79, 83, 84].

3GPP Release 16 Advancements

In 3GPP Release 16, the primary objective is to incorporate NB-IoT within the framework of 5G networks. This integration facilitates the ability of NB-IoT to accommodate new use cases and deployment scenarios that are envisioned for 5G networks. Ongoing enhancements have been implemented to optimize the overall performance of NB-IoT, including its throughput, latency, and reliability. Furthermore, new features have been introduced to support emerging 5G applications, such as Massive IoT deployments and critical (URLLC) IoT. These advancements ensure that NB-IoT remains a relevant and robust technology for the evolving landscape of 5G networks [75, 76].

Industrial IoT: The primary focus of Release 16 is on URLLC IoT for manufacturing. In order to address the needs of Industry 4.0, a new vertical known as IIoT was introduced. This IIoT vertical is comprised of the mMTC and URLLC and offers support for industrial protocols, including Profinet and the implementation of 5G Local Area Networks (LANs) for industrial use [75, 76].

Grant-Free Access: This update incorporated the functionality for devices to send transmissions in RRC-Idle mode through Msg3 (RRC connection request) without an access grant and in RRC-Connected mode without a grant or with a simple control grant. Doing so decreases power consumption and latency by reducing the need for excessive signaling messages [75, 76].

Simultaneous Multi-User Transmission: New techniques, such as Code Division Multiplexing (CDM) and Multi-User MIMO (MU-MIMO), have been implemented to enable simultaneous transmissions by multiple users by sharing resources

in both the time and frequency domains. This was accomplished without the need for additional antennas [75, 76].

Group WUS (GWUS): Release 16 introduced the WUS mechanism enhancement, which includes GWUS. This feature allows a WUS to activate a configurable group of UEs instead of all UEs monitoring the same paging occasion, resulting in more efficient power consumption and extended battery life [75, 76].

Mobile Terminated EDT (MT-EDT): Another feature in Release 16 is MT-EDT, which builds on the EDT introduced in Release 15. This allows UEs to transmit small amounts of data (up to 100 bytes) during the random-access procedure. MT-EDT supports Mobile-Originated (MO) and Mobile-Terminated (MT) data transmission, and when triggered by the MME, an indication in the paging message prompts the UE to initiate random access to resume the connection or start the MO-EDT, depending on the optimization used. This enhances the efficiency of data transmission in scenarios where only small amounts of data need to be sent or received [75, 76].

Preconfigured UL Resources (PUR): This feature was introduced already in 3GPP Release 15 to decrease signaling overhead and power consumption by permitting data transmission during the random-access procedure through Msg3. In Release 16, PUR was enhanced further, allowing for even earlier UL data transmission. By implementing PUR, the random-access preamble and response (Msg1 and Msg2) can be eliminated, enabling data transmission to be completed in just two messages (Msg3 and Msg4), which results in significant improvements in efficiency and reduced latency [75, 76].

3GPP Release 17+ Advancements

One of the main objectives of 3GPP Release 17+ for NB-IoT is the integration of satellite communication, which enables the expansion of 5G IoT NTNs. This addition significantly enhances the performance of existing features, while also addressing connectivity issues in remote and underserved areas. By incorporating satellite communication, the NB-IoT NTN enables IoT devices to operate effectively and dependably in various environments, even in regions with weak or no signal [75, 76].

16-QAM for Unicast in UL and DL: From Release 13 to Release 16, NB-IoT UEs were limited to using QPSK for unicast NPDSCH and QPSK or BPSK for unicast NPUSCH. With the introduction of 16-Quadrature Amplitude Modulation (QAM) in Release 16, this capability has been significantly enhanced. Now, UEs can utilize 16-QAM for unicast NPDSCH with TBS up to 4968 bits in standalone and guard-band deployments, and up to 3624 bits in in-band deployments. For unicast NPUSCH, 16-QAM supports TBS up to 2536 bits, which can be transmitted using

up to half the time-domain resources compared to QPSK. This improvement in spectral efficiency is a significant advancement [75, 76].

Neighbor Cell Measurements and Measurement Triggering Before Radio Link Failure (RLF): In Release 17, a feature was introduced that allows NB-IoT UEs to conduct measurements while in RRC Connected mode. This innovation reduces the time required for RRC connection reestablishment by allowing UEs to evaluate neighboring cells and initiate measurements prior to experiencing a RLF. As a result, connectivity is enhanced and downtime is minimized [75, 76].

Carrier Selection Based on Coverage Level: Release 17 presents coverage-based paging for NB-IoT, a feature that streamlines network resource allocation and minimizes latency. This functionality enables the network to choose carriers based on coverage, thereby ensuring that paging messages are transmitted more effectively, ultimately enhancing network efficiency and user satisfaction [75, 76].

The aforementioned features of the NB-IoT evolution are depicted in Fig. 2.7.

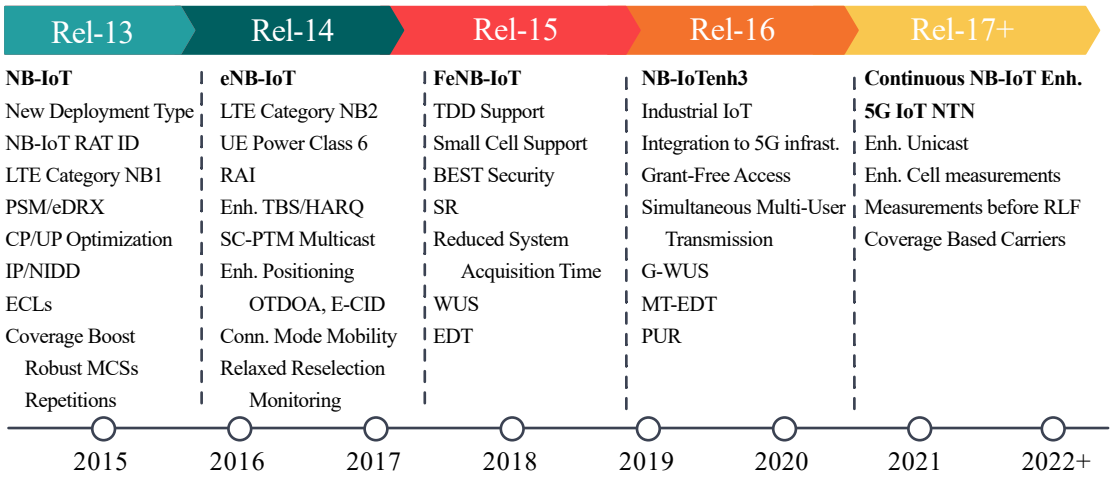


Fig. 2.7: Overview of NB-IoT evolution according to the definition by 3GPP.

2.2.2 Evolution of Enhanced Machine Type Communication

LTE-M or eMTC was introduced in 3GPP Release 13 as an enabler for applications requiring extended battery life and enhanced communication distance compared to legacy LTE categories. Unlike NB-IoT, eMTC is an enabler for applications that demand advanced services, extended network capacity, and higher data throughput while still staying within the eMTC. Thus, eMTC acts as the middle ground between legacy LTE (Cat-1) and the robust NB-IoT. 3GPP Release 13 introduced a new LTE Category M1 UE capable of operating on a bandwidth of 1.08 MHz (i.e., 6 PRB's) within an existing LTE deployment or 1.4 MHz in standalone deployment. This,

along with the extended upper layer message, 1,000 bits TBS, support for QPSK and 16QAM in both UL and DL, and eight HARQ support, enables eMTC to achieve data rates of up to 1 Mbps in full-duplex mode and approximately 300 kbps in half-duplex mode [62, 76, 85].

Coverage Enhancements

In the instance of eMTC, extended coverage is provided through repetitions with the same configuration, with up to 2048 repetitions for both the Physical UL Shared Channel (PUSH) and Physical DL Shared Channel (PDSCH), and 128 for the Public Random Access Channel (PRACH). The eMTC also supports the maximum transmission power classes of 23 dBm and 20 dBm are supported. All mentioned results in a bandwidth to link budget of up to 155.7 dB, which is lower than the NB-IoT Maximum Coverage Limit (MCL) of 164 dB [62, 76, 85].

The eMTC differs from NB-IoT in terms of coverage enhancement levels, which are achieved through the inclusion of two modes: Coverage Enhancement (CE) Mode A and CE Mode B. These modes enhance the versatility and applicability of eMTC for various IoT use cases that require dependable connectivity but have fewer steps than in the case of NB-IoT [62, 76, 85].

CE Mode A: First mode employs capability to retransmit messages up to 32 times. CE Mode A is the default mode and is recommended for situations requiring higher data volume transmission, mobility, and VoLTE usage. It balances extended coverage and power efficiency. Both UL and DL communications benefit from the repetition mechanism, ensuring consistent performance in various signal conditions. This mode is particularly beneficial for devices located in urban and suburban areas with moderate signal challenges, such as urban canyons or suburban homes [62, 76, 85].

CE Mode B: Second mode takes the repetition mechanism further than CE Mode A, allowing retransmission of messages up to 2048 times to achieve maximum coverage extension. This mode is designed for maximum coverage extension, targeting devices located in areas with very weak signal conditions, such as deep indoor environments or remote rural locations. Due to the extensive repetition, CE Mode B may introduce higher latency and lower data rates, a necessary trade-off to achieve the desired coverage enhancements. This mode is ideal for maximizing range, especially when devices are located in heavily obstructed areas or indoors. CE Mode B is suitable for scenarios where the device moves minimally or is stationary, providing robust connectivity despite its higher latency and reduced data throughput. Devices in extremely challenging environments, such as deep within buildings or in remote rural areas, utilize CE Mode B to maintain reliable connectivity [62, 76, 85].

Power Consumption Enhancements

In terms of improving power efficiency, eMTC incorporates eDRX and PSM modes, such as those found in NB-IoT. Furthermore, eMTC and NB-IoT share similar types (naming of legacy LTE channels) of channels and procedures, including those related to random access, data transmission, and power consumption optimization [62, 76].

Data Transmission Enhancements

The eMTC and NB-IoT share similar characteristics, including CP and UP optimization, support for NIDD technology, and reliance on TA-based localization. In contrast to NB-IoT, eMTC technology supports full mobility, including handovers in both TDD and FDD frequencies, and VoLTE capabilities. However, in early eMTC releases, there was no specific RAT type identifier for eMTC, leading to the utilization of legacy LTE identifiers for location update requests to HSS [62, 76, 85].

3GPP Release 14 Advancements

Release 14 brought significant advancements to eMTC, including the introduction of the LTE Category M2. This release focused on increasing throughput, with speeds up to 3 Mbps UL and 1 Mbps DL, achieved by expanding the bandwidth from 1.4 to 5 MHz, increasing the UL TBS to 2,984 bits from 1,000 bits of original and extension of maximum DL HARQ from eight to ten. The expanded bandwidth was utilized only in Mode A, whereas Mode B remained unchanged. VoLTE enhancements were also made to improve the performance in the half-duplex mode, and multicast capability was added for OTA firmware updates [62, 76, 85].

Furthermore, Release 14 includes similar enhancements, such as NB-IoT, in this release. Specifically, (i) E-CID and OTDOA capabilities for positioning, which enable eMTC devices to receive extended Positioning Reference Signal (PRS) transmissions and utilize multiple PRS configurations for increased accuracy; (ii) relaxed reselection monitoring; and (iii) RAI [62, 76, 85].

3GPP Release 15 Advancements

3GPP Release 15 brought substantial improvements to eMTC to increase its efficiency, performance, and suitability for IoT applications. These enhancements consist of new features and optimizations that provide better power efficiency, lower latency, improved access control, and increased spectral efficiency. In terms of eMTC, Release 15 shares new features with NB-IoT, such as the BEST security enhancement, WUS, EDT, and Reduced System Acquisition Time. The following list highlights the unique eMTC enhancements introduced in Release 15 [62, 76, 85].

eMTC Traffic Identifier (RAT Type), Power Class 6: Since Relese 15, eMTC obtained RAT Type (two Releases behind NB-IoT), that allows for the differentiation of traffic types over the LTE network. This feature ensures that eMTC traffic is identified and managed separately from other LTE traffic types, optimizing resource allocation and network performance. Since this release, Power class 6 with maximum output power of 14 dBm is enabled (one Release behind NB-IoT) [76, 85].

Feedback for Early Termination: This feature allows for the early termination of data transmissions when the required quality of service has been met. This feature reduces unnecessary transmissions, saving power and improving overall network efficiency. By providing feedback mechanisms, the network can optimize resource utilization and reduce latency [62, 76, 85].

Spectral Efficiency Improvements in eMTC: Enhanced spectral efficiency in eMTC is accomplished by several methods, including advanced modulation techniques, optimized resource allocation, and minimized inter-cell interference. The use of flexible starting PRB enables efficient scheduling of MTC-related data transmissions alongside other transmissions, allowing PDSCH/PUSCH resource allocation with adaptable starting PRB for UEs in CE mode with a maximum 1.4 MHz channel bandwidth. DL performance is boosted by supporting 64QAM modulation for PDSCH unicast transmission without repetition in CE mode A, which improves spectral efficiency without increasing UE peak rate. A revised CQI table with an extended range helps minimize the need for RRC reconfigurations under changing channel conditions. UL spectral efficiency is enhanced by implementing PUSCH sub-PRB resource allocation, offering new sizes of 1/2 PRB (6 subcarriers) or 1/4 PRB (3 subcarriers), and utilizing $\pi/2$-BPSK modulation to improve UL data coverage and power efficiency. Cat-M1 and Cat-M2 UEs can indicate support for CRS muting outside their narrowband, which helps decrease DL inter-cell interference through frequency-domain CRS muting. Early termination feedback includes a positive HARQ-ACK in an UL DCI over MPDCCH, enabling early cessation of DL monitoring and UL transmission, primarily to enhance power efficiency [62, 76, 85].

3GPP Release 16 Advancements

In a similar vein to NB-IoT, Release 16's eMTC mainly supports IIoT scenarios and includes coexistence with 5G NR via resource sharing through DL/UL resource reservation, DL subcarrier puncturing, and integration into 5G Core. Additionally, eMTC shares several feature enhancements with NB-IoT, such as G-WUS, PUR, and MT-EDT [76, 85].

Mobility enhancements: In 3GPP Release 15, eMTC mobility enhancements introduced the Resynchronization Signal (RSS) and the WUS to improve synchro-

nization and power efficiency. Building on these, Release 16 introduced further mobility enhancements utilizing these signals. RSS-based measurements in Release 16 allow UEs to use RSS for improved synchronization and measurement performance, with configurations signaled for both serving and neighbor cells, enhancing intra-frequency RSRP measurements in both idle and connected modes. Additionally, RRM measurement relaxation was implemented to maximize the power-saving benefits of WUS. While the WUS in Release 15 enabled UEs to sleep through multiple paging cycles, UEs still had to wake up for regular measurements. Release 16 addresses this by allowing UEs to extend their measurement cycles under configurable conditions, significantly reducing power consumption and enhancing battery life in eMTC devices [76, 85].

3GPP Release 17+ Advancements

Commencing with Release 17, in addition to additional improvements of existing functionalities outlined below, eMTC aligns with NB-IoT's objective of facilitating NTN communication within 5G infrastructure to establish comprehensive 5G IoT NTN networks [76, 85].

Additional PDSCH Scheduling Delay for 14-HARQ Processes in DL: This functionality permits HD-FDD eMTC UEs to utilize up to 14 HARQ processes in CE Mode A. By introducing a PDSCH scheduling delay, it guarantees the complete utilization of the accessible Balanced Load/Common Effort (BL/CE) DL and UL subframes, thereby enhancing the overall performance and dependability of the DL [76, 85].

Maximum DL TBS enhancement: Starting from Release 13, the maximum DL TBS for eMTC UEs was 1,000 bits. In Release 17, the maximum DL TBS was extended to 1,736 bits in CE Mode A for HD-FDD Category M1 UEs. Furthermore, UEs that support this feature can handle up to 43008 soft channel bits, thereby significantly increasing DL data capacity and efficiency [76, 85].

The aforementioned features of the eMTC evolution are depicted in Fig. 2.7.

2.3 Deployment of NB-IoT and eMTC

Both selected Mobile IoT technologies, ratified by 3GPP, operate within licensed bands but cater to distinct mMTC scenarios. NB-IoT is designed for scenarios prioritizing power efficiency and extended communication range, accommodating relaxed communication delay requirements. In contrast, eMTC addresses more demanding scenarios necessitating larger data volumes, higher throughput, and lower latency, while maintaining significantly reduced complexity compared to full LTE

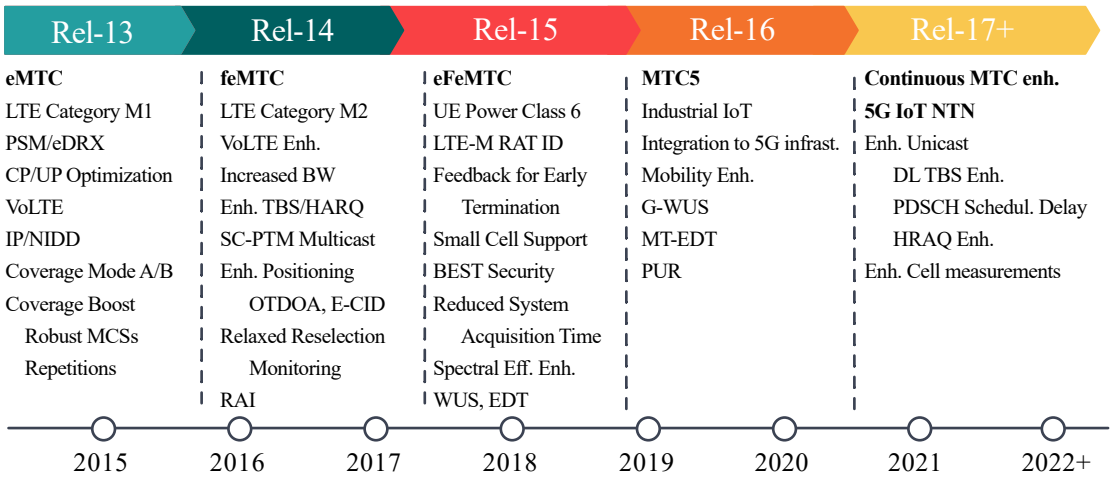


Fig. 2.8: Overview of eMTC evolution according to the definition by 3GPP.

systems. Despite the substantial technological advancements in both Mobile IoT technologies across 3GPP Releases 13 – 17, not all features have achieved sufficient market value or usability, as elaborated in the following text [62, 75, 76, 85].

2.3.1 Recommendation for Mobile IoT Deployments

The GSMA 5G IoT Strategy Group has issued a non-binding guide [76] for the implementation of mobile IoT technologies, focusing on features and enhancements introduced in 3GPP Releases 13-17 aimed at Mobile IoT. The guide provides recommendations that reflect the market traction and practical deployment expertise gathered around the globe. The document serves as the primary source for the following section. The guide offers recommendations to support interoperability among various operators and networks, and it then outlines recommendations for features and enhancements for eMTC and NB-IoT, categorizing them into three groups:

- **Minimum Core Features:** These are widely supported and adopted by operators and device manufacturers, and they are considered deployment minimums. For example, power-saving modes PSM and eDRX are included in this category.
- **New and Emerging Features and Enhancements:** This category includes features for which there has not been sufficient time for widespread implementation or are intended for late releases, typically taking 2 to 3 years to implement. Such features might not necessarily be grouped in future releases but may still have small market momentum and are expected to have more value in the near future. An example of this category is VoLTE for eMTC.

- **Features with Minimal Implementation:** This category includes features and devices that have already been standardized but have not been widely implemented, such as the adoption of eMTC device Category M2.

Minimum Core Features

Support for Bands: Mobile operators' band support varies. It is recommended that operators utilize at least one band around 1 GHz and publish the bands they support in the GSMA Deployment map. This aids users in selecting devices with compatible bands. Typical EU deployments are in Bands 3, 8, and 20 for both technologies.

Data Transmission Options Over the Network: For eMTC, IP data transmission via the UP is the minimum requirement, particularly for roaming scenarios. It is recommended to use Mobile IoT UP optimizations. Data transmission of eMTC via the CP is not supported by most operators and is therefore not recommended for deployment. For NB-IoT, current deployments do not support UP data transmission, so recommendations apply to the CP. While non-IP (NIDD) data transmission is not widely used, if available, data should be routed through the SGi interface.

Support for PSM and eDRX Modes: Implementation of these modes should be without timer restrictions. Data packet buffering should be allowed when the device is in one of these sleep modes. When using both PSM and eDRX, careful timer configuration is crucial for proper paging.

GTP Timer Value Settings: This is crucial for devices in roaming, communicating through a GTP tunnel. The value determines how long the tunnel remains active while the device sleeps. Recommended minimum values are 31 days for NB-IoT and 24 hours for LTE Cat-M.

Use of Mode A for Coverage Extension: This mode is mandatory for devices and highly recommended for operators. Mode B, suitable for low data transmission, is rarely used and had not been implemented by any manufacturers or operators by 2022.

Support for Coverage Enhancement Modes for NB-IoT: Operators should support all three ECLs.

Support for SMS: Users can choose to use SMS or alternatives like UDP or NIDD, though the latter is not widely deployed.

Support and Configuration of C-DRX and C-eDRX Modes: These modes allow devices to enter power-saving modes even during active network connections, with C-DRX allowing up to 2.54 s and C-eDRX up to 10.24 s.

Deactivation of UICC: Should be supported, even when the device is in eDRX.

Support for Different Power Classes: The minimum supported power class should be 3, thus 23 dBm EIRP. The use of lower power classes should be considered by both operators and users, as selecting the wrong class could reduce device costs but lead to coverage issues.

Support for RAT Types: For NB-IoT roaming, implementing support for RAT types specific to NB-IoT on the S6a interface is essential. The same applies to eMTC, which has a specific RAT awaiting broader implementation.

Relaxed Monitoring for Cell Reselection: Recommended mainly for static NB-IoT devices to improve battery life.

Support for Quick Network Disconnection via RAI: When a device finishes UL transmission, it sends RAI to expedite network disconnection, saving energy. RAI support should be available on both the device and network.

Support for Network Access Mechanisms for Coverage Enhancement Modes: Traditional access mechanisms like ACB and EAB cannot distinguish devices in coverage enhancement modes, potentially causing repeated transmissions during high network load. New mechanisms that allow postponing transmissions to less congested times should be supported.

Specific LTE Cat-M Features include:

- Support for half-duplex mode.
- Mobility in connected mode, necessary for VoLTE.
- Support for larger TBS in UL.
- Support for 10 HARQ processes in DL, essential for full-duplex mode but also beneficial for half-duplex mode.
- Support for sending HARQ-ACK feedback for multiple transmitted blocks in half-duplex mode.

Specific NB-IoT Features include:

- Support for all deployment modes.
- Support for the new NB2 device category.
- Capability to report DL channel quality in NB-IoT, allowing eNodeB to optimize the connection, reducing device energy demands.
- Support for measurement improvements.

New and Emerging Features and Enhancements

Support for NIDD Without IP Headers: is currently technically feasible; however, this transmission method is not yet implemented by operators and manufacturers. Therefore, it is uncertain if this function will become widely used.

Implementation of SCEF Functions: allows secure access to network services and capabilities, but support from operators is currently minimal.

UP EPS Optimization for Mobile IoT: UP Mobile IoT EPS Optimization, introduced in 3GPP Release 13, allows user plane data transfer without the Service Request procedure in ECM-IDLE mode (RRC Suspend/Resume). This reduces signaling overhead by approximately 75% compared to legacy LTE, improving state transition efficiency, signaling, and power usage.

Support for BEST Data Security: has the potential to extend device battery life significantly. However, it is too early to recommend this type of security as many manufacturers and operators have yet to implement it.

Support for WUS Signals: for notification of devices with relaxed paging monitoring to resume with monitoring.

EDT: allows sending data sizes from 328 to 1,000 bits using the Msg3 in the random access procedure. Operators and manufacturers should implement this feature as it has significant potential for reducing energy consumption during transmission.

Preconfiguration of EARFCN and Geographic Areas: has the potential to reduce connection time in roaming scenarios.

Specific eMTC Features: These that are not yet widely implemented include VoLTE, support for fast-moving UEs, and improved spectrum utilization.

Specific NB-IoT Features: include modifications for FDD use, storing device information and operational profiles in the MME, increased cell range, mixed standalone mode, and support for small cells.

Additional enhancements, whose implementation is still uncertain, include those introduced in Release 16. These include the possibility of sending a GWUS, EDT triggered by the MME, improved DL signal quality reporting, UL EDT via PUR, connection to the 5G core network, and mobility enhancements for eMTC.

Features with Minimal Implementation

Some of the features were already mentioned above in minimum core features, following text thus completes the list of unlikely to be implemented features.

General Features: Support for Multicast Transmission and Group Messaging is currently minimal.

Specific eMTC Features: Include support for the new Cat-M2 device category, support for wider BWs in eMTC CE modes of 5 MHz or 20 MHz, and eNodeB-controlled transmit antenna selection.

Specific NB-IoT Features: Include mobility in connected mode, paging and random access using a non-anchor carrier, and support for TDD mode.

2.3.2 Comparison of Currently Deployed NB-IoT and eMTC Features

The aforementioned sections delved into the releases of NB-IoT and eMTC, as well as the recommendations for deployment from telco operators and mobile technology vendors. This section focuses on the implementation of features in deployments that were available during the study's competition. Specifically, the deployments in the Czech Republic, which are representative of European deployments[1], were utilized in subsequent sections.

Both technologies are presently being deployed as a hybrid of 3GPP Release 13 and 14, as operators selectively incorporate features without fully committing to Release 14. The settings of the deployed networks are detailed in Appendix A. There are also some nuances to consider, such as the GPT timers, which may be lower than the suggested values, resulting in a low threshold for open TCP connections for both technologies without active communication, which is restrictive. Other deployment-specific aspects include the trend towards global core deployments, which offer the benefit of central management but also have drawbacks, such as increased delay and inconsistent jitter.

Given the implementation of all features, it is clear that both technologies have not yet reached their full potential, as key features are yet to be implemented. The current configuration still relies on outdated legacy features, which can easily lead to congestion, such as random access procedures, particularly in cases of burst traffic. As a result, the current state is not appropriate for the expected 10^6 connected on a single square kilometer, as specified in the IMT-2020 requirements [24].

2.4 Summary

This chapter furnishes a crucial background for **RQ1.1**, elucidating the intricacies of Mobile IoT technologies, including their objectives and prospective advancements. Furthermore, this chapter delivers extensive insights into deployment suggestions and real-world deployment scenarios.

The aforementioned developments present an opportune opening for this study to delve into, as it has been noted that essential enhancements are yet to be realized in the future or are slated for implementation. For example, the utilization of conventional LTE for both signaling and random access processes may result in congestion in specific scenarios. This opening serves as a platform for this work to suggest enhancements that can be executed prior to or concurrent with future features, thus augmenting the performance of Mobile IoT technologies.

[1]See "Mobile IoT Deployment Map" by GSMA, 2023: `www.gsma.com`

3 Energy Grid Smart Metering Specifics and Requirements

This chapter commences with a review of the present state and strategy of smart meter deployments. The discussion initially centers on global deployments of smart meters with remote reading capabilities. Thereafter, the focus shifts to the deployment examples of the EU, including the directive that shapes and guides the phase of smart meter deployment within EU states. The second section delves into a detailed case of smart-grid deployment strategy in the Czech Republic, as it served as the primary source of information and measurement campaigns for this thesis. The primary objective of this section is to conduct a thorough examination of mobile network capacity with respect to smart-grid applications, followed by a detailed exploration of the specific requirements and characteristics of smart metering. The third section then expands upon the smart-grid requirements by incorporating new traffic types that smart electricity meters introduce in contrast to other utility meters, such as water and gas meters with mention of related work regarding this topic. Lastly, the chapter concludes with a summary section.

3.1 Smart Metering Infrastructure Deployment Motivation

As outlined in the introduction, the necessity for the implementation of smart meters equipped with remote data communication is evident. The choice of communication technology is contingent upon the specifications, capabilities, and demands of the smart meter. As previously stated, Mobile IoT technologies are poised to facilitate such use cases. However, there are numerous obstacles that must be taken into account during these deployments, including legislative issues and technical challenges in the form of network capacity constraints or the heterogeneous nature of the deployment environments.

3.1.1 Examples of Global Smart Meter Deployments

The deployment of smart meters has become increasingly widespread globally. In 2018, China initiated large-scale smart meter implementation through hundreds of thousands of smart smoke detectors, streetlamps, and water meters connected via NB-IoT as traditional low-end sensoric devices. The initial deployment of eMTC in conjunction with the NB-IoT was presented by Sony Altair as solution in Japan's smart gas meter implementation [86, 87].

In 2021, Saudi Arabia launched a significant deployment campaign for mid-end smart electricity meters with combined NB-IoT and PLC communication interfaces. This variant of communication technology appears to be the preferred option of distributors. After 12 months, no less than 10 million smart electricity meters were deployed in Saudi Arabia. Australia has commenced a 10-year program to implement four million NB-IoT-enabled smart electricity meters by leveraging the Elestra telco operator network [86, 87].

The European Union is also experiencing ongoing and increasing smart meter deployment. In 2021, the UK deployed 300,000 NB-IoT-enabled watermeters for low-end smart meters. Additionally, in 2022, Sweden's telecom operator Telia, in collaboration with OEM Nordic and Ericsson's support, achieved a massive deployment of 900,000 mid-end smart electricity meters. Telia representatives stated that this deployment faced considerable network capacity issues and optimizations in very diverse environments, which are the aims of this thesis [86].

3.1.2 EU Directive for Smart Meter Deployments

To further enhance the unity and pace of smart meter deployment in Europe, the European Parliament introduced a directive for member states and established the European Smart-Grids Task Force (SGTF) as a responsive institution for deployment and problem solving. The initial discussions began in 2015, when mobile operators began integrating the chosen communication technologies for the IIoT [29]. Since then, member states have had the opportunity to review the specifics of Advanced Metering Infrastructure (AMI) and indicate their deployment strategies for the period between 2019 and 2030 [88]. According to the 2019 report [88], only 17 member states presented their strategies for adopting smart meters, primarily selecting Power Line Communication (PLC) and 2G technologies for remote data transmission, with only four indicating legacy 4G or NB-IoT technologies, which later became the preferred technology for wireless connections [88]. Since 2019, member states have accelerated deployment based on the directive. According to a 2022 report [89], 13 member states have completed the deployment of smart meters, serving more than 80% of their customers through smart meters, and ten more are in the roll-out phase with the aim of completing it after 2024. However, the same report indicated that five member states have not made significant progress in deployment, with the Czech Republic being one of them [89].

The UE Parliament has determined that from 2024 to 2027, intelligent smart meters will start to be implemented in all member states of the EU [36–38]. This definitive timeline for the implementation of these advanced smart meters has significantly spurred collaboration between electricity distributors and telecommunication

operators in recent years [90, 91].

Electricity distributors have introduced a new category of electricity metering, known as Type C or smart metering, which necessitates the utilization of dedicated communication technologies operating in the licensed frequency spectrum. This new type of metering is in line with the requirements outlined in the European Parliament's directive 2019/944, which stipulates that intelligent metering systems must [37, 88, 92, 93]:

- enable access to electricity consumption data in real-time,
- provide verified historical data via a visualization platform,
- and offer access to electricity consumption data without charging through a secured local or remote interface.

For instance, the Czech Republic, a state with a delayed smart meter rollout, has seen the early stages of cooperation between energy distributors and telco operators focused on selecting communication technology for AMI, or more precisely smart-grids, remote data transmission. In this case, NB-IoT and later eMTC were chosen, with the first prototypes of NB-IoT appearing soon after 2016 [94]. A comparison of the key communication parameters for NB-IoT and eMTC is presented in Table A.1.

The aforementioned prototypes evolved in cooperation with electricity distributors, and two years later, remote metering over NB-IoT was tested using real electricity meters [58, 95]. Although the initial trials were successful, the measurement campaign was not up to par with the expectations for the first deployment phase in 2023 or later, before the planned rollout in 2024.

3.2 Energy Grid Specifics and Requirements for Smart Meters in Czechia

3.2.1 Mobile Operator Planing for NB-IoT Driven Smart Meters

Vodafone, one of the three mobile network operators in the Czech Republic, became the first to complete the installation of NB-IoT technology in the country. With over 4,000 base stations (eNodeBs), Vodafone achieved 100% outdoor and 95% indoor communication coverage. In terms of the "deep indoor" extreme use cases such as smart meter applications, NB-IoT technology is currently deployed in the licensed frequency band no. 3 (1800 MHz), and no. 20 (800 MHz)). This technology enables communication with smart meters in locations with poor signal coverage, where the values of the RSRP parameter fall below -120 dBm or/and SINR parameter below -3 dB. The selection of the coverage class based on both RSRP and SINR is displayed in Table 3.1 [29].

Tab. 3.1: Coverage classes for the ECL assignment [2].

Coverage Class	SINR and RSRP Conditions
Class 0	SINR >7 dB and RSRP >−110 dBm
Class 1	−3 dB <SINR <7 dB and RSRP >−133 dBm
Class 2	SINR <−3 dB or RSRP <−133 dBm

The information provided serves as the basis for network capacity planning studies that focus on outlining communication patterns and network loads. Notably, telco operator Vodafone Czech Republic, conducted an extensive study to determine the expected network performance and capacity. All data in this study were sourced from the actual deployment of the NB-IoT network in Hradec Králové, a city in East Bohemia, Czech Republic. Owing to the symmetrical nature of the DLMS protocol utilized for smart meter communication, which operates on a request-response system, a "100 B, 100 B" traffic model was utilized for a rough estimate of the network capacity. This study revealed two scenarios for the distribution of the ECLs (ECL0, ECL1, and ECL2): (i) 60%, 30%, and 10%, and (ii) 80%, 15%, and 5%. Unfortunately, these findings are not publicly available because they result from close collaboration between the research team and the network operator. However, similar studies confirm our findings, such as [96], where the distribution of ECL for water meters is nearly identical to that in the first scenario.

Considering the NB-IoT deployment in the guard-band frequency spectrum the theoretical cell capacity is approximately 10,150 devices per hour for the 60%, 30%, and 10% distributions. As expected, in the case of the second distribution, where the number of smart meters in ECL 0 increases, that is, 80%, 15%, and 5%, the cell capacity is significantly higher; approximately 19,253 devices per hour can be served. Calculations of the delivery time were performed for both groups of ECL distributions. We differentiate between these two situations. When smart meters send data asynchronously, the data from all smart meters are delivered with an end-to-end delay ranging from 250 ms to 10 s, which does not exceed the 15 s threshold defined by the electricity distributors. Alternatively, when smart meters send data synchronously, the data will be delivered with a delay in the range of 250 ms to x (in units of minutes), where x depends on the number of simultaneously connected devices in the cell [88, 92].

As we proceed with the description of the new communication scenario for smart meters, we concentrate on the approximate number of smart meters connected to the eNodeBs. The average number of locations equipped with smart meters and covered by one base station (eNodeB) is 800 in the case of a mid-sized city and 350 in rural areas, as referenced in a conducted study [97, 98]. At this juncture, we

calculate the number of users (smart meters) that can be served within a defined period of time (see Table 3.2). In our calculations, the NB-IoT channel is estimated based on the traffic model and UE distribution model defined in 3GPP TR 45.850, without the use of power saving mode (PSM) by the smart meters. This is because smart meters must be in the IDLE mode with active radio reception all the time for remote control of the load of the smart electricity meter and individual relays of the meter [93].

Tab. 3.2: Results for 100 B / 100 B (UL / DL) traffic model for two ECL distributions [99–102].

Percentage of served users	ECL distribution (60 %, 30 %, 10 %)	ECL distribution (80 %, 15 %, 5 %)
In 5 min	68.9%	86.6%
In 10 min	89.5%	97.7%
In 15 min	96.1%	99.6%

Even though the communication scenarios outlined above may appear strict, electricity distributors have set even more demanding on-demand use cases or situations for recovering communication after a blackout. For instance, in the case of NB-IoT networks, the recovery process after a blackout requires an additional 80 B for the UL attach procedure and 30 B for the DL direction, which includes synchronization and status updates. The outcomes for this scenario are presented in Table 3.3.

Tab. 3.3: Results for the recovery scenario involving two ECL distributions [88, 92].

Percentage of served users	User ECL0:1:2 distribution	
	(60 %, 30 %, 10 %)	(80 %, 15 %, 5 %)
In 5 min	43.2%	62.9%
In 10 min	67.1%	85.2%
In 15 min	80.3%	93.9%
In 20 min	88.3%	97.2%
In 25 min	93.0%	98.7%

In conclusion, the NB-IoT network capacity meets the 3GPP TR 45.820 NB-IoT cell capacity model requirements [103, 104], but these requirements are no longer met in on-demand scenarios or recovery situations. It is crucial to successfully manage communication with the smart meter within 15 minutes for both regular readings and blackout situations. To achieve this, the network capacity can be expanded

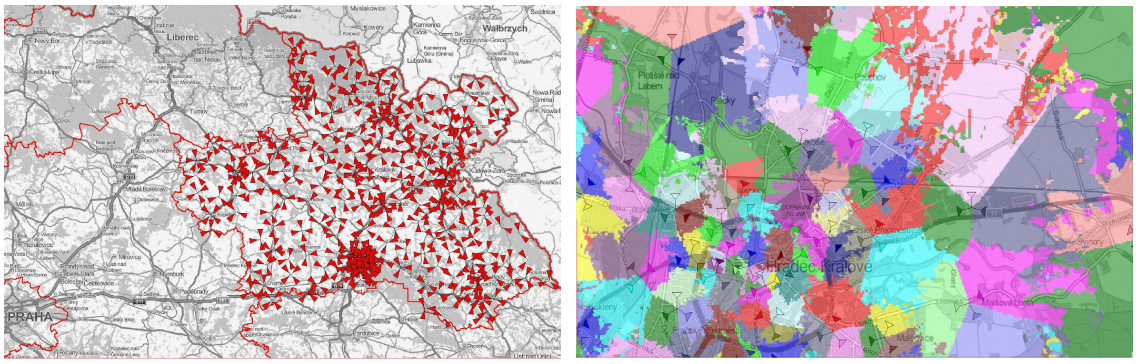


Fig. 3.1: NB-IoT BSs and the anticipated communication coverage in the designated Region 1 [1].

based on factors such as the number of deployed smart meters, radio conditions, and forecasted traffic patterns. This can be done through logical channel optimization or by adding another NB-IoT channel within the frequency spectrum (2^{nd} guard band or stand-alone carrier). To accurately assess the required capacity of the NB-IoT system for servicing both conventional stochastic data traffic and new regular data traffic identified in case of mid-end smart electricity meters, detailed performance evaluation models are necessary [46].

The aforementioned requirements indicate that smart-grid applications are expected to be one of the earliest and most significant use cases for mMTC, and will necessitate close collaboration between distributors and operators. The initial phase of this process involves evaluating network capacity to determine the effects of deploying a specific number of smart meters in a given area. This preliminary data gathering stage is illustrated by the findings from a collaborative measurement initiative in two selected Czech regions (Hradec Králové and České Budějovice), where the potential installation of smart electricity meters is under consideration, as shown in Figs. 3.1 and 3.2. Evaluation revealed that in urban environments, a single cell typically serves an average of 800 apartments, while in rural settings (small towns), one cell covers an average of 350 apartments.

By calculating the average number of EDs per cell in the covered area, we determined the total number of apartments to be 268,578. Two ECL distributions were then established: (i) 60%, 30%, and 10% for ECL0, ECL1, and ECL2, respectively; and (ii) 80%, 15%, and 5% for ECL0, ECL1, and ECL2, respectively. The ECL ratio was determined based on the operator's understanding of the key communication parameters, namely, RSRP, RSSI, RSRQ, SINR, and PER. Calculations were performed to determine the percentage of users serviced within specific time intervals (5, 10, and 15 minutes). Through collaboration with Vodafone Czech Republic, we obtained data on the cell's maximum capacity and channel information, encom-

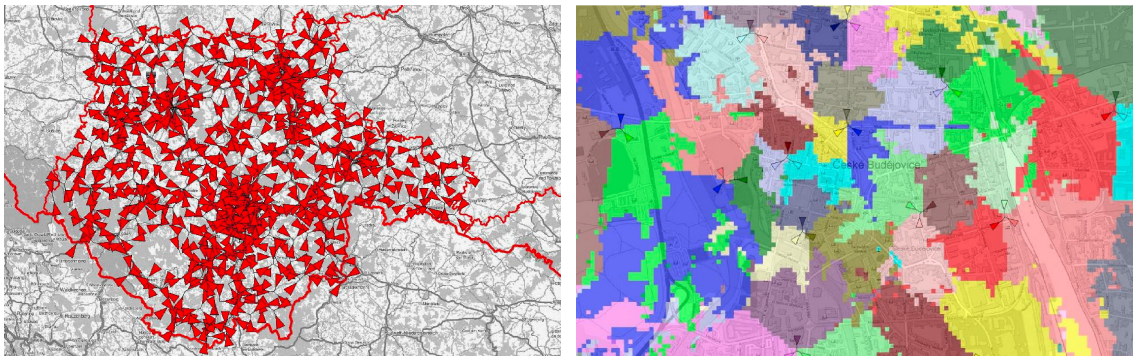


Fig. 3.2: NB-IoT BSs and the anticipated communication coverage in the designated Region 2 [1].

passing NPRACH, NPUSCH, NPDSCH, and NPDCCH. This data considered the procedure: attach + 80 B (UL) + 30 B (DL) (as shown in Table 3.4).

Tab. 3.4: EDs (smart meters) served within five minutes by an NB-IoT cell [1].

Channel Capacity	**User ECL0:1:2 distribution**	
	(60 %, 30 %, 10 %)	(80 %, 15 %, 5 %)
NPRACH	1,185	1,185
NPUSCH	540	1,025
NPDSCH & NPDCCH	195	345
Cell capacity	195	345

3.2.2 Communication Parameter Requirements for Smart-Grids

While ECL distributions were correctly established based on geographical locations and corresponding radio conditions, significant changes in data volume and traffic patterns have occurred in recent years. The initial assumption was that a symmetrical 100 B message would be sent in both the UL and DL directions every six hours, leading to a monthly data usage of approximately 50 kB per ED. However, recent calculations derived from actual register readings have altered this understanding, imposing new and more stringent requirements.

Regular communication is essential for each ED to maintain open connections and transmit telemetry data as needed. Network operators initially determined that a permanent connection would require 600 B of bidirectional UL/DL communication per hour, which they deemed feasible in their capacity planning study for NB-IoT applications. However, smart metering scenarios demand service availability even during data bursts, such as widespread firmware updates, extensive remote readings, or when all EDs experience blackouts followed by simultaneous startups

and communication requests. Recent assessments of smart meter data traffic have shown volumes reaching up to 20 MB monthly.

The second criterion addresses data transmission in smart-grid systems. Currently installed legacy smart meters must create a secure communication link using TLS version 1.2 (or an equivalent security protocol such as DLMS security suite 2), which includes mutual authentication through certificate verification. This process combines the necessity of initiating mobile network registration with the establishment of a secure end-to-end connection, requiring the exchange of multiple messages.

Table 3.5 outlines the projected requirements for smart meters over the next two decades, extending to 2040. These specifications apply to smart meters installed at power supply points consuming up to 6 MWh annually and featuring home-charging stations. To enhance our comprehension and enable more precise evaluations of technology selection and assessment for these devices, we begin by examining the projected quantity of deployed smart meters and the estimated monthly data transmission for individual and collective smart meters within the smart-grid network.

Tab. 3.5: Forecasted requirements from electricity distributors in Czech Republic [1].

Year	2030	2040
Number of EDs (smart meters)	250,000	1,000,000
Minimal committed information rate (CIR) [kbps]	11	39.1
Amount of total transmitted data per month [GB/month]	3,476	12,360
Amount of transmitted data from one ED per month [MB/month]	16	16

The expected data traffic is illustrated in Fig. 3.3, which shows the distribution of the different locations based on the number of citizens. Fig. 3.3 further provides insights into the anticipated communication parameters for four groups of smart-grid operations: (i) telemetry, (ii) reading, (iii) monitoring, and (iv) management. The Round-Trip Time (RTT) is a crucial metric that measures two-way communication delay. Power distribution companies are required to establish a maximum RTT limit of two seconds for smart meters. This limit applies to services with different periodic communication intervals: five minutes for telemetry and reading, one minute for monitoring, and one month for management. In the context of smart metering and remote reading, although each service operates on a distinct communication schedule, they all must adhere to the same two-second RTT requirement.

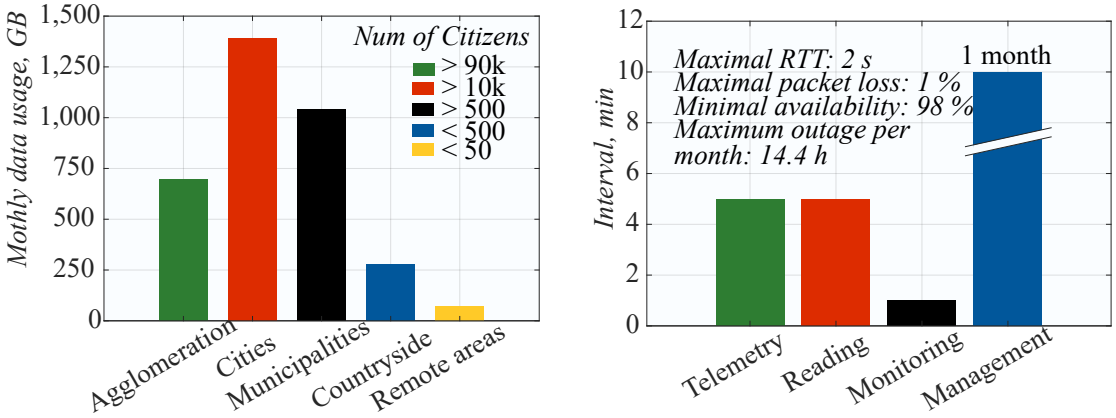


Fig. 3.3: Left: Estimated traffic load across various locations. Right: Specifications for communication parameters [1].

3.3 New Data Traffic Type for Permanently Connected Smart Meters

As stated, the standard reading time for intelligent electricity meters typically varies between 5 and 15 minutes [105–107]. Furthermore, the forthcoming generation of electricity meters is expected to offer advanced features, such as hubs for EV charging, enabling load balancing on power grids, and reducing energy consumption (and consequently, carbon emissions) by incorporating "pause hours" [44, 108, 109]. It is worth noting that such sophisticated systems impose even more demanding requirements on connectivity, with sampling occurring in milliseconds [45, 110]. Permanent connectivity is the only viable solution to satisfy these stringent demands. Additionally, even existing power grid systems may benefit from permanently connected devices, as distributors can utilize smart electricity meters as an auxiliary component of power grid load balancing. Specifically, this can be achieved by integrating it into a Frequency Containment Reserve (FCR), which may require a communication delay as short as 30 s [111].

Although it is important to note that even with a current periodicity of 15 min, permanent connectivity can significantly reduce data consumption, it should be noted that the initialization procedure for data transmission requires considerable overhead. This is because connection-oriented protocols such as DLMS/COSEM and TCP/TLS, which are the de facto standards for electricity metering, are utilized [39]. The re-initialization of the connection every 15 min results in a higher overhead than a permanently connected UE in just 24 h, as shown in Fig. 3.4. This value represents the perspective of transport and application layers in the reference TCP/IP model. However, in the lower layers of the LTE system (physical, medium access, and radio

link control), the overhead can be even higher. For instance, under harsh radio conditions, the reliability of the NB-IoT depends on message repetitions.

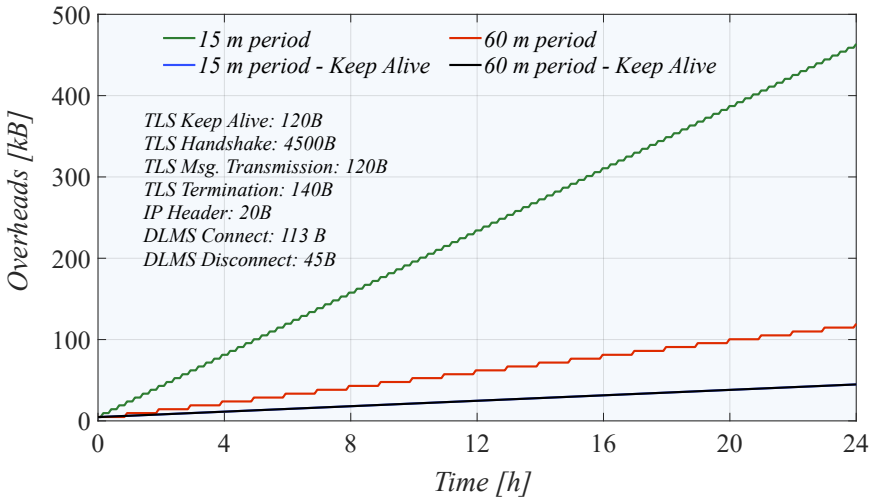


Fig. 3.4: TLS communication overheads for 24 hours.

Permanently connected UEs prevent the use of the PSM, which involves a complete shutdown of the radio part and unavailability of communication. However, it is possible to use the eDRX mode, in which the module listens to incoming messages in regularly recurring windows. By fine-tuning the eDRX cycle (ranging from 23.04 to 175.4 min) and the paging time window (lasting between 2.56 and 40.96 s), we can achieve optimal power consumption while also accepting reasonable communication delays [79, 112].

The communication strategy of newly emerging smart-grid electricity meters has completely transformed the conventional mMTC or smart meter communication paradigm, as previously discussed. Unlike the traditional approach, where each end device acts as the "source" and the central node serves as the "sink," the next-generation communication networks, which are considered enablers for smart-grid scenarios, follow a "request-response" approach that is inherently "client-originated". However, the main concern when transitioning from the legacy "source-sink" model to current cellular IoT technologies, such as NB-IoT and eMTC, is their capacity limitations.

In simpler terms, asynchronous data transmission evolves into more complex synchronous scenarios in smart metering. Applications of this type require smart meters to maintain a persistent connection with the cellular network in a RRC-connected state and receive periodic polling from the network [47]. These activities lead to the accumulation of data packets at the radio access level of the network, thereby causing significant delays at the air interface. This behavior does not fulfill the low latency requirements of 5G-IoT defined by ITU-R M.2410, which specifies up to 10^6 EDs/km^2 with a data transmission intensity of one packet every two hours,

with no more than 10% packet loss [49, 113].

The deployment of smart meters with permanent connectivity together with traditional utility meters can be categorized as Type I and Type II UE, as shown in Fig. 3.5. Type I UE refers to devices that are controlled remotely from an application server and remain active on the radio interface at all times, thus maintaining a "RRC connected" state throughout their operation lifetime, which will be further discussed. These devices represent the progression of the current smart metering technology, primarily utilizing the communication channel in an asymmetrical manner for efficient two-way communication with minimal response time. On the other hand, the Type II UE consists of conventional UEs that undergo a "sleep-awake-transmit" cycle, which may correspond to sensors that perform periodic measurements. The operational cycles of these UEs are assumed to be asynchronous [45, 105, 110].

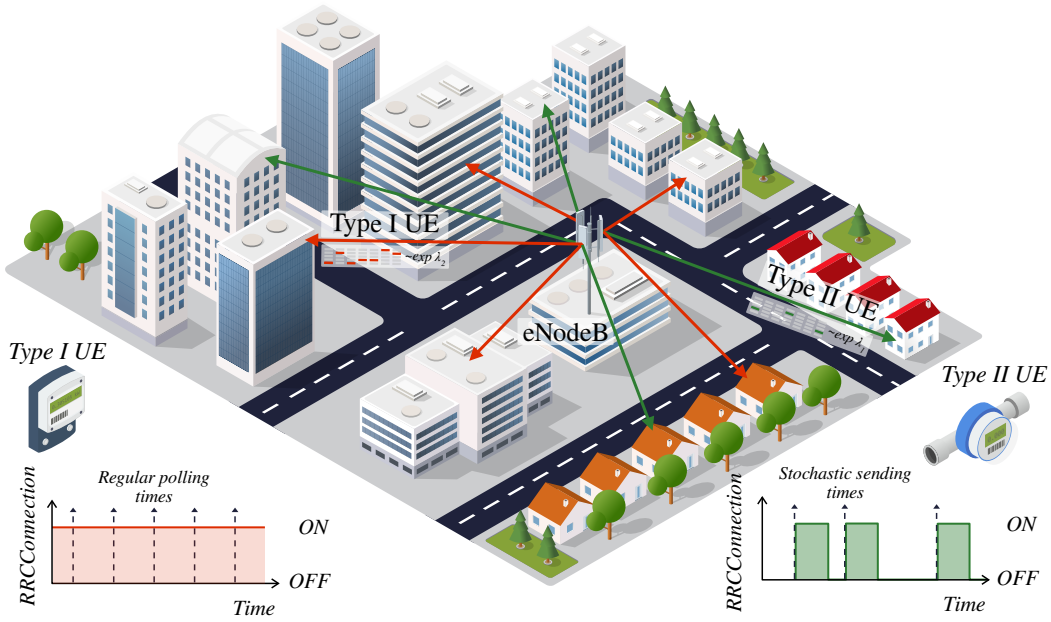


Fig. 3.5: Two UE devices types in NB-IoT network: *Type I* UEs with regular polling times, *Type II* conventional UEs with stochastic arrival pattern [2].

To optimize the delay performance of both types of traffic coexisting at the sole NB-IoT air interface, one of proposed approaches could be based on spreading the message transmission time instants of regular and stochastic traffic. Result would be optimal value of spreading interval for wide range of UEs in the area of coverage between the type of devices utilizing two mentioned traffic types.

Other approach could be load balancing between multiple RATs by proposal of an optimal RAT association algorithm to balance the load between NB-IoT and eMTC interfaces as it was stated, they are preferred way for smart-grids. This algorithm would address both stochastic message arrivals and batch arrivals.

3.4 Summary

Building on the definition of Mobile IoT technologies that are suitable for smart-grid deployments presented in Chapter 2, this chapter delves deeper into the specifics and challenges of smart-grid deployments. In Section 3.1, the global deployment of smart meters is summarized, highlighting the speed of deployment and popularity of the selected technologies, primarily eMTC and NB-IoT. The section also emphasizes the push from the EU directive to accelerate the pace of smart meter AMI deployment.

Section 3.2 provides detailed information on the requirements and challenges of AMI deployment in the Czech Republic, a member state with delayed deployment. Cooperation between telco operators, energy distributors, and industry has resulted in a comprehensive and detailed study of the specific requirements of AMI deployment. It is important to note that, although the issues of AMI deployment in the Czech Republic may seem local, the study is applicable throughout the EU.

Considering the specifics of smart-grid deployment, Section 3.3 identified a new communication traffic type with regular message intervals for permanently connected smart meters. Co-deployment with the traditional stochastic traffic of other utility meters raises questions and challenges regarding potential violations of the presented smart-grid communication requirements. This chapter outlines possible solutions based on spreading the message transmission time instants of regular and stochastic traffic to achieve the optimal value of the spreading interval for a wide range of UEs in the area of coverage between the two types of devices utilizing the aforementioned traffic types addressing the communication latency issues. In addition, it proposes an optimal RAT association algorithm to balance the load between the NB-IoT and eMTC interfaces, which are preferred for smart-grids, thus addressing the communication reliability issues.

According to the research proposed and related work discussed in this study, it has been found out that the identified challenges are significant and require further assessment. Therefore, this chapter provides a comprehensive response to the first research question, **RQ1.1**, which was partially addressed in Chapter 2.

4 Performance Assessment of NB-IoT and eMTC Technologies

The measurement campaign outlined below pertains to observations made in Chapters 2 and 3 with respect to **RQ1**, which focuses on the evaluation of two principal mMTC technologies, eMTC and NB-IoT. These technologies are considered optimal candidates for emerged IoT applications, such as data transfers in smart-grids. The objective of the campaign was to gather communication parameters and features of the currently deployed variants of both technologies, as previously discussed in Chapter 2. Additionally, the campaign aimed to identify potential bottlenecks related to smart-grid requirements presented in Chapter 3, thereby providing a partial and concluding answer to **RQ1** and a complete answer to **RQ2**.

This chapter comprises two sections, which correspond to two distinct stages of a comprehensive measurement campaign. The initial stage of the campaign focused primarily on evaluating the performance of NB-IoT and eMTC technologies in terms of data throughput, delay, jitter, RAN features. The results of this stage are expected to shed light on the behavior of mobile networks in various deployment scenarios, provide input for the numerical model, and determine whether the selected technologies, in their present form, are suitable for the smart-grid use case.

The second stage of the campaign, and hence the second part of this chapter, presents information on the boundaries and performance of NB-IoT in a worst-case scenario, such as a congested radio environment, by subjecting a single cell to the simultaneous demands of 200 EDs.

The methodology employed in the two-stage campaign is as follows:

1. During the first phase, each chosen technology underwent individual evaluation in a controlled laboratory setting. A single ED (equipped with a Quectel module BG77) was connected to one BS to examine fluctuations in user data throughput, delay, and jitter in relation to the signal strength (RSRP). This phase was executed in an isolated environment devoid of external interference to guarantee precise and dependable outcomes.
2. The second phase involved simulating actual data bursts from numerous EDs under a single BS utilizing one RAT (specifically, NB-IoT). This scenario replicates the necessity for all EDs to simultaneously establish connections and transmit data, which is often required in circumstances like smart-grid outages where time constraints apply to service delivery. The aim of this phase was to examine the network's response to batch arrivals (i.e., surges of RA attempts) and assess the impact on service delivery time.

4.1 On the Performance of NB-IoT and eMTC in Stable Environment

The first stage of the measurement campaign was conducted in a laboratory setting to assess the performance of both the NB-IoT and the eMTC across their entire operational range. These technologies were evaluated side by side, and the results pertaining to the RAN communication parameters, network throughput, and jitter under strictly defined radio conditions are presented. Thus, the RSRP levels ranging from −70 dBm to −125 dBm and −120 dBm for the NB-IoT and eMTC, respectively, which represent flawless radio conditions up to coverage edges. It is important to note that the reduced minimal signal levels compared to the 3GPP specifications, as outlined in Chapter 2, are due to the communication module utilized, which will be further discussed in the following subsection.

4.1.1 Measurement Setup and Methodology

The initial phase of the measurement campaign was carried out in a completely stable laboratory environment, with the ED situated in an RF shield box that was directly connected to the mobile operator's core network and remote server. The objective was to employ the *iPerf* tool set to evaluate UL data throughput and jitter in relation to a signal level that was adjusted using step attenuators. The data obtained from this first phase pertain solely to the state of the RRC connected.

This study utilizes a private eNB provided by the Vodafone teleco operator, which is equipped with NB-IoT (NB2) and eMTC (M1) technologies and is connected to public infrastructure. The local Remote Radio Unit (RRU) of the private eNB was then connected to a set of fixed attenuators via a coaxial cable linked to the RF shield box (Rohde&Schwarz CMW-Z10) using a testing device equipped with a selected Mobile IoT module. This module (ED) was operated by a controlling PC, allowing for full data traffic control as well as the extraction of communication parameters such as SIB information and communication time slots, which can be utilized for future reference.

The RF path's fixed attenuators are arranged in sequence to attain the desired signal level for a specific technology. In taking the measurements, two fixed attenuators with a single variable attenuator with 1 dB steps were used to achieve RSRP values between −70 and −140 dBm. This range was determined by considering the transmission power of the eNB of 27 dBm (500 mW). The overall layout of the testbed infrastructure is illustrated in Fig. 4.1.

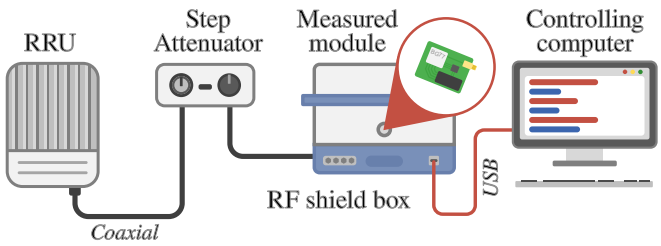


Fig. 4.1: Testbed for the first phase of the measurement campaign [5].

Selected Mobile IoT Module

The module selected for this measurement campaign was the Quectel BG77, which is based on Qualcomm 9205 LTE modem. This module supports both NB-IoT and eMTC standards and features 3GPP Release 14, with support for ED Categories NB2 and M1. The notable drawback of this module is its maximum power output, which is limited to 21 dBm, corresponding to a power class 5, which is common for the NB-IoT/eMTC module. This translates to a decreased link budget of 2 dB compared with the highest power class of 3, as reported in the literature [114].

The module demonstrates a receiving sensitivity of −115 dBm for NB-IoT and −107 dBm for eMTC when tested on the Band 20 (800 MHz) frequency. It's crucial to understand that these measurements are taken without repetitions, a scenario unlikely to occur in real-world networks. Our prior experience indicates that the typical sensitivity threshold is around −127 dBm. This suggests that networks employ 8 or 16 repetitions, as each doubling of repetitions enhances the budget by 3 dB. In this configuration, the repetitions yielded budget improvements of 9 and 12 dB, respectively. These findings align with previous research documented in existing literature [62, 114].

Measurement Methodology

The process began with the UE's registration to the mobile network, guaranteeing a connection to the eNB in the testbed. After successfully connecting, while still in the RRC connected state, the UE initiated UL data transmission to a remote server using the *iPerf* application. To obtain all relevant signaling and other information, the debug port of module BG77 and network capture to the PCAP were utilized. During the *iPerf* test, focus was given on comparing the observations of the throughput, jitter, and delay at various signal levels. To optimize the results, the *iPerf* target bandwidth was adjusted between the measurements to ensure channel saturation without overloading the module's internal buffer, which ultimately prolonged the measurement without providing any additional benefit. The procedure was repeated

for both the NB-IoT and eMTC.

For NB-IoT technology testing, an initial bandwidth of 50 kbps was employed. The speed was gradually reduced, with the bandwidth decreasing to 5 kbps, while RSRP values reached as low as −127 dBm. In the case of eMTC, testing began at 1200 kbps and was reduced to 200 kbps at the technology's limits. UDP transmission was chosen over more sophisticated transmission layer protocols to avoid result interference from additional retransmissions. To determine optimal values, five measurements were conducted for each signal level. The data was subsequently averaged and analyzed.

4.1.2 Measurement Results

The initial measurement outcomes demonstrate the constraints of both technologies regarding RSRP values. Initially, an assessment is made of the RSRP value threshold at which each technology can successfully complete network registration and transmit data to the network.

The application of NB-IoT technology was successful in registering and transmitting data with RSRP values as low as −127 dBm. However, consistent results were not obtained beyond this point. Additionally, as the RSRP values decreased, the latency increased from milliseconds to several seconds, even for small data units, which can be attributed to the severe radio conditions and the multiple repetitions required to successfully transmit the data. In contrast, eMTC was able to achieve registration at −123 dBm, but no data transmission was possible at this point. It is important to note that even in poor radio conditions, eMTC was able to transmit data with a relatively low latency of hundreds of milliseconds. It is crucial to recognize that NB-IoT consistently demonstrated a 10 dB stronger RSRP than eMTC under the same attenuation level, thus providing a 15 dB better MCL than eMTC.

Data Throughput Measurement

To evaluate the performance of each technology in relation to the RSRP measured by the BG77 module, the *iPerf* tool was employed. For NB-IoT technology, good radio conditions (RSRP exceeding −107 dBm) yielded an average speed of about 27.5 kbps. These findings align with the NB-IoT Release 13 specification. Fig. 4.2 displays the NB-IoT throughput measurement outcomes. The median value is represented by a solid line, while the shaded area indicates the 5^{th} and 95^{th} percentiles.

The results indicated that the UL throughput of the NB-IoT remained stable at an RSRP of roughly -106 dBm. As the RSRP value increased beyond this point, the throughput began to decline owing to deteriorating radio conditions, which necessitated the deployment of more robust modulation and coding schemes. This, in

turn, reduced the bandwidth of the transmission channel and increased the number of repetitions required.

Our measurements revealed that throughput decreased at an average rate of 1.18 kbps per 1 dB of RSRP. Notably, this decrease in throughput was consistent throughout the RSRP spectrum, even under severe radio conditions, as indicated by the ECL2 assignment assigned by the network. Our study found that the edge case for the NB-IoT was at an RSRP of −127 dBm, where a speed of 1.82 kbps was achieved. At this point, no data transfer was successful beyond this value.

The results of the eMTC technology measurements demonstrate a throughput of up to 1.02 Mbps, which is in line with the theoretically achievable speed of 1 Mbps defined in eMTC Release 13, under good radio conditions (greater than −103 dBm RSRP). The measured throughput remains stable up to an RSRP of −103 dBm, where the speed begins to decrease, ultimately reaching its limit of 51 kbps at an RSRP of −122 dBm. The slope appeared to be nearly linear, with an average decline of 50 kbps per 1 dB in the RSRP. However, no data transmission was achieved beyond an RSRP of −122 dBm. The outcome of the eMTC technology measurements is depicted in Fig. 4.3, where the solid line represents the median value, and the filled area represents the 5^{th} and 95^{th} percentiles.

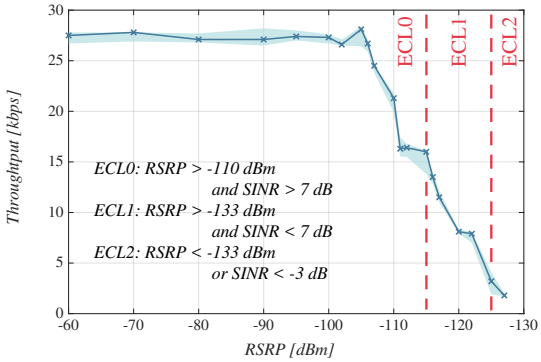


Fig. 4.2: NB-IoT measured data throughput [5].

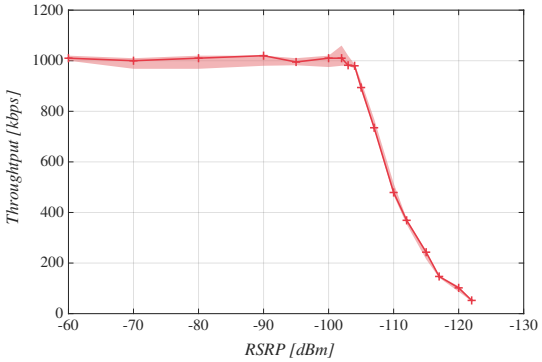


Fig. 4.3: Measured data throughput of eMTC [5].

Jitter and Delay Measurement

By employing *iPerf* and seizing the network interface, an experimental assessment of jitter and delay measurements furnished by the application after each speed evaluation was conveyed. The corresponding average jitter and communication delay values are shown in Figs. 4.4 and 4.5 for NB-IoT and eMTC, respectively.

In accordance with the specifications, the average jitter for the NB-IoT technology in normal radio conditions was approximately 230 ms, with a delay of roughly

one second. However, measurements taken at the technology limits revealed average delay values as high as 1250 ms. A noticeable departure from the expected curve that follows the speed plot can be observed when the device is assigned to a different ECL class, as shown in Fig. 4.4. This deviation is attributable to the BG77 module altering the modulation schemes based on the ECL class assigned to it by the RRU and the network. This modification enhances the signal integrity and decreases the number of repetitions necessary for successful data transmission to the network.

In the context of eMTC, the jitter is directly proportional to the throughput curve. As the speed slows down, the jitter increases, reaching approximately 250 ms, which is the limit in our case. Under normal radio conditions, the average jitter for eMTC is approximately 7 ms. For the average delay, there is a noticeable rise at the −116 dBm RSRP signaling switches to a more robust MCS and an increase in repetitions. However, this did not affect the throughput. Nevertheless, the percentiles show visible variance, indicating more repetitions and retransmissions at the edge of coverage, which is linked to a decreased throughput.

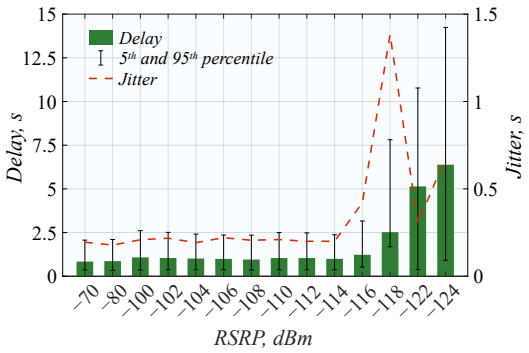


Fig. 4.4: UL delay, jitter of NB-IoT under different radio conditions [1].

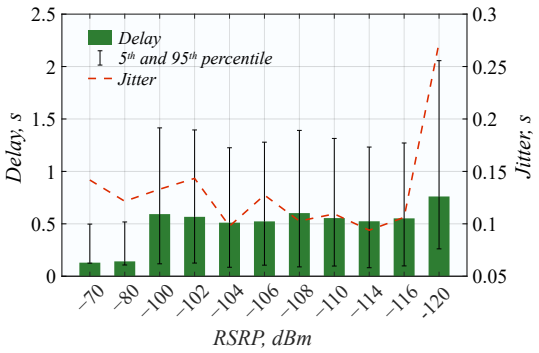


Fig. 4.5: UL delay, jitter of eMTC under different radio conditions [1].

4.1.3 Measurements Conclusion and Lessons Learned

Our analysis revealed that both NB-IoT and eMTC technologies achieved the highest possible throughput across their entire signal range until the point at which the performance of legacy mobile technologies began to decline. Although these technologies are capable of providing communication services beyond this threshold, it is important to consider certain drawbacks when specific communication requirements are necessary. As stated in Chapter 3, the smart-grid use case requires a guaranteed minimum throughput of 11 kbps, as presented in Table 3.5. It is important to note that these criteria are met by both technologies, with eMTC meeting them throughout its entire range and NB-IoT meeting them up to the end of the ECL1

range if the average values are considered. It should be noted that the smart-grid throughput criteria were met for both technologies in roughly the same MCL range owing to a plus 10 dB improvement in RSRP for NB-IoT, as previously mentioned. However, it is essential to emphasize that this assumption is based on perfect radio conditions and unloaded networks.

As outlined in Chapter 3, it's crucial to note that smart-grid specifications for remote smart meter reading communication delay (RTT) require a two-second threshold. This requirement is only fulfilled in the ECL0 zone using NB-IoT, and with eMTC up to a signal strength of $-120\,\text{dBm}$. Given a $+10\,\text{dB}$ advantage for NB-IoT, both technologies satisfy the criteria within the same MCL range. However, the extreme scenario combining the 95^{th} percentile delay with the potential RRC connected state transition time, as seen in the measurement campaign's second phase (refer to Table 4.2), indicates that NB-IoT cannot ensure the RTT criteria across its entire operational spectrum. For eMTC, only under optimal signal conditions above $-90\,\text{dBm}$ does the worst-case scenario meet the requirement. Nonetheless, the broader 10 s communication delay criteria defined by ITU-R M.2410 [24] make both NB-IoT and eMTC more appropriate for use cases adhering to these standards, rather than stringent smart-grid applications. Yet, in NB-IoT's worst-case scenarios below $-122\,\text{dBm}$, even this 10 s limit might be exceeded.

The presented information emphasizes the limitations of each ECL category in terms of delay and jitter for NB-IoT at $-120\,\text{dBm}$ and $-130\,\text{dBm}$. While the delay gradually increases as the ECL changes, the jitter appears to be significantly affected at the boundaries of each ECL category due to the constraints of the employed MCSs and repetitions.

A crucial question was posed based on the understanding of the behavior and performance limitations of an unloaded network. What measures would be taken in the event of a significant number of EDs simultaneously vying for access to limited radio resources?

4.2 Extensive Testing of NB-IoT for Smart-Grid Use-Case

After conducting an assessment of the technological limitations within a controlled environment, we proceeded to execute a real-world scenario with 200 NB-IoT EDs deployed under a single eNodeB and subjected them to a network stress test. The number of EDs corresponds to the estimated maximum number of devices that can be serviced simultaneously by a single cell, as calculated by the network operators.

Our objective in conducting this experiment was to simulate two different communication configurations, which are as follows:

1. *Normal operation:* The first scenario follows the telemetry communication protocol outlined in Section 3.2.1. It resembles the typical sensor data traffic for which NB-IoT was originally developed, but with increased volume and frequency of data transmission.
2. *Blackout:* The second scenario addresses two key smart-grid requirements: (i) the ability to provide services even during a *blackout*, and (ii) the need for robust security measures. Specifically, this configuration was designed to simulate a scenario where all EDs must concurrently establish secure communication channels using TLS when they start up, and register substantial amounts of data with the mobile network.

The substantial volume of burst data noted in the *blackout* scenario indicates that NB-IoT may not be the most suitable technology, given its constrained bandwidth. Furthermore, exposing the NB-IoT network to such conditions could result in radio channel congestion, impacting not only users needing to process this surge of data but also those employing NB-IoT for conventional sensoric purposes. In the following sections, we outline our measurement approach and the findings we obtained.

4.2.1 Measurement Setup and Methodology

In the second phase of our campaign, we utilized a setup comparable to the initial phase of our measurements. Specifically, we replaced the RF shield box with an indoor antenna within the laboratory, creating a private NB-IoT pico cell. This method, characterized by a restricted cell range, minimized interference from other mobile operator customers' EDs in the vicinity, as we operated on a live mobile operator's network. By positioning all 200 of our EDs within the laboratory space, we ensured their connection to the monitored laboratory cell. Consequently, this arrangement allowed for interference between the EDs being tested, imposing a significant load on the monitored laboratory cell. The experimental setup for this measurement campaign is shown in Fig. 4.6.

Before conducting the measurement scenario, we performed a laboratory radio condition check using a single device to establish reference signal conditions in the absence of interference. The test results showed excellent radio conditions, with an average RSRP of $-50\,\mathrm{dBm}$ and an average SINR of $20\,\mathrm{dB}$.

After conducting an initial test and prior to the 200 ED stress testing, a comprehensive examination of the smart-grid use-cases detailed in Chapter 3.3 was carried out. Specifically, two BG77 modules were utilized, with NB-IoT representing both device types. The Type I device simulated the on-demand reading of DLMS protocol

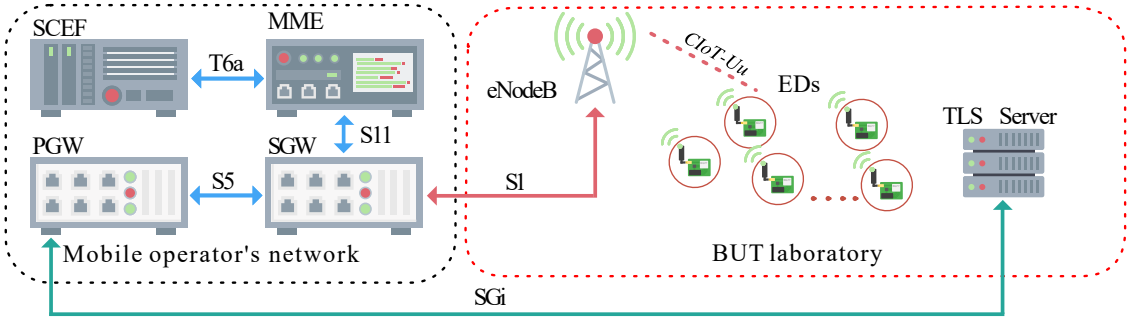


Fig. 4.6: Testbed for the second phase of the measurement campaign [1].

values employed by electricity meters, while the Type II device acted as a traditional sensor, transmitting 200 B messages to a remote server at 1-hour intervals via the UDP on the transport layer. The device then entered a PSM for the remaining time before waking up and performing a RRC connection and TAU prior to transmitting the next message. These actions were taken to mimic the DLMS behavior and are illustrated in Fig. 4.7.

We gathered information about the behavior of the UE by examining its debug serial port, which provided us with precise timing assumptions for our model. We utilized a Blade RF 2.0 Software-Defined Radio (SDR) to measure the radio interface activity for a period of 24 hours. Our spectrum analysis showed that no interference was induced by other devices. The findings from this phase are crucial inputs for the subsequent chapters.

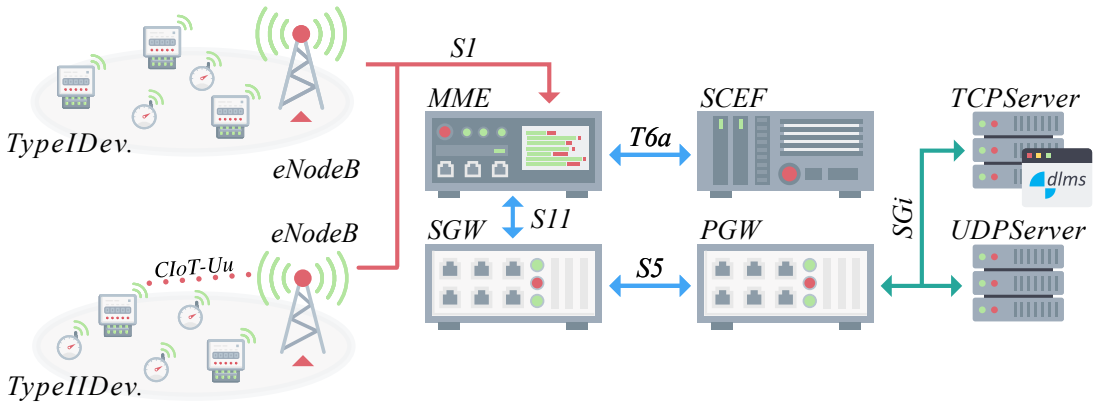


Fig. 4.7: Diagram representing measurement scenario for test of Type 1 and Type 2 devices [2].

After the two initial measurement steps followed experiment enabling us to obtain data about the NB-IoT network, including transmitted service data from EDs to the TLS server and subsequent service data provided by a mobile operator from

the eNodeB. Both *normal* and *blackout* conditions were monitored at different time scales. During *normal* operation, the test duration was seven days, during which we collected hourly averages of processed data from the eNodeB and per-hour samples of measured data from the EDs transmitted as approximately 100 B telemetry messages. Moreover, each ED had to perform at least a single registration to the mobile network on a daily basis, owing to operator-specific mobile network session regulations. Due to the difficulty of emulating a *blackout*, this condition was tested only once for a duration of two hours with the same sample rate as in *normal* operation. Upon initial startup after the power supply *blackout*, all EDs were required to complete mobile network registration, establish TLS, and transmit telemetry data. As in case of *normal* operation test, the results were gathered from the BS statistics and from time of TLS tunnel establishment and from arrival of telemetry data from each ED.

4.2.2 Measurement Results

The main characteristics of the signal obtained from the ED service data after successful transmission and processing by the server are displayed in Table 4.1. The data reveals that in both *normal* and *blackout* situations, the signal demonstrates extensive variation, even though all measurements were conducted in a controlled laboratory environment. This is evident from the significant range of both RSRP and SINR, with more than 30 dB of variation in both conditions. Although the signal level in *normal operations* exhibits a better upper limit in terms of RSRP, the signal characteristics are comparable between the two scenarios. However, the average signal level is about 30 dB lower than the reference signal conditions measured before the test with a single ED. This suggests that there is a similar level of inter-ED interference in both scenarios.

Tab. 4.1: Communication parameters obtained from EDs.

Parameter	**Normal conditions**		**Conditions after blackout**	
	range	**avg.**	**range**	**avg.**
RSSI [dBm]	−59 to −108	−74	−65 to −101	−72
RSRP [dBm]	−44 to −127	−84	−76 to −115	−84
SINR [dB]	19 to −15	7	17 to −7	6
ECL0 EDs [%]	84.4		75	
ECL1 EDs [%]	14.8		19.2	
ECL2 EDs [%]	0.8		5.8	

According to the data from the EDs, the distribution of EDs across ECL classes significantly affected signal variance. In the aftermath of the *blackout*, data bursts

led to network congestion and interference, resulting in a 5% increase in the utilization of robust ECL schemes 1 and 2. This shift could potentially cause longer transmission times, reduced throughput, and overall performance degradation. Typically, ED results would be used to adjust system settings, but an important observation emerged. The signal conditions were recorded after EDs successfully registered with the mobile network, a process that could take several minutes due to radio channel overload. More accurate data was obtained from the eNB, which continuously reported ED ECL levels. Table 4.2 illustrates how delayed network registration led to misleading information about ECL distribution following the *blackout*. Despite apparently good signal conditions in the laboratory, the ECL distribution was severely impacted, with over half of the EDs operating in a more robust ECL, consequently reducing throughput.

Tab. 4.2: Communication parameters obtained from eNB.

Parameter (hourly averages)	**Normal operation**	**Conditions after blackout**
Avg. simultaneously served EDs	22	41
Max. simultaneously served EDs	77	105
RA attempts	5,942	50,236
RA responses	98% (5,809)	28% (14,040)
RA contention resolution	65% (3,844)	16% (8,066)
RRC establish attempts	3,441	4,207
RRC establish success rate	99 – 100%	69% then 100%
ED distribution in ELC 0 / 1 / 2 [%]	86 / 12 / 2	45 / 29 / 26
RRC est. time ECL0 avg./max [s]	0.1 / 14.7	3.1 / 34.1
RRC est. time ECL1 avg./max [s]	0.2 / 10.7	4.3 / 32.2
RRC est. time ECL2 avg./max [s]	0.9 / 6.8	8.8 / 34.9
Interference max. [dBm]	−109	−94
Interference avg. [dBm]	−119	−108

While the ECL distribution after the *blackout* seems to indicate a notable increase in utilized repetitions and more robust MCSs, our findings regarding the NPUSCH and NPDSCH data channels, as shown in Fig. 4.8 and Fig. 4.9, do not support this assumption. Regarding repetitions, there is an increase in the usage of two and 16 repetitions for NPUSCH and NPDSCH, respectively, but this occurs only about 20% of the time. As for MCS settings, both scenarios exhibited similar significant use of robust MCS settings, with values of 7 for NPUSCH and 11 for NPDSCH.

According to the data channel indicators, there was a modest enhancement in the MCS's robustness and repetitions. The ECL distribution presented in Table 4.2

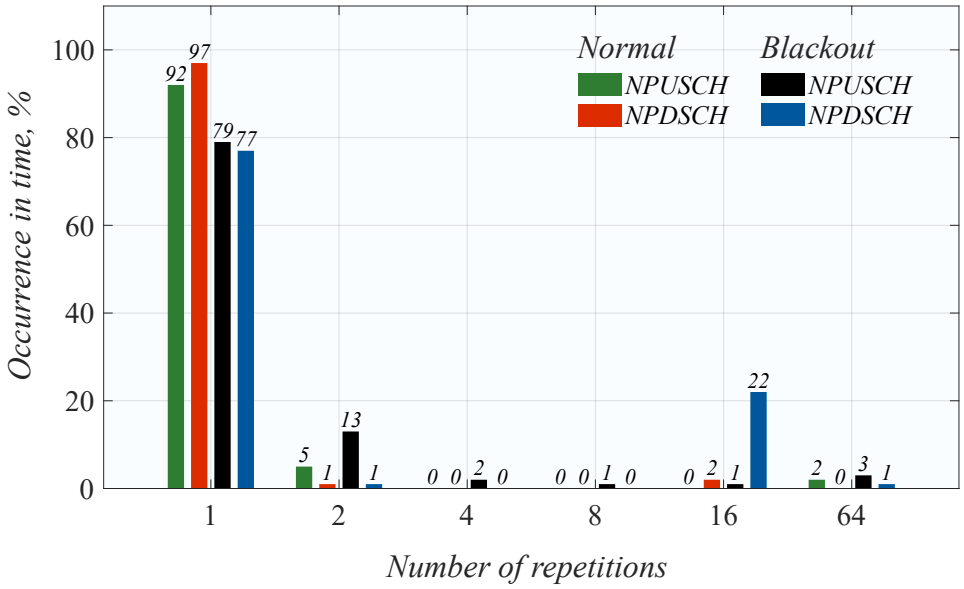


Fig. 4.8: Distribution of NPUSCH/NPDSCH repetitions [1].

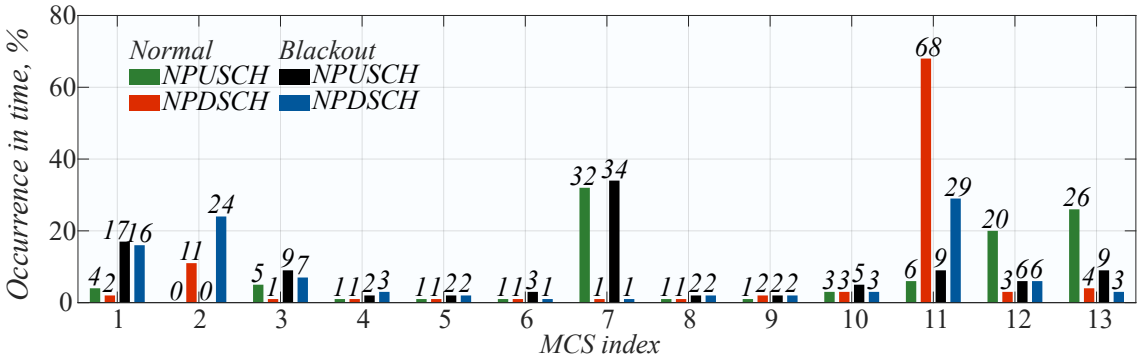


Fig. 4.9: Distribution of NPUSCH/MPDSCH MCS [1].

indicates a more pronounced effect on the RA and RACH communication stages, as well as the DCCH. This theory is corroborated by the almost nine-fold rise in hourly RA attempts between *normal* and *blackout* states. Even under *normal* circumstances, only 65% of RA attempts were successful. Nevertheless, all successful RA attempts were followed by a 100% success rate in RRC establishment and/or data transmission. The surge in RA attempts during the *blackout* period resulted in a drastic reduction of the RA success rate to a mere 16%, accompanied by a decline in the RRC connection attempt success rate to 69% in the initial hour post-*blackout*. This combination led to a substantial increase in the time needed for RRC connection establishment, shifting from hundreds of milliseconds to several seconds on average, with extreme instances reaching 30 s.

The events that transpired following the *blackout* led to significant congestion

of mostly signaling channels, resulting in significant transmission delays. This is demonstrated by Fig. 4.10, which displays the time it took for all EDs to synchronize and establish a TLS connection after the *blackout*. The graph shows the time it took for each ED to initiate the first TLS connection, which typically involves the transmission of roughly 100 Bytes of data. The times of TLS initiation and establishment are also indicated. Unfortunately, we were unable to obtain accurate information about the time required for each ED to register with the network. Therefore, we indicate the time at which all EDs were successfully connected, as indicated by a light-emitting diode (LED) on each device.

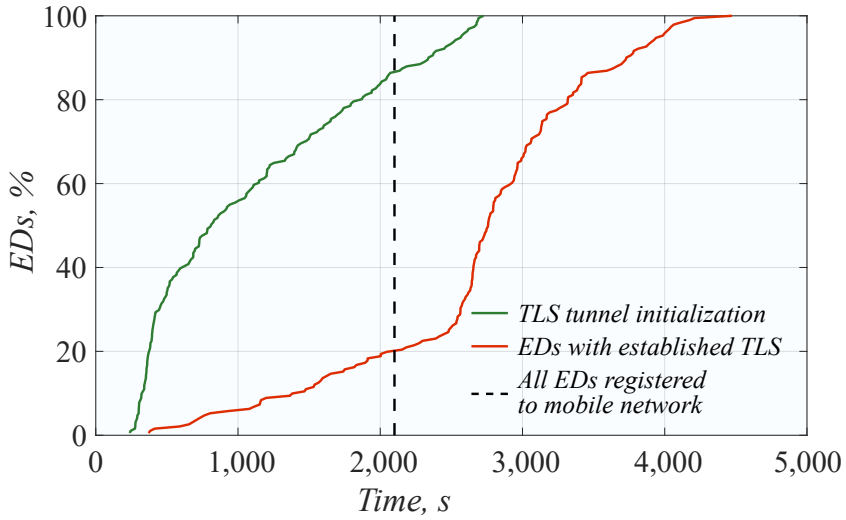


Fig. 4.10: Reconnection time of 200 NB-IoT EDs following a blackout [1].

4.2.3 Measurements Conclusion and Lessons Learned

Analyzing the outcomes of stage two against ideal conditions for a single ED reference test reveals that connecting to the mobile network and initiating a TLS connection within ECL0 range typically requires several seconds to a few tens of seconds. However, data from the *blackout* illustrated in Fig. 4.10 shows a significantly longer process. The first ED took several minutes to register and send a single data message, starting the TLS connection procedure, while 90% of EDs required 40 minutes to accomplish this task. At the point when 90% of EDs had begun the TLS connection process, only 20% had successfully completed it. An additional 45 minutes were necessary for all EDs to finalize their TLS connections. The observed increase in successful TLS connections 45 minutes after the *blackout* suggests a reduction in radio channel load, likely due to the completion of mobile network registrations.

The long intervals appear to be closely related to substantial increases in RA attempts, as previously indicated, which is additionally associated with neither ef-

ficient nor distributed data scheduling from the ED side. Another significant factor that may affect the situation is the best-effort QoS configuration of the mobile network. It is crucial to highlight that the outcomes of both the *normal* and *blackout* scenarios should be comparable to other events where bursts of data occur, such as synchronous remote electricity meter readings or over-the-air firmware updates, as suggested by a substantial amount of data per month (outlined in Chapter 3).

4.3 Summary

The findings and insights of this chapter are based on an examination of **RQ1** and **RQ2**. Based on the initial stage of the measurement campaign, the NB-IoT technology generally exhibits better signal coverage compared to the eMTC technology. This result is anticipated, as NB-IoT technology boasts a narrow bandwidth of 180 kHz (including guard bands) and employs dynamic allocation of robust modulation and coding schemes based on the ECL class designated by the network. Consequently, the NB-IoT technology can maintain data transmission in more adverse radio conditions than the eMTC technology by approximately 15 dB RSRP. Despite this advantage, the NB-IoT fails to fulfill the smart-grid requirements of guaranteed 11 kbps and 2 s communication delay defined in Chapter 3 at RSRP levels surpassing ECL0. In this regard, eMTC technology offers the considerable benefit of higher transmission speeds, reaching up to 1 Mbps. These speeds are attainable due to the wider bandwidth of the technology, which measures 1.08 MHz. However, because of the broader bandwidth, the eMTC technology throughput begins to decrease earlier as the TX power has to be disseminated across the entire spectrum, causing it to reach its capacity sooner than the NB-IoT, although it still surpasses the NB-IoT.

From a broader perspective, it cannot be stated that one technology is inherently better than the other, as the choice of technology is largely contingent on the specific use case. For instance, when it comes to battery-powered devices that are situated in areas with poor radio conditions, such as rural locations or cellars, and where high traffic or low latency is not a requirement, NB-IoT is the preferred choice owing to its wider coverage compared to eMTC. However, in situations where high-traffic, low-latency applications such as VoLTE or low-frequency images are involved, eMTC is the more suitable option, as it provides higher data throughput with low latency (ideal radio conditions with a latency of less than 100 ms).

Initial tests under underloaded eNB conditions with a single ED showed that eMTC meets the requirements for smart-grid applications, making it suitable as the primary interface, NB-IoT is only appropriate as a backup for service data or during eMTC congestion or severe outages. This insight is valuable for network operators

seeking to optimize these technologies within their communication systems. Moreover, the findings could benefit various use case studies or parameter configurations that require an understanding of system performance, even in non-ideal situations. Additionally, this study utilizes data from the initial phase to derive essential parameters for mathematical models, as detailed in subsequent sections. This conclusion addresses RQ1.

This stage of the study aims to determine whether the current technology deployment is adequate for the intended usage or if there are opportunities for optimization to make such scenarios feasible in real-world applications. As previously suggested, the coexistence of multiple interfaces, such as NB-IoT and eMTC, in the form of a Multi-RAT device may be a viable solution for balancing the data load during bursts and addressing the performance gaps of both technologies when used individually. While NB-IoT has a limited coverage gap, eMTC has a more limited radio range compared to NB-IoT.

During the second phase of the measurement campaign, the eNodeB stress test employing NB-IoT further confirmed that the current configuration of NB-IoT is insufficient to handle the burst or batch arrival type of traffic commonly encountered in smart-grid applications. Furthermore, it is unable to manage the increased volume of data transmitted. These findings confirm that eMTC is the preferred option for smart-grid applications. The insights gained from the first phase of the campaign, in conjunction with observations from the second phase, provide a solid foundation for Multi-RAT EDs operating with a selective or parallel design when both NB-IoT and eMTC are available. The results that identify the bottleneck in the RA procedure congestion caused by the batch arrivals of smart electricity meters also raise a second point for consideration. Even in single-RAT deployments, smart meters that generate new traffic types, as defined in Chapter 3 and confirmed in this chapter, must coexist with the traditional IoT sensors. This raises the question of how to balance these two types of devices effectively. The bottleneck and required coexistence address **RQ2** and open the door for further research on **RQ3** and **RQ4** in the subsequent chapters.

5 Latency Improvements for 5G IoT Enabled Smart-Grids

This chapter delves into the modeling and optimization of Single-RAT 5G IoT solutions, with a specific focus on their application within smart-grids, thus addressing **RQ 3** in terms of latency improvements. The integration of smart meters and other IoT devices in energy distribution networks presents unique challenges, owing to their specific traffic patterns and stringent reliability requirements. NB-IoT, with its advantages of low power consumption and extensive coverage, is well-suited for such applications. However, as discussed in Chapters 2 and 3, the theoretical underpinnings of technology selection and the defined traffic types necessitate practical models and optimization strategies to ensure efficient network performance. Chapter 4 further emphasizes the need for such optimization to address the growing demand for smart-grid applications. This chapter aims to present both immediate and future-ready solutions for enhancing the performance of the NB-IoT.

The research methodology involves the development of analytical models and their subsequent validation through extensive simulations. A Markov chain model is employed to characterize the traffic patterns of NB-IoT smart meters, followed by the application of optimization algorithms to improve system performance. The results from the simulations are analyzed to provide insights into the practical implications of the proposed models and optimization strategies.

5.1 Comprehensive Modeling of NB-IoT Smart Metering Stochastic and Regular Traffic Types

Systems with random access have been known since the 70s of the last century. The first wave of studies considered systems with many UEs, each with low message intensity. By tending the number of UEs to infinity while the message intensity per UE to zero such that the overall message intensity is constant, the message arrival process in such a system is approximately Poisson in nature. Such a system with random access, an infinite number of UEs, and a Poisson arrival process was first studied in [115, 116]. In [117], two methods suitable for such systems were proposed: (i) the Markov chain approach and (ii) the diffusion approximation. Several studies have analyzed the performance of NB-IoT and eMTC using various analytical and simulation-based models. The most commonly used approach is the Markov chain model, which provides a tractable framework for characterizing traffic dynamics and access performance. However, alternative methodologies such as queueing theory,

stochastic process models, and machine learning-based approaches have also been explored in the literature [118, 119].

The initial generation of mMTC technologies, such as NB-IoT, was developed based on the assumption of completely random, asynchronous message arrivals from UEs, as specified in the reference test model in ITU-R M.2410. Only recently has the research community become interested in random access systems under batch traffic arrival patterns. These studies were motivated by pure research interest in evaluating the response of random access mechanisms to synchronous access attempts. In contrast, a second wave of interest has arisen from emerging first-generation mMTC technologies, including EC-GSM, SigFox, LoRaWAN, IEEE 802.11ah, and NB-IoT. Notably, the authors in [120–122] investigated different random access mechanisms under batch arrivals and found that the message transmission delay increases exponentially with the size of the batch for the same arrival traffic intensity.

The paper by [123] explored contention-resolution random access algorithms in the context of batch arrivals of messages. The authors developed a model based on the application of the Markov chain modeling approach, which showed a good match with the computer simulations. However, these random access algorithms were not used in the mMTC scenarios. Another study [124] examined a similar random access mechanism and provided guidelines for designing access probabilities and an analytical calculation of error probability. A logical continuation of this work is presented in [125], in which the authors investigated random access in LTE-A technology with NOMA enhancements. They compared the reception of pilot signals for Successive Joint Decoding (SJD) and Interference Cancellation (SIC) during the random access phase.

An accurate representation of batch arrivals can be found in [126], where the authors explicitly assumed batch arrivals from multiple independent sources. The impact of batch arrivals on IEEE 802.11ah systems was explored in [127, 128], while [129] examined a particular instance of batch arrivals in satellite communications. In the context of access barring, the performance of eMTC and NB-IoT technologies under batch arrivals was analyzed in [41]. It is worth noting that most of these studies assumed the presence of only one type of traffic at the air interface, and the number of UEs was often considered to be finite. This assumption enables the formulation of performance evaluation models using Markov chains with finite transition probability matrices, which can then be analyzed using standard methods.

Recently, there has been renewed interest in mMTC access schemes, owing to the emergence of new applications. In particular, the authors of [130] proposed a Delay-Aware Priority Access Classification (DPAC)-based Access Class Barring (ACB) scheme for LTA-A random access preamble congestion control aimed at delay-constrained mMTC devices. The proposed method, with further RL enhancements,

showed an increase in the probability of preamble success compared to the ACB approach. However, in constrained scenarios with limited preambles and a large number of preamble arrivals, the tradeoff resulted in higher delays and increased device drops compared to ACB and other traditional static approaches. In terms of power consumption optimization, the authors of [131] investigated wake-up-based DL access for delay-constrained devices in 5G. The proposed WUS can coexist with DRX schemes and achieve the desired power savings in DRX periods. This approach seems more suitable for mMTC nonperiodic traffic because it is more robust and agnostic by the traffic type. In contrast, the authors in [132] introduce power and resource saving by implementing Mobile Edge Computing (MEC) and implementing a Digital Twin (DT) edge network aided by Deep Reinforced Learning (DRL). This approach is promising in terms of resource offloading and delay optimization during device mobility.

Several studies have examined random access issues in the context of mMTC communication, focusing on the refinement of backoff mechanisms and retransmission limits. The authors of [133] investigated access barring mechanisms to mitigate congestion but assumed a single traffic type. Similarly, the authors of [134] analyzed access probabilities under various retransmission strategies, but did not consider the impact of mixed traffic patterns.

The research presented in [135] showed that even standard LPWAN traffic can display batch-like characteristics due to aggregation. This phenomenon is linked to the occurrence of bursts or clustering in typical Poisson processes [136], as well as in traffic aggregates where sources are not completely asynchronous. Although their model is relatively basic and doesn't consider the specifics of random access procedures, it effectively broadens the applicability of the model proposed in our study. In contrast to [135], our research focuses on burstiness resulting from the specific operation of application layer protocols that regularly poll UEs. Furthermore, we take into account the random access phase and demonstrate that it becomes the limiting factor when a combination of conventional Poisson traffic and bursty traffic is processed at the air interface.

The aforementioned review highlights the fact that systems under mixed stochastic and regular arrivals in Mobile IoT context have not yet been considered in detail. Furthermore, no studies have addressed the performance of the NB-IoT system, recently accepted by ITU-R as an enabling technology for 5G mMTC under such load conditions.

5.1.1 System Model

In this section, we outline the system model by delineating its constituent components, which include deployment, access procedures, and traffic patterns. Subsequently, we present our presuppositions and delineate the performance metrics of focus.

Deployment Model

We analyze an individual NB-IoT cell that has only one RB set aside for operation. The NB-IoT operation is thought to occur in either the guard band or the standalone mode within the licensed frequency spectrum. Moreover, we assume that the telco operator has adequately deployed the network so that there are no other UEs in the cell with extremely poor channel conditions. Consequently, we consider the packet loss due to incorrect reception to be negligible. In these situations, the access phase emerges as the primary cause of performance degradation.

Traffic Types

The considered NB-IoT system accommodates two categories of User Equipment (UE), as illustrated in Fig. 5.1. The first category, Type I UEs, are continuously monitored by a remote application server, maintaining an active radio interface and remaining in a "RRC connected" state throughout their operational lifespan. The second category, Type II UEs, follows a typical "sleep-awake-transmit" asynchronous cycle. In scenarios where the UE count approaches infinity and the message generation probability per UE is near zero, which is typical for asynchronous sensors in Mobile IoT systems, collective Type II UE traffic can be modeled as a homogeneous Poisson process with a rate of λ_2 messages/s [115, 116, 137]. This λ_2 parameter represents the combined load from Type II UEs, which can be interpreted in two ways: it may reflect varying UE densities within the base station (BS) coverage area generating messages as per the ITU-R M.2410 recommendation or it could signify different message generation frequencies for a single UE given a specific UE density.

Random Access Phase in NB-IoT

According to NB-IoT specifications [138], as illustrated in Fig. 5.1, the UE is expected to identify the NB-IoT carrier through measurement of received synchronization signal power in the DL direction, while simultaneously achieving time and frequency synchronization and decoding the CellID. The interval between repetitions of synchronization information can range from 24 to 2604 ms [139]. Following this, the UE decodes the NPBCH, which contains the MIB, over a 640 ms TTI. The

cell characteristics are then transmitted via SIB1-NB over 2560 ms, with additional SIB2-NB information provided by the BS. More comprehensive details about this process can be found in [22].

After synchronization, the UE is able to set up the NPRACH and conduct UL preamble transmissions in accordance with network parameters, ensuring sufficient repetitions and transmit power. Preamble repetitions can range from 1 to 128. Within a repetition unit, a preamble follows a predetermined hopping pattern and is composed of four character groups. Each group contains five characters and a cyclic prefix (66.67 μs or 266.7 μs for 10 or 40 km cell radius, respectively). Consequently, the duration of a random-access attempt spans from 5.6 ms to 819.2 ms. In case of an unsuccessful preamble transmission, the UE adheres to the random backoff timer and retransmission limit set by the network [140]. The BS, upon receiving the signal, can rectify frequency and time offsets, and determine the timing advance (TA) for subsequent transmissions. NB-IoT designates a minimum of $l = 12$ orthogonal preambles (subcarriers) out of 48 available. The NPDCCH initiates the data transfer phase by transmitting DL Control Information (DCI). This signal can be repeated between 1 and 2048 times [79].

After the initial access phase, the transition from the RRC IDLE to the RRC Connected state and subsequent message transmission are identical for both Type I and Type II devices. However, the two device types differ in their subsequent behavior. While the Type I UE maintains a permanent RRC connection, the Type II UE releases the RRC connection after each message transmission. Therefore, if a Type II UE wants to transmit data, it must go through an RRC resume procedure that includes another RAP, followed by requesting UL/DL grants by completing the RAP and message transmission. In contrast, Type I devices only need to complete a single RAP to acquire UL/DL grants with message transmission [79].

Modeling Assumptions

A system that employs slotting, in which time is divided into consecutive intervals, is depicted in Fig. 5.2. Each slot duration is $\delta = 10$ ms, which corresponds to the duration of an LTE frame and represents the minimum time required for random access without repetition. The message transmission process in Mobile IoT systems, including the NB-IoT, comprises three phases: synchronization, random access, and data transmission, as previously outlined. The synchronization phase has a constant duration of t_S. Contemporary Mobile IoT systems, which typically handle small uplink data transmissions as per ITU-R M.2410 specifications, are engineered to prevent bottlenecks during the data transmission phase. Given this design and our focus on a combination of regular and stochastic traffic, we posit

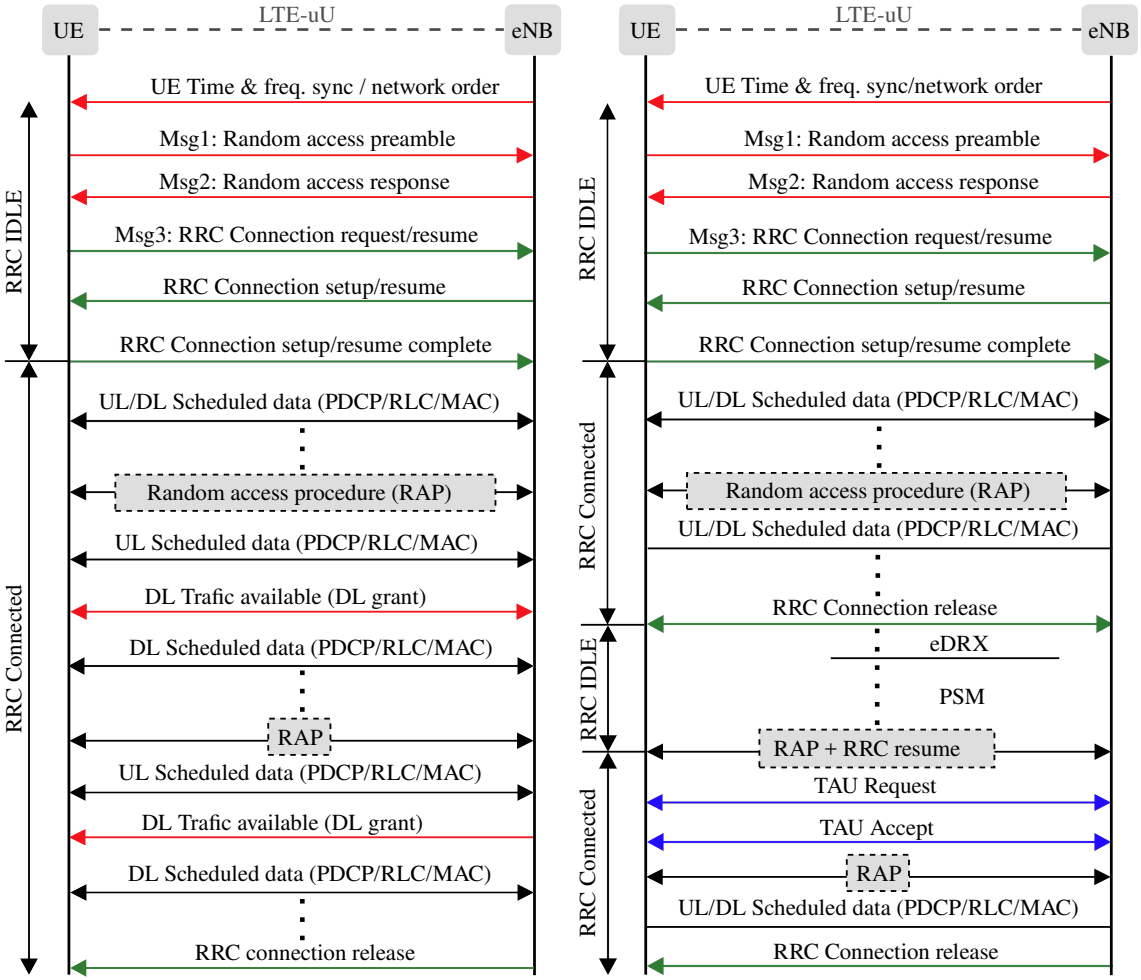


Fig. 5.1: NB-IoT access: left – *Type I* UEs; right – *Type II* UEs.

that the random access phase represents the primary bottleneck. Consequently, we assume a fixed data transmission time, t_{data}, while the random access duration t_R is variable. Accordingly, Fig. 5.15 illustrates only the parameters associated with the random-access phase, excluding the backoff mechanism and the retransmission limit. In addition, we factor the message-processing time t_H. The specific values of these parameters align with both the NB-IoT standard and our empirical measurements in Chapter 4. The values of these parameters are provided in Section 5.1.4.

It is worth noting that there are a fixed number of Type I UEs that fall under cell coverage. These UEs are queried by applications at the beginning of every regular time interval, referred to as the period, which lasts for a duration of Δ. The length of the period depends on the type of application, and can range from minutes to hours. Upon receiving the request, all Type I UEs attempt message transmission by performing the NB-IoT random-access procedure.

When generating a message, both UE categories commence the random-access

process in the following time slot. The hypothesis is that that the probability of beginning the random access procedure is $p = \min(n/k, 1)$, where n signifies the quantity of preambles, and k denotes the number of UEs competing in a single time slot. As demonstrated in [141, 142], this methodology reduces the channel access delay experienced by UEs.

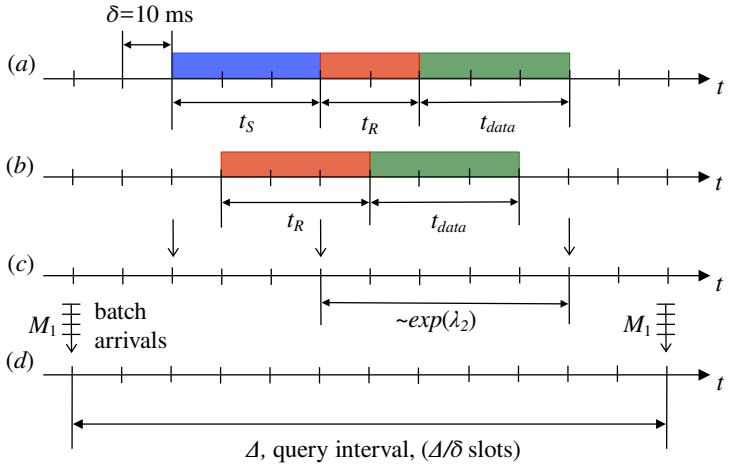


Fig. 5.2: System operation timeline: (a) Type II UE activities, (b) Type I UE activities, (c) Type II UE arrival pattern, and (d) Type I UE arrival pattern [2].

Metrics of Interest

The aim is to evaluate the average delay for both Type I and Type II traffic, specifically $E[\tau_1]$ and $E[\tau_2]$, respectively. The analysis considers complete delay, which encompasses the time required for synchronization, random access, and data transmission.

5.1.2 Performance Analysis

In this section, a two-dimensional Markov chain framework is introduced to model the coexistence of Type I and Type II user equipment in an NB-IoT cell. To simplify the model and reduce its computational complexity, a state aggregation technique is employed. Subsequently, a solution algorithm is presented, followed by an evaluation of the desired metrics.

The Exact Model

The model discussed in Section 5.1.1 is best described by a two-dimensional Markov chain comprising of N_t^1 and N_t^2 which represent the number of Type I and Type II

UEs in the system at time t, respectively. This process occurs over a discrete index set with a slot duration of Δ/δ, where Δ represents the number of slots between packet arrivals from Type I UEs, and δ is the time slot duration.

The structure of the state-transition diagram for the two-dimensional model is shown in Fig. 5.3. Note that the transitions are limited by l states that correspond to the maximum number of available preambles. This is due to the fact that no more than l messages can be processed simultaneously, as there are restrictions on the number of orthogonal channels in NB-IoT. We can establish the following relation for Type I UEs:

$$N_{t+1}^1 = N_t^1 - T^1(N_t^1, N_t^2), \tag{5.1}$$

where the quantity $T^1(N_t^1, N_t^2)$ represents the number of Type I UE messages sent and N_t^2 denotes the number of Type II UEs that are still present in the system at the end of the previous interval Δ.

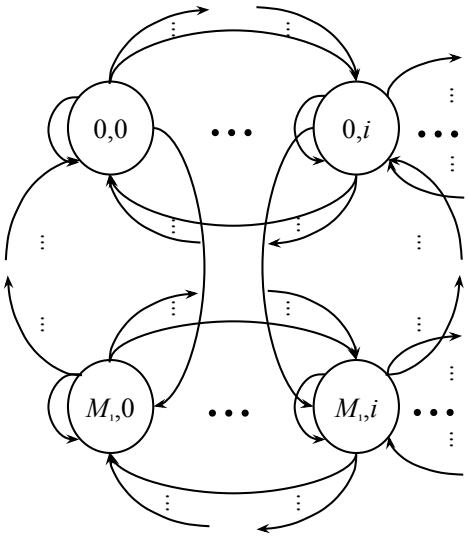


Fig. 5.3: State transition diagram of two-dimensional model [2].

Similarly, for Type II UEs we have

$$N_{t+1}^2 = N_t^2 - T^2(N_t^1, N_t^2) + V_t^2, \tag{5.2}$$

where the number of Type II UEs remaining in the system at the end of the previous time period Δ is denoted by N_t^2. This is to be compared with the number of Type II UEs that become active during the same time period, which is represented by V_t^2. Finally, $T^2(N_t^1, N_t^2)$ represents the total number of Type II UE messages that were successfully transmitted during the time period.

Note that Equation 5.1 demonstrates the connection between the quantity of Type I UEs and the adjacent Markov points necessary for characterizing the Markov

chain in accordance with the standard procedure. Additionally, it is essential to calculate the probability $Pr\{N_t^1 = n \mid N_{t-1}^1 = m\}$ using the relationship in Equation 5.1. This principle also applies to Equation Equation 5.2. The most challenging aspect of determining these probabilities is ascertaining the values of $T^1(N_t^1, N_t^2)$ and $T^2(N_t^1, N_t^2)$, which indicate the number of Type I and Type II messages processed between two Markov time points. Both these values are contingent upon the number of Type I and Type II UEs at the preceding Markov point.

To effectively address the proposed model, one must construct a transition probability matrix, employ it to solve a system of linear equations, and determine the stationary probabilities of states within a two-dimensional Markov chain. These stationary probabilities will further enable us to estimate system parameters, such as the average number of active UEs of both types in the system and the associated average delays. Nevertheless, the state space of this process is quite extensive, which impedes its practical application. Although it is possible to mathematically describe this model, obtaining the final characteristics with a reasonable number of Type I UEs in the system is a challenging task. To overcome this limitation, we apply the state aggregation technique [143, 144] to reduce the dimension of the process and provide a simple numerical algorithm that professionals can utilize in their work.

The state aggregation technique can be employed in two ways: (i) by analyzing the structure of the state transition diagram and identifying sets of states with similar properties, then replacing them with a macro state, or (ii) by leveraging different granularity of the system, that is, by omitting some details of the original system. In our study, we adopt the second approach by considering the aggregated number of Type I and Type II UEs in the system in Section 5.1.2 instead of tracking individual types of UEs in this section. Consequently, the constructed Markov process is defined as the aggregated version of the exact process presented in this section. We then utilize it to develop approximate values for the mean delays of Type I and Type II UEs.

Approximation via State Aggregation

Consider the one-dimensional Markov chain model $\{N_t, t = 1, 2...\delta\}$, which includes the overall number of UEs of both types in the system at each time t. The state space of this process is $N_t \in \{0, 1, \dots\}$. According to our assumptions, the next state of the process depends solely on the current state, thus making it a Markov process.

The state transition diagram of the approximate model is shown in Fig. 5.4. This is a key difference from the two-dimensional model in Fig. 5.3 is that transition probabilities are now easier to derive, as no distinction needs to be made between

the two types of UEs. However, this simplification complicates the calculation of the metrics of interest, which can only be obtained through bounds and approximations, as discussed below.

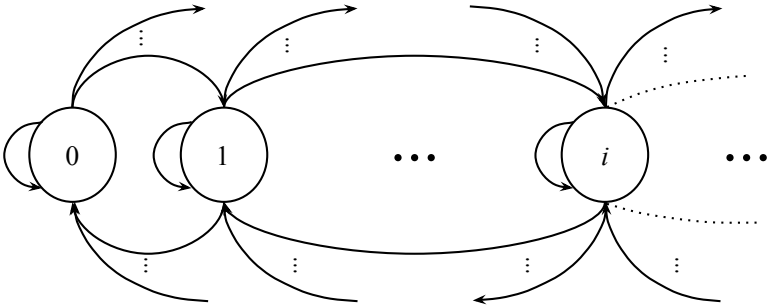


Fig. 5.4: State transition diagram of one-dimensional model [2].

When it comes to parameterizing the proposed model, it is important to note that the number of active Type I and Type II UEs at time slot $t+1$ is directly influenced by the number of active UEs in time slot t. Specifically,

$$N_{t+1} = N_t - T(N_t) + V_t, \tag{5.3}$$

where the number of Type II UEs that become active in a single slot is denoted by V_t, while the number of successful message transmissions when there are N_t messages ready for transmission is represented by $T(N_t)$.

By employing Equation (5.3), it is possible to determine the transition probabilities of the set $\{N_t : t = 1, 2, \ldots, \delta\}$ by considering all transitions from state i to state j. To account for the fact that the system moves from state i to state j when the number of successfully served UEs in the slot is exactly $i - j$, but no more than the number of preambles l, the transition probabilities can be expressed as follows:

$$p_{ij} = p\{j \mid i\} = \sum_{m=\max(0,i-j)}^{l} \frac{\lambda_2^{m-(i-j)}}{(m-(i-j))!} \times e^{-\lambda_2} Pr\{T(i) = m\}, \tag{5.4}$$

where the number of successfully transmitted messages, denoted by m, is the focus of this statement, with its maximum value being 0 when $i - j$ new active UEs enter the system. The upper limit of the sum is determined by l, which is the number of orthogonal channels available for transmission within a slot. The expression under the sum in (5.4) consists of two terms: the unconditional probability, which determines the probability of a specific number of new UEs entering the system in a slot, and the conditional probability, which determines the probability of successfully transmitting a specific number of messages.

In equation (5.4), the term $Pr\{T(i) = m\}$ represents the probability of successfully serving m UEs in a slot when there are i active UEs with messages ready for

transmission in the system. This quantity is equal to 0 when $N_t = 0$. Otherwise, the value is

$$Pr\{T(i) = m\} = \sum_{k=0}^{i} C_i^k p^k (1-p)^{i-k} P(l,k,m), \tag{5.5}$$

where C_i^k is the shortcut for the binomial coefficient and $C_i^k = \binom{k}{i}$.

In (5.5), the probability that an active UE transmits in the current slot is denoted by p. This probability is obtained as follows. Each UE can select a specific preamble with a probability of $1/l$. Therefore, if there are more preambles than active UEs in a slot, each UE transmits with a probability of 1. Conversely, if there are more active UEs than preambles, the likelihood of message transmission from the UE will be less than 1. Following [142], we assume that the UE follows an optimal scheme with the following transmission probability.

$$p = \begin{cases} 0 & \text{if } i = 0 \\ \min(l/i, 1) & \text{otherwise} \end{cases}, \tag{5.6}$$

The final term in equation (5.5), denoted as $P(l,k,m)$, represents the probability of the distribution of k messages across l channels under the condition that m channels are precisely selected by a single UE. This unknown value is determined in the following section as a part of the proposed solution algorithm.

5.1.3 The Solution Algorithm

The calculation of the transition probabilities in equation (5.4) becomes challenging, even when using a one-dimensional approximation. This is because of the presence of intricate binomial coefficients of a high order in $P(l,k,m)$. In the following parts, we outline a straightforward algorithm to compute these probabilities.

The probability $Pr\{T(i) = m\}$ can be broken down into several components. For $i = 0$, there is no probability because k is the total number of messages posted on l channels for any number of m channels selected by the UEs. For $i = 1$, there is one nonzero probability, $P(l,k,m)$. The remaining components, $C_i^k p^k (1-p_i)^{i-k}$, describe the binomial probability in which k out of i messages are successfully transmitted. To consider all possible states, we sum over all values from 0 to i. As i approaches infinity, the following approximation holds

$$Pr\{T(i) = m\} = C_l^m e^{-m} (1-e^{-m})^{l-m}, \tag{5.7}$$

where the number of successful transmissions is denoted by variable m. The justification for equation (5.7) is that as the number of UEs approaches infinity, the probability of a specific preamble being chosen by only one UE is e^{-1}. Consequently,

the probability that m preambles from l are selected by only one UE is governed by binomial law with a success probability of e^{-1}.

The only missing parameter is the probability of the distribution of k messages over l channels, where m channels are selected by exactly one UE. To calculate this probability, we follow the method outlined in [145] and present a solution for estimating the exact values of stationary probabilities $P(l,k,m)$, see the appendices in [2]. This solution is readily available by

$$P(l,k,m) = \frac{(-1)^m l!k!}{l^k m!} \times \sum_{f=m}^{\min(l,k)} \frac{(-1)^f (l-f)^{k-f}}{(f-m)!(l-f)!(k-f)!},\ m \leq k, \tag{5.8}$$

and $P(l,k,m) = 0$, $m > k$.

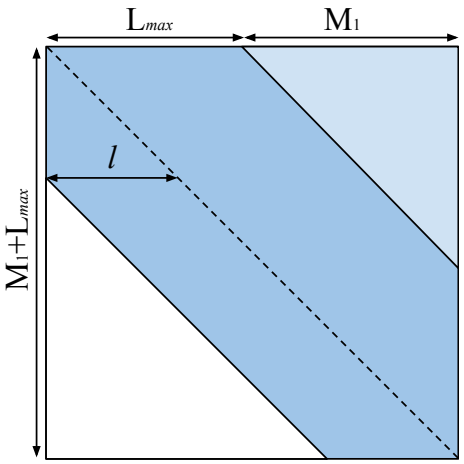


Fig. 5.5: The structure of the transition probability matrix [2].

The process of solving the model includes determining the stationary probabilities of a two-dimensional Markov chain. It is beneficial to express the probabilities represented by (5.4) in the form of a square matrix of transition probabilities over the Markov chain states, as illustrated in Fig. 5.5. For practical calculations, it is necessary to truncate the number of rows and columns to $M_1 + L_{\max}$, where M_1 represents the number of Type I UEs in the system, and $L_{\max}$ is a value selected such that the sum of elements in the first row is approximately 1 with a given precision, such as 10^{-6}.

Metrics of Interest

Our initial model was a two-dimensional Markov model with variables N_t^1 and N_t^2, representing the number of Type I and Type II UEs, respectively. However, this model has a high computational complexity; therefore, we used state aggregation to develop a simplified one-dimensional Markov chain model that tracks the joint variable $N_t = N_t^1 + N_t^2$. This new model allows us to determine the mean aggregated

number of UEs in the system, $E[N]$, and the associated delay, $E[D]$. Now, we will develop bounds on the number of Type I and Type II UEs in the system. Our approach is to consider a system without Type II UEs, calculate the mean number of Type I UEs, and then analyze the system without Type I UEs to determine the mean number of Type II UEs. Both systems form separate one-dimensional Markov chains, as shown in Fig. 5.6.

The transition probability matrix of the two-dimensional Markov chain depends solely on parameters λ_2 and M_1. Given that the distribution of the number of Type I and Type II UEs (N_1^1, N_1^2) is already known in the initial time slot when all Type I UEs are activated, we can proceed with the analysis. Throughout this section, we use Fig. 5.7 to illustrate the behavior of the examined intermediate metrics. This figure specifically compares the intermediate metrics of interest for the average number of Type I and Type II UEs in the system as a function of time against a single computer simulation run with $\lambda_2 = 2.5$ and $M_1 = 100$. It's important to recognize that because Fig. 5.7 relies on a single simulation run, the simulated performance of the metrics may occasionally cross an analytical lower bound.

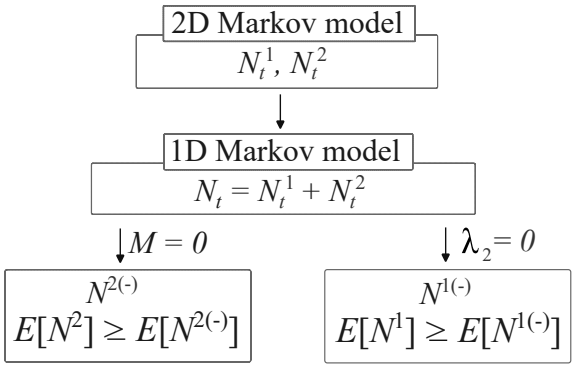


Fig. 5.6: The overall analysis procedure [2].

The transition probability matrix can be utilized to obtain the distribution of the pair (N_t^1, N_t^2) in each subsequent time slot $t = 2, 3, \ldots, \delta$. This allows for the determination of their mean values, $E[N_t^1]$ and $E[N_t^2]$. Consequently, the number of active UEs over the time interval Δ can be calculated.

$$E[N^1] = \sum_{t=1}^{\delta} \frac{E[N_t^1]}{\delta}, \; E[N^2] = \sum_{t=1}^{\delta} \frac{E[N_t^2]}{\delta}. \tag{5.9}$$

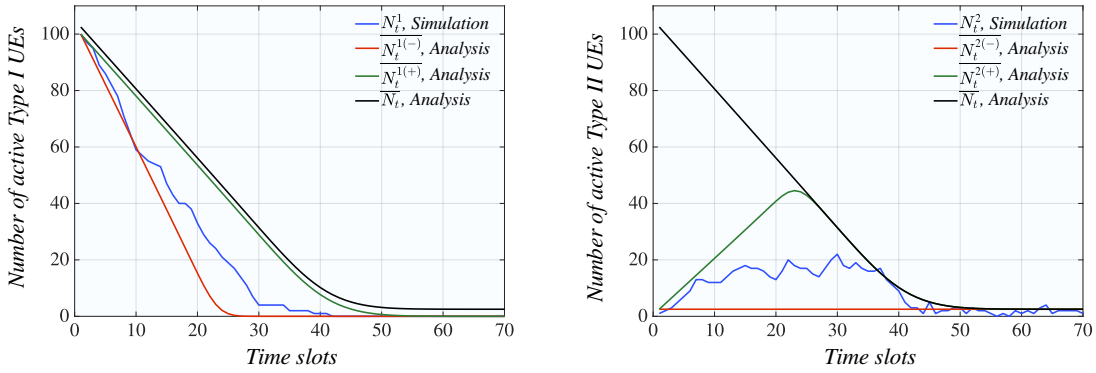
Using Little's result, we can calculate the expected delays for Type I and Type II UEs, denoted as $E[D_1]$ and $E[D_2]$, respectively. The expressions for these delays are given by

$$E[D_1] = \frac{E[N^1]\delta}{M_1}, \; E[D_2] = \frac{E[N^2]}{\lambda_2}. \tag{5.10}$$

According to the two-dimensional Markov model explained in Section 5.1.1, the metrics of interest, $E[\tau_1]$ and $E[\tau_2]$, which are measured in seconds, are calculated as follows:

$$E[\tau_1] = (E[D_1]t_R + t_{data})\, 10^{-3},$$
$$E[\tau_2] = (t_S + E[D_2]t_R + t_{data})\, 10^{-3}. \tag{5.11}$$

The factor 10^{-3} is used to account for the NB-IoT frame duration ($\delta = 10$ ms), which represents the slot duration. The variables t_S, t_R, and t_{data} denote the synchronization duration, random access duration, and data transmission time, respectively. It's important to highlight that the main distinction between Type I and Type II traffic in (5.11) lies in the fact that Type I traffic doesn't require a synchronization phase, as the connection remains active at all times.



a) Mean number of active UEs I Type in the system.

b) Mean number of active UEs II Type in the system.

Fig. 5.7: The number of active UEs in the system during Δ [2].

A method that employs a two-dimensional Markov chain is associated with a high level of computational complexity. To address this issue, we propose upper bounds on the mean delay performance of Type I and Type II UEs. Similar to the two-dimensional scenario, we consider the state of the transition probability matrix at the beginning of the interval. This state depends on the intensity of the packets from Type II UEs, λ_2, and the number of Type I UEs, M_1. Additionally, we can calculate the mean number of active UEs for both types in all subsequent slots, $E[N_t]$. The trajectory of the process $\{N_t, t = 1, 2, \ldots, \delta\}$ is depicted in Fig. 5.7 by black lines. At the beginning of the interval, N_t consists of both types of UEs. However, as time progresses, the number of Type I UEs decreases and only Type II UEs contribute to N_t. As a result, starting from a specific time slot, the distribution of the Type II UEs aligns with the distribution of both types of UEs for the remainder of the

interval Δ. Consequently, the mean value $E[N]$ is obtained as

$$E[N] = \sum_{t=1}^{\delta} \frac{E[N_t]}{\delta}. \tag{5.12}$$

The following statement holds true when a one-dimensional Markov model is expanded to include two-dimensional state aggregation [146]:

$$E[N] = E[N^1] + E[N^2] \tag{5.13}$$

Now, consider a system in which there are no Type II UEs, that is, $\lambda_2 = 0$. The mean number of active UEs in this system is bounded from below by $\overline{N_t^1}$. The behavior of this metric, denoted by $E[N_t^{1(-)}]$, is illustrated in Fig. 5.7 by red lines. The value of $E[N^{1(-)}]$ averaged over the period Δ provides the lower bound for $E[N^1]$. This metric can be obtained by substituting $E[N^{1(-)}]$ into(5.13) to get

$$E[N] > E[N^{1(-)}] + E[N^2], \tag{5.14}$$

and implying that $E[N^2] < E[N] - E[N^{1(-)}] = E[N^{2(+)}]$, that is also illustrated in Fig. 5.7, thus, leading to

$$E[D_2] < \frac{E[N] - E[N^{1(-)}]}{\lambda_2}. \tag{5.15}$$

In contrast, let us consider a system without a Type I UEs, meaning $M_1 = 0$. The mean number of UEs in this system serves as the lower bound for $E[N_t^2]$, as shown in Fig. 5.7. Similar to the previous case, the corresponding mean value over Δ, $E[N^{2(-)}]$, acts as the lower bound for $E[N^2]$. Because Type I UEs are absent, the distribution of N_t throughout the entire interval Δ is given by the defined Markov model and can be determined for any specific λ_2. By substituting the lower bound for $E[N^2]$ into (5.13), we obtain

$$E[N] > E[N^1] + E[N^{2(-)}], \tag{5.16}$$

which leads to $E[N^1] < E[N] - E[N^{2(-)}] = E[N^{1(+)}]$, as shown in Fig. 5.7, and ultimately resulting in the mean delay approximation for Type I UEs in the following form:

$$E[D_1] < \frac{\delta(E[N] - E[N^{2(-)}])}{M_1}. \tag{5.17}$$

By substituting the right-hand sides of (5.15) and (5.17) into (5.11), instead of $E[D_2]$ and $E[D_1]$, we obtain the upper bounds for the metric of interest, that is,

$$\begin{aligned} E[\tau_1] &< \left(\frac{\delta(E[N] - E[N^{2(-)}])}{M_1} t_R + t_{data} \right) 10^{-3}, \\ E[\tau_2] &< \left(t_S + \frac{E[N] - E[N^{1(-)}]}{\lambda_2} t_R + t_{data} \right) 10^{-3}. \end{aligned} \tag{5.18}$$

The resulting computational algorithm is shown in [2].

5.1.4 Numerical Results

In this section, we will delve into the proposed model by examining its effectiveness in the presence of two different types of traffic in an NB-IoT cell. We begin by evaluating the accuracy of the model and then demonstrate the stability region of the system. Subsequently, we will assess the accuracy of the proposed approximation for the mean delay, and finally, we analyze the impact of the system parameters on the performance metrics.

Parametrization Campaign

The creation of the suggested models is contingent upon several time-related assumptions regarding NB-IoT technology. Nevertheless, these presumptions are heavily reliant on the actual rollout of NB-IoT (defined by 3GPP), the configuration of the mobile network (operator-specific), and the UE design (e.g., the selection of subcarriers for random access, the number of repetitions for all communication tasks, and the number of subcarriers utilized for transmission). All these factors affect the time required for synchronization, RRC state transmissions, and message transmissions.

To obtain representative results for practical NB-IoT implementation, we utilized data from the measurements presented in Chapter 4. However, it is essential to acknowledge that the measurement campaign was designed to gather the parameters of non-interfering devices in a network with specific parameter configurations.

Utilizing the analytical model parameters outlined in Table 5.1, we determined the values presented in the table. These values were derived from average measurements of the results, incorporating the following assumptions: (1) synchronization time encompassed the period for initial time and frequency synchronization and gathering information about neighboring cells, (2) random access preambles were configured with two repetitions and selected from 12 subcarriers, (3) the collision-free RAP was defined as the RAP phase extending to Msg3, (4) the 200 B message transmission time was calculated as radio transmissions excluding RAP, TAU, and RRC connection phases, and (5) message handling time represented the duration required for the UE to process messages.

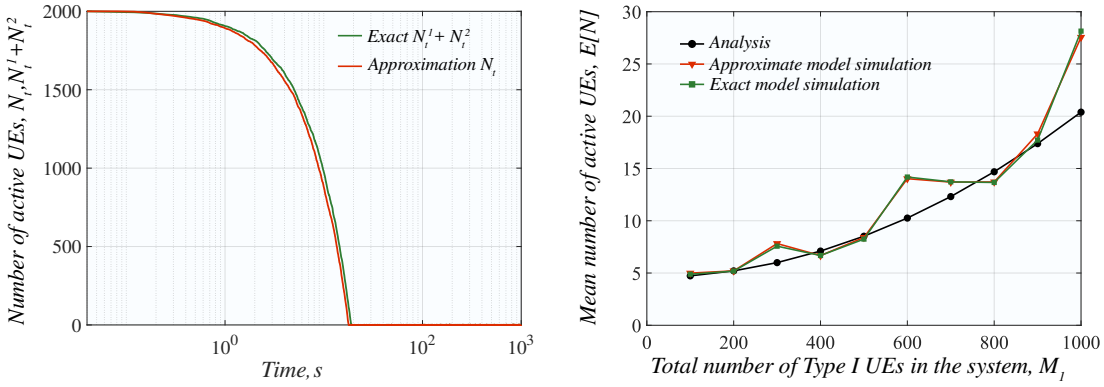
Accuracy Assessment

The initial step was to gauge the precision of the proposed model. It is important to keep in mind that there may be two sources of errors arising from the assumptions we made for analytical convenience and numerical simplicity. The first source of error is connected to substituting the detailed two-dimensional Markov model with a one-dimensional approximation. The second source of error arises from the utilization of delay bounds instead of exact delays. In this section, the focus is on the first

Tab. 5.1: Default system parameters.

Parameter name	Symbol	Value
Time slot duration	δ	10 ms
Number of preambles	l	12
The time to synchronization	t_S	30 slots
Temporary collision-free RAP	t_R	4 slots
Message handling time	t_H	1 slot
Data transfer time	t_{data}	35 slots
Interval between Type I UE arrivals	Δ	15 minutes
Number of Type I UEs	M_1	100-10000
Arrival rate of messages from Type II UEs	λ_2	1-100 per sec

source of error, while in the following sections, we evaluate the accuracy of the approximations for the mean delay of messages.



a) Active UEs during a single cycle

b) Mean number of active UEs in the system

Fig. 5.8: Numerical assessment of model accuracy [2].

To evaluate the effectiveness of using a one-dimensional model to represent a two-dimensional model, it is essential to consider the total number of active UEs in the system. In the first instance, we distinguish between Type I and Type II UEs, whereas in the second instance, we capture the total number of both types of UEs. We used a two-step process for benchmarking. Initially, we ensured that we did not make the mistake of applying the state-aggregation technique by comparing the evolution of the total number of active UEs during a cycle in the precise two-dimensional Markov model ($N_t^1(t) + N_t^2(t)$) and the consolidated number of active UEs in a one-dimensional model ($N_t(t)$). Then, in the second stage, we work with time-averaged metrics that are directly related to the final metric – delay. We compared the mean number of UEs in system $E[N]$ obtained using the simplified

one-dimensional model with those obtained using the precise two-dimensional and approximate one-dimensional models.

Our simulation tool was designed and implemented in MATLAB, following the specifications detailed in Section 5.1.1, and adhering to the timings depicted in Fig. 5.2. The minimum time increment is defined as δ, which corresponds to the duration of the NB-IoT frame. Type II packets were received by the system according to a Poisson process with an intensity parameter of λ_2. Because the system operates on a time-slotted basis, the arrival time is rounded to the nearest slot boundary. The start of the simulations was synchronized with the commencement of the first query interval, which lasted for Δ slots. Type I packets are batched and processed in the initial slot. Each packet underwent an explicit simulation of the random-access procedure, which involved selecting a random preamble. The synchronization and data transmission times were factored into the simulation by skipping the required number of slots. At each slot, a separate variable is utilized to keep track of the numbers of Type I and Type II packets in the system. Statistical data collection was performed only in the steady state. To determine the onset of the steady state, we used the exponentially weighted moving average statistic [147]. Finally, the batch-means method [147] was employed to gather statistical data.

The figure below, specifically Fig. 5.8a, depicts the analytical results of total number of active UEs during a single interval Δ that is obtained from simulating the considered model. This value is obtained by averaging over 100 realizations for 500 Type I UEs and an intensity of $\lambda_2 = 0.1$ messages/s from Type II UEs. It is evident from the figure that the approximate one-dimensional model accurately captures the number of UEs, which confirms the hypothesis that the distributional characteristics of the number of UEs in the system for the two considered models are identical. It is important to note that we primarily focused on the mean delay of Type I and Type II messages as the main metric of interest. However, this delay is directly related to the mean number of active messages or, alternatively, the number of UEs with messages ready for transmission. Therefore, the shape of the delay is similar.

To determine if the previously mentioned hypothesis applies to other system parameters, we conducted an analysis of $E[N]$ using the proposed one-dimensional model and compared it to computer simulations of both exact and one-dimensional models. However, despite averaging the simulation results over multiple runs, stochastic factors still played a role, causing crossings between the analysis and simulated curves, although they followed the same trend. When the number of Type I UEs increases (e.g., to 1,000 or more), the variance of the simulations increases, leading to more visible deviations in Fig. 5.8b. Nevertheless, the simulation data coincide well with the presented data, indicating that there is no error in replacing the more

complex two-dimensional model with a simpler model.

For large gaps between the arrivals of Type I UE messages, there are extended periods when the system is nearly devoid of traffic with only Type II UE messages present. However, the overall quality of service in terms of mean delay (calculated for both time and UEs with messages ready for transmission) for both types of traffic is significantly affected by what transpires at the commencement of these intervals. As a result, we evaluate the so-called "ergodic" metric, which averages the mean number of active UEs and the mean delay over time. At the start of these intervals, the immediate delay for Type II UEs is substantially higher than that at the end, but on average, the delay experienced by Type II UEs is as shown in the steady state. The logic for Type I UEs is similar, but the delay is higher than that for Type II UEs, as they leave the system much earlier than the end of the period between Type I UE arrivals.

System's Stability Region

To ensure optimal performance in practical deployment, it is essential to comprehend the system performance region when the message transmission delay is not always infinite. Our system employs a method in which untransmitted Type I UE messages that have not been sent by the end of the inter-arrival period for Type I UE messages are removed and replaced with new ones. However, Type II UE messages are unique because they can accumulate indefinitely. Nevertheless, Type I messages contribute to the overall intensity of Type II messages when they coexist in a system. Consequently, we conducted an assessment of the stability of the system, considering that the system is considered stable if the maximum output traffic intensity that can be managed by the system exceeds the input traffic intensity. To this end, Fig. 5.9 illustrates the analytical results of the output traffic intensity in messages per second in relation to the number of Type I UEs in the system for various Type II UEs message intensities (also in messages per second) and an interval between Type I UE arrivals set at 15 min. The figure also shows the maximum output traffic intensity that the system can support (black line). At approximately 9,000 Type I UEs and $\lambda_2 = 0.4$ messages/s, we observed that the maximum output traffic intensity was exceeded, implying that the system became unstable from this point onwards. For lower values of λ_2, there are additional crossing points, but they correspond to significantly higher numbers of Type I UEs.

Note that if the transmission time is 40 ms, the Type II message intensity of $\lambda_2 = 4$ messages per second with a standard message inter-arrival time of 2 h specified in ITU-R M.2410 [24] results in 72×10^4 Type II UEs in the coverage area of a single NB-IoT cell. Therefore, the proposed model can be used to assess the

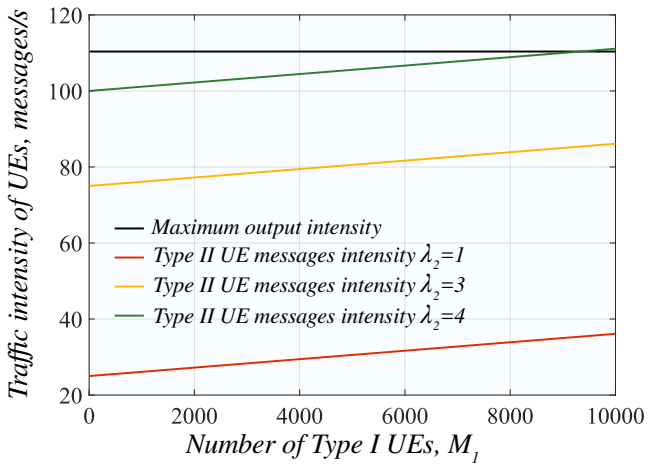


Fig. 5.9: Stability region of the considered system [2].

performance of 5G mMTC services based on NB-IoT, which has a target of 10^4 UEs per square kilometer [148] and up to 9×10^3 Type I UEs requiring permanent connectivity. The stability region for other traffic parameters can be estimated using the proposed model. When designing a network, these constraints must be considered by the network operators. However, we want to emphasize that the system stability does not guarantee that all other performance indicators remain within their specified limits. We now proceed with a delay analysis.

Delay Approximations

After presenting an approximate estimate of the overall number of active UEs in the system using a one-dimensional model and identifying the system stability region, we evaluated the accuracy of the Type I UE message delay approximation based on the proposed model. To accomplish this, Fig. 5.10 depicts the developed approximation for the average delay of both types of UEs as a function of the number of Type I for $\lambda_2 = 1$ messages/s and $\Delta = 15$ min. Upon analyzing the results, it is apparent that the approximation accurately matches the actual average delay experienced by the Type I UEs across the entire range of considered UE numbers. For Type II UEs, the match is good until a relatively high number of Type I UEs, for example, 10^3. Beyond that, the model starts to overestimate Type II UEs' delay. Additionally, it was observed that as the number of Type I UEs increased, the accuracy of the delay approximation for Type I UEs improved, whereas the accuracy for Type II UEs worsened. This is because, under this condition, Type I UEs start to dominate traffic aggregation. This trend also holds true for the increasing Type II UEs' arrival intensity (see Fig. 5.11). However, the impact is less noticeable because the delay experienced by both types of UEs is mainly influenced by the number of Type I UEs.

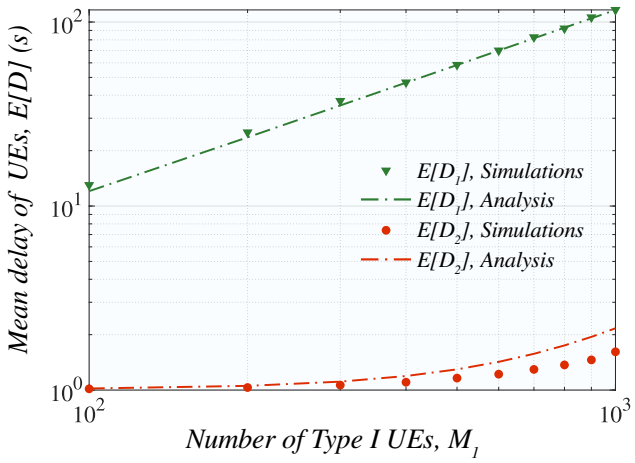


Fig. 5.10: Delay approximations for considered UE types [2].

Performance Analysis

After pinpointing a suitable approximation for delay and verifying the stability region of the system, we evaluated the performance of traffic types in terms of delay. To achieve this, Figs. 5.11 and 5.12 depict simulation-based results of the mean delay of both traffic types as a function of the number of Type I UEs and the intensity of Type II messages in the system for $\Delta = 15$ min. Upon analyzing the data, it becomes evident that the intensity of Type II UEs does not have a significant impact on the performance of either traffic type. This is because the load produced by these UEs is evenly distributed over time. On the other hand, increasing the number of Type I UEs negatively affected the mean delay of both traffic types, as expected. The impact of Type I UEs on the mean delay is linear, whereas the mean delay of Type II UEs increases exponentially with the number of Type I UEs. It is worth noting that this increase becomes significant from approximately 500 UEs, but this value depends on the number of Type I UEs and cycle time values Δ.

Another significant observation is that the presence of Type I UEs affects the compliance of NB-IoT technology with 5G mMTC requirements formalized in Chapter 3. This recommendation sets a delay bound of 10 s for 10^6 traditional sensors, assuming that each generates at most one message in two hours and 2 s for smart meters with regular polling interval. Furthermore, as previously demonstrated in the 'System's Stability Region' subsection, the system remains stable for approximately 9×10^3 and 72×10^4 Type I and Type II UEs, respectively. However, based on the data presented in Fig. 5.12, we see that the mean delay of Type I UEs is already higher than 10 s, even with 100 Type I UEs in the system violating both considered constraints. At the same time, it remains well below for the conventional UEs, even with 10^3 Type I UEs in the system. Thus, we can conclude that the

presence of UEs demanding permanent connectivity questions the applicability of NB-IoT technology in its current form for environments with a mixture of considered traffic types.

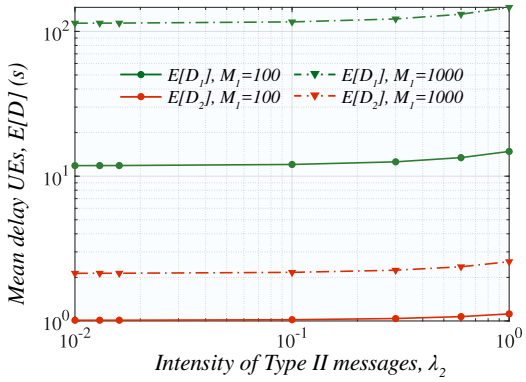


Fig. 5.11: Delay performance as a function of Type II message intensity [2].

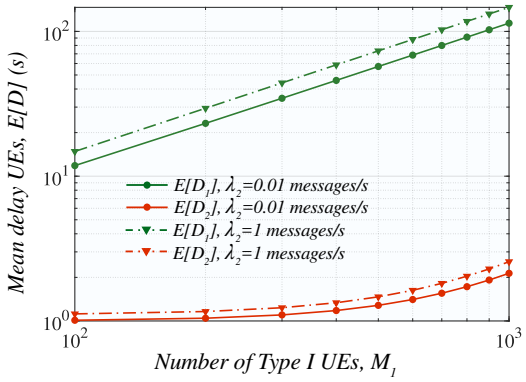
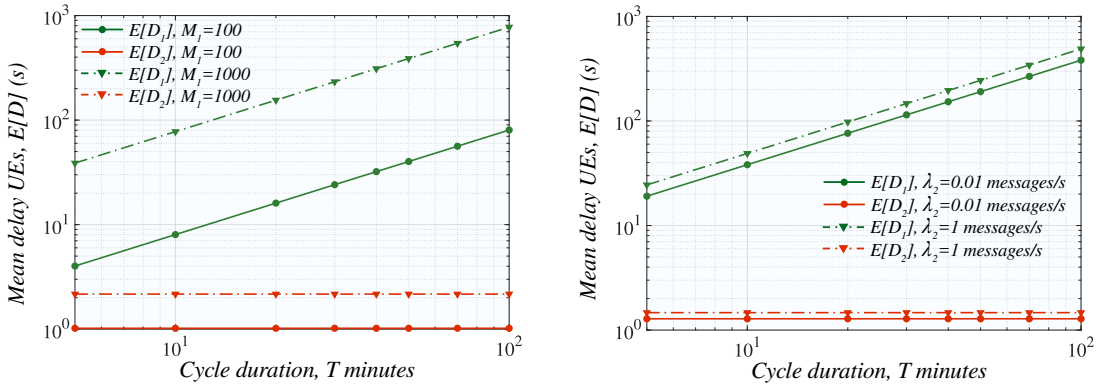


Fig. 5.12: Delay performance as a function of the number of Type I UEs [2].

A factor that may affect the delay performance of various traffic types is cycle time Δ. To investigate this, Fig. 5.13 shows results from simulation of the mean delays of different traffic types as a function of Δ. The figures demonstrate that the duration of the cycle interval does not seem to affect the delay performance of the considered traffic types. This is because the arrival of Type I UEs is so well-timed that it does not affect the stability of the system under the given parameters. In other words, the delay caused by the arrival of Type I UEs is significant but short-lived, and the period during which they are present in the system is much shorter than the cycle duration Δ. Although smaller values of Δ may delay the explosion of the delay, the system remains largely unaffected by any practical duration of Δ.

Finally, we explore a technique to ensure assurances for Type I. To achieve this, Fig. 5.14 depicts the average delay of Type I UEs as a function of their quantity in the system for multiple RBs designated for the NB-IoT network, $\Delta = 15$ m, and a Type II UE arrival intensity of $\lambda_2 = 1$ messages/s. In this case, we assume that both types of UEs are equally allocated between the NB-IoT RBs. The deployment of multiple RBs is achievable by deploying NB-IoT to multiple guard bands or also in inband. However, as stated above, this model considered 12 preambles for the RA procedure out of the 36 available. This division is a typical deployment, where each ECL occupies 12 preambles; thus, this model considering ECL0 expands to the rest of the region. As we can observe, when one, two, or three RBs are utilized, the delay for Type I UEs decreases. Moreover, the observed trend was nonlinear and the average delay decreased significantly when more RBs were added to the system. By employing three RBs, the average delay for the Type I UE satisfies requirement



a) Multiple number of Type I UEs, $\lambda_2 = 0.1$ messages/s.

b) Multiple Type II message arrival intensities, $M_1 = 500$.

Fig. 5.13: Delay performance as a function of cycle time [2].

M.2410 of 10 s for all the considered numbers of Type I UEs. However, considering the 2 s delay defined in Chapter 3, by employing three RBs, this requirement is met for up to approximately 240 Type I UEs. Thus, such an enhancement could overcome the congestion observed in the measurement campaign with 200 UEs, as detailed in Section 4.2.

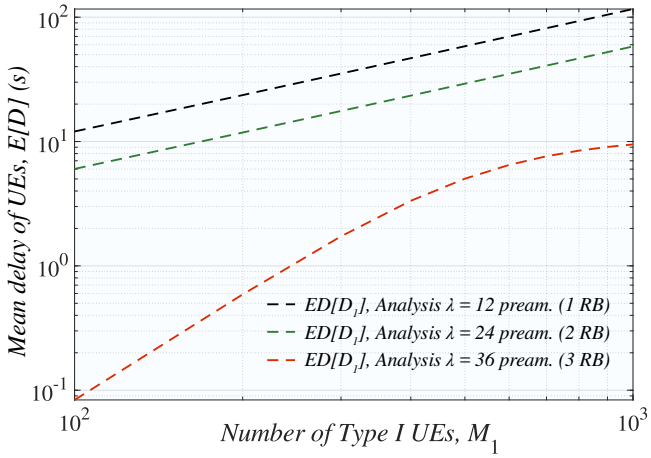


Fig. 5.14: Mean Type I UE delay for different number of RBs [2].

5.2 Optimal Spreading of the NB-IoT Stochastic and Regular Traffic Message Transmission

5.2.1 System Model

As in the previous section, the objective was to develop a system model that encompasses the key components of the system under consideration. Moving forward, we conducted a system analysis to evaluate performance measures of interest. The deployment model and random access procedures from the previous section were implemented, with only a slight adjustment made to the traffic type.

Traffic Types

We focused on two categories of applications that utilize the capabilities of an NB-IoT system. Type I UEs (regular) are those that are managed remotely from an application server and are continuously active, keeping their radio interface operational. As a result, they remain in an RRC-connected state throughout their entire lifecycle, as discussed further. Type II UEs (stochastic) are conventional devices that undergo a sleep-aware-transmit cycle. These devices can be used for periodic measurements, and their operational cycles are assumed to be asynchronous. We assume that the time between message transmissions in this type of traffic follows an exponential distribution, with a mean of λ_2^{-1}.

Modeling Assumptions

We typically use a system that divides time into discrete slots. Each slot is 10 ms long, which corresponds to the duration of one LTE frame and represents the minimum time required for random access without repetitions. There is a fixed number of Type I UEs within the cell coverage, denoted by N_1. Applications query these devices at the start of a regular time interval that lasts for T seconds. The duration of this time interval may vary depending on the type of application, and can range from minutes to hours. In this context, D_1 and D_2 represent delay samples. When requested, all N_1 devices attempt to transmit a message by following the NB-IoT random access procedure. To address this issue, we propose a mechanism that allows devices to delay transmission for a period uniformly distributed over $(T, T + \tau_1)$, as illustrated in Fig. 5.15.

The reasons for this procedure are as follows. When τ_1 is small, the system load is high, leading to significant delays for both Type I and Type II UEs owing to the RACH conflict resolution. As τ_1 increases, the initial RACH requests from Type I UEs become more evenly distributed, which reduces the delay. However, when

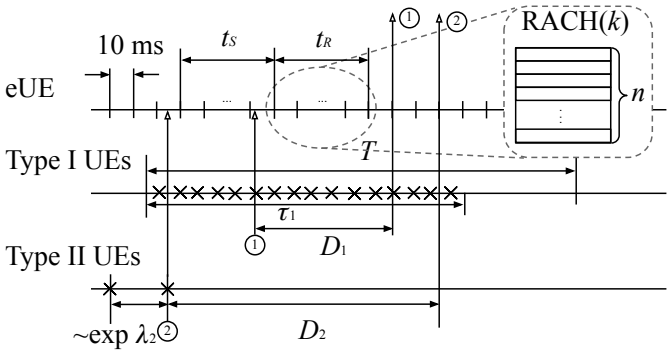


Fig. 5.15: Timing diagram of the system operation [4].

τ_1 becomes too large, the mean delay for the Type I UEs increases again owing to delayed initial RACH access attempts. Moreover, the choice of τ_1 also affects the mean delays for Type II UEs, as both types of devices have an impact on latency. In practical application, τ_1 can be configured by the application server or UEs can be instructed to initiate their first access attempt randomly and uniformly within $(T, T + \tau_1)$.

When generating a message, both types of UEs commence a random-access procedure in the following time slot. We assume that the probability of starting this procedure is $p = \min(n/k, 1)$, where n is the number of preambles and k is the number of UEs competing in a single slot. According to the studies referenced in [141], and [142], this approach reduces the delay experienced by UEs.

Metrics

We aim to address two key questions: first, what is the average delay for Type I and Type II traffic, specifically $E[D_1]$ and $E[D_2]$, respectively, and second, what is the ideal selection of τ_1 that reduces the message transmission delay for Type I UEs to its lowest?

5.2.2 System Analysis

We examine the system at discrete intervals for a duration of t_R. The initial time slot corresponds to the start of the query interval for Type I UEs. Let's denote by N_t the number of active UEs of both types that have messages ready for transmission, but have not yet been granted access. As outlined in Subsection 5.2.1, the number of active UEs in the subsequent time interval $t + 1$ depends on the number of active UEs in the current time interval t as

$$N_{t+1} = \begin{cases} N_t - T(N_t) + V_t^1(N_t) + V_t^2 & t \cdot t_R \leq \tau_1 \\ N_t - T(N_t) + V_t^2 & t \cdot t_R > \tau_1 \end{cases}, \tag{5.19}$$

where $T(N_t)$ represents the number of successful transmissions when the system has N^t active UEs, $V_t^1(N_t)$ and V_t^2 are the amounts of Type I and Type II UEs, respectively, that become active in slot t, provided that N_t represents the number of active UEs at the beginning of slot t.

A plot depicting the trajectory of this probabilistic process, which was obtained through computer simulations, is illustrated in Fig. 5.16. It is worth mentioning that this process does not adhere to the Markov property, rendering its precise description challenging. To grasp the essential features of this process, we used approximation techniques. Instead of explicitly following the stochastic development of $\{N_t, t = 0, 1, \dots\}$ according to (5.19), we present and examine the attributes of a related deterministic process that represents the state-space reduction of the number of active UEs, $\overline{N_t}$, over time.

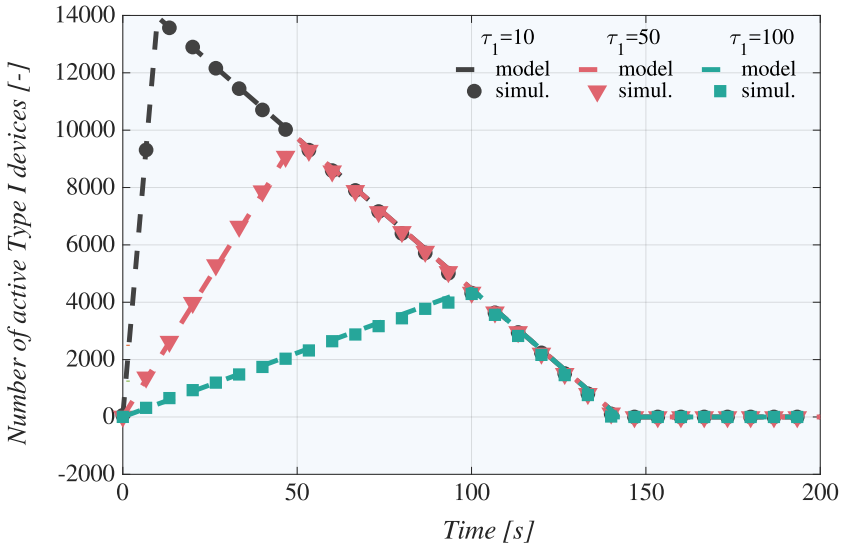


Fig. 5.16: Number of active UEs of the Type I [4].

Given that active UEs are assumed to transmit in a slot with a probability of $p = \min(n/k, 1)$, as stated in Subsection 5.2.1, the following can be approximated

$$E[T(N_t)] \approx ne^{-1}. \tag{5.20}$$

Moreover, considering that Type I UEs are instructed to choose the transmission slot uniformly over the interval from 1 to τ_1, we can express the expectation of $V_t^1(N_t)$ as follows:

$$E[V_t^1(N_t)] = m_1 t_R/\tau_1. \tag{5.21}$$

Ultimately, it is important to recognize that the arrival rate of Type II user equipment remains consistent over time, suggesting

$$E[V_t^2] = \lambda_2 t_R. \tag{5.22}$$

Utilizing the results (5.20) – (5.22), we can apply the state-space reduction, as discussed in Kleinrock's theory [149]. With this, we can write the recurrent relation for $\overline{N_{t+1}}$, which approximates the evolution of the mean value of the stochastic process $\{N_t, t = 0, 1, \dots\}$ as follows

$$\overline{N_{t+1}} = \begin{cases} \overline{N_t} - ne^{-1} + \frac{m_1 t_R}{\tau_1} + \lambda_2 t_R & t \cdot t_R \leq \tau_1 \\ \overline{N_t} - ne^{-1} + \lambda_2 t_R & t \cdot t_R > \tau_1 \end{cases}. \tag{5.23}$$

The development of the initial stochastic process involving the number of active UEs $\{N_t, t = 0, 1, \dots\}$ and its corresponding deterministic average process is depicted in Fig. 5.16 for the scenario of $ne^{-1} < m_1 t_R/\tau_1$ and various spreading time values τ_1. It can be seen that the process $\{\overline{N_t}, t > 0\}$ closely resembles the original process, suggesting that the former can be employed to accurately approximate the average message transmission delay for both Type I and Type II UE devices. We now calculate $E[D_1]$ and $E[D_2]$.

Let us denote T' as the time instant when all Type I UEs have successfully left the system. Additionally, let's consider d, which is the mean message transmission delay. These two metrics are related as follows

$$d = \left(\sum_{t=1}^{T'} \overline{N_t} \right) \Big/ \left(m_1 + \lambda_2 T' \right), \tag{5.24}$$

where the numerator represents the only unknown value.

Consider term $\sum_{t=1}^{T'} E[N_t]$. An examination of the $\{\overline{N_t}, t > 0\}$ structure depicted in Fig. 5.16 suggests that the desired sum is equivalent to the area of a triangular shape. This triangle has its base on the interval $(0, T')$ as shown in Fig. 5.16. The triangle's height corresponds to the peak value of $E[N_t]$, which is represented by the following expression

$$m_1 - ne^{-1}\tau_1 + \lambda_2 \tau_1. \tag{5.25}$$

The numerical value of the triangle base can then be expressed as $T' = \tau_1 + X$, where X is determined by the following equation

$$X = \frac{m_1 - ne^{-1}\tau_1 + \lambda_2 \tau_1}{ne^{-1}} = \frac{m_1}{ne^{-1}} - \tau_1 + \frac{\lambda_2 \tau_1}{ne^{-1}}. \tag{5.26}$$

Solving equation 5.26 with respect to T', we obtain the following expression

$$T' = \tau_1 + X = \frac{m_1}{ne^{-1}} + \frac{\lambda_2 \tau_1}{ne^{-1}}. \tag{5.27}$$

The area of the triangle is calculated by the following formula

$$\sum_{t=1}^{T'} E[N_t] = \left(m_1 - ne^{-1}\tau_1 + \lambda_2 \tau_1 \right) \times \frac{1}{2} \left(\frac{m_1}{ne^{-1}} + \frac{\lambda_2 \tau_1}{ne^{-1}} \right). \tag{5.28}$$

By inserting (5.28) into (5.24), we derive

$$d = \frac{1}{2ne^{-1}} \left(m_1 - ne^{-1}\tau_1 + \lambda_2 \tau_1 \right). \tag{5.29}$$

After determining the value of d, we can calculate the expected transmission delays for Type I and Type II UEs, $E[D_1]$ and $E[D_2]$, by adding up the delay components as

$$\begin{aligned} E[D_1] &= \tau_1/2 + d, \\ E[D_2] &= T'(d + t_S) + (T - T')(t_S + t_R)/T. \end{aligned} \tag{5.30}$$

Substituting equation (5.29) into equation (5.30) and differentiating between two cases, $t \cdot t_R \leq \tau_1$ and $t \cdot t_R > \tau_1$, in equation (5.23), we obtain the final result for Type I UE mean message transmission delay in the form of equation (5.31), which is given by

$$E[D_1] = \begin{cases} \frac{1}{2}\left(\frac{m_1}{ne^{-1}} + \frac{\lambda_2 \tau_1}{ne^{-1}} \right) & ne^{-1} < \frac{m_1 t_R}{\tau_1} + \lambda_2 \\ \frac{1}{2}\tau_1 & ne^{-1} > \frac{m_1 t_R}{\tau_1} + \lambda_2 \end{cases}. \tag{5.31}$$

To convey the final expression for the mean message transmission delay of Type II UEs, we adopt the same strategy by substituting (5.29) into (5.30). The mean delay is considered insignificant when $ne^{-1} < m_1 t_R/\tau_1$. Consequently, we derive

$$E[D_2] = \frac{T'(d + t_S) + (T - T')(t_S + t_R)}{T}, \tag{5.32}$$

for $ne^{-1} < m_1 t_R/\tau_1 + \lambda_2$ and $E[D_2] = 0$ in all the other cases.

Numerical Results

This section employs the parameters obtained from the measurement campaign described in Table 5.2 to assess the delay performance of Type I and Type II UEs. Our primary objective is to determine the optimal spreading interval τ_1 that minimizes the delay for Type I UE messages. To accomplish this, we begin by comparing our model's results with computer simulations, followed by an examination of how various system parameters influence the optimal τ_1 value. It is important to note that the simulation tool used in this study was developed and implemented in MATLAB, adhering to the specifications detailed in Section 5.2.1 and following the timings shown in Fig. 5.15.

A comparison between the model and simulation data is shown in Fig. 5.17. The simulation data was generated using a specially developed Discrete Event Simulation (DES) system that specifically modeled the message transmission process for Type I and Type II UEs in accordance with NB-IoT access and data transmission

protocols. The simulation parameters were consistent with those obtained during the measurement campaign outlined in Chapter 4. To address potential correlations between consecutive query intervals, which might cause discrepancies between the model and computer simulations, we calculated the average of simulation results across 100 query intervals. Furthermore, in line with ITU-R M.2412 guidelines, we assumed that Type II UEs generate messages every two hours. Consequently, the λ_2 value of 5 msg/s used in the study corresponds to 3,600 Type II UEs.

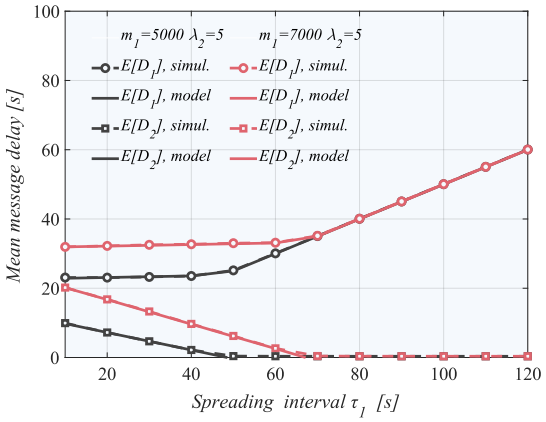


Fig. 5.17: Mean message delay for different m_1 [4].

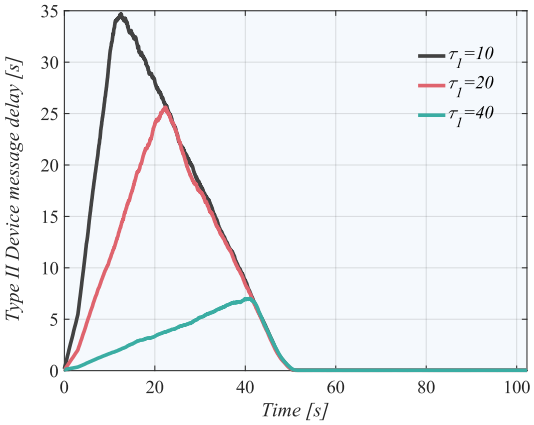


Fig. 5.18: Smoothed message delay [4].

Tab. 5.2: Input parameters for analytical model.

Parameter name	**Symbol**	**Value**
Number of preambles	n	12
The time to synchronization	t_S	30 slots
Temporary collision-free RAP	t_R	4 slots
Message handling time	t_H	1 slot
Data transfer time	t_{data}	35 slots
Message period for I Type	T	1 - 60 min
Number of user equipment of I Type	m_1	100-10000
Input flow rate of II Type	λ_2	1-100 per sec

Examining the outcomes shown in Fig. 5.17, it becomes evident that the model accurately predicts the complete average message transmission delay for Type I UEs. However, initially, the suggested method does not seem to enhance this metric, as the delay remains constant before increasing. This can be attributed to the fact that the advantage of distributing a Type I message by sending requests over τ_1 is offset by the increased delay in the UE's initial transmission attempts. Another

noteworthy observation is the relatively low delay experienced by Type II UEs when compared to the delays of Type I UEs.

We aim to explain the statements made above by analyzing the instantaneous delay of the Type II messages depicted in Fig.5.18. In order to present a clear visual representation of the trends in the data, we employed a moving average filter. The results indicate that Type I UEs have a significant influence on the delay of Type II messages throughout the spreading interval τ_1. After τ_1 has ended, the delay gradually decreases and eventually reaches small values, which is typical for a system with only Type II UEs. Additionally, we observe that this trend is also applicable to the actual delay of Type I UEs during the spreading interval τ_1 (excluding $\tau_1/2$).

With the advancement in the value of τ_1, it is possible to reduce the immediate delay for both Type I and Type II UEs. The developed analytical model offers a straightforward approach for identifying the optimal spreading time τ_1 that minimizes the delay for both UE types by pinpointing the point at which the slope of the mean delay function fluctuates.

After evaluating the performance of the developed model, we investigated the influence of the number of UEs on the mean message delay for various values of τ_1, as shown in Figs. 5.19 and 5.20. Initially, in Fig. 5.19, we observe that the number of Type I UEs has a negligible impact on the mean delay of Type II UEs, resulting in a minimal increase. Comparable results were obtained for τ_1. The reason behind this is that τ_1 is substantially smaller than the considered query interval T, which was set to 15 min in our simulations. Nevertheless, for a specific value of τ_1, the number of Type I UEs significantly affects the mean delay of the Type II UEs. On the other hand, we observed that in Fig. 5.20, Type II traffic did not produce any significant effect on the delay performance of Type I UEs.

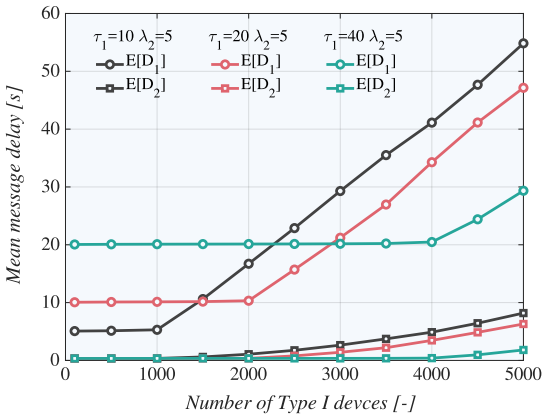


Fig. 5.19: Delay as a function of Type I devices [4].

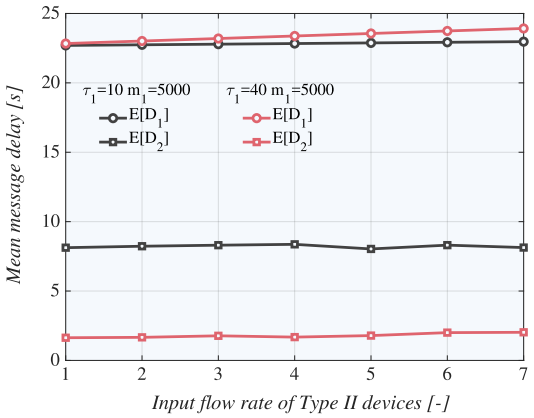


Fig. 5.20: Delay as a function of Type II devices [4].

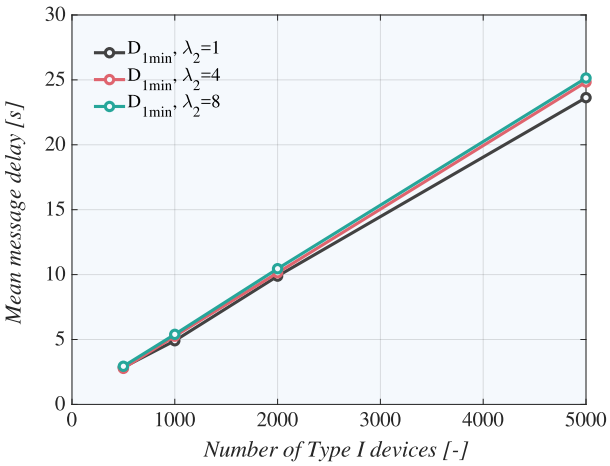


Fig. 5.21: Type I device optimal message delay [4].

Eventually, we display the typical Type I UE message delay for various values of Type I UEs and the intensity of Type II traffic corresponding to the optimal spreading interval τ_1 in Fig. 5.21. It is evident that the desired metric scales proportionally with the number of Type I UEs, and its inclination is nearly insensitive to Type II traffic intensity λ_2. These traits facilitate the estimation of the expected Type I UE delay associated with the optimal spreading interval under diverse traffic conditions. The corresponding optimal spreading intervals and delays are presented in Table 5.3. It is apparent that τ_1 also increased linearly.

Tab. 5.3: Optimal values of spreading interval τ_1.

Type I UEs [-]	500	1,000	2,000	5,000
Optimal τ_1 **[s]**	5	10	20	45
Mean D_1 **[s]**	2.5	5	10	23.7

5.3 NB-IoT Transmission Delay Reduction by Early Data Trasmissions

Data Transmission Mechanisms

The reduction in EDT signaling overhead can be observed in Fig. 5.22. The eNB communicates DL data transmission to the UE using a DCI message in the NPDCCH. Nevertheless, NB-IoT UEs possess limited computational capabilities, so a 4-ms time offset is implemented between the end of NPDCCH and the start of NPDSCH. Once

the data in NPDSCH are received, 4 ms of decoding is anticipated with good radio conditions, MCS of 10, and a TB of 680 bits. After a minimum of 12 ms, a HARQ message, including an acknowledgment of the transmission, was sent via NPUSCH. If more data are to be sent, the UE must wait for 3 ms before listening to the NPDCCH. The entire process of combining the DCI and NPDCCH is shown in Fig. 5.23. Thus, the projected DL throughput can be calculated as

$$\begin{aligned} T &= TB_{Max}/(\text{NPDCCH} + \text{NPDSCH} + \text{NPUSCH} + Off) \\ &= 680/(1 + 4 + 2 + (4 + 12 + 3)) = 26.15\ kbps. \end{aligned} \tag{5.33}$$

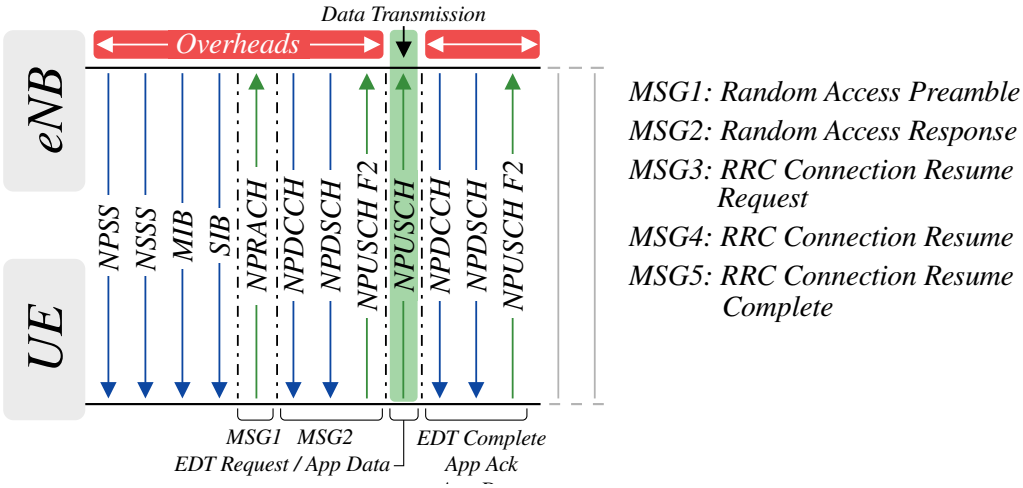


Fig. 5.22: NB-IoT: Early Data Transmission [3].

This calculation is valid only for NB-IoT in Release 13, as newer releases allow for a larger TB and two HARQ processes, resulting in a maximum throughput of up to 127 kbps [82, 150]. However, it should be noted that the maximum throughput may vary depending on the specific release and implementation.

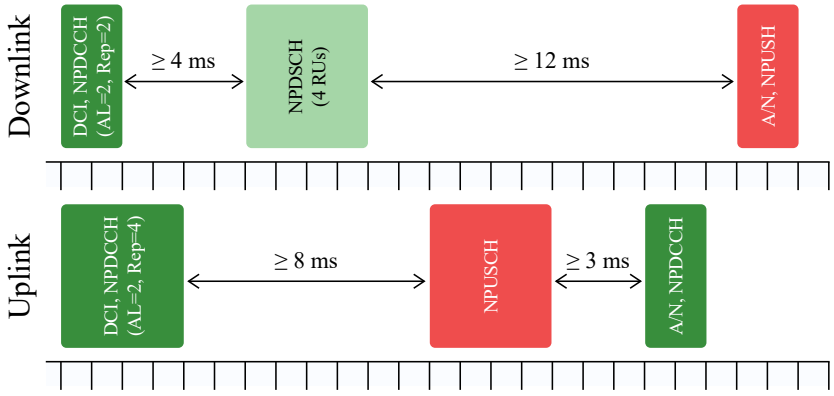


Fig. 5.23: NB-IoT: Transmission Timing [3].

For UL, the UE receives scheduling information through DCI on the NPDCCH. This allows the device to transmit UL data via the NPUSCH. The UE requires a minimum of 8 ms to prepare and transmit UL data on NPUSCH. Under favorable radio conditions with a maximum MCS of 12, a TB of 1,000 bits, and all 12 tones available, data transmission through NPUSCH takes 4 ms. However, single-tone transmissions need at least 48 ms. Following UL data transmission, the UE waits a minimum of 3 ms to receive an acknowledgment from the eNB via NPDCCH, potentially accompanied by the next grant. Using these parameters, it is possible to calculate the maximum throughput for both single and multi-tone as

$$\begin{aligned} T &= TB_{Max}/(\text{NPDCCH} + \text{NPUSCH} + Off) \\ T_{Multi} &= 1000/(1 + 4 + (8 + 3)) = 62.5\ kbps. \\ T_{Single} &= 1000/(1 + 48 + (8 + 3)) = 16.7\ kbps. \end{aligned} \tag{5.34}$$

Similar to the DL, the initial NB-IoT release supports the aforementioned values. Subsequent versions have enhanced throughput capabilities up to 159 kbps owing to the utilization of larger TBs and two HARQ mechanisms [82, 83, 150].

Given the impracticality of estimating SINR in simulations, we chose to rely solely on RSRP values. This approach was adopted to assess NB-IoT performance in scenarios involving numerous devices simultaneously interacting with a remote server. To accomplish this, we selected the discrete-event network simulator NS-3, which provides a comprehensive suite of tools for analyzing wireless cellular technologies. Specifically, we employed the LENA-NB module, developed at the Technical University Dortmund. LENA-NB stands out among other available frameworks by offering a complete implementation of the NB-IoT protocol, encompassing power states and cutting-edge NB-IoT features like EDT [151].

In order to assess the performance of NB-IoT, we devised two simulation scenarios. In the first scenario, a single eNB served 50 to 1,000 devices uniformly distributed over a circular area with a diameter of 3 km, with the base station located at the center to ensure that the communication distance did not exceed 1.5 km. This setup is representative of a typical suburban or rural area, with an eNB density comparable to that of the scenario. In this scenario, all UEs were assumed to have an exact height of 1 m above ground level, allowing them to operate under optimal conditions. Consequently, all UEs were expected to use ECL0 with the highest MCS, resulting in the fastest data rates. The second scenario was designed to be more realistic, with 20% of the deployed UEs located 1.5 m underground, representing deep-indoor conditions. It was anticipated that most of these devices would fall into ECL1, which is characterized by a lower MCS and increased repetitions.

In both instances, the simulation duration was consistent, ranging from 15 to 30 min, depending on the number of devices used. This time frame is fixed, considering the time required for all UEs to connect to the network and send a message of 64 B in size. To minimize the overhead from transmission on higher layers, we chose UDP because it provides connectionless communication. The message size of 64 B is determined by EDT communication, which can transmit only a limited amount of data with reduced signaling. We evaluated the performance of the EDT by comparing it with the remaining UP and CP optimizations.

5.3.1 Simulator Parameters

To obtain optimal realism in our study, we employed the parameters from the LENA-NB framework, which was derived from a genuine NB-IoT network in Germany, in conjunction with our observations from the Vodafone network available in the Czech Republic. The comprehensive list of parameters for all three ECLs can be located in Table 5.4.

Tab. 5.4: Simulator radio resource configuration [151].

	Parameter	ECL0	ECL1	ECL2
	RSRP Threshold [dBm]	-	-110	-133
NPRACH	**Periodicity [ms]**	320	640	2560
	Subcarriers Offset	36	24	12
	Subcarriers Number	12	12	12
	MSG3 Range Start	2/3	2/3	2/3
	Max Preamble Attempts	10	10	10
	Repetitions Per Pream. Att.	1	8	32
NPDCCH	**Repetitions Number**	8	64	512
	Start Subframe	2	1.5	4
	Starting Subframe Offset	0	0	0

5.3.2 Results and Discussion

We derived two sets of outcomes from our simulation scenarios: one using exclusively ECL0, and another employing a mixture of 90% ECL0 and 10% ECL1. It's worth noting that both these scenarios incorporate transmission delay, RAP delay, and collision frequency.

Transmission Delay

Considering the transmission delay, it is evident that the EDT method provides at least a 50% reduction compared with both UP and CP optimizations, as illustrated simulation-based results in Figs. 5.24 and 5.25. The height of the bars represents the average value, whereas the top and bottom edges of the error line indicate the 85th and 15th percentiles, respectively. It is important to note that the impact of CP optimization on transmission delay is minimal, improving it by approximately 10 – 20ms. However, this outcome is expected because the only difference between the two optimizations is that the CP data are transmitted in MSG 5, which is sent in *RRC Connection Resume Complete*. In contrast, the UP data are sent after acknowledgment of this message.

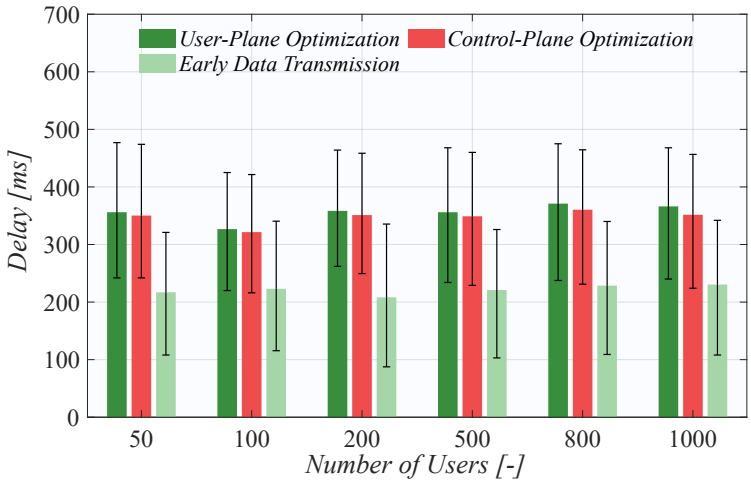


Fig. 5.24: Transmission delay in first scenario [3].

It is evident that the transmission delay remains consistent for the first scenario, even when the number of UEs increases. In contrast, the second scenario experienced a significant increase in delay values of approximately 60 ms. In addition, the transmission delay displays a linear relationship with the increase in the number of users for the UP and CP optimizations. However, for the EDT case, the situation aligns with that of the first scenario. The most significant observation is that the presence of only a small number of ECL1 UEs (approximately 10%) can significantly affect the network performance. Hence, it is crucial to consider this when planning and evaluating network capacity.

Random Access Delay

The duration of the random access procedure is a crucial factor that affects the transmission delay. As shown in simulation-based results in Figs. 5.26 and 5.27,

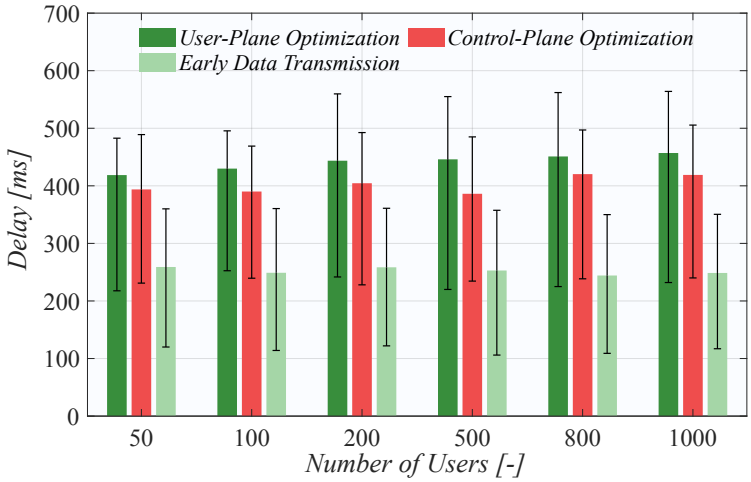


Fig. 5.25: Transmission delay in second scenario [3].

EDT accounts for most of the delay in the radio access network.

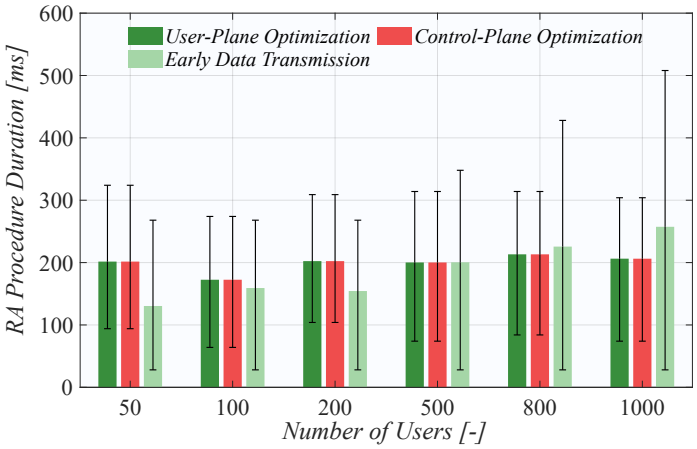


Fig. 5.26: Random access delay in first scenario [3].

When examining the transmission delay results in comparison to the standard RAP, it becomes apparent that EDT does not yield lower delay times. This trend, however, was not seen in the UP and CP optimizations. A comparable pattern is noticeable in both ECL0 and the combined scenarios. Given that the fundamental mechanism of RAP remains consistent across all three approaches (with EDT potentially utilizing a dedicated pool in NPRACH), there is no apparent reason for this discrepancy. Therefore, the only plausible explanation for the increased RAP delay is the high frequency of collisions occurring at the Medium Access Control (MAC) layer. Although this seems counterintuitive for EDT, subsequent findings confirmed this hypothesis.

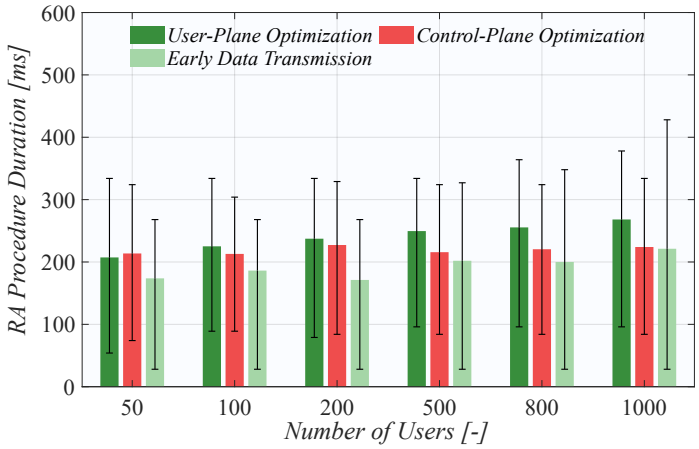


Fig. 5.27: Random access delay in second scenario [3].

Number of Collisions

The simulation outcomes corroborated the hypothesis presented in the preceding section, as demonstrated by the simulation results in Figs. 5.28 and 5.29. To enhance clarity, both figures employ a logarithmic scale for their respective y axes.

The next interesting finding was that collisions occurred more frequently in the ECL0 pure scenario. This holds true even for the UP and CP optimizations, which can be attributed to the shorter inter-repetition delays in NPRACH. Since devices reveal their present/anticipated ECL, the scheduler can rearrange these values in response to actual network conditions.

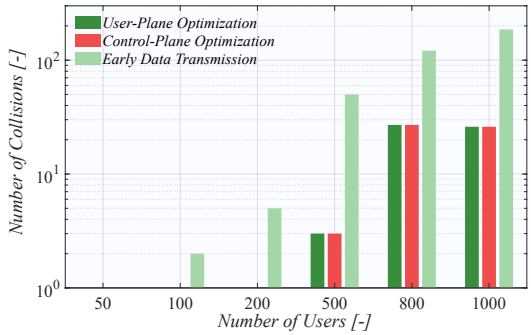


Fig. 5.28: RAP collisions in first scenario [3].

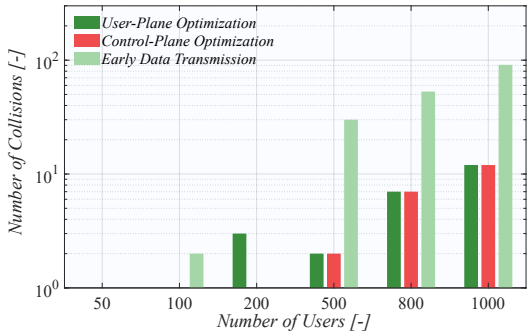


Fig. 5.29: RAP collisions in second scenario [3].

In order to study the number of collisions more thoroughly, we created a graph, as shown in Fig. 5.30 which illustrates the number of random access procedures and collisions over time. This was performed using the results from a simulation scenario with 1,000 users. The results for the number of EDT RAPs showed that there were

twice the actual number of UEs. This finding is logical, as it explains the increased number of collisions that lead to longer RAP times. However, it is unclear why twice the number of RAPs is necessary.

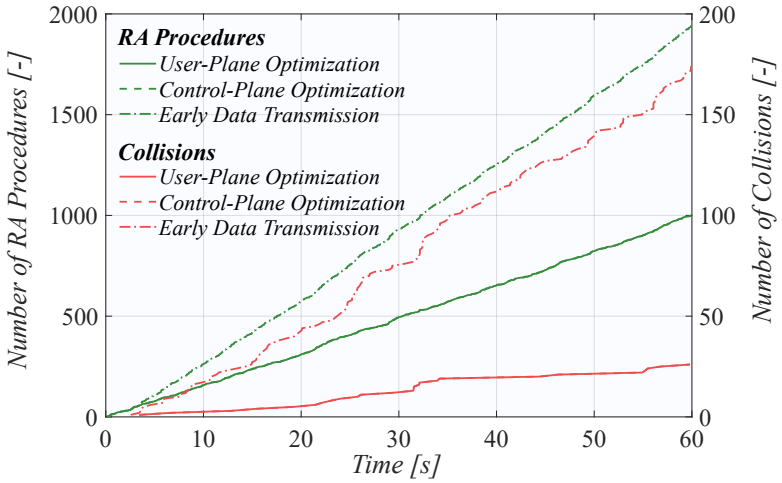


Fig. 5.30: Number of RAPs and collisions over time [3].

Our subsequent examination of the MAC layer logs revealed that each UE underwent a second RAP about 30 ms after successful UL transmission. Interestingly, this RAP was not accompanied by UL message transmission. The reason for this behavior remained unclear until we inspected the DL transmission. In TBS settings according to Release 13, the EDT message can fill the entire TB, which is 1,000 bits (NPUSCH) for UL and 680 bits (NPDSCH) for DL. The need for an additional RAP became evident when considering our 64 B UDP payload, echoed back to the UE from the remote server. Combining a 64 B user payload with an 8 B UDP header and at least 20 B from the IP header exceeded the maximum TB size with a 57 B user payload. Our 64 B message necessitated a TB larger than 736 bits.

Extending the maximum TB to 2536 bits in Release 14 eliminates the need for an additional RAP. However, in this study, we present the results from the default configuration of the LENA-NB modules to highlight that even an improper network configuration can have severe consequences on its performance. This is also important for real-world deployments, as network operators often select specific parts of newer NB-IoT releases based on their potential benefits and deployment costs.

5.4 Summary

This chapter addressed **RQ 3** by focusing on the modeling and optimization of single-RAT 5G IoT solutions for smart-grid applications, specifically using NB-IoT.

The first section presented a Markov chain model for NB-IoT smart metering traffic. This model accounted for both stochastic and regular traffic types, reflecting the diverse nature of data transmission in smart-grids as described in Chapter 3.

The proposed model in the first section reveals the stability region for up to 72×10^4 conventional NB-IoT UEs per NB-IoT cell (using sectoral antenna) and 9×10^3 UEs requiring permanent connectivity. It has been observed that conventional traffic does not have a substantial impact on the delay experienced by UEs that require uninterrupted connectivity. On the other hand, the average delay behavior of both types of traffic is primarily influenced by the number of Type I UEs and remains unaffected by the cycle time Δ.

The results indicate that delay performance of 10 s specified in ITU-R M.2410 is met for conventional UEs even under a high number of permanently connected UEs reaching 10^3. However, both delay bounds of 2 s and 10 s for the latter UEs is violated, even for 10^2 permanently connected UEs, necessitating a redesign of the NB-IoT channel access mechanism or increasing the available resources. As observed, the final analysis considering the extension of deployment to 3 RBs proved that the delay requirements of 2 s are met for up to 240 UEs with regular traffic providing a solution and answer to **RQ3** and performance enhancement as compared to the results from testing of 200 UEs in Chapter 4. In general, the emergence of UEs requiring permanent connectivity raises doubts about the suitability of NB-IoT technology for 5G mMTC in its present state.

The second section introduced a simpler model focused on optimizing the spreading of regular UEs messages considering 15 minute reading intervals. The analysis showed that optimizing the spreading intervals can lead to better load distribution and reduced congestion, improving overall network efficiency. For example, an optimal spreading interval of around 7.5 minutes was found to minimize the delay for Type I messages under certain traffic conditions. This model is particularly relevant for scenarios where rapid deployment and immediate performance gains are required. We introduced the concept of a spreading interval to optimize the mean delay performance of regular traffic, and utilized our model to determine the optimal values for this interval, resulting in minimal delays for regular traffic.

Our findings in the second section suggest that the number of UEs that generate regular traffic has a substantial impact on the average delay for both types of traffic. However, the practical values of conventional sporadic traffic do not appear to have a discernible influence on regular traffic. Moreover, the optimal spreading intervals

significantly reduce the delay associated with regular traffic. The optimal values for this interval and the corresponding ideal delay for the regular traffic type depend linearly on the number of UEs that generate regular traffic. This method allows the determination of the maximum number of Type I UEs that can be supported at an NB-IoT base station given a specific level of delay guarantee. With respect to the requirements defined in Chapter 3, the optimal τ_1 message spreading interval is approximately 5 s for less than 500 Type I UEs considering 15 min message periods. This approach showed that even the effective spreading of messages within the reading period could optimize the currently deployed NB-IoT network, so it meets the 2 s delay requirement. This provides a solution and answer to **RQ3**.

The third section explored future optimization strategies that leverage advanced features not yet deployed in current NB-IoT systems, notably EDT defined in 3GPP Release 15. This strategy aim at addressing the limitations of existing implementations and preparing for future advancements in 5G technology.

By utilizing EDT, we can effectively decrease transmission delay and enhance spectrum efficiency with minimal overhead. Our findings indicate that even with our unfortunate message size selection, the transmission delay is reduced by over 50%. This highlights the significant improvement in latency provided by EDT, particularly for smaller messages in Release 13, which have a size of less than 97 B in the UL and 57 B in the DL. For larger messages, the payload may be segmented or a standard RRC procedure may be initiated, resulting in reduced spectral efficiency. On the other hand, when the predefined TBS for EDT is overestimated, a substantial amount of padding data is transmitted.

It is important to recognize that an NB-IoT network has the ability to efficiently manage 1,000 UEs within a 15 min window, even with fundamental UP optimization. Additionally, there is a substantial buffer available to accommodate thousands of devices distributed between 90% and 10% for ECL0 and ECL1, respectively.

The results of the simulation highlight the significance of configuring the network correctly, as demonstrated by the doubling of RAP collisions when transmitting a flawless 64-byte message with EDT. This section can act as a guide for network operators to attain peak network capacity and optimal spectrum efficiency.

In conclusion, Chapter 5 successfully addressed **RQ 3** by developing and validating models and optimization techniques for NB-IoT in smart-grid applications. The 2-dimensional Markov chain model provided a detailed framework for understanding and improving traffic management, while the simpler spreading interval optimization offered practical solutions for immediate deployment. Additionally, the exploration of future optimization strategies underscores the ongoing potential for enhancing NB-IoT performance in the evolving landscape of 5G IoT.

6 Reliability Improvements for 5G IoT Enabled Smart-Grids

This chapter provides a comprehensive comparative analysis and performance evaluation of NB-IoT and eMTC technologies, focusing on their applicability and optimal load balancing in smart-grid environments, thus addressing communication reliability issues. Building on the theoretical foundations and requirements defined in Chapters 2 and 3 and the optimization needs highlighted in Chapter 4, this chapter addresses **RQ4**.

The increasing integration of IoT devices in energy grids necessitates robust communication networks capable of handling diverse and demanding requirements. Both NB-IoT and eMTC offer distinct advantages in terms of coverage, power efficiency, and cost-effectiveness. However, understanding their performance under various conditions is essential for informed decision-making and effective deployment.

The primary objectives of this chapter are to provide practical recommendations for optimal load balancing between NB-IoT and eMTC in smart-grid applications based on empirical data. Furthermore, exploring future enhancements necessary for the synergetic combination of these technologies in Multi-RAT deployments.

6.1 Load-Balancing Optimization of NB-IoT and eMTC for Permanently Connected Smart Meters

Our framework for the optimal association of Multi-RAT EDs is developed based on the results of the stress test described in Section 4.2. This framework employs stochastic geometry and Markov chains to account for the spatial and random access specifics of the system, considering the timing and batch traffic patterns of EDs operating in polling mode for both NB-IoT and eMTC. Using this framework, we are able to evaluate the best strategies for ED association in order to minimize access delay at the radio interface.

To enhance the productivity of Multi-RAT implementations, we propose a method for selecting the most suitable RAT for each device to balance the workload between the NB-IoT and eMTC connections. Our algorithm was designed to handle two scenarios: (i) purely stochastic message arrivals, and (ii) batch arrivals. The latter could be caused by polling mode operations or power outages, and has been shown to negatively impact system performance, as detailed in [49]. Our approach begins with a description of the system model and assumptions, followed by a detailed explanation of the proposed method. We utilized the mean delay as the KPI.

6.1.1 System Model

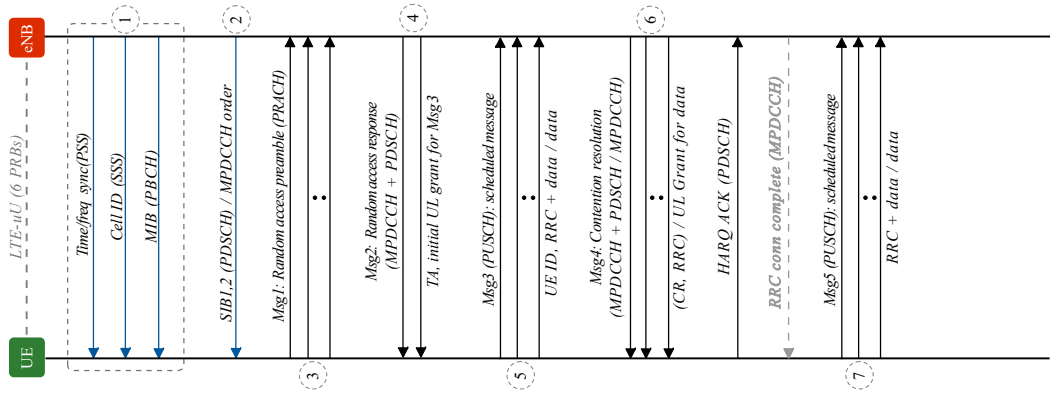
NB-IoT and eMTC RA

The process of establishing initial network connectivity in eMTC technology is illustrated in Fig.6.1(a). This method focuses on increasing access retry attempts to expand the coverage range of contemporary cellular communication systems. The main goal is to achieve UL synchronization and gain authorization to connect. To accomplish this, the ED interprets a fundamental configuration block, known as SIB2, transmitted by the BS. This block contains crucial parameters such as signal strength, carrier frequency, and configurations for PUCCH and PUSCH. Using these details, the ED chooses and sends a RA preamble, referred to as Msg1. If multiple EDs select the same preamble, resulting in a collision, the access resolution fails. Conversely, if no more than one ED picks a given preamble, the access resolution succeeds, and the BS allocates resources to the ED in the PUSCH through the RA response (Msg2). Subsequently, the ED transmits an RRC Connection Request Message to the BS (Msg3), followed by the BS sending a contention resolution message (Msg4) in response.

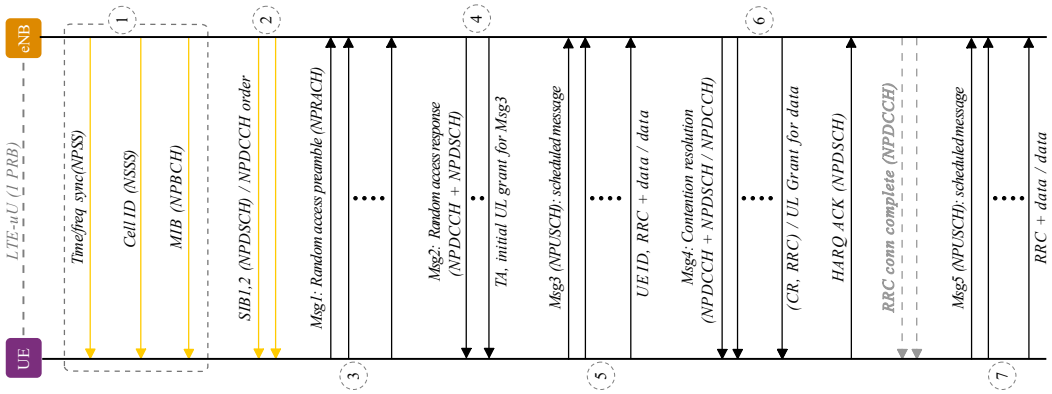
A timing diagram for the NB-IoT RAP is shown in Fig. 6.1(b). The process begins with the ED detecting the NB-IoT carrier by assessing the strength of received synchronization signals in the DL. It then achieves time and frequency synchronization and decodes the BS identifier (Cell ID). Subsequently, the ED decodes the NPBCH, which contains the MIB. Following synchronization, the ED sets up NPRACH resources and initiates UL preamble transmissions based on network configurations, ensuring adequate repetitions and transmission power. The NB-IoT protocol mandates a minimum of 12 out of 48 available orthogonal preambles (refer to Table 6.1). The final stage involves data transfer, which starts on NPDCCH, used for transmitting DL control information. This signal can be repeated between 1 and 2,048 times. Table A.1 outlines the primary features of the LPWAN technologies under consideration, while Table 6.1 presents the main parameters for the system model in this study, derived from the initial phase of the measurement campaign detailed in Chapter 4.

Considered Scenario and Assumptions

To describe the proposed system, we introduce several assumptions. We posit the existence of multiple RAT EDs, denoted by U, capable of utilizing both NB-IoT and eMTC technologies. We also assume the presence of only two BSs providing coverage: BS_1 for NB-IoT and BS_2 for eMTC. While each ED in set U can access either BS, it must select only one. The primary objective is to determine the opti-



(a) eMTC random access procedure.



(b) NB-IoT random access procedure.

Fig. 6.1: Timing diagrams for the random access procedures of NB-IoT and eMTC [1].

Tab. 6.1: Comparison of key random access and delay parameters of NB-IoT and eMTC in ECL0 according to 3GPP Release 13.

Technology	**NB-IoT**	**eMTC**
Number of preambles	12	36
Synchronization	300 ms	300 ms
Temporary collision-free RAP	40 ms	20 ms
RRC estab. time	200 ms	70 ms
Message handling time	10 ms	10 ms
200 B transfer time (from CONNECTED)	80 ms	10 ms
200 B transfer time (from IDLE)	300 ms	80 ms

mal distribution of EDs between the NB-IoT and eMTC BSs (subsets U_1 and U_2, respectively) to minimize the average access delay.

An ALOHA-based multi-channel random multiple access system can be used to

describe RAPs for both NB-IoT and eMTC systems, as shown in [142]. Several assumptions have been made in this regard.

- *Assumption 1: Preambles and slots.* BS_1 and BS_2 are related to random multiple access channels, named l_1 and l_2, representing the preambles. The channel transmission time is divided into equal slots. The durations of each slot are $t_{w,1}$ and $t_{w,2}$ for BS_1 and BS_2, respectively. The slot with index t is referred to as slot t. The ED cannot send a message until the start of the next window.
- *Assumption 2: Feasible event in a single slot.* Three possible scenarios can occur in a slot:
  - One ED (emergency department) sends information during the slot: denoted as S (success).
  - No EDs send information during the slot: denoted as E (empty).
  - Two or more EDs send information simultaneously during the slot: denoted as C (collision).
- *Assumption 3: Transmission probability.* Through observation of the channel output at the end of the slot, the BS gains knowledge of the event that occurred during the slot for all EDs. The BS then communicates this information to the EDs through a dedicated signaling channel. When EDs have messages to send, they use this information to decide whether to transmit a message based on the probability $P_t = \min(1, l/N_t)$, where N_t is the number of active EDs at the beginning of slot t and l is the number of preambles. This algorithm has been proven optimal in terms of mean delay, as demonstrated by Galinina [142].
- *Assumption 4: Ideal coverage.* Based on the assumption, the coverage areas of the NB-IoT and eMTC base stations are expected to be adequate enough to minimize data losses resulting from the weak signal reception.
- *Assumption 5: Arrival models.* We evaluated two models for message delivery: (i) periodic polling of EDs and (ii) random activity among EDs.
  - *Periodic polling.* The system employs a model that features periodic polling of EDs with a total of M EDs. At regular intervals of Δ units of time, a new message for transmission is simultaneously presented to all EDs. Upon generating the message, each ED attempts to relay it to the BS. If an ED fails to transmit a message before a new one appears, the untransmitted message is discarded. The two key parameters of the model are the number of EDs (M) and length of the polling period (Δ). The objective is to allocate EDs M_1 and M_2 to BS_1 and BS_2, respectively, in order to minimize the average access delay.
  - *Random ED activity.* A model featuring a random ED activity can potentially have an endless number of EDs. The message arrival process

was modeled using a homogeneous Poisson process with an intensity of Λ. Each event in this process corresponds to a new ED that is available to transmit a message. Then, the ED attempts to send messages to the BS. Once transmission is successful, the ED leaves the system. Therefore, the model with random ED activity is entirely determined by a single parameter, namely, the intensity of the input stream, Λ. In this scenario, the objective is to divide the intensity Λ into two separate intensities, Λ_1 and Λ_2, for BS_1 and BS_2, respectively, in order to minimize the average waiting time.

- *Assumption 6.* We assume that the time interval between the polling of EDs is sufficient such that by the end of the polling period, no active EDs remain. This is applicable to the model with periodic polling of EDs.

To compare the two models mentioned earlier, we assume that the average number of messages received by the system per unit time is equal, which is represented by the equation $\Lambda = M/\Delta$.

6.1.2 Multi-RAT Association Method

The Mean Delay for the Random Activity Model

A model with random ED activity can be represented using a Markov chain $\{N_t, t > 0\}$, where N_t denotes the number of EDs that possess a message that is ready for transmission. This is demonstrated in Section 5.2. The transition probabilities of this model are expressed as follows:

$$p_{ij} = p\{j \mid i\} = \sum_{m=\max(0,i-j)}^{l} \frac{\lambda_2^{m-(i-j)}}{(m-(i-j))!} \times e^{-\lambda_2} Pr\{T = m \mid i\}, \tag{6.1}$$

where m denotes the number of messages that are successfully transmitted.

Note that the lower bound of m in Equation (6.1) is defined as $\max(0, i-j)$, as it signifies the number of new active EDs entering the system. The upper bound of the sum is determined by l, which represents the maximum number of messages that can be transmitted in a single slot over available orthogonal channels. When conducting practical calculations, the number of rows and columns of Markov chain transition probability matrix can be truncated to a sufficiently large value such as $M + L_{\max}$. Here, $L_{\max}$ represents the chosen value, which means that the sum of the elements in the first row is approximately equal to 1 with a specified level of precision.

The method for calculating the mean delay for this model is based on calculating the number of active EDs at time $t - N_t$. To this aim, we utilize the transition probability matrix $P(\lambda)$ of $\{N_t, t > 0\}$, which is fully determined by λ. Let D represent

a random variable that indicates the number of time slots between when a message is generated and when it is successfully transmitted. The following algorithm establishes the relationship between delay and the arrival rate, expressed as $d(\lambda)$:

- *Step 1.* For a fixed value of $\lambda < le^{-1}$, utilize $P(\lambda)$ to obtain the stationary state distribution of $\{N_t, t > 0\}$

$$P(\lambda) \to \vec{\pi} = \left(\pi^0, \pi^1, \dots \pi^{n-1}\right). \tag{6.2}$$

- *Step 2.* Based on the stationary state distribution, calculate the mean number of EDs in the system as

$$E[N] = \sum_{i=0}^{n-1} i\pi^i. \tag{6.3}$$

- *Step 3.* Calculate $d(\lambda)$

$$d(\lambda) = \frac{E[N]}{\lambda}. \tag{6.4}$$

Considering the duration of the slot, the mean delay is given by the

$$\tau(\Lambda) = d(\Lambda t_w)t_w + t_{syn} + t_{data}. \tag{6.5}$$

To ascertain the optimal number of EDs associated with NB-IoT and eMTC technologies, it is crucial to determine the intensities Λ_1 and Λ_2, $\Lambda_1 + \Lambda_2 = \Lambda$ of the Poisson arrival process, in order to minimize the mean delay. This can be achieved by solving the following optimization problem:

$$\begin{aligned} &\max(\tau_1(\Lambda_1), \tau_2(\Lambda_2)) \to \min \\ &\text{s.t. } \Lambda_1 + \Lambda_2 = \Lambda. \end{aligned} \tag{6.6}$$

Thus, the equality $\tau_1(\Lambda_1) = \tau_2(\Lambda_2)$ that holds for the optimal distribution of EDs between the two technologies is straightforward. After obtaining $d_1(\lambda)$ and $d_2(\lambda)$ for both technologies using equation (6.4), the solution to equation (6.6) can be easily obtained by numerically solving the following equation

$$t_{s,1} + d_1(\Lambda_1 t_{w,1})t_{w,1} + t_{d,1} = t_{s,2} + d_2(\Lambda_2 t_{w,2})t_{w,2} + t_{d,2}, \tag{6.7}$$

with the following conditions

$$\begin{cases} \Lambda_1 + \Lambda_2 = \Lambda, \\ \Lambda_1 t_{w,1} < l_1 e^{-1}, \\ \Lambda_2 t_{w,2} < l_2 e^{-1}. \end{cases} \tag{6.8}$$

It should be noted that according to constraint (6.8), the Markov chain characterizing a system exhibiting random ED activity is ergodic. Consequently, it is possible to precisely determine the functions $d_1(\lambda)$ and $d_2(\lambda)$ for this type of system.

The Mean Delay for the Polling Model

The three methods for calculating the mean delay of a model with periodic polling of EDs vary in terms of accuracy and complexity.

The first approach is an exact method that employs the Markov model presented in Section 5.2. This Markov chain represents the number of active EDs at time t, assuming that M EDs with the measure for transmission arrive in the system at the polling time instant. The dimension of the matrix for this Markov chain is $(M+1) \times (M+1)$, and the transition probabilities are calculated in a manner similar to that in Subsection 6.1.2, but for $\lambda = 0$. We refer to this matrix as $P(0)$.

The following algorithm establishes the relationship among B, M, and δ, represented by $b(M, \delta)$. B is a random variable that describes the number of slots between the polling time instant and the time when it is successfully delivered to the BS, whereas $\delta = \Delta / t_w$ represents the number of slots in the polling period. The algorithm consists of four steps.

- *Step 1.* Form the initial state vector:

$$\vec{p_0} = (0, 0, \ldots 1). \tag{6.9}$$

- *Step 2.* For $t = 1, 2, \ldots \delta$ calculate the probability mass function (pmf) of the number of active EDs in slot t

$$\vec{p}_{t+1} = \vec{p}_t P(0). \tag{6.10}$$

- *Step 3.* For $t = 1, 2, \ldots \delta$ calculate the mean number of active EDs in the slot t as

$$E[N_t] = \sum_{i=0}^{n-1} i p_t^{(i)}. \tag{6.11}$$

- *Step 4.* Calculate $b(M, \delta)$ as

$$b(M, \delta) = \frac{\delta}{M - N_\delta} \sum_{i=0}^{\delta} \frac{E[N_t]}{\delta}. \tag{6.12}$$

The time it takes for a message to be delivered to the BS from the polling time instant is calculated in a similar manner to the model with random ED activity, but without considering the synchronization time because all EDs are in the RRC connected state. This is expressed by the equation

$$\tau(M, \Delta) = b(M, \Delta)/t_w) t_w + t_{data}. \tag{6.13}$$

The method described previously is computationally demanding owing to the large transition probability matrices that determine the number of active EDs in

slot t. To address this issue, we propose an alternative approach that utilizes a recursive equation to capture the gradual transitions of EDs from an active state to a state without messages ready for transmission. This approach allows for an approximate calculation of the mean delay using the recursive formula for the average number of active EDs

$$\hat{E}[N_{t+1}] = \hat{E}[N_t] - \max\left[0, l\hat{E}[N_t]\frac{P_t}{l}\left(1 - \frac{P_t}{l}\right)^{\hat{E}[N_t]-1}\right], \tag{6.14}$$

where $\hat{\cdot}$ denotes the approximation $P_t = \min(1, l/N_t)$, and the number of active EDs in the first slot is $\hat{E}[N_1] = M$.

The approximation in Equation 6.14 can be summarized as follows. At each time moment t and for each channel $i = 0, 1, \ldots, l$, messages can be successfully transmitted. Therefore, the second term is the maximum between 0 and the output flow intensity. The output flow rate was determined by multiplying the probability of successfully transmitting a message over l channels by P_t. Dividing this by l provides the probability of successfully transmitting a single message. The first factor represents the probability that one of the N_t active EDs chooses one of l channels. The second multiplier with power $\hat{E}[N_t]-1$ determines the probability of all possible events when the EDs do not transmit on a given channel. In addition, the mean delay is calculated using Equation 6.12.

Neither of the methods discussed previously offer a direct, closed-form solution for the mean delay, which is a useful feature for practical applications. To address this, we propose an approximation for (6.14) as (6.15)

$$\hat{\hat{E}}[N_{t+1}] = \left(\hat{\hat{E}}[N_t] - le^{-1}\right) I\left(\hat{\hat{E}}[N_t] \geq le^{-1}\right), \tag{6.15}$$

where $\hat{\hat{\cdot}}$ denotes an additional approximation of the number of active EDs in the system at time slot t.

The primary objective of this method is to use an indicator function instead of the second multiplier in Equation (6.14). This strategy allows for a more rapid calculation of the average number of active EDs, which is useful when dealing with large time intervals between batch arrivals to the system, as opposed to using the Markov model and Equation (6.14).

As shown in (6.15), all EDs vacate the system within the time frame of e^{-1}. This results in the following approximate representation for $b(M, \delta)$

$$\hat{\hat{b}}(M, \delta) = \begin{cases} M/2le^{-1} & \text{if } M/le^{-1} \geq \delta, \\ M/le^{-1} - \delta/2 & \text{otherwise.} \end{cases} \tag{6.16}$$

To demonstrate the precision of the methods discussed, the graph in Fig. 6.2 depicts the change in the average number of active emergency department personnel

over time, obtained through the following techniques: (i) computer modeling, (ii) employing the Markov chain method via equations (6.9)–(6.11), (iii) an estimate based on the repeating equation (6.14), and (iv) a closed-form approximation based on equation (6.15). It is evident from the graph that all approaches examined are consistent with the simulation results.

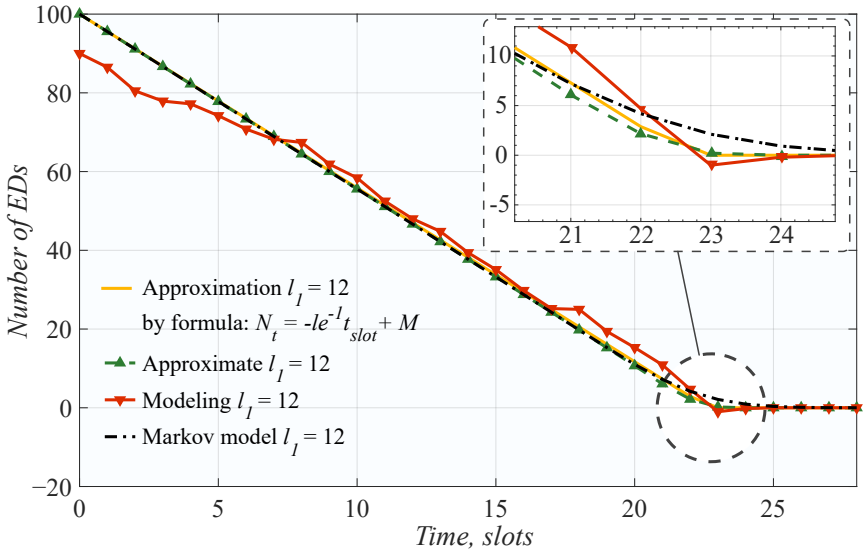


Fig. 6.2: The mean number of active EDs in the system [1].

To find the best distribution of the number of NB-IoT and eMTC technologies, M_1 and M_2, such that $M_1 + M_2 = M$, one must solve the following optimization problem

$$\begin{aligned} &\max(\tau_1(M_1, \Delta), \tau_2(M_2, \Delta)) \to \min \\ &\text{s.t. } M_1 + M_2 = M. \end{aligned} \tag{6.17}$$

We find that the equality $\tau_1(M_1) = \tau_2(M_2)$ holds true for the optimal values of M_1 and M_2. Consequently, once $b_1(M_1, \delta)$ and $b_2(M_2, \delta)$ are obtained using equation (6.16), the problem simplifies to solving the numerical equation

$$b_1(M_1, \Delta/t_{w,1})t_{w,1} + t_{d,1} = b_2(M_2, \Delta/t_{w,2})t_{w,2} + t_{d,2}, \tag{6.18}$$

subject to the following conditions

$$\begin{cases} M_1 + M_2 = M, \\ M_1/(l_1 e^{-1}) < \Delta/t_{w,1}, \\ M_2/(l_2 e^{-1}) < \Delta/t_{w,2}. \end{cases} \tag{6.19}$$

Note that under the constraints defined in Equation (6.19), all EDs will successfully transmit their messages by the end of the polling period Δ. To further

simplify the analysis, instead of using the precise values of $b_1(M_1, \delta)$ and $b_2(M_2, \delta)$, we can rely on estimates based on Equation (6.16) and obtain explicit solutions. To illustrate this point, consider scenarios where $M_1/l_1e^{-1} < \delta$ and $M - M_1/l_2e^{-1} < \delta$. In these cases, the delays are primarily caused by

$$\begin{aligned}\tau_1 &= \frac{M_1}{2l_1e^{-1}}t_{w,1} + t_{d,1},\\ \tau_1 &= \frac{M - M_1}{2l_2e^{-1}}t_{w,2} + t_{d,2}.\end{aligned} \tag{6.20}$$

In order to calculate M_1, the following equation must be solved:

$$\frac{M_1}{2l_1e^{-1}}t_{w,1} + t_{d,1} = \frac{M - M_1}{2l_2e^{-1}}t_{w,2} + t_{d,2}. \tag{6.21}$$

provided that (6.19) is met.

The solution for equation (6.21) is expressed as follows

$$M_1 = \frac{Ml_1t_{w,2} + l_1l_2e^{-1}(t_{d,2} - t_{d,1})}{l_1t_{w,2} + l_2t_{w,1}}. \tag{6.22}$$

At the optimal point, the delay for both technologies is the same and is equal to

$$\tau_O = \frac{Ml_1t_{w,2} + l_1l_2e^{-1}(t_{d,2} - t_{d,1})}{(l_1t_{w,2} + l_2t_{w,1})2l_1e^{-1}}t_{w,1} + t_{d,1}. \tag{6.23}$$

6.1.3 Numerical Results

This section presents our numerical results. We examine the optimal allocation of EDs between NB-IoT and eMTC technologies for both random ED activity and polling models. Our findings are compared with computer simulation outcomes. Furthermore, we illustrate how the optimal distribution varies with the number of EDs in the system and showcase the optimal message delay. The numerical evaluation utilizes standard NB-IoT and eMTC parameters outlined in Table 6.1.

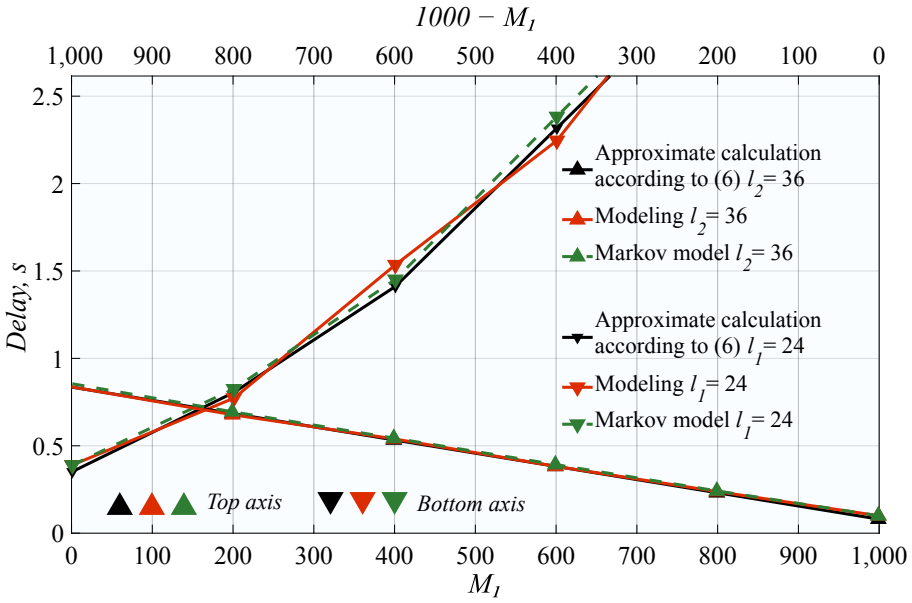
Our investigation begins by examining the optimal distribution of EDs between NB-IoT and eMTC technologies. To achieve this, we use computer simulations, the Markov model specified by Equations (6.9)–(6.11), and a closed-form approximation in Equation (6.15). The upper axis of the graph in Fig. 6.3(a) shows the number of EDs associated with the eMTC technology, M_2, while the lower axis shows the number of EDs associated with the NB-IoT technology, M_1. From this analysis, we can see that the approximation in equation (6.15) provides an accurate representation of the mean delay experienced by messages for both technologies across a wide range of numbers of EDs in the system. This aligns with the time-dependent evolution of the number of EDs in the system after the polling time instant, as shown in Fig. 6.2.

Examination of Fig. 6.3(a) indicates a convergence point between NB-IoT and eMTC technologies. The lower and upper sections of the x axes display the distribution of EDs (from a total of 1,000) that should be allocated to each technology. The observed distribution was roughly 180:820, with NB-IoT having 24 preambles and eMTC having 36. For this ED allocation, message delays were approximately 700 ms, which is considerably less than the 10 s threshold outlined in ITU-R M.2410 [24] and aligns with the 2 s limit established in Chapter 3 for which even distribution of 500:500 EDs satisfies this given constraint. It's worth noting that this distribution is also influenced by the timing characteristics of each technology (see Table 6.1).

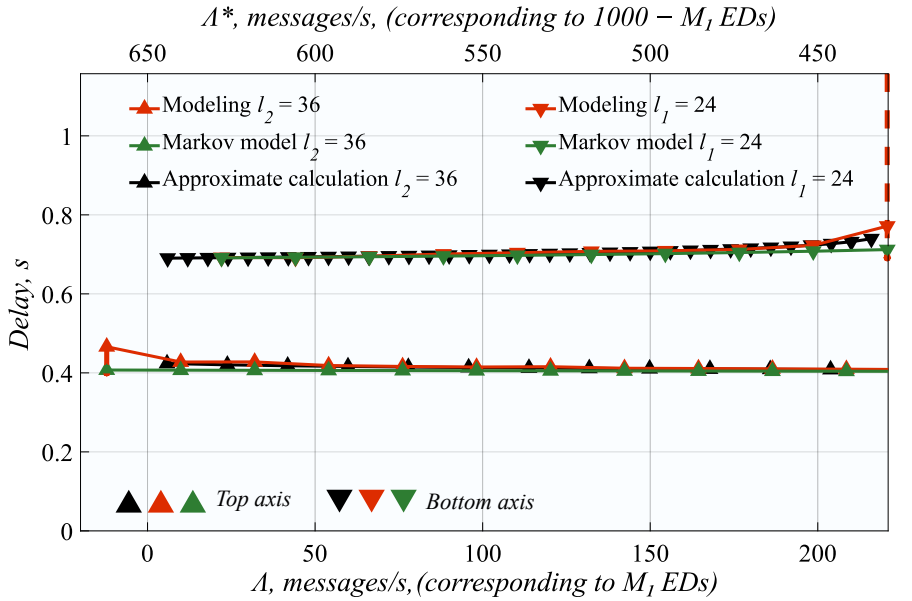
Examine the graph depicting a model with random ED activity, where message arrivals follow a Poisson process. Fig. 6.3(b) illustrates an overall intensity Λ of 1,000 EDs surveyed every quarter-hour. Analysis of the data revealed no intersection point indicating an optimal distribution of EDs between the technologies under consideration. The delay associated with eMTC technology is notably shorter than that of NB-IoT, which can be attributed to two factors. Firstly, eMTC offers a substantially larger selection of preambles for EDs compared to NB-IoT. More crucially, EDs in random activity mode must first achieve RRC connected state before transmitting a random preamble. For NB-IoT, this process takes significantly longer than for eMTC (refer to Table 6.1). These parameters are explicitly incorporated in equations (6.5) and (6.13), respectively. The impact of this timing difference is substantial and overshadows the effect of available preamble quantity. Even when the number of preambles for NB-IoT was increased to impractical levels (36 and subsequently 48), eMTC alone still demonstrated superior performance compared to the combination of these technologies. Consequently, the implementation of Multi-RAT EDs is advisable only when eMTC technology reaches complete overload.

Following the identification of the model's optimal point using periodic ED polling, we evaluated the ideal number of EDs that should be linked to NB-IoT and eMTC technologies. We also analyzed their proportions and average delays associated with the optimal ED distribution. To achieve this, both analytical modeling and simulation-based evaluation were employed. The analytical results, utilizing equations (6.5) and (6.13), provide theoretical predictions for the optimal ED distribution, while simulation results validate these findings under practical conditions.

We started by investigating the optimal fraction of EDs connected to eMTC technology in relation to the total number of EDs in the system, using a 15-minute polling interval, as shown in Fig. 6.4. The results indicate that as the quantity of active EDs grows, there is a corresponding linear increase in the number of EDs that need to be linked to eMTC technology. In contrast, the proportion of EDs that should be associated with eMTC technology exhibits a nonlinear decrease as the total number of EDs rises.



(a) Periodic polling model.



(b) Random EDs activity model.

Fig. 6.3: Optimal association point for the two models considered [1].

The numerical analysis in this research unveiled the correlation between the optimal allocation of EDs across various technologies and the resulting delay. As illustrated in Fig. 6.5, the delay initially remains minimal. However, as the quantity of simultaneously polled EDs increased, the delay exhibited a linear growth. Comparing the outcomes presented in Fig. 6.5 with those detailed in Section 5.2, it becomes evident that the optimized use of Multi-RAT EDs enables a substantial in-

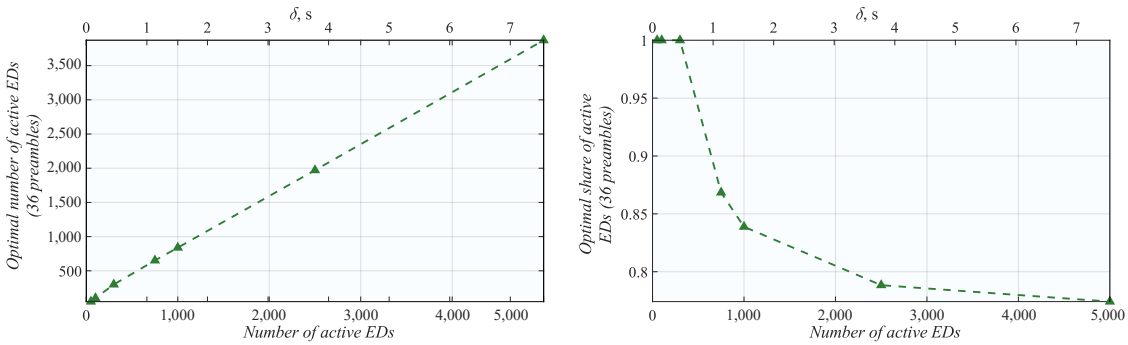


Fig. 6.4: Optimal number and fraction of EDs at eMTC for polling model [1].

crease in the number of supported EDs operating in polling-based mode, compared to a single NB-IoT interface. Notably, for 1,000 EDs, the average delay for a single NB-IoT technology exceeded 10 s, which fails to meet the minimum requirements outlined in ITU-R M.2410 [24]. In contrast, our study demonstrated that even with 5,000 EDs, the average delay remained under 3 s, and 3,000 EDs still satisfied the 2 s requirement specified in Chapter 3. These improvements surpass linear gains, as the average delay escalates more rapidly than a linear function (refer to Fig. 6.3).

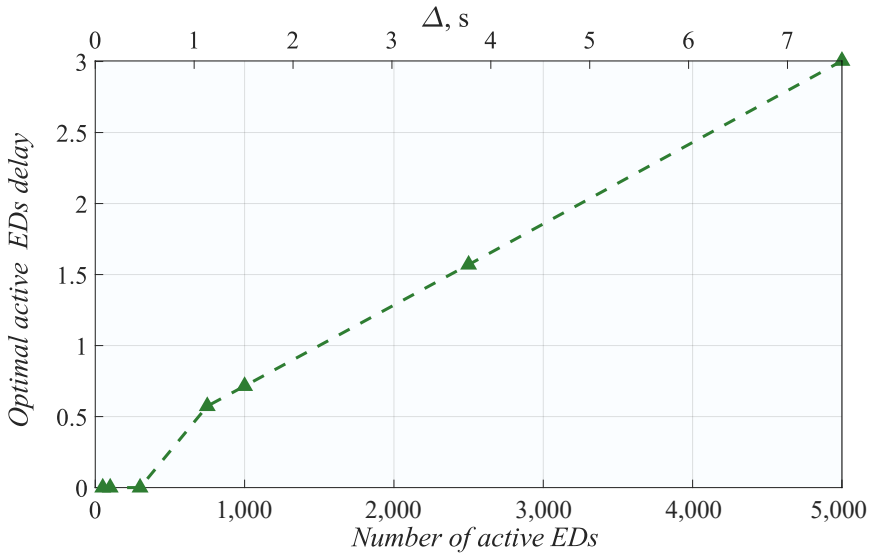


Fig. 6.5: Delay related to the optimal deployment of EDs [1].

6.2 Summary

This chapter addressed **RQ4** by conducting a comparative analysis and performance evaluation of NB-IoT and eMTC for smart-grid applications. To improve the dependability of mMTC communication technologies, a theoretical model and corresponding algorithm were developed to balance the traffic load between NB-IoT and

eMTC with the aim of minimizing the mean access delay, and primarily to increase communication reliability. The results show that optimizing the use of Multi-RAT EDs substantially boosts the quantity of EDs functioning in polling-based mode. When over 500 EDs employ single eMTC technology, the average access delay surpasses two seconds, which is incompatible with smart-grid application requirements. In contrast, Multi-RAT EDs implementing the newly developed algorithm improve service capacity by a factor of six (accommodating up to 3,000 EDs) while still meeting the two-second latency threshold.

The key takeaways can be summarized as follows:

- Theoretical frameworks and a closed-form solution facilitate optimal linkages to Multi-RAT EDs using NB-IoT and eMTC systems, considering the scheduled polling and random activity of EDs in terms of average access delays. The framework includes several elements: the NB-IoT and eMTC base station, a group of EDs with BS access, a defined quantity of preambles and time slots, the likelihood of transmission under ideal coverage conditions, and two arrival patterns: (i) periodic polling of EDs and (ii) spontaneous ED activity.
- Observations: (i) An optimal balance in load distribution between NB-IoT and eMTC technologies can always be achieved when operating within standard parameter values in a RRC connected state. (ii) For EDs with purely random arrivals, which typically alternate between idle and connected states, eMTC is consistently the preferred technology. NB-IoT should only be employed when eMTC reaches full capacity. (iii) The optimized use of Multi-RAT EDs can support up to six times more EDs in polling-based mode compared to a single NB-IoT system.

Chapter 6 successfully answered **RQ4**. The findings demonstrate that eMTC should be the primary technology for smart-grid applications due to its superior performance in terms of data rates, latency, and reliability. NB-IoT, while not as performant, provides significant benefits in terms of coverage and power efficiency, making it an ideal backup in a Multi-RAT system. The optimal load balancing strategy proposed ensures efficient handling of diverse traffic types and enhances overall network resilience. The results from this chapter, in relation to the observations in Chapter 4, demonstrate potential for real-world enhancements through the implementation of the proposed method in both EDs and the network.

7 Conclusions and future steps

This chapter provides a comprehensive overview of the key results and discoveries that emerged from the research while also examining their relevance in light of existing knowledge. This chapter also provides possible paths for future research.

7.1 Summary of doctoral thesis outcomes

The shift from communication networks designed for human connectivity to industry-driven scenarios has led to the emergence of 5G mMTC technologies, which require reevaluation of how devices interact with networks. This transition from legacy IoT use cases, which followed irregular traffic patterns, to new applications demanding permanent connectivity presents challenges, even for Mobile IoT or LPWAN technologies, such as NB-IoT and LTE-M, also known as eMTC, which were initially designed for sporadic communication. Among the emerging use cases, smart-grids stand out, driven by requirements such as permanent connectivity, guaranteed bit rates, and substantial data transmission, all of which are further motivated by the European Parliament Directive 2019/944 mandating smart meter deployments starting in 2024. As utilities seek long-term reliability and resilience in communication networks, these new demands push telco operators to adapt their offerings to support AMI and to meet the evolving needs of smart-grid applications.

Based on these facts and requirements, this doctoral thesis aims to propose a solution to cover smart-grid requirements by deploying mMTC technologies NB-IoT and eMTC that are identified as capable of the task. Both technologies have proven sustainable as they continue to evolve, even in 5G networks, as 5G IoT technologies. The aim of this thesis is not only to propose a technology, but also to enhance communication reliability and service provision, as both technologies are unable to meet all the requirements in every situation. These proposals have not been sufficiently covered by the research community and thus could greatly benefit both researchers and the industry in upcoming and future smart meter deployments.

Gaining a deep understanding of the evolution and capabilities of mMTC technologies is crucial for suggesting dependable solutions for emerging use cases that have specific demands, such as smart metering and other industrial applications. Therefore, the first part of this thesis provides a comprehensive overview of potential solutions, taking into account the evolution of mobile IoT technologies, primarily NB-IoT and eMTC, and examines the advantages of employing these technologies in the aforementioned sectors, with a focus on smart-grid requirements and expectations. Afterward, extensive measurements were carried out to investigate the performance of both technologies under ideal conditions and network congestion in

the RAN. Based on the results of these measurements, and considering the future technological advancements of both technologies, mathematical models and simulations were developed. These were designed to (i) decrease communication delay on a single interface, taking into account both stochastic and regular traffic types generated by traditional sensors and smart meters, and (ii) enhance communication reliability and decrease congestion by deploying multiple RATs of NB-IoT and eMTC with optimized load balancing.

A summary of the research outcomes obtained by answering the research questions formulated in Section 1.2 are provided at the end of each Chapter, that is in Sections 2.4, 3.4, 4.3, 5.4, and 6.2. The following list summarizes key findings and primary observations to illustrate how this thesis fulfills the established research objectives.

- Comprehensive overview of Mobile IoT technologies, focusing on the evolution and deployment of NB-IoT and eMTC for smart-grid applications. The chapters also highlighted the challenges and specific requirements for AMI deployments in the Czech Republic and the broader European context.
- Measurement campaigns were conducted to evaluate the performance of technologies NB-IoT and eMTC under various network conditions. The results showed that NB-IoT generally had better signal coverage than eMTC, with an approximate 15 dB difference in RSRP. However, NB-IoT did not meet the smart-grid requirements of guaranteed 11 kbps and 2 s communication delay at RSRP levels above ECL0, whereas eMTC was able to do so until its operational limits. The choice between these technologies depends on the specific use cases: NB-IoT is better for battery-powered devices under poor radio conditions with low traffic requirements, whereas eMTC is more suitable for high-traffic, low-latency applications. The study found that eMTC is ideal for use as the primary interface for smart-grid applications, whereas NB-IoT can serve as a backup link. The second phase of the campaign's stress test demonstrated that the NB-IoT struggles with burst traffic typical of smart-grids, further confirming that eMTC is the preferred option.
- Developed a generalized Markov chain model tailored specifically for NB-IoT traffic, with the aim of optimizing the coexistence of stochastic and regular traffic. The proposed solution demonstrated that model for the NB-IoT network can remain stable for up to 72×10^4 conventional UEs per cell (using a sectoral antenna) and 9×10^3 UEs requiring permanent connectivity. While permanently connected UEs have a linear impact on their delay performance, they significantly affect the conventional UEs. The ITU-R M.2410 delay performance requirement of 10 s was met for the conventional UEs, even with up to 10^3 permanently connected UEs. However, the delay for permanently

connected UEs exceeded acceptable levels of 2 s defined by smart-grids even with 100 such UEs, indicating a need to redesign the NB-IoT channel access mechanism or increase the available resources. Simulations showed that extending the deployment to three RBs met the 2 s delay requirement for up to 240 UEs with regular traffic. This finding raises questions regarding the suitability of the current NB-IoT technology for 5G mMTC, particularly for UEs requiring uninterrupted connectivity.

- Introduced a streamlined model that focuses on optimizing the dissemination of routine UE messages, considering reading intervals of 15 min. The analysis revealed an optimal spreading interval of approximately 7.5 min the delay for Type I messages under specific traffic conditions. This approach is essential in situations that require swift deployment and immediate performance improvements. The results showed that the optimal spreading interval substantially reduces the delay, and this interval, along with the corresponding ideal delay for routine traffic, depends linearly on the number of UEs generating regular traffic. For instance, with fewer than 500 Type I UEs, an optimal message spreading interval of approximately 5 s could meet the existing 2 s delay requirement in NB-IoT networks that are currently in use.
- Examined EDT strategies in NB-IoT to decrease delays, particularly under heavy network traffic. By employing the EDT, transmission delays were reduced by more than 50%, significantly enhancing the spectrum efficiency with minimal added overhead. This improvement was especially noticeable for smaller messages that were less than 97 B in size in the UL and 57 B in size in the DL. NB-IoT networks demonstrated their ability to effectively manage 10^3 UEs within a 15 min window, even with basic UP optimization. However, the correct network configuration is critical, as simulations have shown that RAP collisions double when transmitting a 64 B message with an EDT. This emphasizes the importance of proper network tuning to achieve peak capacity and optimal efficiency.
- A study comparing NB-IoT and eMTC for use in smart-grid applications was conducted, and a Multi-RAT approach for optimal load balancing was proposed. The results showed that distributing 1,000 devices with a ratio of 180 NB-IoT to 820 eMTC devices achieved a delay of approx. 700 ms, and ratio satisfying the 2-second constraint up until 500 to 500 UEs. Moreover, the results indicate that up to 3,000 distributed UEs following these ratios maintain a communication delay of no more than 2 seconds. This is a considerable improvement over single-RAT deployments, where delays could exceed 10 s, even for 1,000 UEs. This chapter emphasizes the advantages of eMTC in handling data-intensive tasks, while NB-IoT serves as a secondary interface, provid-

ing better coverage and power efficiency, making the combination suitable for smart-grid applications.

The primary objective of this thesis is to offer beneficial contributions to both corporate sector and the research community. By conducting a performance analysis of the technologies explored, this thesis aims to aid companies in making informed deployment decisions. Additionally, the research community can use the findings of this study to gain insights into the optimization of Mobile IoT systems in real-world situations. The findings of this study suggest that the optimal use of the NB-IoT, eMTC, or a combination of both is a viable solution for smart meter deployment. This conclusion is not only advantageous for electricity distributors, but also serves as an essential lesson for them to avoid relying solely on default network configurations and to carefully optimize communication patterns and parameters.

7.2 Future research orientation

5G-IoT Enhancements for IIoT Use Cases Presented by 3GPP Release 16+

In accordance with Chapter 2, the emphasis placed on enhancing connectivity for industrial purposes in Release 16 is noteworthy. A number of the mentioned features, such as optimized security, multicast support, grant-free access, and wake-up signals, offer significant potential for improving RAN operations for demanding use cases, such as smart-grids with unexpected requirements. By conducting a performance analysis on these new features, it may be possible to establish optimal parameter associations for industrial use cases, which could potentially build upon the findings of this study.

5G-IoT Performance Analysis and Optimizations for NTN Deployments

To address the pressing issue of remote connectivity in challenging environments, the current focus is on NTN support for the 5G-IoT, as outlined in Chapter 2. At present, trials of the NB-IoT over NTN deployments are underway, with both industry and academia working diligently to advance this technology. The NB-IoT pilot is a promising solution owing to its low complexity and narrow bandwidth, which can be supported by CubeSats. Building on previous research that primarily focused on the RAN part of terrestrial Mobile IoT, this work could be extended to include low-orbit satellite coverage and address unsolved problems. However, there are several challenges associated with handovers between satellites and terrestrial stations as well as between the satellites themselves. While previous studies have investigated critical challenges, such as random access, there is a lack of real-world examples and optimization strategies for deployment [152–154].

References to the main author's publications

[1] **R. Mozny**, P. Masek, D. Moltchanov, M. Stusek, P. Mlynek, Y. Koucheryavy, and J. Hosek. "Characterizing optimal LPWAN access delay in massive multi-RAT smart grid deployments". In: *Internet of Things* 25 (2024), p. 101001. ISSN: 2542-6605. DOI: https://doi.org/10.1016/j.iot.2023.101001. URL: https://www.sciencedirect.com/science/article/pii/S2542660523003244.

[2] P. Masek, D. Moltchanov, M. Stusek, **R. Mozny**, Y. Koucheryavy, and J. Hosek. "Quantifying NB-IoT Performance in 5G-IoT Use-Cases with Mixture of Regular and Stochastic Traffic". In: *IEEE Internet of Things Journal* (2024). Accepted for publication.

[3] M. Stusek, P. Masek, R. Dvorak, T. L. Dinh, **R. Mozny**, K. Zeman, A. Ometov, P. Cika, P. Mlynek, and J. Hosek. "Exploiting NB-IoT Network Performance and Capacity for Smart-Metering Use-Cases". In: *2023 15th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. 2023, pp. 193–199. DOI: 10.1109/ICUMT61075.2023.10333294.

[4] M. Stusek, N. Stepanov, D. Moltchanov, P. Masek, **R. Mozny**, A. Turlikov, and J. Hosek. "Optimizing NB-IoT Communication Patterns for Permanently Connected mMTC Devices". In: *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. 2022, pp. 1413–1418. DOI: 10.1109/WCNC51071.2022.9771847.

[5] M. Mikulasek, R. Dvorak, M. Stusek, P. Masek, **R. Mozny**, P. Mlynek, and J. Hosek. "NB-IoT vs LTE Cat M1: Demystifying Performance Differences under Varying Radio Conditions". In: *2022 14th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. 2022, pp. 133–138. DOI: 10.1109/ICUMT57764.2022.9943485.

[6] **R. Mozny**, M. Stusek, P. Masek, K. Mikhaylov, and J. Hosek. "Unifying Multi-Radio Communication Technologies to Enable mMTC Applications in B5G Networks". In: *2020 2nd 6G Wireless Summit (6G SUMMIT)*. 2020, pp. 1–5. DOI: 10.1109/6GSUMMIT49458.2020.9083791.

[7] **R. Mozny**, P. Masek, M. Stusek, K. Zeman, A. Ometov, and J. Hosek. "On the Performance of Narrow-band Internet of Things (NB-IoT) for Delay-tolerant Services". In: *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*. 2019, pp. 637–642. DOI: 10.1109/TSP.2019.8768871.

Other references

[8] M. Stusek, **R. Mozny**, P. Masek, P. Kriz, K. Molnar, M. Palenska, T. Dinh Le, R. Dvorak, D. Moltchanov, P. Mlynek, P. Cika, and J. Hosek. "Current State of Localization Methods in 5G networks: Performance Assessment in Real Large-Scale Deployments". In: *Internet of Things* (2024). Under review. ISSN: 2542-6605.

[9] K. Mikhaylov, M. Stusek, P. Masek, R. Fujdiak, **R. Mozny**, S. Andreev, and J. Hosek. "On the Performance of Multi-Gateway LoRaWAN Deployments: An Experimental Study". In: *2020 IEEE Wireless Communications and Networking Conference (WCNC)*. 2020, pp. 1–6. DOI: 10.1109/WCNC45663.2020.9120655.

[10] P. Masek, M. Stusek, K. Zeman, **R. Mozny**, A. Ometov, and J. Hosek. "A Perspective on Wireless M-Bus for Smart Electricity Grids". In: *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*. 2019, pp. 730–735. DOI: 10.1109/TSP.2019.8768840.

[11] **R. Mozny**, P. Masek, M. Stusek, K. Molnar, M. Palenska, D. Moltchanov, and J. Hosek. "Experimental Quality Assessment of Cellular Networks and their Utilization for UAV Services". In: *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*. 2023, pp. 1–6. DOI: 10.1109/VTC2023-Spring57618.2023.10200310.

[12] M. Babela, P. Musil, **R. Mozny**, M. Mahut, and M. Rusz. "Monitoring Vehicles Using OBD II and LPWA Network". In: *2023 15th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. 2023, pp. 17–21. DOI: 10.1109/ICUMT61075.2023.10333099.

[13] R. Dvorak, L. Jabloncik, M. Mikulasek, M. Stusek, P. Masek, **R. Mozny**, A. Ometov, P. Mlynek, P. Cika, and J. Hosek. "LWM2M for Cellular IoT: Protocol Implementation and Performance Evaluation". In: *2023 15th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. 2023, pp. 212–218. DOI: 10.1109/ICUMT61075.2023.10333286.

[14] **R. Mozny**, P. Ilgner, P. Dzurenda, and P. Cika. "Design of Physical Security for Constrained End Devices within the IoT Ecosystem". In: *2022 14th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. 2022, pp. 85–89. DOI: 10.1109/ICUMT57764.2022.9943338.

[15] P. Masek, E. Younesian, M. Bahna, **R. Mozny**, M. Mikulasek, M. Stusek, A. Ometov, J. Hosek, R. Fujdiak, and P. Mlynek. "Performance Analysis of Different LoRaWAN Frequency Bands for mMTC Scenarios". In: *2022 45th International Conference on Telecommunications and Signal Processing (TSP)*. 2022, pp. 417–420. DOI: 10.1109/TSP55681.2022.9851311.

[16] M. Mikulasek, P. Masek, M. Stusek, L. Novak, **R. Mozny**, and J. Hosek. "Optimizing the Switching Speed of the Current Probe Utilizing the FPGA for Input Signal Processing". In: *2021 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. 2021, pp. 174–181. DOI: 10.1109/ICUMT54235.2021.9631579.

[17] **R. Mozny**, P. Masek, M. Stusek, M. Mikulasek, A. Ometov, and J. Hosek. "Pitfalls of LPWA Power Consumption: Hands-On Design of Current Probe". In: *2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. 2020, pp. 13–19. DOI: 10.1109/ICUMT51630.2020.9222449.

[18] K. Mikhaylov, M. Stusek, P. Masek, R. Fujdiak, **R. Mozny**, S. Andreev, and J. Hosek. "Communication Performance of a Real-Life Wide-Area Low-Power Network Based on Sigfox Technology". In: *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*. 2020, pp. 1–6. DOI: 10.1109/ICC40277.2020.9148645.

[19] N. Mahmood, O. Lopez, O. S. Park, I. Noerman, K. Mikhaylov, E. Mercier, A. Munari, F. Clazzer, S. Bocker, and H. Bartz. *Massive Machine Type Communication Towards 6G*. [White paper]. (6G Research Visions, No. 11). University of Oulu. June 2020.

[20] V. Cisco. "Cisco Visual Networking Index: Forecast and Trends, 2017–2022". In: *White Paper* (2019).

[21] *3GPP Low Power Wide Area Technologies*. White Paper. GSMA, May 2018.

[22] O. Liberg, M. Sundberg, E. Wang, J. Bergman, and J. Sachs. *Cellular Internet of Things: Technologies, Standards, and Performance*. Academic Press, 2017.

[23] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang. "6G Wireless Communication Systems: Applications, Requirements, Technologies, Challenges, and Research Directions". In: *IEEE Open Journal of the Communications Society* 1 (2020), pp. 957–975. DOI: 10.1109/OJCOMS.2020.3010270.

[24] M. Series. "Minimum Requirements Related to Technical Performance for IMT-2020 Radio Interface (s)". In: *Report* (2017), pp. 2410–.

[25] Ericsson. *Ericsson Mobility Report*. 2024.

[26] A. ElNashar and M. A. El-Saidny. *Practical Guide to LTE-A, VoLTE and IoT: Paving the way towards 5G*. John Wiley & Sons, 2018.

[27] Y. Li, X. Cheng, Y. Cao, D. Wang, and L. Yang. "Smart Choice for the Smart Grid: Narrowband Internet of Things (NB-IoT)". In: *IEEE Internet of Things Journal* 5.3 (2018), pp. 1505–1515.

[28] K. Mekki, E. Bajic, F. Chaxel, and F. Meyer. "Overview of Cellular LPWAN Technologies for IoT Deployment: Sigfox, LoRaWAN, and NB-IoT". In: *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE. 2018, pp. 197–202.

[29] P. Masek, M. Stusek, K. Zeman, J. Hosek, K. Mikhaylov, S. Andreev, Y. Koucheryavy, O. Zeman, J. Votapek, and M. Roubicek. "Tailoring NB-IoT for Mass Market Applications: A Mobile Operator's Perspective". In: *2018 IEEE Globecom Workshops (GC Wkshps)*. IEEE. 2018, pp. 1–7.

[30] B. Martinez, F. Adelantado, A. Bartoli, and X. Vilajosana. "Exploring the Performance Boundaries of NB-IoT". In: *arXiv preprint arXiv:1810.00847* (2018).

[31] M. Kamel, W. Hamouda, and A. Youssef. "Uplink Coverage and Capacity Analysis of mMTC in Ultra-Dense Networks". In: *IEEE Transactions on Vehicular Technology* 69.1 (2020), pp. 746–759. DOI: 10.1109/TVT.2019.2954233.

[32] S. Persia and L. Rea. "Next Generation M2M Cellular Networks: LTE-MTC and NB-IoT Capacity Analysis for Smart Grids Applications". In: *2016 AEIT International Annual Conference (AEIT)*. 2016, pp. 1–6. DOI: 10.23919/AEIT.2016.7892789.

[33] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu. "5G Network Capacity: Key Elements and Technologies". In: *IEEE Vehicular Technology Magazine* 9.1 (2014), pp. 71–78. DOI: 10.1109/MVT.2013.2295070.

[34] D. Han, H. Minn, U. Tefek, and T. J. Lim. "Network Dimensioning, QoE Maximization, and Power Control for Multi-Tier Machine-Type Communications". In: *IEEE Transactions on Communications* 67.1 (2019), pp. 859–872. DOI: 10.1109/TCOMM.2018.2875735.

[35] M. Stusek, D. Moltchanov, P. Masek, J. Hosek, S. Andreev, and Y. Koucheryavy. "Learning-Aided Multi-RAT Operation for Battery Lifetime Extension in LPWAN Systems". In: *2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. 2020, pp. 26–32. DOI: 10.1109/ICUMT51630.2020.9222440.

[36] A. Nouicer and L. Meeus. "The EU Clean Energy Package". In: *Florence School of Regulation* (2019).

[37] E. Union. "Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on Common Rules for the Internal Market for Electricity and Amending Directive 2012/27". In: *J. Eur. Union* 158 (2019), pp. 125–199.

[38] NAP 5G. *AMM Metering Using the Non-PLC Technologies*. Accessed 2024-03-04. Jan. 2020. URL: https://www.mpo.cz/cz/energetika/strategicke-a-koncepcni-dokumenty/narodni-akcni-plan-pro-chytre-site/jednani-think-tanku-nap-sg-18--1-2021--259047/.

[39] N. Myoung, Y. Kwon, M. Park, and C. Eun. "Data Interworking Model and Analysis for Harmonization of Smart Metering Protocols in IoT-Based AMI System". In: *Sensors* 23.6 (2023). ISSN: 1424-8220. URL: https://www.mdpi.com/1424-8220/23/6/2903.

[40] C. Alaton and F. Tounquet. In: *Tractebel Impact: Antwerpen, Belgium* (2020).

[41] O. Vikhrova, S. Pizzi, A. Molinaro, and G. Araniti. "Paging Group Size Distribution for Multicast Services in 5G Networks". In: *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE. 2020, pp. 484–489.

[42] EU. *Advanced Technologies for Industry: Product Watch IoT Components in Connected and Autonomous Vehicles*. 2020.

[43] R. Germanà, E. De Santis, F. Liberati, and A. Di Giorgio. "On the Participation of Charging Point Operators to the Frequency Regulation Service using Plug-in Electric Vehicles and 5G Communications". In: *2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*. IEEE. 2021, pp. 1–6.

[44] M. Kemal, R. Sanchez, R. Olsen, F. Iov, and H.-P. Schwefel. "On the Trade-Off Between Timeliness and Accuracy for Low Voltage Distribution System Grid Monitoring Utilizing Smart Meter Data". In: *International Journal of Electrical Power & Energy Systems* 121 (2020), p. 106090. ISSN: 0142-0615.

DOI: https://doi.org/10.1016/j.ijepes.2020.106090. URL: https://www.sciencedirect.com/science/article/pii/S0142061519316801.

[45] G. Giaconi, D. Gunduz, and H. V. Poor. "Privacy-Aware Smart Metering: Progress and Challenges". In: *IEEE Signal Processing Magazine* 35.6 (2018), pp. 59–78. DOI: 10.1109/MSP.2018.2841410.

[46] Y. Perez and W. Arowolo. "Integration of Electromobility with the Electric Power Systems: The Key Challenges". In: *Annales des Mines-Enjeux Numériques* (2021).

[47] D. Ghose, L. Tello-Oquendo, V. Pla, and F. Y. Li. "On the Behavior of Synchronous Data Transmission in WuR enabled IoT Networks: Protocol and Absorbing Markov Chain based Modeling". In: *IEEE Transactions on Wireless Communications* (2022).

[48] A. Samuylov, D. Moltchanov, J. Pirskanen, J. Numminen, Y. Koucheryavy, and M. Valkama. "Performance Assessment of DECT-2020 NR and Classic DECT Coexistence Mechanisms". In: *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*. IEEE. 2023, pp. 1–7.

[49] N. Stepanov, D. Moltchanov, and A. Turlikov. "Modeling the NB-IoT Transmission Process with Intermittent Network Availability". In: *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. Springer, 2020, pp. 241–254.

[50] C. Bockelmann, N. K. Pratas, G. Wunder, S. Saur, M. Navarro, D. Gregoratti, G. Vivier, E. De Carvalho, Y. Ji, Č. Stefanović, et al. "Towards Massive Connectivity Support for Scalable mMTC Communications in 5G Networks". In: *IEEE access* 6 (2018), pp. 28969–28992.

[51] M. Stusek, K. Zeman, P. Masek, J. Sedova, and J. Hosek. "IoT Protocols for Low-power Massive IoT: A Communication Perspective". In: *2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. IEEE. 2019, pp. 1–7.

[52] N. Ali, W. Saad, and D. (Eds.) Steinbach. *White Paper on Machine Learning in 6G Wireless Communication Networks*. [White paper]. (6G Research Visions, No. 7). University of Oulu. 2020.

[53] R. Sanchez-Iborra, J. Santa, J. Gallego-Madrid, S. Covaci, and A. Skarmeta. "Empowering the Internet of Vehicles with Multi-RAT 5G Network Slicing". In: *Sensors* 19.14 (2019), pp. 1–16.

[54] H. Tabassum, M. Salehi, and E. Hossain. "Mobility-Aware Analysis of 5G and B5G Cellular Networks: A Tutorial". In: *arXiv preprint:1805.02719* (2018).

[55] K. Mikhaylov, J. Petäjäjärvi, M. Mäkeläinen, A. Paatelma, and T. Hänninen. "Extensible Modular Wireless Sensor and Actuator Network and IoT Platform with Plug&Play Module Connection". In: *Proc. 14th Int. Conf. Inf. Proces. in Sensor Netw.* Apr. 2015, pp. 386–387.

[56] O. Galinina, A. Pyattaev, S. Andreev, M. Dohler, and Y. Koucheryavy. "5G multi-RAT LTE-WiFi Ultra-dense Small Cells: Performance Dynamics, Architecture, and Trends". In: *IEEE J. Selected Areas in Commun.* 33.6 (2015), pp. 1224–1240.

[57] A. Ometov, N. Daneshfar, A. Hazmi, S. Andreev, L. F. D. Carpio, P. Amin, J. Torsner, Y. Koucheryavy, and M. Valkama. "System-level Analysis of IEEE 802.11ah Technology for Unsaturated MTC Traffic". In: *International Journal of Sensor Networks* 26.4 (2018), pp. 269–282.

[58] K. Mikhaylov, M. Stusek, P. Masek, V. Petrov, J. Petajajarvi, S. Andreev, J. Pokorny, J. Hosek, A. Pouttu, and Y. Koucheryavy. "Multi-RAT LPWAN in Smart Cities: Trial of LoRaWAN and NB-IoT Integration". In: *2018 IEEE International Conference on Communications (ICC)*. IEEE. 2018, pp. 1–6.

[59] J. Yun, Y. Goh, W. Yoo, and J.-M. Chung. "5G Multi-RAT URLLC and eMBB Dynamic Task Offloading with MEC Resource Allocation Using Distributed Deep Reinforcement Learning". In: *IEEE Internet of Things Journal* 9.20 (2022), pp. 20733–20749.

[60] M. Saily, G. Sebire, and E. Riddington. *GSM/EDGE: Evolution and Performance*. John Wiley & Sons, 2011. ISBN: 9781119972976.

[61] T. Halonen, J. Romero, and J. Melero. *GSM, GPRS and EDGE Performance: Evolution Towards 3G/UMTS*. John Wiley & Sons, 2004. ISBN: 9780470866962.

[62] O. Liberg, M. Sundberg, Y.-P. E. Wang, J. Bergman, J. Sachs, and G. Wikström. *Cellular Internet of Things: From Massive Deployments to Critical 5G Applications*. Second Edition. Academic Press, 2020. ISBN: 978-0-08-102902-2.

[63] E. Dahlman, S. Parkvall, and J. Skold. *4G: LTE/LTE-advanced for mobile broadband*. Academic press, 2013. ISBN: 978-0-12-419985-9.

[64] H. Holma, A. Toskala, and J. Reunanen. *LTE Small Cell Optimization: 3GPP Evolution to Release 13*. John Wiley & Sons, 2016. ISBN: 9781118912577.

[65] S. Ahmadi. *5G NR: Architecture, Technology, Implementation, and Operation of 3GPP New Radio Standards*. Academic Press, 2019.

[66] Qualcomm. *Expanding the 5G NR ecosystem 5G NR roadmap in 3GPP Release 16 and beyond*. Online; accessed: 15-April-2021. 2018. URL: https://www.qualcomm.com/media/documents/files/expanding-the-5g-nr-ecosystem-and-roadmap-in-3gpp-rel-16-beyond.pdf.

[67] Z. Qin, F. Y. Li, G. Y. Li, J. A. McCann, and Q. Ni. "Low-power Wide-area Networks for Sustainable IoT". In: *IEEE Wireless Communications* 26.3 (2019), pp. 140–145.

[68] Y.-P. E. Wang, X. Lin, A. Adhikary, A. Grovlen, Y. Sui, Y. Blankenship, J. Bergman, and H. S. Razaghi. "A Primer on 3GPP Narrowband Internet of Things". In: *IEEE Communications Magazine* 55.3 (2017), pp. 117–123. DOI: 10.1109/MCOM.2017.1600510CM.

[69] P. Andres-Maldonado, P. Ameigeiras, J. Prados-Garzon, J. Navarro-Ortiz, and J. M. Lopez-Soler. "Narrowband IoT Data Transmission Procedures for Massive Machine-Type Communications". In: *IEEE Network* 31.6 (2017), pp. 8–15. DOI: 10.1109/MNET.2017.1700081.

[70] F. Muteba, K. Djouani, and T. Olwal. "A Comparative Survey Study on LPWA IoT Technologies: Design, Considerations, Challenges and Solutions". In: *Procedia Computer Science* 155 (2019), pp. 636–641.

[71] M. Vaezi, A. Azari, S. R. Khosravirad, M. Shirvanimoghaddam, M. M. Azari, D. Chasaki, and P. Popovski. "Cellular, Wide-Area, and Non-Terrestrial IoT: A Survey on 5G Advances and the Road Toward 6G". In: *IEEE Communications Surveys & Tutorials* 24.2 (2022), pp. 1117–1174. DOI: 10.1109/COMST.2022.3151028.

[72] S. Painuly, S. Sharma, and P. Matta. "Future Trends and Challenges in Next Generation Smart Application of 5G-IoT". In: *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. 2021, pp. 354–357. DOI: 10.1109/ICCMC51019.2021.9418471.

[73] S. Ugwuanyi, G. Paul, and J. Irvine. "Survey of IoT for Developing Countries: Performance Analysis of LoRaWAN and Cellular NB-IoT Networks". In: *Electronics* 10.18 (2021), p. 2224.

[74] S. W. H. Shah, A. N. Mian, A. Aijaz, J. Qadir, and J. Crowcroft. "Energy-Efficient MAC for Cellular IoT: State-of-the-Art, Challenges, and Standardization". In: *IEEE Transactions on Green Communications and Networking* 5.2 (2021), pp. 587–599. DOI: 10.1109/TGCN.2021.3062093.

[75] GSMA. *NB-IoT Deployment Guide to Basic Feature set Requirements*. White Paper. GSMA, June 2019.

[76] *Mobile IoT Deployment Guide*. White Paper. GSMA, Oct. 2022.

[77] Y. Chen, Y. A. Sambo, O. Onireti, and M. A. Imran. "A Survey on LPWAN-5G Integration: Main Challenges and Potential Solutions". In: *IEEE Access* 10 (2022), pp. 32132–32149. DOI: 10.1109/ACCESS.2022.3160193.

[78] V. Petrov, A. Samuylov, V. Begishev, D. Moltchanov, S. Andreev, K. Samouylov, and Y. Koucheryavy. "Vehicle-based Relay Assistance for Opportunistic Crowdsensing Over Narrowband IoT (NB-IoT)". In: *IEEE Internet of Things journal* 5.5 (2018), pp. 3710–3723.

[79] M. Kanj, V. Savaux, and M. Le Guen. "A Tutorial on NB-IoT Physical Layer Design". In: *IEEE Communications Surveys & Tutorials* (2020).

[80] Third Generation Partnership Project. "Technical Specifications 24.301 v13.8.0, Technical Specification Group Core Network and Terminals; Non-access-Stratum (NAS) Protocol for Evolved Packet System (EPS)". In: (2016).

[81] Third Generation Partnership Project. "Technical Specifications 24.301 v13.5.0, Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 Application Protocol (S1AP)". In: (2017).

[82] T.-Y. Wu, R.-H. Hwang, A. Vyas, C.-Y. Lin, and C.-R. Huang. "Persistent Periodic Uplink Scheduling Algorithm for Massive NB-IoT Devices". In: *Sensors* 22.8 (2022). ISSN: 1424-8220. DOI: 10.3390/s22082875. URL: https://www.mdpi.com/1424-8220/22/8/2875.

[83] F. Yassine, M. El Helou, S. Lahoud, and O. Bazzi. "Energy-Efficient Uplink Scheduling in Narrowband IoT". In: *Sensors* 22.20 (2022). ISSN: 1424-8220. DOI: 10.3390/s22207744. URL: https://www.mdpi.com/1424-8220/22/20/7744.

[84] A. P. Matz, J.-A. Fernandez-Prieto, J. Cañada-Bago, and U. Birkel. "A Systematic Analysis of Narrowband IoT Quality of Service". In: *Sensors* 20.6 (2020). ISSN: 1424-8220. DOI: 10.3390/s20061636. URL: https://www.mdpi.com/1424-8220/20/6/1636.

[85] GSMA. *LTE-M Deployment Guide to Basic Feature Set Requirements*. White Paper. GSMA, June 2019.

[86] Enterprise IoT Insights. *NB-IoT - what has gone wrong, and when will it go right?* [Report]. Oct. 2021.

[87] GSMA. *Smart Meters Fact Sheet*. [Report]. Aug. 2023. URL: https://www.gsma.com/iot/resources/smart-meters-fact-sheet/.

[88] European Commission. *Benchmarking Smart Metering Deployment in the EU-28*. Accessed 30 June 2023. Mar. 2020. URL: https://energy.ec.europa.eu/publications/benchmarking-smart%5Cbreak%20-metering-deployment-eu-28%5C_en.

[89] European Union Agency for the Cooperation of Energy Regulators and the Council of European Energy Regulators. *Energy Retail and Consumer Protection 2023 Market Monitoring Report*. [Report]. Sept. 2023. URL: https://www.acer.europa.eu/sites/default/files/documents/Publications/2023_MMR_Energy_Retail_Consumer_Protection.pdf.

[90] TZB. *E.ON installs the first smart electricity meters in its network. It will enable remote readings*. Accessed 30 June 2023. Oct. 2020. URL: https://www.tzb-info.cz/ceny-paliv-a-energii/127058-e-on-%5Cbreak%20instaluje-do-sve-site-prvni-chytre-elektromery-umozni-%5Cbreak%20dalkove-odecty.

[91] MPO. *New decree on electricity metering No. 359/2020 Coll.* Accessed 30 June 2023. Jan. 2021. URL: https://www.mpo.cz/assets/cz/energetika/strategicke-a-%5Cbreak%20koncepcni-dokumenty/narodni-akcni-plan-pro-chytre-site/2021/%5Cbreak1/Prezentace-MPO-vyhlaska-o-mereni-elektriny.pdf.

[92] European Commission. *Energy Performance of Buildings Directive*. Accessed 30 June 2023. Jan. 2018. URL: https://energy.ec.europa.eu/topics/energy-efficiency/energy-%5Cbreak%20efficient-buildings/energy-performance-buildings-directive%5C_en.

[93] O. Bularca, M. Florea, and A.-M. Dumitrescu. "Smart Metering Deployment Status Across EU-28". In: *2018 International Symposium on Fundamentals of Electrical Engineering (ISFEE)*. IEEE. 2018, pp. 1–6.

[94] Vodafone Czech Republic. *Vodafone will light up a nationwide network for NB-IoT this year*. Accessed 30 June 2023. July 2017. URL: https://www.vodafone.cz/en/about-vodafone/press-releases/%5Cbreak%20message-detail/vodafone-letos-rozsviti-celonarodni-sit-pro-%5Cbreak%20nb-iot/.

[95] EG.D. *HES System for the Smart Metering Scenarios*. Accessed 30 June 2023. Jan. 2020. URL: https://ezak.eon.cz/contract%5C_display%5C_257.html.

[96] X. Chang, J. Zhan, G. Xing, J. Huang, B. Chen, and L. Zhou. "Measurement-Based Optimization of Cell Selection in NB-IoT Networks". In: *ACM Trans. Sen. Netw.* 18.4 (Nov. 2022). ISSN: 1550-4859. DOI: 10.1145/3544017. URL: https://doi.org/10.1145/3544017.

[97] Y.-M. Kim, D. Jung, Y. Chang, and D.-H. Choi. "Intelligent Micro Energy Grid in 5G Era: Platforms, Business Cases, Testbeds, and Next Generation Applications". In: *Electronics* 8.4 (2019), p. 468.

[98] Y. Shen, W. Fang, F. Ye, and M. Kadoch. "EV Charging Behavior Analysis Using Hybrid Intelligence for 5G Smart Grid". In: *Electronics* 9.1 (2020), p. 80.

[99] Vodafone Czech Republic. *Narrowband-IoT for Massive IoT: Capacity Study for Selected Areas*. Accessed: 2023-11-15. URL: https://www.vut.cz/www_base/vutdisk.php?i=324298a408.

[100] Vodafone Czech Republic. *Vodafone is Going to Use NB-IoT to Extend Implementation of the Internet of Things*. Accessed: 2023-11-15. URL: https://www.vodafone.cz/en/about-vodafone/press-releases/message-detail/vodafone-si-pro-rozsireni-sluzby-internet-veci-vyb/.

[101] Vodafone Czech Republic. *Vodafone to Light Up a Nationwide NB-IoT Network*. Accessed: 2023-11-15. URL: https://www.vodafone.cz/en/about-vodafone/press-releases/message-detail/vodafone-letos-rozsviti-celonarodni-sit-pro-nb-iot/.

[102] Vodafone Czech Republic. *Vodafone's IoT Network Coverage of the Czech Republic Completed*. Accessed: 2023-11-15. URL: https://www.vodafone.cz/en/about-vodafone/press-releases/message-detail/vodafone-dokoncil-pokryti-ceske-republiky-siti-pro/.

[103] A. Abou El Hassan, A. El Mehdi, and M. Saber. "NarrowBand-IoT and eMTC Towards Massive MTC: Performance Evaluation and Comparison for 5G mMTC". In: *Networking, Intelligent Systems and Security*. Springer, 2022, pp. 177–195.

[104] A. Chehri, H. Chaibi, R. Saadane, E. M. Ouafiq, and A. Slalmi. "On the Performance of 5G Narrow-Band Internet of Things for Industrial Applications". In: *Networking, Intelligent Systems and Security*. Springer, 2022, pp. 275–286.

[105] Office of Electricity Delivery and Energy Reliability. *Advanced Metering Infrastructure and Customer Systems*. Accessed: 2024-03-04. 2016. URL: https://www.energy.gov/sites/prod/files/2016/12/f34/AMI%20Summary%20Report_09-26-16.pdf.

[106] Dr. Robert Riemann. *TechDispatch #2: Smart Meters in Smart Homes*. Accessed: 2024-03-04. 2019. URL: https://edps.europa.eu/data-protection/our-work/publications/techdispatch/techdispatch-2-smart-meters-smart-homes_en.

[107] Pavel Solc. *Smart Metering in the Czech Republic, Viewed by ČEZ Distribuce*. Accessed: 2024-03-04. 2017. URL: https://www.cez.cz/edee/content/file-other/distribucni-sluzby/konference-2017/06_solc_dso_en.pdf.

[108] A. Mäkivierikko, H. Siepelmeyer, H. Shahrokni, D. Enarsson, and O. Kordas. "Reducing electricity peak loads through 'pause hours' - a community-based behavioural demand response approach". In: *Journal of Cleaner Production* 408 (2023), p. 137064. ISSN: 0959-6526. DOI: https://doi.org/10.1016/j.jclepro.2023.137064. URL: https://www.sciencedirect.com/science/article/pii/S0959652623012222.

[109] F. Farokhi. "Review of Results on Smart-Meter Privacy by Data Manipulation, Demand Shaping, and Load Scheduling". In: *IET smart grid* 3.5 (2020), pp. 605–613. ISSN: 2515-2947. DOI: 10.1049/iet-stg.2020.0129. URL: doi:10.1049/iet-stg.2020.0129.

[110] S. Lee, S. H. Nengroo, H. Jin, Y. Doh, C. Lee, T. Heo, and D. Har. "Anomaly Detection of Smart Metering System for Power Management with Battery Storage System/Electric Vehicle". In: *ETRI Journal* 45.4 (2023), pp. 650–665.

[111] Z. A. Obaid, L. M. Cipcigan, L. Abrahim, and M. T. Muhssin. "Frequency control of future power systems: reviewing and evaluating challenges and new control methods". In: *Journal of Modern Power Systems and Clean Energy* 7.1 (2019), pp. 9–25. DOI: 10.1007/s40565-018-0441-1.

[112] J. Schlienz and D. Raddino. "Narrowband Internet of Things Whitepaper". In: *White Paper, Rohde&Schwarz* (2016), pp. 1–42.

[113] A. K. Bachkaniwala, V. Dhanwani, S. S. Charan, D. Rawal, and S. K. Devar. "IMT-2020 Evaluation of EUHT Radio Interface Technology". In: *2020 IEEE 3rd 5G World Forum (5GWF)*. IEEE. 2020, pp. 631–636.

[114] L. Quectel Wireless Solutions Co. *BG77 Hardware Design*. (Accessed on 2022-08-01). Shanghai, 2021. URL: https://www.quectel.com/wp-content/uploads/2021/03/Quectel_BG77_Hardware_Design_V1.2.pdf.

[115] B. S. Tsybakov and V. A. Mikhailov. "Free Synchronous Packet Access in a Broadcast Channel with Feedback". In: *Problemy peredachi informatsii* 14.4 (1978), pp. 32–59.

[116] J. Capetanakis. “Tree Algorithms for Packet Broadcast Channels”. In: *IEEE transactions on information theory* 25.5 (1979), pp. 505–515.

[117] L. Kleinrock and S. Lam. “Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation”. In: *IEEE transactions on Communications* 23.4 (1975), pp. 410–423.

[118] R. Harwahyu, R.-G. Cheng, C.-H. Wei, and R. F. Sari. “Optimization of Random Access Channel in NB-IoT”. In: *IEEE Internet of Things Journal* 5.1 (2018), pp. 391–402. DOI: 10.1109/JIOT.2017.2786680.

[119] C.-H. Wei, G. Bianchi, and R.-G. Cheng. “Modeling and Analysis of Random Access Channels With Bursty Arrivals in OFDMA Wireless Networks”. In: *IEEE Transactions on Wireless Communications* 14.4 (2015), pp. 1940–1953. DOI: 10.1109/TWC.2014.2377121.

[120] B. Van Houdt and C. Blondia. “Throughput of Q-ary Splitting Algorithms for Contention Resolution in Communication Networks”. In: *Communications in information and systems* 4.2 (2005), pp. 135–164.

[121] B. Van Houdt and C. Blondia. “Stability and Performance of Stack Algorithms for Random Access Communication Modeled as a Tree Structured QBD Markov Chain”. In: *Stochastic Models* 17.3 (2001), pp. 247–270.

[122] R. Block and B. Van Houdt. “Design and analysis of multiple access algorithms”. PhD thesis. University of Antwerp, 2017.

[123] C. Stefanovic, P. Popovski, and D. Vukobratovic. “Frameless ALOHA Protocol for Wireless Networks”. In: *IEEE Communications Letters* 16.12 (2012), pp. 2087–2090.

[124] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic. “Random Access for M2M Communications With QoS Guarantees”. In: *IEEE Transactions on Communications* 65.7 (2017), pp. 2889–2903.

[125] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic. “A Novel Analytical Framework for Massive Frant-free NOMA”. In: *IEEE Transactions on Communications* 67.3 (2018), pp. 2436–2449.

[126] G. C. Madueno, N. K. Pratas, Č. Stefanović, and P. Popovski. “Massive M2M Access with Reliability Guarantees in LTE Systems”. In: *2015 IEEE International Conference on Communications (ICC)*. IEEE. 2015, pp. 2997–3002.

[127] M. S. Haghighi. “Critical Study of Markovian Approaches for Batch Arrival Modeling in IEEE 802.15. 4-based Networks”. In: *arXiv preprint arXiv:2003.00829* (2020).

[128] B. Yu, X. Chi, and H. Sun. “Delay Analysis for Aggregate Traffic Based on Martingales Theory”. In: *IET Communications* 14.5 (2020), pp. 760–767.

[129] D. T. C. Wong, Q. Chen, X. Peng, and F. Chin. “Multi-Channel Pure Collective Aloha MAC Protocol with Decollision Algorithm for Satellite Uplink”. In: *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*. IEEE. 2018, pp. 251–256.

[130] M. R. Chowdhury and S. De. “Delay-Aware Priority Access Classification for Massive Machine-Type Communication”. In: *IEEE Transactions on Vehicular Technology* 70.12 (2021), pp. 13238–13254. DOI: 10.1109/TVT.2021.3120280.

[131] S. Rostami, S. Lagen, M. Costa, M. Valkama, and P. Dini. “Wake-Up Radio Based Access in 5G Under Delay Constraints: Modeling and Optimization”. In: *IEEE Transactions on Communications* 68.2 (2020), pp. 1044–1057. DOI: 10.1109/TCOMM.2019.2954389.

[132] W. Sun, H. Zhang, R. Wang, and Y. Zhang. “Reducing Offloading Latency for Digital Twin Edge Networks in 6G”. In: *IEEE Transactions on Vehicular Technology* 69.10 (2020), pp. 12240–12251. DOI: 10.1109/TVT.2020.3018817.

[133] R.-G. Cheng, J. Chen, D.-W. Chen, and C.-H. Wei. “Modeling and Analysis of an Extended Access Barring Algorithm for Machine-Type Communications in LTE-A Networks”. In: *IEEE Transactions on Wireless Communications* 14.6 (2015), pp. 2956–2968. DOI: 10.1109/TWC.2015.2398858.

[134] R. Harwahyu, R.-G. Cheng, W.-J. Tsai, J.-K. Hwang, and G. Bianchi. “Repetitions Versus Retransmissions: Tradeoff in Configuring NB-IoT Random Access Channels”. In: *IEEE Internet of Things Journal* 6.2 (2019), pp. 3796–3805.

[135] F. Metzger, T. Hoßfeld, A. Bauer, S. Kounev, and P. E. Heegaard. “Modeling of Aggregated IoT Traffic and its Application to an IoT Cloud”. In: *Proceedings of the IEEE* 107.4 (2019), pp. 679–694.

[136] C. Correia, A. C. M. Freitas, and J. M. Freitas. “Cluster Distributions for Dynamically Defined Point Processes”. In: *Physica D: Nonlinear Phenomena* 457 (2024), p. 133968.

[137] N. Stepanov, A. Turlikov, and V. Begishev. “Balancing the Data Transmission and Random Access Phases in 6G mMTC Radio Technologies”. In: *IEEE Communications Letters* (2023).

[138] 3GPP. *LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); LTE Physical Layer;* TS 36.201 V14.1.0. ETSI, Apr. 2017.

[139] A. Adhikary, X. Lin, and Y.-P. E. Wang. "Performance Evaluation of NB-IoT Coverage". In: *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall).* IEEE. 2016, pp. 1–5.

[140] L. Feltrin, G. Tsoukaneri, M. Condoluci, C. Buratti, T. Mahmoodi, M. Dohler, and R. Verdone. "Narrowband IoT: A Survey on Downlink and Uplink Perspectives". In: *IEEE Wireless Communications* 26.1 (2019), pp. 78–86.

[141] M. Koseoglu. "Lower Bounds on the LTE-A Average Random Access Delay Under Massive M2M Arrivals". In: *IEEE Transactions on Communications* 64.5 (2016), pp. 2104–2115.

[142] O. Galinina, A. Turlikov, S. Andreev, and Y. Koucheryavy. "Stabilizing Multichannel Slotted Aloha for Machine-type Communications". In: *2013 IEEE International Symposium on Information Theory.* IEEE. 2013, pp. 2119–2123.

[143] B. N. Feinberg and S. S. Chiu. "A Method to Calculate Steady-state Distributions of Large Markov Chains by Aggregating States". In: *Operations Research* 35.2 (1987), pp. 282–290.

[144] G. Franceschinis and R. R. Muntz. "Bounds for Quasi-lumpable Markov Chains". In: *Performance Evaluation* 20.1-3 (1994), pp. 223–243.

[145] W. Szpankowski. "Statistic Analysis of Multiaccess Systems with Random Access and Feedback". PhD thesis. University of Gdansk, 1980.

[146] J. G. Kemeny and J. L. Snell. *Finite Markov Chains: with a New Appendix "Generalization of a Fundamental Matrix".* Springer, 1983.

[147] H. G. Perros. *Computer simulation techniques: The definitive introduction!* 2009.

[148] S. A. Gbadamosi, G. P. Hancke, and A. M. Abu-Mahfouz. "Building Upon NB-IoT Networks: A Roadmap Towards 5G New Radio Networks". In: *IEEE Access* 8 (2020), pp. 188641–188672.

[149] L. Kleinrock. "Theory". In: *Queueing systems* 2 (1975).

[150] A. A. E. Hassan, A. E. Mehdi, and M. Saber. "NB-IoT and LTE-M Towards Massive MTC: Complete Performance Evaluation for 5G mMTC". In: *Indonesian Journal of Electrical Engineering and Computer Science* (2021). URL: `https://api.semanticscholar.org/CorpusID:237770694`.

[151] Jörke, Pascal and Gebauer, Tim and Wietfeld, Christian. "From LENA to LENA-NB: Implementation and Performance Evaluation of NB-IoT and Early Data Transmission in NS-3". In: *Proc. of the 2022 Workshop on Ns-3*. WNS3 '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 73–80. ISBN: 9781450396516. DOI: 10.1145/3532577.3532600. URL: https://doi.org/10.1145/3532577.3532600.

[152] K. Mikhaylov, H. Alves, and Y.-P. M. Hoyhtya. *Integration of MTC and Satellites for IoT toward 6G Era*. First Edition. Wiley-IEEE Press, 2024. ISBN: 978-1-119-98214-2.

[153] A. Sattarzadeh, Y. Liu, A. Mohamed, R. Song, P. Xiao, Z. Song, H. Zhang, R. Tafazolli, and C. Niu. "Satellite-Based Non-Terrestrial Networks in 5G: Insights and Challenges". In: *IEEE Access* 10 (2022), pp. 11274–11283. DOI: 10.1109/ACCESS.2021.3137560.

[154] L. Ma, C. Park, X. Wang, P. Gaal, and A. R. Alvarino. "Uplink Time Synchronization for Non-Terrestrial Networks". In: *IEEE Communications Magazine* 61.7 (2023), pp. 114–118. DOI: 10.1109/MCOM.001.2200626.

Symbols and abbreviations

3GPP	3rd Generation Partnership Project
ACB	Access Class Barring
AKA	Authentication and Key Agreement
AMI	Advanced Metering Infrastructure
AR	Augmented Reality
AS	Access Stratum
B5G	Beyond 5G
BEST	Battery Efficiency Security for low Throughput
BSs	Base Stations
BPSK	Binary-Phase Shift Keying
BW	Bandwidth
CA	Carrier Aggregation
CDM	Code Division Multiplexing
CE	Coverage Enhancement
CIR	Committed Information Rate
CIoT	Cellular Internet of Things
COSEM	Companion Specification for Energy Metering
CP	Control Plane
DCI	Downlink Control Information
DMRS	Demodulation Reference Signal
DL	Downlink
DLMS	Device Language Message Specification
DPAC	Delay-Aware Priority Access Classification
DRL	Deep Reinforced Learning
DT	Digital Twin
EDGE	GSM Evolution
EDs	End Devices
EDT	Early Data Transmission
ECL	Enhanced Coverage Level
E-CID	Enhanced Cell-ID
eDRX	extended Discontinuous Reception
eMBB	Enhanced Mobile Broadband
eMTC	Enhanced Machine-Type Communication
eNB	eNodeB
EPS	Evolved Packet System

EIRP	Equivalent Isotropically Radiated Power
EV	Electric Vehicle
FCR	Frequency Containment Reserve
FDD	Frequency-Division Duplex
FR	Frequency Range
GF	Grant-Free
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communication
GWUS	Group WUS
HARQ	Hybrid Automatic Repeat Request
H2H	Human-to-Human
HES	Head-End System
IAB	Integrated Access Backhaul
IIoT	Industrial Internet of Things
IP	Internet Protocol
IoT	Internet of Things
ITU-R	International Telecommunication Union Radiocommunication Sector
KPIs	Key Performance Indicators
LAN	Local Area Network
LoRaWAN	Long Range Wide Area Network
LPWA	Low-Power Wide-Area
LPWAN	Low-Power Wide Area Network
LTE	Long-Term Evolution
LTE-M	Long-Term Evolution Machine Type Communication
M2M	Machine-to-Machine
MCSs	Modulation and Coding Schemes
MEC	Mobile Edge Computing
MIB	Main Information Block
MIMO	Multiple-Input Multiple-Output
ML	Machine learning
MCL	Maximum Coverage Limit
MME	Mobility Management Entity
mMTC	Massive Machine-Type Communications
MTC	Machine-Type Communication
MT-EDT	Mobile Terminated EDT
MU-MIMO	Multi-User MIMO

Multi-RAT	Multi-Radio Access Technology
NAS	Non-Access Stratum
NaaS	Network as a Service
N3IWF	Non-3GPP InterWorking Function
NB-IoT	Narrow Band Internet of Things
NIDD	Non-IP Data Delivery
NOMA	Non-Orthogonal Multiple Access
NPBCH	Narrowband Physical Broadcast Channel
NPDCCH	Narrowband Physical Downlink Control Channel
NPDSCH	Narrowband Physical Downlink Shared Channel
NPRACH	Narrowband Physical Random Access Channel
NPSS	Narrowband Primary Synchronization Signal
NPUCCH	Narrowband Physical Uplink Control Channel
NPUSCH	Narrowband Physical Uplink Shared Channel
NPNs	Non-Public Networks
NR	New Radio
NRS	Narrowband Reference Signal
NRSRP	Narrowband Reference Signal Received Power
NRSRQ	Narrowband Reference Signal Received Quality
NSA	Non-Standalone
NSPS	Security and Public Safety
NTN	Non-Terrestrial Network
OTDOA	Observed Time Difference of Arrival
PAPR	Peak-to-Average Power Ratio
PRACH	Physical Random Access Channel
PRB	Physical Resource Block
PRS	Positioning Reference Signal
PUF	Physical Unclonable Function
PUR	Preconfigured Uplink Resources
PUSCH	Physical Uplink Shared Channel
PSK	Phase Shift Keying
PSM	Power Saving Mode
QoS	Quality of Service
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
RA	Random Access
RAI	Release Assistance Information

RAP	Random Access Procedure
RAT	Radio Access Technology
RedCap	Reduced Capability
RF	Radio Frequency
RLF	Radio Link Failure
RO	Research Objective
RQ	Research Question
RRC	Radio Resource Control
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSS	Resynchronisation Signal
RSSI	Received Signal Strength Indicator
RTT	Round-Trip Time
SC-PTM	Single-cell Point-to-Multipoint
SIB	System Information Block
SIB2	System Information Block Type 2
SIC	Interference Cancellation
SINR	Signal to Interference and Noise Ratio
SDR	Software-Defined Radio
SG	Smart Grids
SGTF	Smart Grids Task Force
SR	Scheduling Request
TA	Timing advance
TAU	Tracking Area Update
TCP	Transmission Control Protocol
TDD	Time-Division Duplex
TLS	Transport Layer Security
TTI	Transmission Time Interval
UL	Uplink
UP	User Plane
URRLC	Ultra-Reliable Low-Latency Communications
UE	User Equipment
VR	Virtual Reality
WUS	Wakeup Signal
XR	Mixed Reality

List of appendices

A Currently deployed variants of NB-IoT and eMTC

Tab. A.1: Comparison of NB-IoT and eMTC parameters in current deployments.

Technology	NB-IoT	eMTC
Frequency	700–2,100 MHz	700–2,600 MHz
Bandwidth	200 kHz	1.4 MHz[1]
Link budget	164 dB	157.7 dB
Max. EIRP	23 dBm	23 dBm
Max. PDCP payload	1,600 B	8,188 B
Max. TBS	1000 bits UL / 680 bits DL	1000 bits UL/DL
Max. HARQ	1	8
UL data rate	0.3 – 62.5 kbps[1] 0.3 – 159 kbps[2]	HD: 375[1]kbps FD: 1[1]Mbps
DL data rate	0.5 – 27.2 kbps[1] 0.5 – 127 kbps[2]	HD: 300[1]kbps FD: 0.8[1]Mbps
Coverage enhancements	Level 0 – 2	Mode A only
(N)PDSCH/(N)PUSCH repetitions	1 – 2,048/1 – 128	1 – 32/1 – 32 (Mode A)
(N)PRACH rep.	1–128	1–128
(N)PRACH BW	180 kHz	1.05 MHz
(N)PRACH TTI [ms]	5.6[1,3] 6.4[1,4] 19.2[5]	1[3] 2[4,5] 3[6]
Category NB2/M2	Yes (NB2)[2]	No (M2)[2]
Multicast	No[2]	No[2]
CP/UP Optimization	CP[1]	UP[1]
PSM/eDRX	Yes[1]	Yes[1]
NIDD	No[1]	No[1]
VoLTE	-	No[1]
OTDOA/E-CID Pos.	No[2]	No[2]
UE Power Class 6	Yes[2]	-
Conn. mode mobility	No[2]	Yes[1]
RAI	Yes[2]	Yes[2]
Relaxed Resel. Mon.	No[2]	No[2]

[1] 3GPP Rel. 13. [2] 3GPP Rel. 14. [3] (N)PRACH format 0.
[4] (N)PRACH format 1. [5] (N)PRACH format 2. [6] PRACH format 3.